National Technical University of Athens
School of Civil Engineering
Department of Transportation Planning and Engineering

Doctoral Dissertation

The Driver Behavior Telematics Feedback Mechanism

Armira Kontaxi

Civil Transportation Engineer, NTUA

Supervising Committee:

George Yannis, Professor NTUA
Eleni Vlahogianni, Professor NTUA
Eleonora Papadimitriou, Assistant Professor NTUA

Athens, February 2025

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Πολιτικών Μηχανικών
Τομέας Μεταφορών και Συγκοινωνιακής Υποδομής

Διδακτορική Διατριβή

Μηχανισμός Ανατροφοδότησης της Συμπεριφοράς του Οδηγού Μέσω Τηλεματικής

Αρμίρα Κονταξή
Πολιτικός Μηχανικός, Συγκοινωνιολόγος ΕΜΠ

Συμβουλευτική Επιτροπή:

Γιώργος Γιαννής, Καθηγητής ΕΜΠ
Ελένη Βλαχογιάννη, Καθηγήτρια ΕΜΠ
Ελεονώρα Παπαδημητρίου, Επίκουρη Καθηγήτρια ΕΜΠ

Αθήνα, Φεβρουάριος 2025

Contact Details

Colleagues, students, or anyone interested in obtaining further information about this PhD dissertation, please feel free to contact me using the following contact details:

Armira Kontaxi

National Technical University of Athens (NTUA)
Department of Transportation Planning and Engineering
5 Heroon Polytechniou Str., Zografou Campus, GR-15773, Athens, Greece

Professional e-mail: akontaxi@mail.ntua.gr
Personal e-mail: amkontaxi@gmail.com

Phone (NTUA): +30.210.772.1575

Researcher website: https://www.nrso.ntua.gr/akontaxi

In memory of my beloved father, Ardi

Acknowledgments

As I reach the completion of this doctoral dissertation, I would like to express my deepest gratitude to those who have supported and guided me throughout this journey.

First and foremost, I extend my sincere appreciation to my supervisor, Professor George Yannis, for granting me the opportunity to embark on this remarkable research path. His support, insightful guidance, and inspiring mentorship have been instrumental in shaping both my academic and professional growth. Beyond his invaluable advice, his wisdom and encouragement have provided me with lifelong lessons that extend far beyond research.

I am equally grateful to my co-supervisors, Professor Eleni Vlahogianni and Assistant Professor Eleonora Papadimitriou, for their continuous support and invaluable contributions throughout every stage of this dissertation. I also sincerely thank the esteemed members of my examination committee, Professor Constantinos Antoniou, Associate Professor Stergios Mavromatis, Assistant Professor Konstantinos Gkiotsalitis, and Assistant Professor Athanasios Theofilatos, for their constructive feedback, academic insight, and thoughtful discussions, which greatly enriched my work.

I would also like to express my sincere gratitude to OSeven Telematics and its team for providing the naturalistic driving data that formed the foundation of this research, especially to Dr. Petros Fortsakis for our seamless collaboration throughout these years.

A special acknowledgment goes to my colleague and friend, Dr. Apostolos Ziakopoulos, whose unwavering belief in my potential has been truly invaluable. His guidance, covering everything from statistical modeling to research concepts, methodologies and even life hacks, has played a vital role in my academic journey and the completion of this dissertation.

I am incredibly fortunate to have been surrounded by an inspiring community of colleagues from my beloved NRSO team and the other research groups at the Department of Transportation Planning and Engineering. Their collaboration, support, and encouragement have enriched both my academic journey and daily work life. A heartfelt thank you to Julia and Dimitris, with whom I started this journey, as well as to every colleague who has contributed in one way or another to its completion and become a dear friend along the way. Special thanks to my 201 office mates, past and present, for the laughter, engaging discussions, and meaningful or even silly conversations during our breaks. To all of you, thank you for making this experience a little more fun, enriching and rewarding.

I am also deeply grateful to my dearest lifelong friends, Aliki, Yota, and Eirini, for standing by my side through every high and low.

Finally, I would like to express my gratitude to my family, my mother, Mirela, and my brothers, Agron and Antzelo for their unconditional love and support.

This PhD thesis has been performed within the framework of the following research project carried out by the Department of Transportation Planning and Engineering of the School of Civil Engineering of the National Technical University of Athens.

- “BeSmart – Multi-modal driver behavior and safety support system on the basis of smartphone applications” co-financed by the European Union – European Regional Development Fund (ERDF) and Greek national funds through the Operational Program “Competitiveness, Entrepreneurship and Innovation” (EPAnEK) of the National Strategic Reference Framework (NSRF) – Research Funding Program (project code: T1EΔK-03405).

Table of Contents

List of Figures

List of Tables

Abstract

Road safety remains a critical global concern, with traffic crashes claiming millions of lives annually and imposing significant emotional and economic burdens. Human error accounts for 95% of road crashes, underscoring the pivotal role of driver behavior in enhancing safety. This dissertation investigates the impact of driver telematics feedback mechanism on driving behavior, addressing gaps in understanding feedback's lifecycle effects, including pre-feedback, feedback, and post-feedback phases. Employing a multi-modal approach, the study analyzes diverse user groups, namely car drivers, motorcyclists, and professional drivers, across urban, rural, and highway environments to evaluate feedback mechanism comprehensively.

A 21-month naturalistic driving experiment involving 230 drivers across six feedback phases generated a robust dataset of 106,776 trips, covering 1.3 million kilometers. The tailored feedback interventions concerned scorecards, gamification, and peer comparisons. Advanced statistical and machine learning models, including Generalized Linear Mixed-Effects Models (GLMMs), Structural Equation Models (SEMs), and Survival Analysis methods (e.g., Weibull AFT, Cox-PH with frailty, and Random Survival Forests), were utilized to analyze behavioral mΣetrics such as speeding, mobile phone use, harsh braking, and accelerations which demonstrated substantial impacts on reducing risky behaviors.

The overall impact of feedback significantly improved driving behavior, with notable variations across user groups and driving contexts. Urban environments demonstrated the most substantial reductions in mobile phone use and harsh events, likely driven by the heightened complexity and demands of navigating urban settings. Feedback features also influenced outcomes differently; scorecards were particularly effective in reducing risky behaviors like speeding, while gamification elements motivated sustained engagement among professional drivers. Despite these successes, survival analyses revealed significant relapse tendencies once feedback was removed, with survival probabilities for maintaining improved behaviors, such as reduced speeding and harsh braking, falling below 50% within 150 trips post-feedback. These findings highlight the need for continuous and adaptive engagement strategies, incorporating diverse features tailored to the specific needs of different user groups and driving contexts, to ensure long-term effectiveness and sustained safety improvements.

This dissertation uniquely approaches feedback as a holistic system, examining its impacts across multiple phases and user groups, offering a comprehensive framework for evaluating telematics-based feedback via an integrated suite of three-layer models. Findings provide actionable insights for policymakers, technology developers, and road safety advocates, supporting the development of scalable solutions to improve driver behavior and enhance road safety sustainably.

Σύντομη Περίληψη

Η οδική ασφάλεια παραμένει ένα κρίσιμο παγκόσμιο πρόβλημα, με τα οδικά ατυχήματα να στοιχίζουν εκατομμύρια ζωές ετησίως και να επιφέρουν σημαντικές συναισθηματικές και οικονομικές επιβαρύνσεις. Το ανθρώπινο λάθος ευθύνεται για το 95% των οδικών ατυχημάτων, γεγονός που υπογραμμίζει τον καθοριστικό ρόλο της συμπεριφοράς των οδηγών στην ενίσχυση της ασφάλειας. Η παρούσα διατριβή διερευνά τον αντίκτυπο της ανατροφοδότησης του οδηγού μέσω τηλεματικής στην οδηγική συμπεριφορά, αντιμετωπίζοντας τα κενά στην κατανόηση των επιπτώσεων του κύκλου ζωής της ανατροφοδότησης, συμπεριλαμβανομένων των φάσεων πριν την ανατροφοδότηση, της ανατροφοδότησης και της μετά την ανατροφοδότηση. Χρησιμοποιώντας μια πολυτροπική προσέγγιση, η μελέτη αναλύει διαφορετικές ομάδες χρηστών οδού, δηλαδή οδηγούς αυτοκινήτων, μοτοσικλετιστές και επαγγελματίες οδηγούς, σε αστικά και υπεραστικά περιβάλλοντα, και αυτοκινητόδρομους για να αξιολογήσει συνολικά τον μηχανισμό ανατροφοδότησης.

Ένα πείραμα φυσικής οδήγησης διάρκειας 21 μηνών, στο οποίο συμμετείχαν 230 οδηγοί σε έξι φάσεις ανατροφοδότησης, δημιούργησε ένα ισχυρό σύνολο δεδομένων 106.776 διαδρομών, που κάλυψαν 1,3 εκατομμύρια χιλιόμετρα. Οι εξατομικευμένες παρεμβάσεις ανατροφοδότησης αφορούσαν κάρτες βαθμολογίας, παιχνίδισμα και συγκρίσεις μεταξύ των οδηγών. Χρησιμοποιήθηκαν προηγμένα στατιστικά μοντέλα και μοντέλα μηχανικής μάθησης, συμπεριλαμβανομένων γενικευμένων γραμμικών μοντέλων μικτών αποτελεσμάτων (GLMM), μοντέλων δομικών εξισώσεων (SEM) και μεθόδων ανάλυσης επιβίωσης (π.χ. Weibull AFT, Cox-PH και Random Survival Forests), για την ανάλυση μετρήσεων συμπεριφοράς όπως η υπερβολική ταχύτητα, η χρήση κινητού τηλεφώνου, τα απότομα συμβάντα, οι οποίες κατέδειξαν σημαντικές επιπτώσεις στη μείωση των επικίνδυνων συμπεριφορών.

Η συνολική επίδραση της ανατροφοδότησης βελτίωσε σημαντικά τη συμπεριφορά οδήγησης, με αξιοσημείωτες διαφοροποιήσεις μεταξύ ομάδων χρηστών και οδηγικών πλαισίων. Οι αστικές περιοχές εμφάνισαν τις πιο σημαντικές μειώσεις στη χρήση κινητού τηλεφώνου και σε απότομα συμβάντα, πιθανώς λόγω της αυξημένης πολυπλοκότητας και των απαιτήσεων που συνδέονται με την πλοήγηση σε αστικά περιβάλλοντα. Τα χαρακτηριστικά της ανατροφοδότησης επηρέασαν επίσης τα αποτελέσματα διαφορετικά, οι βαθμολογίες αποδείχθηκαν ιδιαίτερα αποτελεσματικές στη μείωση επικίνδυνων συμπεριφορών, όπως η υπερβολική ταχύτητα, ενώ στοιχεία παιχνιδοποίησης ενίσχυσαν τη διαρκή δέσμευση των επαγγελματιών οδηγών στη βελτίωση της οδηγικής τους συμπεριφοράς. Παρά τις επιτυχίες αυτές, οι αναλύσεις επιβίωσης αποκάλυψαν σημαντικές τάσεις υποτροπής μετά την αφαίρεση της ανατροφοδότησης, με τις πιθανότητες επιβίωσης για τη διατήρηση βελτιωμένων συμπεριφορών, να πέφτουν κάτω από 50% μέσα σε 150 διαδρομές μετά την ανατροφοδότηση. Τα ευρήματα αυτά υπογραμμίζουν την ανάγκη για συνεχείς και προσαρμοστικές στρατηγικές δέσμευσης, που να ενσωματώνουν ποικίλα χαρακτηριστικά προσαρμοσμένα στις συγκεκριμένες ανάγκες διαφορετικών ομάδων χρηστών και οδηγικών πλαισίων, ώστε να διασφαλίζεται η μακροπρόθεσμη αποτελεσματικότητα και η διαρκής βελτίωση της οδικής ασφάλειας.

Αυτή η διατριβή προσεγγίζει μοναδικά την ανατροφοδότηση ως ένα ολιστικό σύστημα, εξετάζοντας τις επιπτώσεις της σε πολλαπλές φάσεις και ομάδες χρηστών, προσφέροντας ένα ολοκληρωμένο πλαίσιο για την αξιολόγηση της ανατροφοδότησης μέσω τηλεματικής με μια ολοκληρωμένη σειρά τριστρωματικών μοντέλων. Τα ευρήματα παρέχουν πρακτικές πληροφορίες για υπεύθυνους χάραξης πολιτικής, τεχνολογικούς προγραμματιστές και υποστηρικτές της οδικής ασφάλειας, υποστηρίζοντας

την ανάπτυξη λύσεων μεγάλης κλίμακας για τη βελτίωση της συμπεριφοράς οδηγών και την ενίσχυση της οδικής ασφάλειας με βιώσιμο τρόπο.

Summary

Objectives and Methodology

Road safety is a critical public health and societal issue, as road traffic crashes claim millions of lives and cause severe injuries globally every year. Beyond the tragic loss of life, these incidents impose immense emotional and economic burdens on families and communities. **Research attributes approximately 95% of road crashes to human error**, underscoring the critical role of driver behavior in accident prevention. Understanding and addressing risky driving behaviors—such as distracted driving, speeding, and harsh events—are pivotal to enhancing road safety.

Based on the above, **the primary aim of this dissertation is to investigate the driver telematics feedback mechanism** under the framework of driving behavior and road safety. Despite growing interest from automotive manufacturers and transportation researchers in driver behavior, limited research exists on quantifying the comprehensive impact of driver feedback across its entire lifecycle—encompassing the pre-feedback, feedback, and post-feedback phases. To address this gap, this dissertation adopts a **holistic approach to evaluate the effectiveness of feedback** on modifying driving behavior and ultimately enhancing road safety.

To achieve these objectives, **a series of methodological steps** were carefully implemented. These steps are outlined and visually depicted in Figure I. The methodological framework provides a structured approach to achieving the objectives of this dissertation.

As a first step, a **systematic literature review** was conducted to evaluate the effectiveness of driver feedback within naturalistic driving studies. Feedback methods have evolved from in-vehicle devices and paper-based reports to sophisticated, user-friendly smartphone applications. These advancements enable the collection of high-resolution driving data and the delivery of personalized, data-driven feedback. Systems such as real-time alerts, post-trip summaries, and performance reports have demonstrated potential for improving driver behavior.

However, **significant gaps remain regarding the long-term sustainability of these effects and the differential impacts of feedback features**. While studies highlight the effectiveness of feedback in reducing speeding, harsh braking, and mobile phone use, the influence of feedback type, frequency, and incentives on behavior remains underexplored. Additionally, many studies observe a relapse into risky behaviors once feedback is removed, necessitating further investigation.

Based on the results of the systematic literature review, the following **research questions** were formulated:

1. How does feedback influence driver speeding and distracted behavior in terms of the percentage of trip time during which the speed limit was exceeded and mobile phone was used while driving?

2. How does feedback influence harsh driving events, in terms of the number of harsh accelerations and harsh brakings?

3. Do different feedback features (e.g., scorecards, maps, peer comparisons, motivations, gamification, rewards) have different effects on driver behavior? Which feature demonstrates the most significant impact?

4. How does the post-feedback effect influence long-term driver behavior, and to what extent are the changes sustained after the feedback is removed?

5. How can advanced statistical techniques be applied to understand the mechanisms of driver feedback and develop more individualized, data-driven approaches for driving behavior change?

To answer the research questions, a **robust methodological framework** was developed, combining theoretical approaches and experimental design principles. This included the application of advanced modeling techniques, such as Generalized Linear Mixed Effects Models (GLMMs), Structural Equation Models (SEMs), and Survival Analysis Models, alongside the design of a naturalistic driving experiment.

A **21-month naturalistic driving experiment involving 230 drivers** was conducted. The participants were divided into three groups (car drivers, professional van drivers, and motorcyclists), and their driving behavior was monitored across **six distinct feedback phases**. These phases were defined as follows:

- Phase 1: Basic trip data and characterization were accessible to drivers.
- Phase 2: Introduction of scorecards with trip-level scoring.
- Phase 3: Addition of maps and highlights for further trip insights.
- Phase 4: Peer comparisons enabled for driver performance benchmarking.
- Phase 5: Competitions and challenges introduced with rewards for safe driving.
- Phase 6: Reversion to Phase 1, removing all additional feedback.

High-resolution data were collected from 106,776 trips, covering a total of 1,317,573 kilometers and 30,532 hours of driving. Behavioral metrics were captured using non-intrusive smartphone sensors, ensuring a seamless and accurate recording of driving behaviors. This sensor data was complemented by self-reported information from participants, providing a holistic understanding of driver perceptions, habits, and behavioral changes. The **experimental design adhered to strict ethical standards**, having been approved by the Research Ethics and Conduct Committee of NTUA, and ensured full compliance with General Data Protection Regulation (GDPR) guidelines. Continuous communication with participants was maintained throughout the study to address any technical issues, sustain engagement, and monitor the smooth execution of the experiment.

Extensive data processing and cleaning were carried out to ensure the quality and reliability of the dataset. Invalid or incomplete trip data were systematically identified and excluded, while key behavioral metrics such as speeding, mobile phone use while driving, harsh accelerations, and braking events were standardized for analysis. Data preprocessing steps included the conversion of raw sensor outputs into meaningful variables and the integration of self-reported data for cross-validation. **This rigorous approach to data management enabled the creation of a robust dataset**, facilitating detailed statistical analyses and ensuring the accuracy of the findings presented in this dissertation.

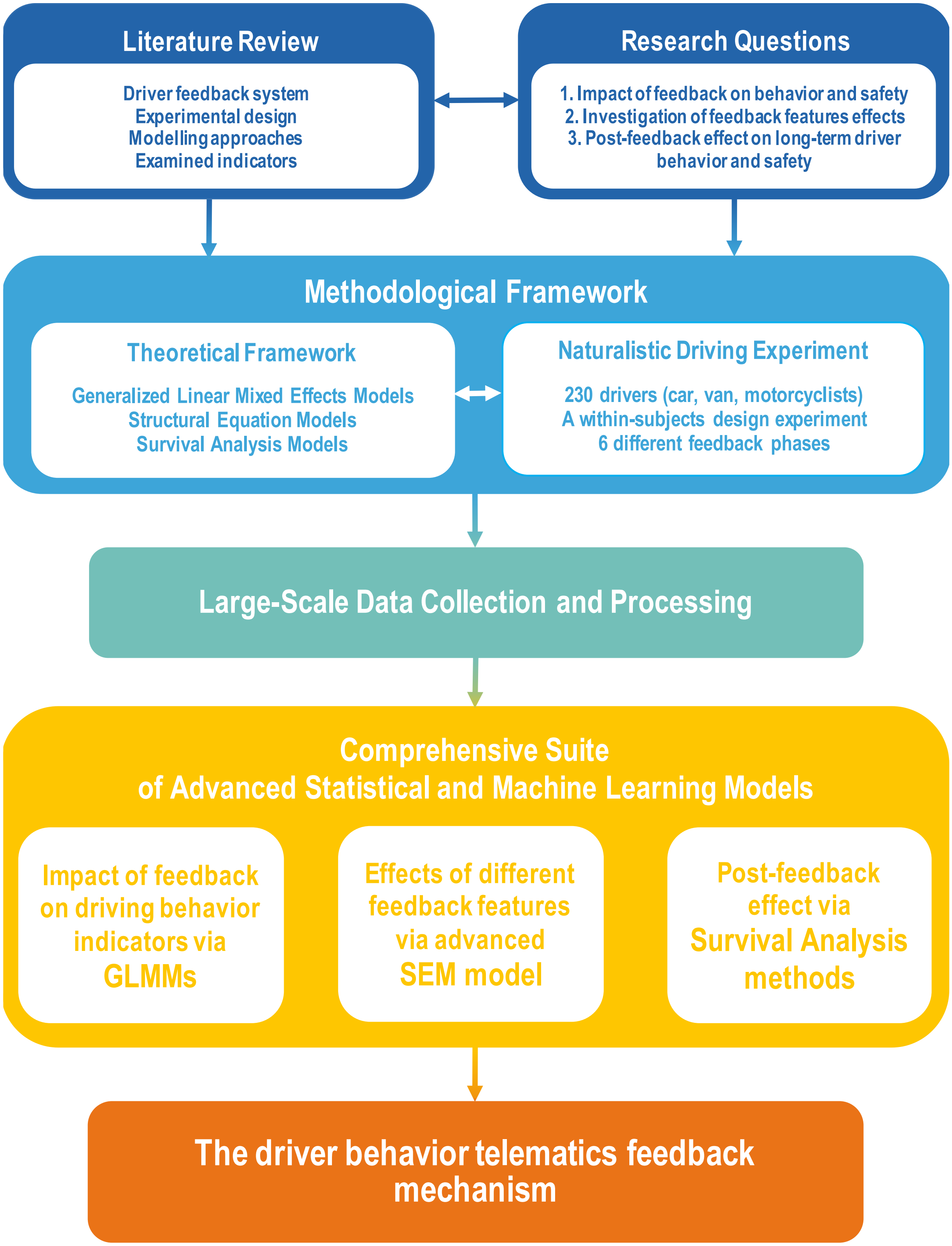


Figure I: Graphical representation of the overall methodological framework of the doctoral dissertation

Advanced statistical techniques were employed to analyze feedback effects on critical driving indicators, such as speeding, mobile phone use, harsh accelerations, and harsh brakings. The analysis unfolded in **three key pillars**:

1. **Impact of Feedback**: This pillar assessed the immediate effects of feedback on i) driver speeding and distracted behavior, focusing on speeding among motorcyclists and distraction due to mobile phone use while driving in car drivers, and ii) driving harsh evenrs, focusing on harsh braking and harsh accelerations among car drivers, and professional drivers on highways. Generalized Linear Mixed-Effects Models (GLMM) were then employed in all cases to evaluate the effects of feedback while accounting for individual differences and contextual factors.

2. **Effects of Different Feedback Features**: A Structural Equation Model (SEM) was developed to explore the complex relationships between feedback features (e.g. scorecards, maps, peer comparisons, motivations, gamification, rewards) and driver behavior, exposure metrics and safety outcomes, allowing for the simultaneous analysis of multiple variables and their interactions.

3. **Post-Feedback Effects**: Effects on long-term driver behavior, with a particular emphasis on understanding the relapse of driving behaviors following the withdrawal of feedback telematics during the last phase of the experiment. Survival analysis methods were employed to investigate relapse patterns across various indicators, including harsh accelerations, harsh braking, speeding behavior, and mobile phone use while driving. These analyses leverage Kaplan-Meier curves, Cox-PH models with frailty, Weibull AFT models with clustered heterogeneity, and Random Survival Forests to evaluate and compare the predictive power and insights offered by each model.

Ultimately, the synthesis of all the analyses carried out within the framework of this doctoral dissertation resulted in a driver behavior telematics feedback mechanism with numerous original and interesting results, which are discussed below.

Main findings

Feedback Impact on Driver Behavior

The investigation of feedback impacts on driver behavior yielded significant findings across different user groups, driving environments, and behavioral metrics. Overall, during the two phases of the experiment a **large dataset of 3,537 trips from a sample of 13 motorcyclists** were recorded and analysed. Using Generalized Linear Mixed-Effects Models with random intercepts and random slopes for total trip duration revealed that providing motorcyclists with feedback about their riding performance during experiment Phase 2 led to a remarkable **decrease in speeding percentage over a trip.** Particularly, in the developed models rider feedback seems to decrease speeding percentage, having a risk ratio of $\exp(\beta=-0.145) = 0.865$ for the overall model (13.5% decrease), and $\exp(\beta=-0.031) = 0.970$ and $\exp(\beta=-0.420) = 0.657$ for urban (3.0% decrease) and rural (34.3%) road types respectively. These results highlight the effectiveness of feedback in targeting high-risk behaviors, offering a foundation for scalable interventions in rider training and policy design.

Similarly, **distracted driving, particularly mobile phone use, was examined across urban, rural, and highway** contexts via GLMM models for **65 car drivers over 21,167 trips**. Feedback emerged as a strong restrictive in using the mobile phone while driving overall ($\beta = -0.4276$, $p < 2e-16$), particularly in urban settings ($\beta = -0.3687$, $p < 2e-16$), while its impact was notably weaker in rural environments ($\beta = -0.1180$, $p < 2e-16$) and unexpectedly positive on highways ($\beta = 0.5490$, $p < 2e-16$), suggesting compensatory behaviors or a perceived lower risk of distraction on high-speed roads. The substantial variability in random intercepts (SD = 1.4024 overall) highlights **notable individual differences in baseline behavior**, while random slopes for trip duration (SD = 0.2827 overall) show diverse responses to prolonged trips.

In the domain of harsh events such as accelerations and brakings, feedback mechanisms demonstrated significant behavioral improvements, as well. Results from the **analysis of 65 car drivers** during the first two phases of the experiment, revealed a **significant reduction in harsh accelerations (12%) and harsh brakings (10%)**, both changes being statistically significant ($p < 0.001$). The GLMM models further reinforce these findings, as feedback was consistently associated with reduced frequencies of harsh events, particularly in urban and rural environments. Notably, the relative risk ratios for speeding duration and trip duration indicate strong positive associations with harsh events, though feedback appears to mitigate these effects to a degree in Phase 2 of the experiment. Importantly, driver-specific variability, captured through random intercepts and slopes, underscores the need for tailored feedback mechanisms to address unique behavioral traits.

Professional drivers, due to their prolonged driving hours and distances, were also a key area of exploration. Using GLMMs calibrated on a dataset of 5,345 trips from 19 professional drivers, the analysis revealed that participation in a **social gamification scheme with incentives led to notable improvements in harsh events of professional drivers**. During the competition phase, the likelihood of harsh accelerations was reduced by a factor of 0.348 ($p < 0.001$), while harsh brakings decreased by a factor of 0.404 ($p < 0.001$), indicating the efficacy of gamification in promoting safer driving practices. Additionally, trip duration showed a positive association with harsh events, with a 1-second increase in driving time raising the odds of harsh accelerations and harsh brakings by factors of 1.558 and 1.564, respectively, highlighting the cumulative effects of extended driving. The inclusion of random intercepts in the models underscored substantial variability in baseline driver behavior, emphasizing the importance of personalized interventions.

Feedback Different Features Effects on Driver Behavior

The **Structural Equation Model (SEM) analysis** provided significant insights into the impact of effects of feedback features on driver behavior, specifically speeding, harsh braking, and harsh acceleration events. The dataset, comprising **73,869 trips from 175 car drivers over 21 months**, offered a robust basis for modeling. The SEM results identified two latent variables, namely feedback and exposure as critical influences. The model exhibited excellent goodness-of-fit measures, with Comparative Fit Index (CFI) = 0.940, Tucker–Lewis Index (TLI) = 0.944, Root Mean Square Error Approximation (RMSEA) = 0.049, and Standardized Root Mean Square Residual (SRMR) = 0.025, indicating a robust and well-specified structure. The inclusion of covariances among variables, guided by residual correlation analysis, further improved the model fit and highlighted critical relationships, such as those between speeding and harsh braking behaviors.

Among the feedback features analyzed, the **scorecard emerged as the most influential feature**, with the highest positive estimate ($\beta = 2.076$, $p < 0.001$), demonstrating its powerful role in promoting safer driving habits by immediately altering risky behaviors in comparison with the baseline phase. This result can be attributed to the clear, concise, and actionable nature of scorecards, which **provide drivers with straightforward insights into their performance** and specific areas for improvement, making it easier to adjust their behavior. Similarly, the maps feature showed a strong impact ($\beta = 1.646$, $p < 0.001$), emphasizing the importance of spatial awareness in enhancing driving practices. The compare feature allowed drivers to assess their performance relative to peers, positively influencing behavior ($\beta = 1.215$, $p < 0.001$). Additionally, the **competition & challenges feature proved highly effective** ($\beta = 2.053$, $p < 0.001$) by motivating drivers to adopt safer driving behaviors through gamified elements and rewards for safe driving.

In terms of driving behavior metrics, driver telematics feedback significantly **reduced the percentage of speeding time** ($\beta = -0.214$, $p < 0.001$) **and harsh braking events per 100km** ($\beta = -0.027$, $p < 0.001$). However, an **increase in harsh accelerations** per 100km ($\beta = 0.026$, $p < 0.001$) suggests the need for further refinement of feedback systems to address unintended consequences. Exposure factors also played a key role in shaping driver behavior, with morning peak exposure correlating with increased risk-taking ($\beta = 2.473$, $p < 0.001$), likely driven by time pressure during commuting hours. Conversely, afternoon peak exposure was associated with less aggressive behavior ($\beta = -1.360$, $p < 0.001$), providing insights into temporal variations in driving patterns.

Regression analysis confirmed these findings, highlighting the **interplay between exposure and feedback features**. While exposure positively influenced speeding ($\beta = 0.326$, $p < 0.001$), feedback features effectively mitigated this behavior. The competition & challenges feature, in particular, showed promise in moderating harsh accelerations ($\beta = -0.001$, $p < 0.001$). Harsh braking incidents were also significantly reduced by feedback, reinforcing the role of feedback in promoting safer driving practices. Covariance analysis further revealed strong interrelationships between risky behaviors, such as speeding and harsh braking, **underscoring the complexity of driver behavior patterns**. These findings suggest that speeding often necessitates sudden corrections, like harsh braking, and both behaviors may stem from underlying traits such as risk-taking tendencies or aggressive driving habits.

The **practical implications of these findings** are substantial. Feedback features, particularly those leveraging personalized scorecards, spatial tools, and gamification elements, hold great promise for improving driver safety. Tailored interventions targeting specific behaviors and times of day could further enhance the efficacy of these systems. However, limitations such as the exclusion of mobile phone use from the final model and potential selection biases due to the voluntary nature of participation should be addressed in future research.

Post-Feedback Effect on Long-Term Driver Behavior

Survival analysis techniques were applied to a dataset of 24,904 trips from 31 car drivers, each contributing at least 20 trips in the post-feedback phase, to investigate the long-term effects of driver telematics feedback on driving behavior. The analysis focused on relapse patterns in mobile phone use, speeding, harsh braking, and harsh accelerations. The methods utilized included Kaplan-Meier curves, Cox-PH models with frailty, Weibull Accelerated Failure Time (AFT) models incorporating clustered heterogeneity, and Random Survival Forests. The findings demonstrate the effectiveness of feedback interventions in achieving significant short-term behavioral improvements during the feedback phase. However, the **post-feedback phase reveals varied relapse tendencies**, emphasizing the need for sustained interventions to maintain these improvements over time.

The Kaplan-Meier survival analysis emphasized **relapse trends**, showing a steady decline in improved behavior over successive trips in the post-feedback phase. For harsh accelerations, **survival probabilities dropped from 84.8% at 50 trips to 49.2% by 150 trips**. Similar trends were observed for harsh braking and speeding, with survival probabilities declining to approximately 40.3% and 46.8%, respectively, by the 150-trip mark. These patterns underscore the **transient nature of feedback effects** and the need for continuous reinforcement mechanisms. Mobile phone use showed slightly greater resilience, with survival probabilities remaining above 80% at 100 trips, but the gradual relapse was evident over time.

Among the survival analysis models applied, **the Weibull Accelerated Failure Time (AFT) model consistently emerged as a robust performer** across the examined indicators, balancing predictive accuracy and interpretability. The concordance index (C-index) values ranged between 0.677 and 0.773, with the model achieving the highest predictive ability for mobile phone use relapse (C-index = 0.773), indicating strong discriminative capacity in identifying drivers most at risk of relapse. Key predictors such as age group [35-54] ($\beta = 0.165$, $p = 0.041$), trip duration ($\beta = -0.022$, $p < 0.001$), and self-reported aggressiveness (approaching significance at $p = 0.089$) were highlighted, providing actionable insights into relapse behavior. The model also **captured heterogeneity across drivers by incorporating random effects**, with frailty effects showing significant variability in survival times.

For speeding relapse, the Weibull AFT model achieved a C-index = 0.700 with significant predictors including trip duration ($\beta = -0.022$, $p < 0.001$) and morning peak hours ($\beta = -0.096$, $p = 0.004$). **Trip duration, in particular, emerged as the dominant predictor, consistently reducing survival time across all relapse indicators**, underscoring the role of prolonged driving in behavioral regression. Similarly, in the analysis of harsh braking relapse, the model achieved a moderate predictive accuracy (C-index = 0.724) with significant contributions from variables such as age group [35-54] ($\beta = 0.360$, $p = 0.010$) and vehicle engine capacity (>1400cc) ($\beta = -0.508$, $p = 0.012$). These findings highlight that younger age groups and drivers of larger-engine vehicles are more prone to relapse.

The **Random Survival Forest (RSF) model demonstrated superior predictive performance in some examined indicators**, excelling in capturing non-linear interactions and complex relationships between predictors. With Root Mean Squared Error (RMSE) values as low as 85.87 and out-of-bag (OOB) prediction errors of 24.3% for mobile phone use relapse, RSF identified critical predictors such as trip duration, aggressive driving tendencies, and vehicle engine size. Its **flexibility in handling diverse predictors** and uncovering nuanced dynamics makes RSF an invaluable tool for predictive analyses. However, the model's "black box" nature and reliance on larger datasets limit its interpretability and applicability for explanatory purposes.

Overall, comparing the models**, the Weibull AFT model stands out for balancing interpretability and predictive accuracy**, making it particularly suited for contexts requiring actionable insights into survival dynamics. The Cox model offers a useful compromise with its interpretability and ability to handle frailty, however repeatedly failed to meet model assumptions. The RSF model is most appropriate for predictive tasks where capturing non-linear relationships and complex interactions is critical, though its lack of transparency limits its utility in understanding the underlying behavioral mechanisms. These findings emphasize the **importance of aligning model configuration with research objectives**. For studies focused on understanding behavioral dynamics and guiding intervention design, the Weibull AFT model provides robust insights. Conversely, when predictive accuracy is paramount, RSF offers a superior alternative.

While this study offers valuable insights into the dynamics of driver feedback and relapse, limitations such as the relatively small sample size, exclusion of traffic conditions, and macroscopic focus should be noted. Future research could incorporate more granular data, such as traffic dynamics and moment-to-moment driver decisions, to provide a deeper understanding of behavioral patterns. Employing advanced modeling techniques, such as random parameters with heterogeneity-in-means, could further enhance the analysis by accounting for driver-specific variability.

Innovative Scientific Contributions

The innovative contributions of this doctoral dissertation consist of five original scientific contributions, as described below, and illustrated in Figure II.

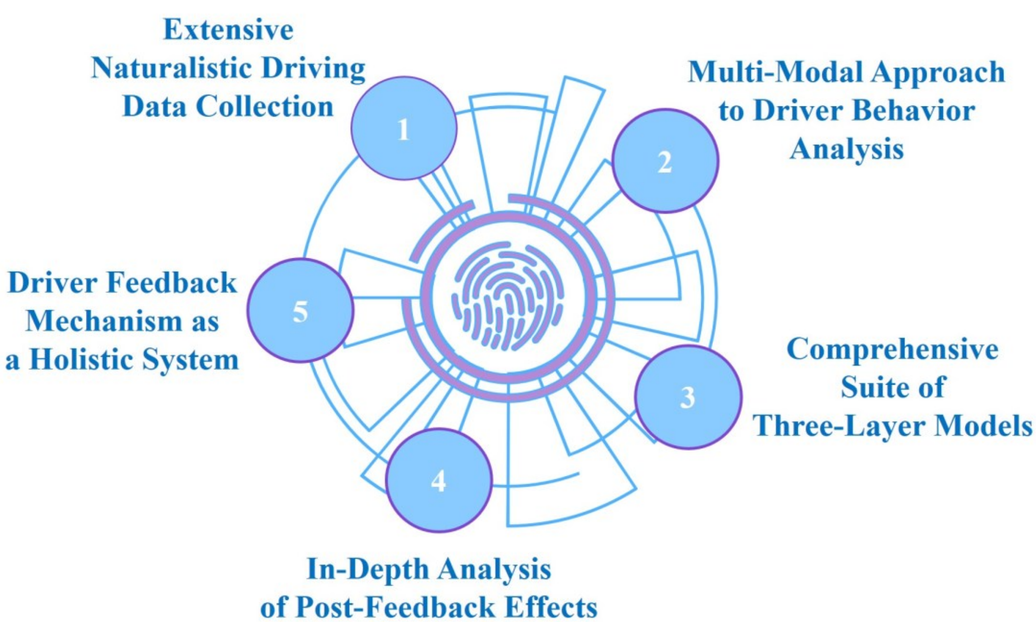


Figure II: Innovative contributions of the doctoral dissertation

Extensive Naturalistic Driving Data Collection

The present dissertation represents a **significant step forward in naturalistic driving (ND) research** by leveraging non-intrusive data collection methods that rely on smartphone sensors. Unlike traditional approaches, this methodology minimizes disruption to participants, enabling the unobtrusive capture of real-world driving behaviors. **The data spans a large sample size of drivers (230) across diverse road environments and vehicle types**, including car drivers, motorcyclists, and professional van drivers. This inclusivity ensures that findings are not only representative but also account for variations across driver demographics and vehicle categories. The dataset's rich temporal resolution provides detailed insights into driving behaviors at the trip level, offering a granular perspective on driver behavior dynamics.

Moreover, the long-term data collection 21-month period, spanning multiple feedback phases and covering various road environments, adds unique value. By capturing behavior changes over time, **the study bridges a critical gap in existing ND research**, which often relies on short-term observations. This long-term perspective enables the assessment of sustained behavior modifications and relapse tendencies, providing a robust foundation for developing adaptive and sustainable interventions to improve road safety. The **methodology sets a new benchmark for ND experiments**, paving the way for more scalable, cost-effective, and technologically advanced driving behavior studies.

Multi-Modal Approach to Driver Behavior Analysis

This dissertation takes a multi-modal approach, emphasizing the importance of understanding driving behaviors across diverse road user groups and environments. By including **car drivers, motorcyclists, and professional van drivers**, the research recognizes the critical need to study vulnerable road users, such as motorcyclists, who face heightened risks, and professional drivers, who spend extended hours on the road. This i**nclusive focus ensures a comprehensive evaluation of driver telematics feedback**, highlighting their relevance across varying risk profiles and exposure levels.

The investigation also considers the influence of urban, rural, and highway environments, acknowledging the distinct challenges posed by each road type. This contextual approach reveals that **feedback effectiveness is not uniform**; behaviors like mobile phone use or speeding respond differently to interventions **depending on the driving environment**. For instance, motorcyclists may benefit more from feedback targeting situational awareness, while professional drivers might require tailored interventions addressing fatigue and repetitive exposure to high-risk scenarios.

By integrating this diversity of user groups and contexts, the dissertation provides **actionable insights for policymakers**, road safety advocates, and technology developers. It emphasizes the importance of developing tailored feedback systems that cater to the unique needs of vulnerable road users, such as motorcyclists, and professional drivers, who contribute significantly to road traffic activity. This comprehensive approach supports the creation of adaptive, context-sensitive interventions, ultimately improving road safety for all users.

Comprehensive Suite of Three-Layer Models

This dissertation employs a **comprehensive suite of advanced statistical and machine learning models,** tailored to address the multifaceted nature of driving behavior analysis. By incorporating Generalized Linear Mixed-Effects Models, Structural Equation Models, and Survival Analysis techniques (e.g., Weibull AFT, Cox-PH with frailty, and Random Survival Forest), the study provides a rigorous analytical framework capable of uncovering both linear and non-linear relationships between variables. Each model is carefully selected to align with the research objectives, balancing predictive accuracy with interpretability to ensure actionable insights.

This model suite also enables the **exploration of complex phenomena**, such as the interplay between feedback features, driving behaviors, and contextual factors like time of day or road type. For example, survival models uniquely capture relapse dynamics, offering novel insights into post-feedback behavioral tendencies. Machine learning techniques further enhance the study by capturing nuanced, non-linear interactions, ensuring that the models are equipped to handle the complexity of real-world driving data. This **innovative analytical framework** not only elevates the scientific rigor of the research but also demonstrates the potential of combining traditional statistical methods with state-of-the-art machine learning approaches for driver behavior studies.

In-Depth Analysis of Post-Feedback Effects

This dissertation is **among the first to analyze thoroughly post-feedback effects** on driver behavior using advanced statistical and machine learning techniques, addressing a critical gap in existing research. Through survival analysis methods, such as Weibull AFT and Random Survival Forest, the study evaluates long-term behavior changes and relapse patterns after feedback withdrawal. These techniques enable a detailed exploration of the factors influencing relapse in risky behaviors like speeding, harsh events, and mobile phone use, providing actionable insights for the design of sustained intervention strategies.

The findings reveal the importance of adaptive feedback systems that can maintain behavior improvements over time. For example, survival analysis showed that trip duration and time of day significantly influence relapse dynamics, emphasizing the need for context-aware feedback mechanisms. This innovative focus on the post-feedback phase provides a **novel framework for understanding the longevity of feedback-induced improvements**, allowing for more durable and impactful road safety interventions. It also sets a precedent for future research to integrate long-term perspectives into the evaluation of driving behavior modification strategies.

Feedback Mechanism as a Holistic System

This dissertation uniquely approaches the **feedback mechanism as a holistic system, examining its full lifecycle through a multiparametric analytical framework**. By systematically analyzing the pre-feedback, feedback, and post-feedback phases, the study offers a comprehensive understanding of how feedback influences driver behavior across time. The integration of diverse feedback features, such as scorecards, maps, comparison tools, and competition elements, enables the evaluation of their individual and combined impacts on behavior modification. This multi-phase perspective not only captures immediate behavior changes but also sheds light on long-term patterns and relapse tendencies.

Furthermore, the holistic framework provides valuable insights into the synergies and trade-offs between different feedback features. For instance, while scorecards and competition elements are highly effective in reducing speeding, their impact on other behaviors like harsh accelerations requires further refinement. This systemic approach advances the field by moving beyond isolated feedback evaluations, offering a **scalable, data-driven framework for designing and implementing telematics-based interventions**. The findings emphasize the potential of adaptive feedback systems to improve driving behavior sustainably, ultimately contributing to safer road environments.

Εκτεταμένη Περίληψη

Στόχοι και Μεθοδολογία

Η οδική ασφάλεια αποτελεί κρίσιμο ζήτημα δημόσιας υγείας και κοινωνίας, καθώς τα οδικά ατυχήματα κοστίζουν εκατομμύρια ζωές και προκαλούν σοβαρούς τραυματισμούς παγκοσμίως κάθε χρόνο. Πέρα από την τραγική απώλεια ζωών, αυτά τα περιστατικά επιβαρύνουν συναισθηματικά και οικονομικά τις οικογένειες και τις κοινότητες. Έρευνες δείχνουν ότι **το ανθρώπινο λάθος ευθύνεται για περίπου το 95% των οδικών ατυχημάτων**, υπογραμμίζοντας τον κρίσιμο ρόλο της συμπεριφοράς των οδηγών στην πρόληψη των ατυχημάτων. Η κατανόηση και αντιμετώπιση επικίνδυνων οδηγικών συμπεριφορών—όπως η απόσπαση προσοχής, η υπερβολική ταχύτητα και οι απότομοι χειρισμοί—είναι ζωτικής σημασίας για τη βελτίωση της οδικής ασφάλειας.

Με βάση τα παραπάνω, **ο κύριος στόχος αυτής της διατριβής είναι η διερεύνηση του μηχανισμού ανατροφοδότησης της συμπεριφοράς των οδηγών μέσω τηλεματικής** στο πλαίσιο της οδηγικής συμπεριφοράς και της οδικής ασφάλειας. Παρά το αυξανόμενο ενδιαφέρον από κατασκευαστές αυτοκινήτων και ερευνητές μεταφορών για τη συμπεριφορά των οδηγών, υπάρχει περιορισμένη έρευνα σχετικά με την ποσοτικοποίηση της συνολικής επίδρασης της ανατροφοδότησης κατά τη διάρκεια ολόκληρου του κύκλου ζωής της—συμπεριλαμβανομένων των φάσεων πριν την ανατροφοδότηση, κατά τη διάρκεια της ανατροφοδότησης και μετά την ανατροφοδότηση. Για να καλυφθεί αυτό το κενό, η διατριβή υιοθετεί μια **ολιστική προσέγγιση για την αξιολόγηση της αποτελεσματικότητας της ανατροφοδότησης** στην τροποποίηση της οδηγικής συμπεριφοράς και στη βελτίωση της οδικής ασφάλειας.

Για την επίτευξη αυτών των στόχων, εφαρμόστηκε **μια σειρά μεθοδολογικών βημάτων**. Τα βήματα αυτά περιγράφονται και απεικονίζονται γραφικά στο Σχήμα Ι. Το μεθοδολογικό πλαίσιο παρέχει μια δομημένη προσέγγιση για την επίτευξη των στόχων της διατριβής.

Ως πρώτο βήμα, πραγματοποιήθηκε μια **συστηματική ανασκόπηση βιβλιογραφίας** για την αξιολόγηση της αποτελεσματικότητας της ανατροφοδότησης των οδηγών στο πλαίσιο φυσικών πειραμάτων οδήγησης. Οι μέθοδοι ανατροφοδότησης έχουν εξελιχθεί από συσκευές εντός οχήματος και αναφορές σε χαρτί σε προηγμένες, φιλικές προς τον χρήστη εφαρμογές για κινητά τηλέφωνα. Αυτές οι εξελίξεις επιτρέπουν τη συλλογή δεδομένων υψηλής ανάλυσης και την παροχή εξατομικευμένης ανατροφοδότησης βασισμένης σε δεδομένα.

Ωστόσο, **παραμένουν σημαντικά κενά όσον αφορά τη μακροπρόθεσμη βιωσιμότητα αυτών των αποτελεσμάτων αλλά και τις διαφοροποιημένες επιδράσεις των χαρακτηριστικών** της ανατροφοδότησης. Ενώ οι μελέτες αναδεικνύουν την αποτελεσματικότητα της ανατροφοδότησης στη μείωση της υπερβολικής ταχύτητας, των απότομων συμβάντων και της χρήσης κινητού τηλεφώνου, η επίδραση του τύπου ανατροφοδότησης, της συχνότητάς της και των κινήτρων στη συμπεριφορά παραμένει ανεξερεύνητη. Επιπλέον, πολλές μελέτες παρατηρούν υποτροπή στις επικίνδυνες συμπεριφορές μόλις αφαιρεθεί η ανατροφοδότηση, απαιτώντας περαιτέρω διερεύνηση.

Με βάση τα αποτελέσματα της συστηματικής ανασκόπησης βιβλιογραφίας, διατυπώθηκαν τα ακόλουθα **ερευνητικά ερωτήματα**:

1. Πώς επηρεάζει η ανατροφοδότηση τη συμπεριφορά του οδηγού όσον αφορά στην υπερβολική ταχύτητα και τη χρήση κινητού τηλεφώνου κατά την οδήγηση;

2. Πώς επηρεάζει η ανατροφοδότηση τη συμπεριφορά του οδηγού όσον αφορά στα απότομα συμβάντα, όπως οι απότομες επιταχύνσεις και τα απότομα φρεναρίσματα;

3. Τα διαφορετικά χαρακτηριστικά της ανατροφοδότησης (π.χ. βαθμολογίες, χάρτες, συγκρίσεις με άλλους οδηγούς, κίνητρα, παιχνιδοποίηση, επιβραβεύσεις) έχουν διαφορετικές επιδράσεις στη συμπεριφορά και την ασφάλεια του οδηγού; Ποιο χαρακτηριστικό έχει τον σημαντικότερο αντίκτυπο;

4. Πώς επηρεάζει η επίδραση της ανατροφοδότησης μετά την ολοκλήρωση της φάσης της την μακροπρόθεσμη συμπεριφορά και ασφάλεια των οδηγών, και σε ποιο βαθμό διατηρούνται οι αλλαγές αφού αφαιρεθεί η ανατροφοδότηση;

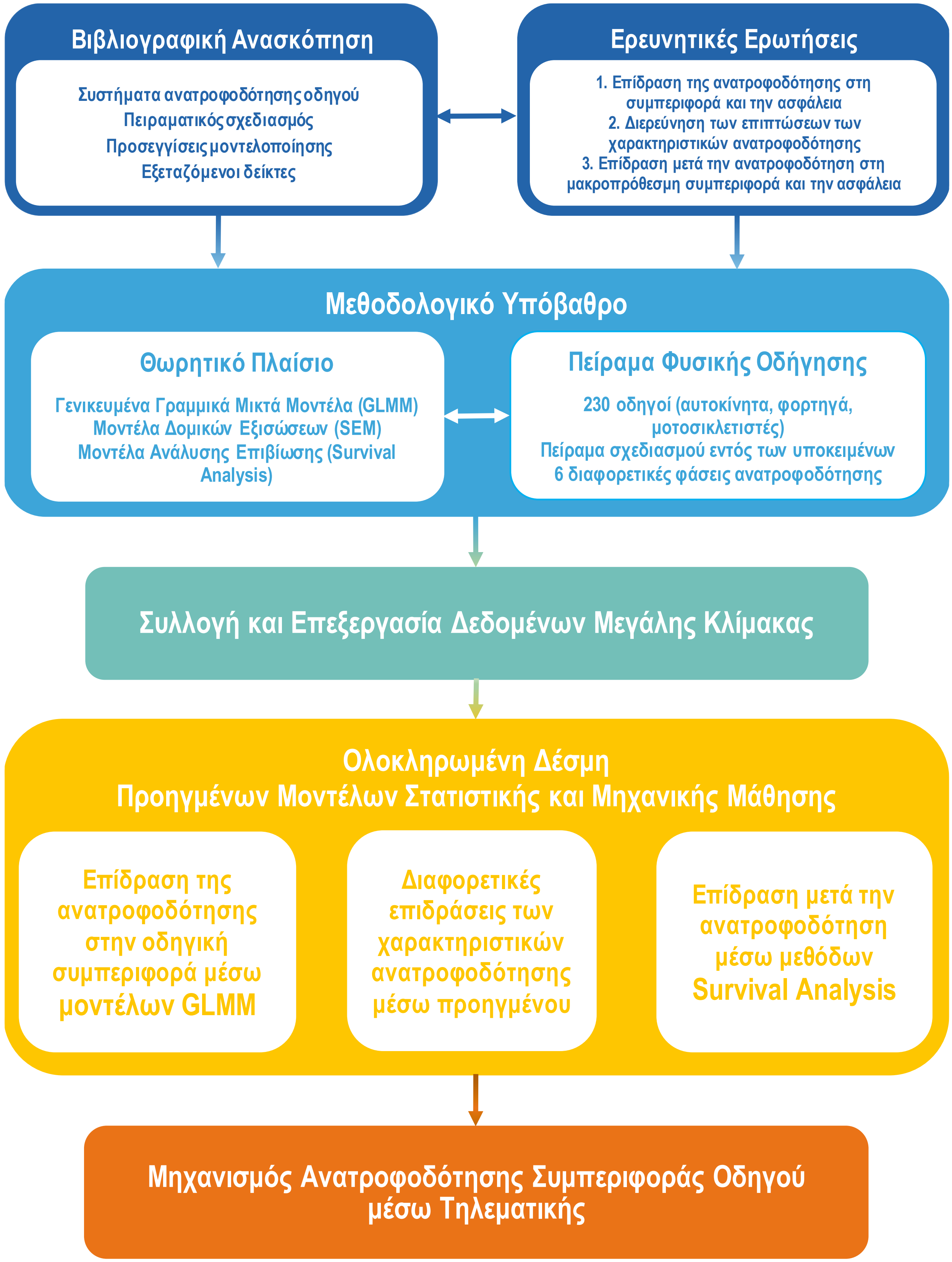
5. Πώς μπορούν να εφαρμοστούν προηγμένες στατιστικές τεχνικές για την κατανόηση των μηχανισμών της ανατροφοδότησης και την ανάπτυξη πιο εξατομικευμένων, βασισμένων σε δεδομένα προσεγγίσεων για την αλλαγή της οδηγικής συμπεριφοράς;

Για την απάντηση αυτών των ερωτημάτων αναπτύχθηκε ένα ισχυρό μεθοδολογικό πλαίσιο, το οποίο συνδύασε θεωρητικές προσεγγίσεις και αρχές πειραματικού σχεδιασμού. Αυτό περιλάμβανε την εφαρμογή προηγμένων τεχνικών μοντελοποίησης, όπως Γενικευμένα Γραμμικά Μοντέλα μικτών επιδράσεων (GLMMs), Μοντέλα Δομικών Εξισώσεων (SEM) και Μοντέλα Ανάλυσης Επιβίωσης (Survival Analysis), καθώς και τον σχεδιασμό ενός πειράματος φυσικής οδήγησης.

Πραγματοποιήθηκε ένα πείραμα φυσικής οδήγησης διάρκειας 21 μηνών, στο οποίο συμμετείχαν **230 οδηγοί**. Οι συμμετέχοντες χωρίστηκαν σε τρεις ομάδες (οδηγοί αυτοκινήτων, επαγγελματίες οδηγοί βαν και μοτοσικλετιστές) και η οδηγική τους συμπεριφορά παρακολουθήθηκε σε έξι διακριτές φάσεις ανατροφοδότησης. Αυτές οι φάσεις ορίστηκαν ως εξής:

- Φάση 1: Πρόσβαση σε βασικά δεδομένα διαδρομών και χαρακτηριστικά.
- Φάση 2: Εισαγωγή βαθμολογιών με σκορ ανά διαδρομή.
- Φάση 3: Προσθήκη χαρτών και σημαντικών στιγμών για περαιτέρω πληροφορίες διαδρομής.
- Φάση 4: Ενεργοποίηση συγκρίσεων με συναδέλφους για την αξιολόγηση της απόδοσης των οδηγών.
- Φάση 5: Εισαγωγή διαγωνισμών και προκλήσεων με επιβραβεύσεις για ασφαλή οδήγηση.
- Φάση 6: Επιστροφή στη Φάση 1, με την αφαίρεση όλων των πρόσθετων χαρακτηριστικών ανατροφοδότησης.

Συλλέχθηκαν **δεδομένα υψηλής ανάλυσης από 106.776 διαδρομές**, καλύπτοντας συνολικά **1.317.573 χιλιόμετρα** και **30.532 ώρες οδήγησης**. Οι μετρήσεις συμπεριφοράς καταγράφηκαν χρησιμοποιώντας μη παρεμβατικούς αισθητήρες smartphone, διασφαλίζοντας μια αδιάκοπη και ακριβή καταγραφή των οδηγικών συμπεριφορών. Αυτά τα δεδομένα αισθητήρων συμπληρώθηκαν από αυτό-αναφερόμενες πληροφορίες των συμμετεχόντων, παρέχοντας μια ολιστική κατανόηση των αντιλήψεων, συνηθειών και αλλαγών στη συμπεριφορά των οδηγών. **Ο πειραματικός σχεδιασμός τηρήθηκε αυστηρά στις ηθικές προδιαγραφές**, καθώς είχε εγκριθεί από την Επιτροπή Ερευνών και Δεοντολογίας του ΕΜΠ, και διασφάλιζε την πλήρη συμμόρφωση με τις κατευθυντήριες γραμμές του Γενικού Κανονισμού για την Προστασία των Δεδομένων. Η συνεχής επικοινωνία με τους συμμετέχοντες διατηρήθηκε καθ' όλη τη διάρκεια της μελέτης για την επίλυση τεχνικών ζητημάτων, τη διατήρηση της συμμετοχής και την παρακολούθηση της ομαλής εκτέλεσης του πειράματος.



Εικόνα Ι: Γραφική αναπαράσταση του συνολικού μεθοδολογικού πλαισίου της διδακτορικής διατριβής

Ακολούθως, πραγματοποιήθηκε **επεξεργασία και καθαρισμός δεδομένων** εκτενώς για να διασφαλιστεί η ποιότητα και η αξιοπιστία του συνόλου δεδομένων. Τα άκυρα ή ελλιπή δεδομένα διαδρομών εντοπίστηκαν συστηματικά και εξαιρέθηκαν, ενώ βασικές μετρήσεις συμπεριφοράς όπως η υπερβολική ταχύτητα, η χρήση κινητού τηλεφώνου κατά την οδήγηση, οι απότομες επιταχύνσεις και τα απότομα φρεναρίσματα τυποποιήθηκαν για την ανάλυση. Τα βήματα προ επεξεργασίας δεδομένων περιλάμβαναν τη μετατροπή των πρωτογενών δεδομένων των αισθητήρων σε ουσιαστικές μεταβλητές και την ενσωμάτωση των αυτό-αναφερόμενων δεδομένων για τη διασταύρωση των αποτελεσμάτων. Αυτή η αυστηρή προσέγγιση στη διαχείριση δεδομένων επέτρεψε τη δημιουργία ενός ισχυρού συνόλου δεδομένων, διευκολύνοντας λεπτομερείς στατιστικές αναλύσεις και διασφαλίζοντας την ακρίβεια των ευρημάτων που παρουσιάζονται στη διατριβή.

Προηγμένες στατιστικές τεχνικές εφαρμόστηκαν για την ανάλυση των επιδράσεων της ανατροφοδότησης σε κρίσιμους δείκτες οδήγησης, όπως η υπερβολική ταχύτητα, η χρήση κινητού τηλεφώνου, οι απότομες επιταχύνσεις και τα απότομα φρεναρίσματα. Η ανάλυση πραγματοποιήθηκε σε **τρεις βασικούς πυλώνες**:

1. **Επίδραση της Ανατροφοδότησης**: Αυτός ο πυλώνας αξιολόγησε τις άμεσες επιδράσεις της ανατροφοδότησης i) στη συμπεριφορά του οδηγού, με έμφαση στην υπερβολική ταχύτητα για τους μοτοσικλετιστές και την απόσπαση προσοχής λόγω χρήσης κινητού τηλεφώνου για οδηγούς αυτοκινήτων, και ii) στην ασφάλεια των οδηγών, με έμφαση στα απότομα φρεναρίσματα και επιταχύνσεις για οδηγούς αυτοκινήτων και επαγγελματίες οδηγούς σε αυτοκινητόδρομους. Χρησιμοποιήθηκαν Γενικευμένα Γραμμικά Μοντέλα Μικτών Επιδράσεων (GLMMs) για την αξιολόγηση των επιδράσεων της ανατροφοδότησης, λαμβάνοντας υπόψη τις ατομικές διαφορές και τους συμφραζόμενους παράγοντες.

2. **Επιδράσεις των Διαφορετικών Χαρακτηριστικών Ανατροφοδότησης**: Αναπτύχθηκε ένα Μοντέλο Δομικών Εξισώσεων (SEM) για την εξερεύνηση των πολύπλοκων σχέσεων μεταξύ χαρακτηριστικών ανατροφοδότησης (π.χ. βαθμολογίες, χάρτες, συγκρίσεις με άλλους οδηγούς, κίνητρα, παιχνιδοποίηση, επιβραβεύσεις), μετρικών έκθεσης και αποτελεσμάτων ασφάλειας. Το μοντέλο επέτρεψε την ταυτόχρονη ανάλυση πολλών μεταβλητών και των αλληλεπιδράσεών τους.

3. **Επιδράσεις Μετά την Ανατροφοδότηση**: Μελέτη των μακροπρόθεσμων αλλαγών στη συμπεριφορά και ασφάλεια των οδηγών, με ιδιαίτερη έμφαση στην υποτροπή των συμπεριφορών οδήγησης μετά την απόσυρση της ανατροφοδότησης κατά την τελευταία φάση του πειράματος. Χρησιμοποιήθηκαν μέθοδοι Ανάλυσης Επιβίωσης για την εξέταση μοτίβων υποτροπής σε διάφορους δείκτες, όπως οι απότομες επιταχύνσεις, τα απότομα φρεναρίσματα, η υπερβολική ταχύτητα και η χρήση κινητού τηλεφώνου κατά την οδήγηση. Αυτές οι αναλύσεις βασίστηκαν σε καμπύλες Kaplan-Meier, μοντέλα Cox-PH με παραγοντική τυχαιότητα, μοντέλα Επιταχυνόμενου Χρόνου Αποτυχίας (Weibull AFT) με ετερογένεια, και Τυχαία Δάση Επιβίωσης για την αξιολόγηση και σύγκριση της προβλεπτικής ικανότητας και των ευρημάτων κάθε μοντέλου.

Η σύνθεση όλων των αναλύσεων που πραγματοποιήθηκαν στο πλαίσιο αυτής της διατριβής κατέληξε σε έναν μηχανισμό ανατροφοδότησης συμπεριφοράς του οδηγού με πολυάριθμα πρωτότυπα και ενδιαφέροντα αποτελέσματα, τα οποία συζητούνται παρακάτω.

Κύρια Ευρήματα

Επίδραση Ανατροφοδότησης στη Συμπεριφορά των Οδηγών

Η διερεύνηση των επιπτώσεων της ανατροφοδότησης στη συμπεριφορά και την ασφάλεια των οδηγών απέδωσε σημαντικά ευρήματα σε διαφορετικές ομάδες χρηστών, περιβάλλοντα οδήγησης και μετρικές συμπεριφοράς. Συνολικά, κατά τη διάρκεια των δύο φάσεων του πειράματος καταγράφηκε και αναλύθηκε ένα μεγάλο σύνολο δεδομένων **3.537 διαδρομών από ένα δείγμα 13 μοτοσικλετιστών**. Η χρήση Γενικευμένων Γραμμικών Μοντέλων Μεικτών (ΓΓΜΜ) αποτελεσμάτων με τυχαίες παρεμβολές και τυχαίες κλίσεις για τη συνολική διάρκεια του ταξιδιού αποκάλυψε ότι η παροχή ανατροφοδότησης στους μοτοσικλετιστές σχετικά με τις επιδόσεις οδήγησής τους κατά τη διάρκεια της φάσης 2 του πειράματος οδήγησε σε **στατιστικά σημαντική μείωση του ποσοστού υπερβολικής ταχύτητας κατά τη διάρκεια ενός ταξιδιού**. Ειδικότερα, στα μοντέλα που αναπτύχθηκαν η ανατροφοδότηση του αναβάτη φαίνεται να μειώνει το ποσοστό υπερβολικής ταχύτητας, έχοντας λόγο κινδύνου $\exp(\beta=-0{,}145) = 0{,}865$ για το συνολικό μοντέλο (μείωση 13,5%) και $\exp(\beta=-0{,}031) = 0{,}970$ και $\exp(\beta=-0{,}420) = 0{,}657$ για τους αστικούς (μείωση 3,0%) και αγροτικούς (34,3%) τύπους δρόμων αντίστοιχα. Τα αποτελέσματα αυτά αναδεικνύουν την αποτελεσματικότητα της ανατροφοδότησης στη στόχευση συμπεριφορών υψηλού κινδύνου, προσφέροντας μια βάση για επεκτάσιμες παρεμβάσεις στην εκπαίδευση των αναβατών και στο σχεδιασμό πολιτικής.

Αντίστοιχα, η **απόσπαση προσοχής, ιδιαίτερα λόγω χρήσης κινητού τηλεφώνου,** εξετάστηκε σε αστικά και υπεραστικά περιβάλλοντα και αυτοκινητόδρομους μέσω μοντέλων ΓΓΜΜ για **65 οδηγούς αυτοκινήτων σε 21.167 διαδρομές**. Η ανατροφοδότηση προέκυψε ως ισχυρός περιοριστικός παράγοντας στη χρήση κινητού τηλεφώνου κατά την οδήγηση συνολικά ($\beta = -0.4276$, $p < 2e-16$), ιδιαίτερα σε αστικά περιβάλλοντα ($\beta = -0.3687$, $p < 2e-16$). Ωστόσο, η επίδρασή της ήταν ασθενέστερη σε υπεραστικά περιβάλλοντα (Estimate = -0.1180, $p < 2e-16$) και απροσδόκητα θετική σε αυτοκινητόδρομους ($\beta = 0.5490$, $p < 2e-16$), υποδεικνύοντας αντισταθμιστικές συμπεριφορές ή μια μειωμένη αντιληπτή επικινδυνότητα απόσπασης σε οδικούς άξονες υψηλής ταχύτητας. Η σημαντική μεταβλητότητα στους τυχαίους συντελεστές (SD = 1.4024 συνολικά) υπογραμμίζει τις σημαντικές ατομικές διαφορές στη βασική συμπεριφορά, ενώ οι τυχαίες κλίσεις για τη διάρκεια διαδρομής (SD = 0.2827 συνολικά) δείχνουν ποικίλες αντιδράσεις σε παρατεταμένες διαδρομές.

Στον τομέα των απότομων συμβάντων, όπως οι επιταχύνσεις και τα φρεναρίσματα, οι μηχανισμοί ανατροφοδότησης έδειξαν επίσης σημαντικές βελτιώσεις στη συμπεριφορά. Τα αποτελέσματα από την **ανάλυση 65 οδηγών αυτοκινήτων** κατά τις πρώτες δύο φάσεις του πειράματος αποκάλυψαν **σημαντική μείωση στις απότομες επιταχύνσεις (12%) και στα απότομα φρεναρίσματα (10%)**, και οι δύο αλλαγές ήταν στατιστικά σημαντικές ($p < 0.001$). Τα μοντέλα ΓΓΜΜ ενισχύουν περαιτέρω αυτά τα ευρήματα, καθώς η ανατροφοδότηση συνδέθηκε σταθερά με μειωμένες συχνότητες απότομων γεγονότων, ιδιαίτερα σε αστικά και υπεραστικά περιβάλλοντα. Σημαντικό είναι ότι οι δείκτες σχετικού κινδύνου για τη διάρκεια της ταχύτητας και της διαδρομής υποδεικνύουν ισχυρές θετικές συσχετίσεις με τα απότομα γεγονότα, αν και η ανατροφοδότηση φαίνεται να μετριάζει αυτές τις επιδράσεις σε κάποιο βαθμό κατά τη Φάση 2 του πειράματος.

Οι επαγγελματίες οδηγοί, λόγω των παρατεταμένων ωρών και αποστάσεων οδήγησης, αποτέλεσαν επίσης βασικό πεδίο εξερεύνησης. Χρησιμοποιώντας μοντέλα ΓΓΜΜ που βαθμονομήθηκαν σε ένα σύνολο δεδομένων από 5.345 διαδρομές από 19 επαγγελματίες οδηγούς, η ανάλυση αποκάλυψε ότι η συμμετοχή σε ένα **κοινωνικό πλαίσιο παιχνιδοποίησης με κίνητρα οδήγησε σε αξιοσημείωτες βελτιώσεις στα απότομα συμβάντα των επαγγελματιών οδηγών.** Κατά τη φάση του διαγωνισμού, η πιθανότητα απότομων επιταχύνσεων μειώθηκε κατά έναν συντελεστή (β= 0.348p < 0.001), ενώ τα απότομα φρεναρίσματα μειώθηκαν κατά έναν συντελεστή (β= 0.404, p < 0.001), υποδεικνύοντας την αποτελεσματικότητα της παιχνιδοποίησης στην προώθηση ασφαλέστερων πρακτικών οδήγησης. Επιπλέον, η διάρκεια της διαδρομής παρουσίασε θετική συσχέτιση με τα απότομα γεγονότα, καθώς η αύξηση της διάρκειας οδήγησης κατά 1 δευτερόλεπτο αύξησε τις πιθανότητες απότομων επιταχύνσεων και φρεναρισμάτων κατά 1.558 και 1.564, αντίστοιχα, αναδεικνύοντας τις σωρευτικές επιδράσεις της παρατεταμένης οδήγησης. Η ενσωμάτωση τυχαίων μεταβλητών στα μοντέλα ανέδειξε σημαντική μεταβλητότητα στη βασική οδηγική συμπεριφορά, υπογραμμίζοντας την ανάγκη για εξατομικευμένες παρεμβάσεις.

Επιδράσεις των Διαφορετικών Χαρακτηριστικών Ανατροφοδότησης στη Συμπεριφορά των Οδηγών

Η **ανάλυση μέσω του Μοντέλου Δομικών Εξισώσεων** (ΜΔΕ) παρείχε σημαντικές γνώσεις για την επίδραση των διαφοροποιημένων χαρακτηριστικών ανατροφοδότησης στη συμπεριφορά και την ασφάλεια των οδηγών, συγκεκριμένα στην υπερβολική ταχύτητα, τα απότομα φρεναρίσματα και τις απότομες επιταχύνσεις. Το σύνολο δεδομένων, που περιλάμβανε **73.869 διαδρομές από 175 οδηγούς αυτοκινήτων για διάστημα 21 μηνών**, προσέφερε μια ισχυρή βάση δεδομένων για τη μοντελοποίηση. Τα αποτελέσματα του ΜΔΕ εντόπισαν δύο λανθάνουσες μεταβλητές, δηλαδή την ανατροφοδότηση και την έκθεση ως κρίσιμες επιρροές. Το μοντέλο παρουσίασε άριστα μέτρα καλής προσαρμογής, με δείκτη συγκριτικής προσαρμογής (CFI) = 0,940, δείκτη Tucker-Lewis (TLI) = 0,944, μέσο τετραγωνικό σφάλμα προσέγγισης (RMSEA) = 0,049 και τυποποιημένο μέσο τετραγωνικό υπόλοιπο (SRMR) = 0,025, υποδεικνύοντας μια ισχυρή και καλά καθορισμένη δομή.

Μεταξύ των χαρακτηριστικών ανατροφοδότησης που αναλύθηκαν, η **Κάρτα αποτελεσμάτων αναδείχθηκε ως το πιο επιδραστικό χαρακτηριστικό**, με την υψηλότερη θετική εκτίμηση (β= 2.076, p < 0.001), αποδεικνύοντας τον ισχυρό του ρόλο στην προώθηση ασφαλέστερων συνηθειών οδήγησης μέσω της άμεσης τροποποίησης επικίνδυνων συμπεριφορών. Παρομοίως, το χαρακτηριστικό των χαρτών έδειξε ισχυρή επίδραση (β = 1.646, p < 0.001), τονίζοντας τη σημασία της χωρικής επίγνωσης για τη βελτίωση των πρακτικών οδήγησης. Η λειτουργία σύγκρισης με άλλους οδηγούς (peer comparison) επηρέασε θετικά τη συμπεριφορά (β = 1.215, p < 0.001), ενώ **τα χαρακτηριστικά των διαγωνισμών και προκλήσεων αποδείχθηκαν ιδιαίτερα αποτελεσματικά** (β = 2.053, p < 0.001) ενισχύοντας τους οδηγούς να υιοθετήσουν ασφαλέστερες οδηγικές συμπεριφορές μέσω παιχνιδοποίησης.

Σε ό,τι αφορά τις μετρικές οδηγικής συμπεριφοράς, η ανατροφοδότηση μέσω τηλεματικής **μείωσε σημαντικά τον χρόνο υπερβολικής ταχύτητας** (β = -0.214, p < 0.001) **και τα απότομα φρεναρίσματα** (β = -0.027, p < 0.001). Ωστόσο, **μια αύξηση στις απότομες επιταχύνσεις** (β =

0.026, $p < 0.001$) υποδηλώνει την ανάγκη περαιτέρω βελτίωσης των συστημάτων ανατροφοδότησης για την αντιμετώπιση απρόβλεπτων συνεπειών. Οι παράγοντες έκθεσης διαδραμάτισαν επίσης σημαντικό ρόλο στη διαμόρφωση της συμπεριφοράς των οδηγών, με την έκθεση στην πρωινή αιχμή να συσχετίζεται με αυξημένη ανάληψη κινδύνου ($\beta = 2{,}473$, $p < 0{,}001$), πιθανότατα λόγω της πίεσης του χρόνου κατά τις ώρες μετακίνησης. Αντίθετα, η έκθεση στην απογευματινή αιχμή συσχετίστηκε με λιγότερο επιθετική συμπεριφορά ($\beta= -1{,}360$, $p < 0{,}001$), παρέχοντας πληροφορίες σχετικά με τις χρονικές διακυμάνσεις στα πρότυπα οδήγησης.

Η ανάλυση παλινδρόμησης επιβεβαίωσε αυτά τα ευρήματα, αναδεικνύοντας την **αλληλεπίδραση μεταξύ των μεταβλητών έκθεσης και των χαρακτηριστικών ανατροφοδότησης**. Ενώ η έκθεση επηρέασε θετικά την υπερβολική ταχύτητα ($\beta = 0{,}326$, $p < 0{,}001$), οι μηχανισμοί ανατροφοδότησης μετρίασαν αποτελεσματικά αυτή τη συμπεριφορά. Το χαρακτηριστικό «ανταγωνισμός και προκλήσεις», ειδικότερα, έδειξε υποσχόμενο να μετριάσει τις απότομες επιταχύνσεις ($\beta = -0{,}001$, $p < 0{,}001$). Τα περιστατικά απότομου φρεναρίσματος μειώθηκαν επίσης σημαντικά από την ανατροφοδότηση, ενισχύοντας το ρόλο της ανατροφοδότησης στην προώθηση ασφαλέστερων οδηγικών πρακτικών. Η ανάλυση συνδιακύμανσης αποκάλυψε περαιτέρω ισχυρές αλληλεπιδράσεις μεταξύ επικίνδυνων συμπεριφορών, όπως η υπερβολική ταχύτητα και το απότομο φρενάρισμα, υπογραμμίζοντας την ανάγκη για ολοκληρωμένα συστήματα ανατροφοδότησης που αντιμετωπίζουν πολλαπλές πτυχές της συμπεριφοράς του οδηγού.

Οι πρακτικές επιπτώσεις αυτών των ευρημάτων είναι ουσιαστικές. Τα χαρακτηριστικά ανατροφοδότησης, ιδίως εκείνα που αξιοποιούν εξατομικευμένους πίνακες αποτελεσμάτων, χωρικά εργαλεία και στοιχεία παιχνιδοποίησης, υπόσχονται πολλά για τη βελτίωση της ασφάλειας των οδηγών. Οι εξατομικευμένες παρεμβάσεις που στοχεύουν σε συγκεκριμένες συμπεριφορές και ώρες της ημέρας θα μπορούσαν να ενισχύσουν περαιτέρω την αποτελεσματικότητα αυτών των συστημάτων. Ωστόσο, περιορισμοί όπως ο αποκλεισμός της χρήσης κινητού τηλεφώνου από το τελικό μοντέλο και οι πιθανές μεροληψίες επιλογής λόγω της εθελοντικής φύσης της συμμετοχής θα πρέπει να αντιμετωπιστούν σε μελλοντικές έρευνες.

<u>Μετά την Ανατροφοδότηση Επιδράσεις στη Μακροπρόθεσμη Συμπεριφορά των Οδηγών</u>

Οι **τεχνικές Ανάλυσης Επιβίωσης** εφαρμόστηκαν σε ένα **σύνολο δεδομένων 24.904 διαδρομών από 31 οδηγούς αυτοκινήτων**, με κάθε οδηγό να έχει υλοποιήσει τουλάχιστον 20 διαδρομές κατά τη φάση μετά την ανατροφοδότηση, για τη διερεύνηση των μακροπρόθεσμων επιδράσεων της τηλεματικής ανατροφοδότησης στη συμπεριφορά του οδηγού. Η ανάλυση επικεντρώθηκε σε μοτίβα υποτροπής όσον αφορά στη χρήση κινητού τηλεφώνου, την υπερβολική ταχύτητα, τα απότομα φρεναρίσματα και τις απότομες επιταχύνσεις. Οι μέθοδοι που χρησιμοποιήθηκαν περιλάμβαναν καμπύλες Kaplan-Meier, μοντέλα Cox-PH με frailty, μοντέλα Weibull AFT με ετερογένεια σε ομάδες, και Random Survival Forests. Τα ευρήματα καταδεικνύουν την αποτελεσματικότητα των παρεμβάσεων ανατροφοδότησης στην επίτευξη σημαντικών βραχυπρόθεσμων βελτιώσεων συμπεριφοράς κατά τη φάση της ανατροφοδότησης. Ωστόσο, **η φάση μετά την ανατροφοδότηση αποκάλυψε ποικίλες τάσεις υποτροπής**, υπογραμμίζοντας την ανάγκη για συνεχιζόμενες παρεμβάσεις για τη διατήρηση αυτών των βελτιώσεων με την πάροδο του χρόνου.

Η **ανάλυση επιβίωσης Kaplan-Meier τόνισε τις τάσεις υποτροπής**, δείχνοντας μια σταδιακή μείωση της βελτιωμένης συμπεριφοράς στις διαδοχικές διαδρομές κατά τη φάση μετά την ανατροφοδότηση. Για τις απότομες επιταχύνσεις, **οι πιθανότητες επιβίωσης μειώθηκαν από 84,8% στις 50 διαδρομές στο 49,2% στις 150 διαδρομές**. Παρόμοιες τάσεις παρατηρήθηκαν για τα απότομα φρεναρίσματα και την υπερβολική ταχύτητα, με τις πιθανότητες επιβίωσης να μειώνονται περίπου στο 40,3% και 46,8%, αντίστοιχα, έως τη διαδρομή 150. Αυτά τα μοτίβα υπογραμμίζουν τη **μεταβατική φύση των επιδράσεων της ανατροφοδότησης** και την ανάγκη για μηχανισμούς συνεχούς ενίσχυσης. Η χρήση κινητού τηλεφώνου παρουσίασε ελαφρώς μεγαλύτερη ανθεκτικότητα, με τις πιθανότητες επιβίωσης να παραμένουν πάνω από 80% στις 100 διαδρομές, αλλά η σταδιακή υποτροπή ήταν εμφανής με την πάροδο του χρόνου.

Μεταξύ των μοντέλων ανάλυσης επιβίωσης που εφαρμόστηκαν, το **μοντέλο Επιταχυνόμενου Χρόνου Αποτυχίας (Weibull AFT) με ετερογένεια αναδείχθηκε ως ιδιαίτερα ισχυρό εργαλείο** για την ανάλυση των δυναμικών υποτροπής, ισορροπώντας την προβλεπτική ακρίβεια και την ερμηνευσιμότητα. Οι τιμές C-index κυμάνθηκαν μεταξύ 0.677 και 0.773, με το μοντέλο να επιτυγχάνει τη μεγαλύτερη προβλεπτική ικανότητα για την υποτροπή στη χρήση κινητού τηλεφώνου (C-index = 0.773), υποδεικνύοντας ισχυρή διακριτική ικανότητα στον εντοπισμό οδηγών με αυξημένο κίνδυνο υποτροπής. Κύριοι προβλεπτικοί παράγοντες όπως η ηλικιακή ομάδα [35-54] ($\beta = 0.165$, $p = 0.041$), η διάρκεια διαδρομής ($\beta = -0.022$, $p < 0.001$), και η αυτοαναφερόμενη επιθετικότητα (πλησιάζοντας τη στατιστική σημαντικότητα στο $p = 0.089$) αναδείχθηκαν, παρέχοντας χρήσιμες πληροφορίες για τη συμπεριφορά υποτροπής. Το μοντέλο επίσης **κατέγραψε ετερογένεια μεταξύ οδηγών μέσω ενσωμάτωσης τυχαίων παραμέτρων**, με τα αποτελέσματα ετερογένειας να δείχνουν σημαντική μεταβλητότητα στους χρόνους επιβίωσης.

Για την υποτροπή στην υπερβολική ταχύτητα, το μοντέλο Weibull AFT πέτυχε C-index = 0.700, με σημαντικούς προβλεπτικούς παράγοντες, όπως η διάρκεια διαδρομής ($\beta = -0.022$, $p < 0.001$) και οι πρωινές ώρες αιχμής ($\beta = -0.096$, $p = 0.004$). **Η διάρκεια διαδρομής αναδείχθηκε ως ο κυρίαρχος προβλεπτικός παράγοντας, μειώνοντας σταθερά τους χρόνους επιβίωσης σε όλους τους δείκτες υποτροπής**, υπογραμμίζοντας τον ρόλο της παρατεταμένης οδήγησης στη συμπεριφορική υποτροπή. Αντίστοιχα, στην ανάλυση υποτροπής των απότομων φρεναρισμάτων, το μοντέλο πέτυχε μέτρια προβλεπτική ακρίβεια (C-index = 0.724) με σημαντικές συνεισφορές από μεταβλητές όπως η ηλικιακή ομάδα [35-54] ($\beta = 0.360$, $p = 0.010$) και η χωρητικότητα κινητήρα (>1400cc) ($\beta = -0.508$, $p = 0.012$). Αυτά τα ευρήματα δείχνουν ότι οι νεότερες ηλικιακές ομάδες και οι οδηγοί οχημάτων με μεγαλύτερους κινητήρες είναι πιο επιρρεπείς σε υποτροπή.

Το **μοντέλο Τυχαίων Δασών Επιβίωσης (Random Survival Forest) παρουσίασε ανώτερη προβλεπτική απόδοση** σε ορισμένους εξεταζόμενους δείκτες, διακρινόμενο στην καταγραφή μη γραμμικών αλληλεπιδράσεων και πολύπλοκων σχέσεων μεταξύ προβλεπτικών παραγόντων. Με τιμές Root Mean Squared Error (RMSE) έως 85.87 και σφάλματα πρόβλεψης εκτός δείγματος (OOB) της τάξης του 24,3% για την υποτροπή στη χρήση κινητού τηλεφώνου, το RSF εντόπισε κρίσιμους προβλεπτικούς παράγοντες, όπως η διάρκεια διαδρομής, οι τάσεις επιθετικής οδήγησης, και το μέγεθος του κινητήρα του οχήματος. Η **ευελιξία του μοντέλου στην επεξεργασία ποικίλων προβλεπτικών παραγόντων** και η αποκάλυψη λεπτομερών δυναμικών το καθιστούν πολύτιμο εργαλείο για προβλεπτικές αναλύσεις. Ωστόσο, η «μαύρο κουτί» φύση του μοντέλου και η εξάρτησή του από μεγαλύτερα σύνολα δεδομένων περιορίζουν την ερμηνευσιμότητά του και τη χρησιμότητά του για επεξηγηματικούς σκοπούς.

Συνολικά, συγκρίνοντας τα μοντέλα, **το Weibull AFT ξεχωρίζει για την ισορροπία ερμηνευσιμότητας και προβλεπτικής ακρίβειας**, καθιστώντας το ιδιαίτερα κατάλληλο για περιβάλλοντα που απαιτούν εφαρμόσιμες γνώσεις σχετικά με τη δυναμική της επιβίωσης. Αντίθετα, το μοντέλο Τυχαίων Δασών Επιβίωσης είναι πιο κατάλληλο για καθαρά προβλεπτικούς στόχους όπου η καταγραφή μη γραμμικών σχέσεων και πολύπλοκων αλληλεπιδράσεων είναι κρίσιμη, αν και η έλλειψη διαφάνειας περιορίζει τη χρησιμότητά του στην κατανόηση των υποκείμενων μηχανισμών συμπεριφοράς. Αυτά τα ευρήματα υπογραμμίζουν τη **σημασία της ευθυγράμμισης της επιλογής μοντέλου με τους ερευνητικούς στόχους**. Για μελέτες που επικεντρώνονται στην κατανόηση των δυναμικών της συμπεριφοράς και στον σχεδιασμό παρεμβάσεων, το μοντέλο Weibull AFT παρέχει αξιόπιστες πληροφορίες. Αντίθετα, όταν η ακρίβεια πρόβλεψης είναι πρωταρχικής σημασίας, το RSF αποτελεί ανώτερη εναλλακτική λύση.

Παρόλο που η δατριβή παρέχει πολύτιμες γνώσεις για τη δυναμική της ανατροφοδότησης και της υποτροπής των οδηγών, περιορισμοί όπως το σχετικά μικρό δείγμα, η απουσία δεδομένων κυκλοφοριακών συνθηκών και η μακροσκοπική εστίαση πρέπει να σημειωθούν. Μελλοντική έρευνα θα μπορούσε να ενσωματώσει πιο λεπτομερή δεδομένα, όπως τις κυκλοφοριακές δυναμικές και τις αποφάσεις των οδηγών κατά τη διάρκεια της οδήγησης, για την παροχή βαθύτερης κατανόησης των συμπεριφορικών μοτίβων. Η εφαρμογή προηγμένων τεχνικών μοντελοποίησης, όπως οι τυχαίες παράμετροι με ετερογένεια-μέση, θα μπορούσε να ενισχύσει περαιτέρω την ανάλυση, λαμβάνοντας υπόψη τη μεταβλητότητα μεταξύ οδηγών.

Καινοτόμες Επιστημονικές Συνεισφορές

Οι καινοτόμες συνεισφορές της παρούσας διδακτορικής διατριβής αποτελούνται από πέντε πρωτότυπες επιστημονικές συνεισφορές, όπως περιγράφονται παρακάτω και απεικονίζονται στο Σχήμα II.

Εικόνα II: Καινοτόμες συνεισφορές της διδακτορικής διατριβής

Εκτεταμένη Συλλογή Δεδομένων Φυσικής Οδήγησης

Η παρούσα διατριβή αποτελεί ένα **σημαντικό βήμα προς τα εμπρός στην έρευνα πειραμάτων φυσικής οδήγησης (Naturalistic Driving - ND)**, αξιοποιώντας μη παρεμβατικές μεθόδους συλλογής δεδομένων που βασίζονται σε αισθητήρες smartphone. Σε αντίθεση με τις παραδοσιακές προσεγγίσεις, αυτή η μεθοδολογία ελαχιστοποιεί τις παρεμβάσεις στους συμμετέχοντες, επιτρέποντας την ανεμπόδιστη καταγραφή πραγματικών οδηγικών συμπεριφορών. **Το σύνολο δεδομένων καλύπτει ένα μεγάλο δείγμα οδηγών (230), σε ποικίλα οδηγικά περιβάλλοντα και τύπους οχημάτων**, συμπεριλαμβανομένων οδηγών αυτοκινήτων, μοτοσικλετιστών και επαγγελματιών οδηγών βαν. Αυτή η ποικιλία διασφαλίζει ότι τα ευρήματα είναι όχι μόνο αντιπροσωπευτικά, αλλά λαμβάνουν επίσης υπόψη τις διαφοροποιήσεις μεταξύ δημογραφικών χαρακτηριστικών οδηγών και κατηγοριών οχημάτων.

Η υψηλή χρονική ανάλυση των δεδομένων προσφέρει λεπτομερείς πληροφορίες για τη συμπεριφορά οδήγησης σε επίπεδο διαδρομής, παρέχοντας μια βαθιά κατανόηση της δυναμικής της οδηγικής συμπεριφοράς. Επιπλέον, η μακροπρόθεσμη συλλογή δεδομένων για 21 μήνες, που καλύπτει πολλές φάσεις ανατροφοδότησης και διαφορετικά οδικά περιβάλλοντα, προσδίδει μοναδική αξία στη μελέτη. Καταγράφοντας αλλαγές στη συμπεριφορά με την πάροδο του χρόνου, **η διατριβή γεφυρώνει ένα κρίσιμο κενό στην υπάρχουσα έρευνα ND**, η οποία συχνά βασίζεται σε βραχυπρόθεσμες παρατηρήσεις. Αυτή η μακροπρόθεσμη προοπτική επιτρέπει την αξιολόγηση της διατηρησιμότητας των αλλαγών στη συμπεριφορά και των τάσεων υποτροπής, προσφέροντας μια ισχυρή βάση για την ανάπτυξη προσαρμοστικών και βιώσιμων παρεμβάσεων για τη βελτίωση της οδικής ασφάλειας. Η **μεθοδολογία αυτή θέτει νέα πρότυπα για τα πειράματα ND**, ανοίγοντας τον δρόμο για πιο κλιμακούμενες, οικονομικά αποδοτικές και τεχνολογικά προηγμένες μελέτες οδηγικής συμπεριφοράς.

Πολυτροπική Προσέγγιση Ανάλυσης Οδηγικής Συμπεριφοράς

Αυτή η διατριβή υιοθετεί μια πολυτροπική προσέγγιση, δίνοντας έμφαση στη σημασία της κατανόησης της οδηγικής συμπεριφοράς σε διαφορετικές ομάδες χρηστών οδού και διαφορετικά περιβάλλοντα. **Συμπεριλαμβάνοντας οδηγούς αυτοκινήτων, μοτοσικλετιστές και επαγγελματίες οδηγούς βαν**, η έρευνα αναγνωρίζει την ανάγκη μελέτης ευάλωτων χρηστών οδού, τους οι μοτοσικλετιστές, οι οποίοι αντιμετωπίζουν αυξημένους κινδύνους, καθώς και επαγγελματιών οδηγών, οι οποίοι περνούν εκτεταμένες ώρες οδηγώντας. Αυτή η **περιεκτική προσέγγιση διασφαλίζει μια πλήρη αξιολόγηση της ανατροφοδότησης μέσω τηλεματικής**, αναδεικνύοντας τη σημασία της σε διαφορετικά προφίλ κινδύνου και επίπεδα έκθεσης.

Η έρευνα εξετάζει την επίδραση των αστικών και υπεραστικών περιβαλλόντων, αλλά και αυτοκινητόδρομων, λαμβάνοντας υπόψη τις διαφορετικές προκλήσεις που θέτει κάθε τύπος οδού. Αυτή η περιληπτική προσέγγιση αποκαλύπτει ότι **η αποτελεσματικότητα της ανατροφοδότησης δεν είναι ομοιόμορφη**· συμπεριφορές όπως η χρήση κινητού τηλεφώνου ή η υπερβολική ταχύτητα αντιδρούν διαφορετικά στις παρεμβάσεις, **ανάλογα με το οδικό περιβάλλον**. Για παράδειγμα, οι μοτοσικλετιστές μπορεί να επωφεληθούν περισσότερο από ανατροφοδότηση που στοχεύει στην επίγνωση τους κατάστασης, ενώ οι επαγγελματίες οδηγοί ενδέχεται να απαιτούν εξατομικευμένες παρεμβάσεις που αντιμετωπίζουν την κόπωση και την επαναλαμβανόμενη έκθεση σε καταστάσεις υψηλού κινδύνου.

Ενσωματώνοντας αυτήν την ποικιλία ομάδων χρηστών και συμφραζομένων, η διατριβή παρέχει **πρακτικές πληροφορίες για υπεύθυνους χάραξης πολιτικής**, υποστηρικτές τους οδικής ασφάλειας και προγραμματιστές στον κλάδο της τεχνολογίας. Τονίζει τη σημασία της ανάπτυξης εξατομικευμένων συστημάτων ανατροφοδότησης που ανταποκρίνονται στις μοναδικές ανάγκες ευάλωτων χρηστών οδού, τους μοτοσικλετιστές, και επαγγελματίες οδηγούς, οι οποίοι συμβάλλουν σημαντικά στη δραστηριότητα οδικής κυκλοφορίας. Αυτή η περιεκτική προσέγγιση υποστηρίζει τη δημιουργία προσαρμοστικών, ευαίσθητων στο πλαίσιο παρεμβάσεων, βελτιώνοντας τελικά την οδική ασφάλεια για όλους τους χρήστες.

<u>Ολοκληρωμένη Δέσμη Μοντέλων Τριών Πυλώνων</u>

Αυτή η διατριβή χρησιμοποιεί μια **ολοκληρωμένη σειρά προηγμένων στατιστικών και τεχνικών μηχανικής μάθησης,** προσαρμοσμένων για την αντιμετώπιση της πολυδιάστατης φύσης της ανάλυσης οδηγικής συμπεριφοράς. Ενσωματώνοντας Γενικευμένα Γραμμικά Μικτά Μοντέλα (GLMMs), Μοντέλα Δομικών Εξισώσεων (SEMs) και Τεχνικές Ανάλυσης Επιβίωσης (όπως καμπύλες Kaplan-Meier, μοντέλα Cox-PH με παραγοντική τυχαιότητα, μοντέλα Επιταχυνόμενου Χρόνου Αποτυχίας (Weibull AFT) με ετερογένεια, και Τυχαία Δάση Επιβίωσης), η διατριβή παρέχει ένα αυστηρό αναλυτικό πλαίσιο ικανό να αποκαλύψει τόσο γραμμικές όσο και μη γραμμικές σχέσεις μεταξύ των μεταβλητών. Κάθε μοντέλο επιλέγεται προσεκτικά ώστε να ευθυγραμμίζεται με τους ερευνητικούς στόχους, εξισορροπώντας την προβλεπτική ακρίβεια με την ερμηνευσιμότητα για να διασφαλίσει πρακτικές πληροφορίες.

Αυτή η δέσμη μοντέλων επιτρέπει την **εξερεύνηση πολύπλοκων φαινομένων**, όπως η αλληλεπίδραση μεταξύ χαρακτηριστικών ανατροφοδότησης, οδηγικών συμπεριφορών και παραγόντων πλαισίου, όπως η ώρα της ημέρας ή ο τύπος δρόμου. Για παράδειγμα, τα μοντέλα επιβίωσης καταγράφουν μοναδικά τη δυναμική της υποτροπής, προσφέροντας νέες γνώσεις για τις τάσεις συμπεριφοράς μετά την αφαίρεση της ανατροφοδότησης. Οι τεχνικές μηχανικής μάθησης ενισχύουν περαιτέρω την ανάλυση, συλλαμβάνοντας λεπτές, μη γραμμικές αλληλεπιδράσεις, διασφαλίζοντας ότι τα μοντέλα είναι εξοπλισμένα για την αντιμετώπιση της πολυπλοκότητας των πραγματικών δεδομένων οδήγησης. Το **καινοτόμο αυτό αναλυτικό πλαίσιο** όχι μόνο αυξάνει τη μεθοδολογική αυστηρότητα της έρευνας αλλά και καταδεικνύει το δυναμικό της συνδυαστικής χρήσης παραδοσιακών στατιστικών μεθόδων και τεχνολογιών μηχανικής μάθησης για μελέτες οδηγικής συμπεριφοράς.

<u>Σε Βάθος Ανάλυση Μετά την Ανατροφοδότηση Επιδράσεων</u>

Αυτή η διατριβή είναι **από τις πρώτες που αναλύει σε βάθος τις επιδράσεις μετά την ανατροφοδότηση στη συμπεριφορά του οδηγού**, χρησιμοποιώντας προηγμένες στατιστικές και τεχνικές μηχανικής μάθησης, αντιμετωπίζοντας ένα κρίσιμο κενό στην υπάρχουσα έρευνα. Μέσω μεθόδων ανάλυσης επιβίωσης, όπως τα Weibull AFT και Random Survival Forest, η μελέτη αξιολογεί μακροπρόθεσμες αλλαγές στη συμπεριφορά και μοτίβα υποτροπής μετά την απόσυρση της ανατροφοδότησης. Αυτές οι τεχνικές επιτρέπουν τη λεπτομερή εξερεύνηση των παραγόντων που επηρεάζουν την υποτροπή σε επικίνδυνες συμπεριφορές, όπως η υπερβολική ταχύτητα, τα απότομα γεγονότα και η χρήση κινητού τηλεφώνου, παρέχοντας πρακτικές πληροφορίες για τον σχεδιασμό βιώσιμων στρατηγικών παρέμβασης.

Τα ευρήματα αποκαλύπτουν τη σημασία προσαρμοστικών συστημάτων ανατροφοδότησης που μπορούν να διατηρήσουν τις βελτιώσεις στη συμπεριφορά με την πάροδο του χρόνου. Για παράδειγμα, η ανάλυση επιβίωσης έδειξε ότι η διάρκεια της διαδρομής και η ώρα της ημέρας επηρεάζουν σημαντικά τη δυναμική υποτροπής, υπογραμμίζοντας την ανάγκη για μηχανισμούς ανατροφοδότησης που να λαμβάνουν υπόψη το πλαίσιο. Αυτή η καινοτόμος εστίαση στη φάση μετά την ανατροφοδότηση παρέχει ένα **νέο πλαίσιο για την κατανόηση της διατηρησιμότητας των βελτιώσεων που προκύπτουν από την ανατροφοδότηση**, επιτρέποντας πιο ανθεκτικές και αποτελεσματικές παρεμβάσεις οδικής ασφάλειας. Επιπλέον, καθορίζει ένα πρότυπο για τη μελλοντική έρευνα ώστε να ενσωματώσει μακροπρόθεσμες προοπτικές στην αξιολόγηση στρατηγικών τροποποίησης οδηγικής συμπεριφοράς.

Μηχανισμός Ανατροφοδότησης Οδηγού ως Ολιστικό Σύστημα

Η παρούσα διατριβή προσεγγίζει μοναδικά **τον μηχανισμό ανατροφοδότησης ως ένα ολιστικό σύστημα, εξετάζοντας ολόκληρο τον κύκλο ζωής του μέσω ενός πολυπαραμετρικού αναλυτικού πλαισίου.** Αναλύοντας συστηματικά τις φάσεις πριν την ανατροφοδότηση, κατά τη διάρκεια της ανατροφοδότησης και μετά την ολοκλήρωσή της, η διατριβή παρέχει μια ολοκληρωμένη κατανόηση του πώς η ανατροφοδότηση επηρεάζει τη συμπεριφορά των οδηγών με την πάροδο του χρόνου. Η ενσωμάτωση διαφόρων χαρακτηριστικών ανατροφοδότησης, όπως οι βαθμολογίες, οι χάρτες, τα εργαλεία σύγκρισης και τα στοιχεία παιχνιδοποίησης, επιτρέπει την αξιολόγηση των επιμέρους και συνδυαστικών τους επιδράσεων στην τροποποίηση της συμπεριφοράς.

Αυτή η πολυφασική προοπτική δεν εστιάζει μόνο στις άμεσες αλλαγές στη συμπεριφορά αλλά ρίχνει φως και στα μακροπρόθεσμα μοτίβα και τις τάσεις υποτροπής. Για παράδειγμα, ενώ οι βαθμολογίες και τα στοιχεία διαγωνισμού είναι ιδιαίτερα αποτελεσματικά στη μείωση της υπερβολικής ταχύτητας, η επίδρασή τους σε άλλες συμπεριφορές, όπως οι απότομες επιταχύνσεις, απαιτεί περαιτέρω βελτίωση. Το συστημικό αυτό πλαίσιο προχωρά πέρα από τις απομονωμένες αξιολογήσεις της ανατροφοδότησης, προσφέροντας ένα **κλιμακούμενο, βασισμένο σε δεδομένα πλαίσιο σχεδιασμού και εφαρμογής παρεμβάσεων μέσω τηλεματικής του οδηγού**. Τα ευρήματα τονίζουν το δυναμικό προσαρμοστικών συστημάτων ανατροφοδότησης για τη βιώσιμη βελτίωση της οδηγικής συμπεριφοράς, συμβάλλοντας τελικά σε ασφαλέστερα οδικά περιβάλλοντα.

1 Introduction

1.1 Road Safety Overview

1.1.1 Road safety statistics

Road safety is a critical public health and societal issue, as road traffic crashes claim millions of lives and cause severe injuries globally every year. Beyond the tragic loss of life, these incidents impose immense emotional and economic burdens on families and communities. Vulnerable road users, such as pedestrians, cyclists, and motorcyclists, are particularly at risk, necessitating the implementation of comprehensive safety measures, infrastructure improvements, and awareness campaigns.

Over the last decade, although substantial advances have been made in enhancing road safety, road crashes continue to pose a significant public health challenge worldwide. Road crashes, resulting in injuries and fatalities, are the 12th leading cause of death globally across all age groups, with young people aged 5-29 being at the greatest risk. In 2021, road crashes led to approximately 1.19 million fatalities globally (World Health Organization; 2023.), equating to a mortality rate of 15 deaths per 100,000 people, while the European Commission published that in 2023, the fatal crashes in Europe were 20,400 (European Road Safety Observatory, 2024).

In 2022, Greece experienced a total of 10,487 road crashes resulting in death or injury, marking a 0.3% increase compared to 2021, when 10,454 such incidents were recorded (Figure 1.1). Specifically, the casualties in 2022 included 654 fatalities, 664 serious injuries, and 11,961 minor injuries, reflecting increases of 4.8%, 8.9%, and 1.8%, respectively, compared to 2021 (624 fatalities, 610 serious injuries, and 11,746 minor injuries).

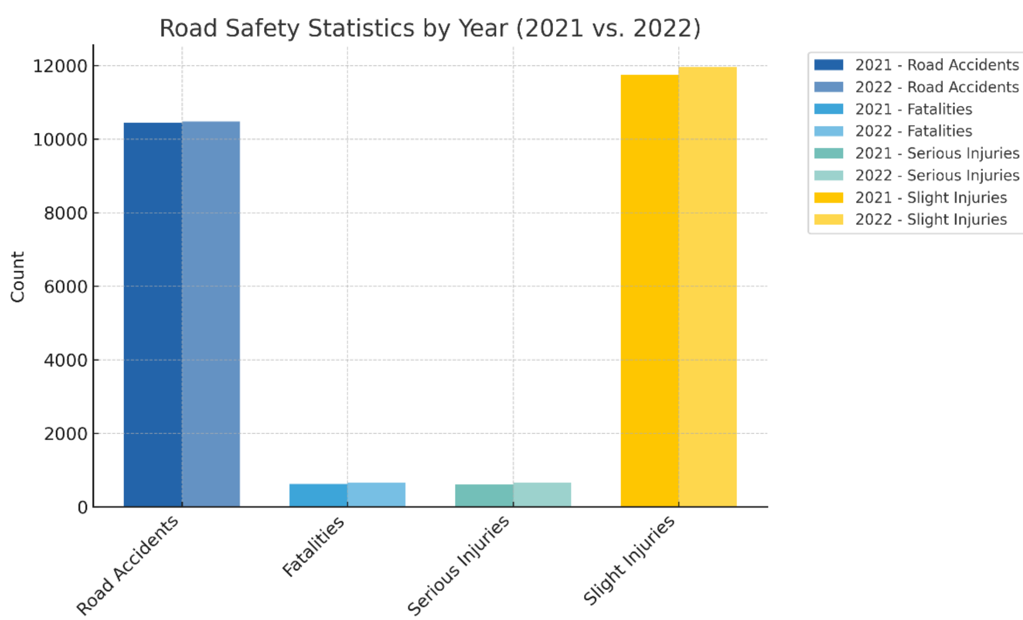


Figure 1.1: Numbers of road crashes in Greece during 2021-2022
[Source: Hellenic Statistical Authority, 2024]

Notably, of the 654 fatalities, 257 (39.3%) were in passenger vehicles, 212 (32.4%) involved two-wheel vehicles (including mopeds), and 112 (17.1%) were pedestrians. These figures underscore the significant risks faced by vulnerable road users and highlight the urgent need for targeted safety measures to protect them.

1.1.2 Human risk factors

There are a significant number of risk factors identified in literature, which affect accident probability. The most important risk factors recognized in literature are human factors (speeding, distracted driving, driving under the influence of alcohol and other psychoactive substances etc.), unsafe road infrastructure, unsafe vehicles and inadequate law enforcement of traffic laws (WHO, 2023). Human factors are likely to be the most important cause of road traffic fatalities and injuries every year and therefore the importance of quantifying the influence of driver's behavior on crash risk is extremely high. It has been observed that a large percentage of crashes is due to human error (exclusively or not), a percentage cited to be as high as 95% (Singh, 2015).

Distracted driving, as one of the most important human factors that influence crash risk, has been attracting the attention of researchers in the past decades. Mobile phone use (handheld or hands-free) and complex conversation (at mobile phone or with passengers) appear to be the most critical in-vehicle distraction factors (Papantoniou et al., 2017). Given that mobile usage is an inevitable part of the everyday driving process and is expected to increase over the years (Charlton, 2009), its impact on driving behavior in traffic and road safety is particularly crucial and merits further investigation. The literature so far has shown that when drivers are using mobile phones while driving, several impacts manifest on their behavior expressed in terms of loss of control, response to incidents, or crash occurrence (Bellinger et al., 2009; Fitch et al., 2015).

Speeding has also been the subject of extensive research in the transportation field. Excess and inappropriate speed are responsible for a high proportion of the mortality and morbidity that result from road crashes (Elvik et al., 2004; Nilsson, 2004). In high-income countries, speed contributes to about 30% of deaths on the road, while in some low-income and middle-income countries, speed is estimated to be the main contributory factor in about half of all road crashes (ITF, 2018). Elvik et al. (2019) engaged in relevant research by first questioning whether speed is still as important for road safety as it was in the past, taking into account the penetration of constantly evolving vehicle safety systems in the global automotive market. After reviewing recent research studies regarding the impact of speed on road safety, they conclude that speed remains an important risk factor both for crash occurrence and for injury severity.

Furthermore, apart from speeding and distracted driving, more recent studies highlight the importance of investigating the phenomenon of harsh events in greater detail, as they have been associated with driving risk assessment, risk level correlation and classification (Bonsall et al., 2005; Gündüz et al., 2018; Vaiana et al., 2014). This is because harsh driving events, such as harsh accelerations and harsh brakings, indicate an overall aggressive and unsafe driving behavior, unsafe distance from adjacent vehicles, possible near misses, lack of concentration, increased reaction time, poor driving judgement or low level of experience. From a research scope, during recent years harsh or safety-critical events have been adopted as crash surrogate measures and the understanding of related opportunities and challenges in their interpretation is under increased examination (e.g. Tarko, 2018; Johnsson et al., 2018).

Although both harsh accelerations and harsh brakings are associated with unsafe driving behavior, they constitute two different types of events and should therefore examined as such. More specifically, harsh acceleration events may reveal high levels of anxiety and anger while driving (Stephens et al., 2009; Roidl et al., 2014) leading to a risky driving behavior characterized by drivers' involvement in situations of high risk. Harsh brakings may indicate driver struggle to anticipate the occurrence of a critical situation, which most of the time would not have occurred at the first place if it were not for driver's inattention, high speed development, inadequate distances from adjacent vehicles and other unsafe behavior indicators. As a result, harsh breaking events harsh are often used to locate safety critical events in Naturalistic Driving (ND) data (Hanowski et al., 2005; Jansen & Wesseling, 2018).

Additionally, given the strong correlation between harsh events and driving risk, it is not surprising why harsh accelerations and harsh brakings have been investigated by insurance industry in the context of usage-based motor insurance (UBI) schemes (Boquete et al., 2010; Paefgen et al., 2013), allowing for more behavioral parameters being used in UBI models. Harsh events, in combination with other driving behavioral indicators such as speeding and distracted driving are being increasingly used by Pay How You Drive (PHUD) Usage Based Insurance schemes as the critical risk factor indicators in terms of driving behavior (Tselentis et al., 2017).

1.1.3 Smartphone data exploitation

Technological advancements during recent decades have led to the development of a wide array of tools and methods in order to record driving behavior and measure various aspects of driving performance. The most common methodology applied included driving simulators (Papantoniou et al. 2014), questionnaires (Matthews et al., 1998) combined with simulators and naturalistic driving experiments (Toledo et al., 2008; Birrell et al., 2014), while the most common method of monitoring driving measures included recorders that relate to the car engine (Zaldivar et al., 2011; Backer-Grøndahl & Sagberg, 2011) and smartphones (Vlahogianni & Barmpounakis, 2017). As shown from previous research (Ziakopoulos et al., 2020), smartphones and their sensors are increasingly used as informative devices for monitoring driver behavior because they present many advantages due to high market penetration rates, Internet of Things (IOT) connectivity as well as low cost and ease of use in data collection. The exploitation of the various sensors for the purpose of transport and safety research allows for continuous, inexpensive and fast data collection, with plenty of studies confirming and even improving the reliability of smartphone measurement data implementing state-of-the-art machine learning and big data algorithms (Ghose et al., 2016, Tselentis et al., 2018).

In that environment, the high penetration rate of smartphones and social networks nowadays provide new opportunities and features to monitor and analyze driver behavior. Apart from the wide smartphone application capabilities and the low cost and ease of use in data collection, experiments under naturalistic conditions with the use of smartphones allow for drivers to be recorded under normal driving conditions and without any influence from external parameters, resulting in being considered as one of the most appropriate methods for the assessment of driving behavior.

Smartphones are equipped with a variety of sensors, such as motion sensors (e.g. accelerometer and gyroscope), position sensors (e.g. magnetometer), global navigation satellite system (GNSS) receivers, environmental sensors (barometers, photometers, and thermometers), microphones, cameras, etc. As a result, the exploitation of the various sensors for the purpose of transportation and road safety research allows for continuous, inexpensive and fast data collection, with plenty of studies confirming and even improving the reliability of smartphone measurement data implementing state-of-the-art machine learning and big data algorithms (Ghose et al., 2016; Tselentis et al., 2018). Vlahogianni and Barmpounakis (2017) examined the use of smartphones as an alternative for driving behavior analysis and they concluded that the smartphone-based algorithms may accurately detect four distinct patterns (braking, acceleration, left cornering and right cornering) with an average accuracy comparable to other popular detection approaches based on data collected using a fixed position device.

Many studies have shown promising results using data collected through smartphone sensors under naturalistic driving conditions. By conducting naturalistic driving experiments by means of mobile phone, researchers aim either at examining the effect of various driving behavior indicators on driver performance and cash risk (Tselentis et al., 2018, Yannis et al., 2017) or analyzing and modeling driver profiles (Castignani et al., 2015; Mantouka et al., 2019). These studies also investigate unsafe behaviors such as speeding (Kontaxi et al., 2023; Richard et al., 2013), mobile phone use (Papantoniou et al., 2021; Ziakopoulos et al., 2023), harsh driving events and driver aggressiveness (Frantzola et al., 2022; Precht et al., 2017) and driver fatigue (Dingus et al., 2006; Wang et al., 2022). Additionally, researchers have developed technologies and machine learning algorithms to detect these behaviors (Shahverdy et al., 2020; Shi et al., 2019; Zhang et al., 2022) and technologies that provide feedback to drivers (Braun et al., 2015; Gu et al., 2019).

1.1.4 Driver feedback in road safety studies

Feedback to drivers has been shown to be a highly effective method for enhancing road safety. Feedback itself has long been acknowledged as a powerful tool for shaping behavior in diverse areas, including education, healthcare, and human resource management (Archer, 2010; Hattie & Timperley, 2007; London & Smither, 2002). In traffic and road safety, the importance of feedback can be highlighted through several key points such as behavior modification, enhanced awareness, reduction in crash rates, stress and fatigue management, integration with advanced technologies and promotion of a road safety culture.

Many studies have examined the effect of feedback, however there is very little research that quantify the exact effect on driver behavior, as in many cases the drivers were recorded after the feedback system/mechanism had been applied, without monitoring a baseline period. A naturalistic driving experiment (Meuleners et al., 2023) was conducted for 57 car drivers with a control and intervention group for 11 weeks through a smartphone application and the drivers received a text message after the completion of the trip with personalized feedback about the participant's risky driving behavior. Four separate Generalized Estimating Equations (GEE) linear regression models were developed for each driving indicator and the results showed that the treatment effects for feedback were consistently in the expected positive direction. Another recent study (Chen & Donmez, 2022) conducted a 16-week ND experiment including 3 phases (i.e. baseline, different types of feedback, follow-up without feedback) and provided real-time and post-drive feedback to drivers.

Results showed that real-time feedback alone and in conjunction with financial incentives were effective in raising speed limit compliance. It is also interesting to note that the effects did not sustain when feedback and incentives were removed. The post feedback effect is an aspect that should be further investigated as the few studies that have dealt with the matter have not come to conclusive results (Peer et al., 2020; Soleymanian et al., 2019), while some showing both positive (Ghamari et al., 2022a) and negative (Merrikhpour et al., 2014) effects.

The methods used in studies examining the effect of driver feedback vary widely. After establishing the context and research questions, methodologies employed in these studies, are also important to be discussed. Many studies (Hari et al., 2012; Husnjak et al., 2015; Mazureck & Hattem, 2006) initially focus on basic correlation tests, presenting critical summary statistics that compare the feedback and non-feedback phases or groups. These basic statistical comparisons serve as a foundation for understanding the immediate effects of driver feedback.

However, relying solely on basic correlation tests can be limiting, as these methods do not account for the complexity and multifaceted nature of driving behavior. As a result, several studies employ these basic methods as a preliminary step before moving on to more advanced statistical modeling. For instance, (Kontaxi et al., 2021a; Toledo & Lotan, 2006) use initial correlation analyses as a data exploration phase. This phase is crucial for identifying key patterns and relationships, which then inform the development of sophisticated statistical models that can better capture the nuances of driver behavior and the impact of feedback.

Advanced methodologies, such as multivariate regression models, machine learning algorithms, and time-series analyses, are increasingly being used to understand the effects of driver feedback in a more comprehensive manner (Aidman et al., 2015; Bell et al., 2017a) . These methods allow researchers to control for various confounding factors, explore interactions between multiple variables, and predict outcomes based on complex data patterns. For example, the use of machine learning techniques like supervised learning algorithms (e.g., XGBoost) has proven effective in modeling the contributions of different driving features to the decision to engage in risky behaviors, such as mobile phone use while driving (Ziakopoulos et al., 2023).

1.2 Objectives of the Dissertation

Taking the previous into consideration, the primary objective of this dissertation is to examine the driver behavior telematics feedback mechanism. The dissertation adopts a comprehensive approach to explore feedback's role in modifying driving behavior, with a focus on three key pillars:

1. Assessing the impact of feedback on driving behavior of different road user groups (car drivers, professional van drivers, and motorcyclists) in various road environments.

2. Investigating the effects of different feedback features across the experimental phases.

3. Analyzing the long-term, post-feedback effects on driving behavior.

To achieve these objectives, a 21-month naturalistic driving experiment was thoroughly designed and implemented, involving 230 drivers across six distinct feedback phases. High-resolution data were collected using precise, non-intrusive smartphone sensors, complemented by self-reported data to provide a holistic understanding of driver perceptions and behavioral changes. A multiparametric analytical framework was employed to comprehensively examine the function of the feedback mechanism, encompassing the pre-feedback, feedback, and post-feedback phases during the naturalistic driving experiment. This approach provided valuable insights into the entire feedback mechanism, while also exploring the sustainability of feedback-induced improvements over time.

A variety of advanced statistical methods are employed to address the research objectives, including Generalized Linear Mixed Effects Models, Structural Equation Models, and Survival Analysis Models. Specifically, survival analysis techniques such as Kaplan-Meier curves, the Cox Proportional Hazards Model with Frailty, the Weibull Accelerated Failure Time Model with Clustered Heterogeneity, and Random Survival Forests are used to examine feedback effects over time. By leveraging these methods, this dissertation aims to enrich academic understanding while offering practical applications for enhancing road safety. The findings from this thesis have the potential to inform the development of more effective driver feedback systems that could significantly improve safety outcomes and enhance the overall road safety and insurance domains.

1.3 Methodology of the Dissertation

To achieve the scientific objectives of this doctoral dissertation, a series of methodological steps were carefully implemented. These steps are outlined in this subsection and visually depicted in Figure 1.2. The methodological framework provides a structured approach to achieving the objectives of this dissertation.

The methodological framework began with an extensive literature review, which explored key aspects of driver feedback systems under naturalistic driving conditions. This review focused on experimental design, feedback types, modeling approaches, and key indicators for evaluation. This phase guided the formulation of research questions, including the impact of feedback on behavior and safety, the effects of feedback features, and the post-feedback influence on long-term driver behavior.

Building on the findings of the literature review, a comprehensive methodological background was developed, combining theoretical modelling approaches and experimental design principles. This included the application of advanced modeling techniques, such as Generalized Linear Mixed Effects Models (GLMMs), Structural Equation Models (SEMs), and Survival Analysis Models, alongside the design of a naturalistic driving experiment. The experimental setup involved a cohort of 130 drivers, encompassing car drivers, van professionals, and motorcyclists, evaluated over six feedback phases using a within-subjects design.

The research utilized data from the BeSmart Research Project, focusing on key driving indicators such as speeding and mobile phone use as measures of risky behavior, and harsh accelerations and harsh brakings as proxies for safety-critical events. The analysis unfolded in three key pillars:

1. Impact of feedback: This phase assessed the immediate effects of feedback on driving behavior (e.g., speeding, mobile phone use, harsh events frequency) across the three driver groups.

2. Effects of different feedback features: A Structural Equation Model (SEM) was developed to explore the relationship between feedback features and driving behavior factors, revealing the differential impact of feedback elements.

3. Post-feedback effects: A survival analysis was conducted to examine the long-term influence of feedback mechanisms. Models such as Cox Proportional Hazards, Accelerated Failure Time (AFT), and Random Survival Forests (RSFs) were applied and compared to identify the best-fitting model.

By employing this comprehensive methodology, the driver behavior telematics feedback mechanism is effectively achieved. The integration of an extensive literature review, advanced modeling techniques, and a well-structured naturalistic driving experiment provided a robust framework to investigate the impact of feedback on driver behavior. The research focused on critical driving indicators such as speeding, mobile phone use, harsh accelerations, and harsh braking events, ensuring a holistic assessment of the impact of driver feedback on critical driver risk factors. The systematic analysis of immediate feedback effects, the different influence of feedback features, and the long-term impact of feedback mechanisms culminated in actionable insights for designing and implementing telematics feedback systems. These findings lay the foundation for a scalable and data-driven feedback mechanism aimed at improving driving behavior, reducing risky behaviors, and enhancing road safety.

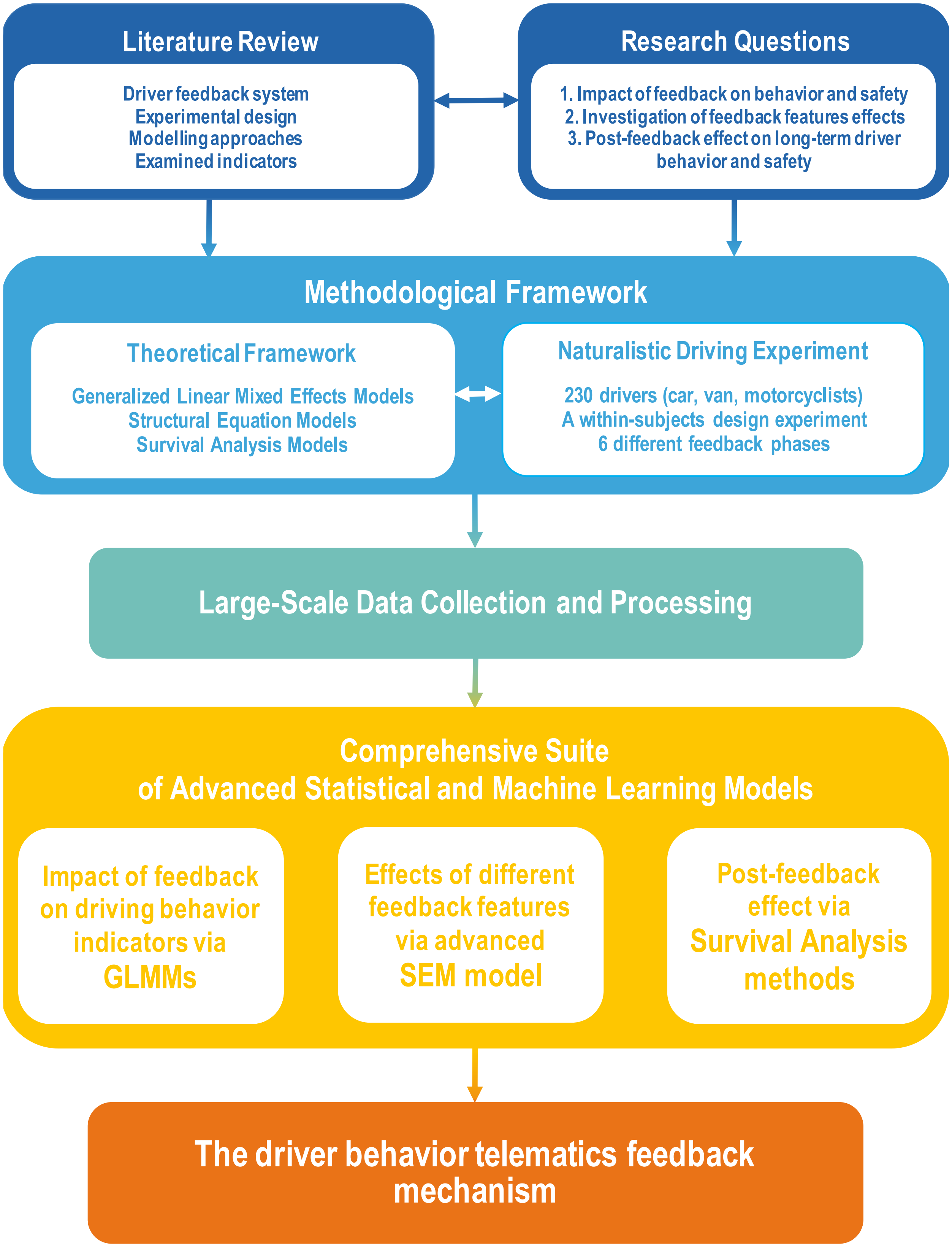


Figure 1.2: Graphical representation of the overall methodological framework of the doctoral dissertation

1.4 Structure of the Dissertation

The remainder of this doctoral dissertation is organized in nine sections which are briefly described within this subsection.

Section 2 provides a review of the scientific literature of naturalistic driving studies that investigate the effect of feedback on driving behavior. The review synthesizes the different types of feedback utilized in naturalistic driving studies and the technologies employed to deliver driver feedback. Subsequently, the methodologies used by researchers to design and evaluate the effectiveness of driver feedback (i.e. experimental design and statistical analysis) are examined, and finally evidence-based findings on the impact of feedback on driver behavior are discussed.

Section 3 describes the overall methodological framework employed to achieve the objectives of this doctoral dissertation and delves into the theoretical foundations of the analytical methods and models utilized throughout the dissertation. Specifically, it provides an in-depth overview of descriptive and inferential statistical approaches and advanced modeling techniques, such as Generalized Linear Mixed-Effects Models (GLMM), Structural Equation Models (SEM) and Survival Analysis methods. These methods are complemented by discussions on interpreting coefficients and evaluating model goodness-of-fit.

Section 4 outlines the naturalistic driving experiment conducted to achieve the objectives of the dissertation. It begins by detailing the experimental design, including the recruitment process of participants and the sample distribution and characteristics of participants drivers. The six distinct experimental phases during the naturalistic driving process are also presented. Following this, the section introduces the smartphone application used in the experiment, covering its data collection system, data processing and feature engineering, and the metadata and driver behavior indicators generated for analysis. Additionally, the structure and content of the carefully designed questionnaire are discussed, accompanied by statistical summaries of the responses. Lastly, the section addresses big data processing, focusing on data integration, organization, and the structure of the developed database.

Section 5 investigates the impact of feedback on driver behavior, focusing on speeding among motorcyclists and distraction using mobile phone use while driving in car drivers. Both analyses begin with an introduction to the respective topics, followed by descriptive statistics and preliminary analysis to provide a comprehensive understanding of the datasets. Generalized Linear Mixed-Effects Models are then employed in both cases to assess the effects of feedback on speeding and distraction, leveraging their ability to account for variability among participants and contextual factors. The findings from these models are then discussed, highlighting the shared insights into how feedback mechanisms influence driver behavior, while also addressing the specific nuances of each case.

Section 6 examines the impact of feedback on driver safety, specifically focusing on harsh braking and harsh accelerations among car drivers and professional drivers on highways. Generalized Linear Mixed-Effects Models are applied in both analyses to evaluate the effects of feedback on the frequency of harsh driving events, accounting for individual differences and contextual factors. The results from these models are discussed, emphasizing the effectiveness of feedback mechanisms in reducing unsafe driving behaviors for both driver groups.

Section 7 investigates the effects of different feedback features on driver behavior, providing a comprehensive analysis of how different feedback characteristics influence driving outcomes. The section begins with an introduction to the topic, followed by descriptive statistics and preliminary analysis to establish the context and key trends in the data. Structural Equation Models (SEMs) are then employed to explore the complex relationships between feedback features, driver behavior, exposure metrics and safety outcomes, allowing for the simultaneous analysis of multiple variables and their interactions. The results are discussed in detail, highlighting the critical feedback features that most significantly impact driver behavior.

Section 8 focuses on the post-feedback effects on long-term driver behavior, with a particular emphasis on understanding the relapse of driving behaviors following the withdrawal of feedback phases. Survival analysis methods are employed to investigate relapse patterns across various indicators, including harsh accelerations, harsh braking, speeding behavior, and mobile phone use while driving. These analyses leverage Kaplan-Meier curves, Cox-PH models with frailty, Weibull AFT models with clustered heterogeneity, and Random Survival Forests to evaluate and compare the predictive power and insights offered by each model. The results provide a detailed understanding of the sustainability of feedback-induced behavior improvements and highlight key factors influencing relapse.

Section 9 presents the conclusions of the thesis and discusses the contribution to knowledge, the limitations as well as the recommendations for further research.

Lastly, a complete list of the bibliographical references is provided.

2 Literature Review

2.1 Introduction

Over the last decade, substantial advances have been made in enhancing road safety, yet road crashes continue to pose a significant public health challenge worldwide. In 2021, road crashes led to approximately 1.19 million fatalities globally (World Health Organization; 2023.), equating to a mortality rate of 15 deaths per 100,000 people. Numerous studies have explored the key factors contributing to traffic crashes, consistently identifying human behavior as the leading cause.

Research indicates that human error is responsible for approximately 95% of road crashes (Singh, 2015) highlighting the crucial role of driver behavior in accident prevention. Understanding these behaviors allows for the development of tailored interventions that address common risky practices, such as distracted driving, speeding, and driving under the influence, which are critical to improving road safety. In this context, driver feedback appears to be an essential tool for improving driver behavior, ultimately reducing road crashes. Additionally, considering that drivers generally perceive their performance to be better than they actually drive, feedback seems a required tool that can lead to proper self-assessment on the part of the drivers (Amado et al., 2014).

2.1.1 Definition of driver feedback

Feedback has long been recognized as an effective method for influencing behavior across various fields such as education, health interventions, and human resource management (Archer, 2010; Hattie & Timperley, 2007; London & Smither, 2002). This technique has proven to be instrumental in advancing learning, enhancing task performance, and fostering positive behavioral changes (DiClemente et al., 2001; Thurlings et al., 2013).

But what exactly is meant by driver feedback? In the context of driving, driver feedback refers to information provided about driver performance, typically focused on the status of the driver-vehicle system (Donmez et al., 2007). However, the concept of feedback extends beyond merely presenting information; it inherently includes an evaluative dimension. This evaluative aspect is crucial because it provides a benchmark against which the receiver can measure their actions or performance. For example, driver feedback does not merely inform drivers about their actions, such as speeding or harsh braking patterns, but also evaluates these actions against safe driving standards or performance benchmarks.

Driver feedback can be distinguished in terms of delivery method, context and timing. Feedback may be delivered through smartphones, in-vehicle devices, dashboards, etc., providing drivers with personalized information on their driving behavior, score ranking, comparison with peers, instructions and recommendations on safe driving, as well as motivations and rewards (Feng & Donmez, 2013). Moreover, a common separation is between real-time vs. post-trip feedback (Choudhary et al., 2021; Zhao & Wu, 2012). When providing drivers with real-time feedback, they can correct their behavior instantly, e.g. get complied with the speed limits when the speed warning is activated inside the vehicle. On the other hand, post-trip feedback typically concerns a summary of the overall driving performance after the trip, allowing drivers to reflect on their behavior and gradually improve it. In the following sections, driver feedback will be thoroughly explored, through the analysis of the reviewed studies.

2.1.2 Background and knowledge gaps

Many studies have been carried out on driving behavior and naturalistic studies which mainly: (i) examine the recording of behavior and proceed to analyses and models for driver profiling (Castignani et al., 2015; Mantouka et al., 2019) or (ii) examine specific unsafe behaviors such as speeding (Kontaxi et al., 2023; Richard et al., 2020), mobile phone use (Akritidou et al., 2023; Papadimitriou et al., 2019; Papantoniou et al., 2021), aggressiveness (Precht et al., 2017), harsh events (Frantzola et al., 2022; Ziakopoulos et al., 2022a) and fatigue (Dingus et al., 2006; J. Wang et al., 2022). Additionally, an array of technologies and machine learning algorithms have been developed to identify such behaviors (Shahverdy et al., 2020; Shi et al., 2019; X. Wang et al., 2022) and provide feedback to drivers (Braun et al., 2015; Gu et al., 2019), although these studies often lack baseline comparisons.

The landscape of naturalistic driving studies has been shaped significantly by several key review papers, each contributing unique insights into driver behavior, technological integration, and psychological aspects of driving safety. A thorough systematic review on analyzing driver behavior under naturalistic driving conditions was conducted by (H. Singh & Kathuria, 2021a), in which the authors laid the groundwork for understanding driver behavior outcomes derived from naturalistic driving studies, contributing to outlining broad patterns in driver behavior. Another systematic review using the PRISMA method was conducted by (Ahmed et al., 2022) highlighting that driver behavior is the most studied topic using naturalistic driving concluding that naturalistic studies could be effectively utilized to refine driving behavior enhancement.

It is worth noting that, despite growing interest from automotive manufacturers and transportation researchers in driver behavior, limited research exists on quantifying the direct impact of driver feedback on road safety by comparing driver performance before and after receiving said feedback. A study focusing on the impact of telematics on road safety and the transformative effects it has had on driver safety behavior through different methods of incentives (Ziakopoulos et al., 2022b). While feedback was considered, it was not the exclusive focus of the study. On a related note, (Boylan et al., 2024) in their more recent review, highlight that while telematics has been effective in monitoring driving behavior and assessing insurer risk through variables like speeding, braking, and distance, the effects of telematics-based feedback on driver behavior remain largely unknown. Despite the dominance of machine learning in analyzing telematics data, there are still gaps in understanding how different types of feedback influence behavior change.

In (Michelaraki et al., 2021) a critical overview was conducted on the advancements in online gaming platforms and tools that provide safety support to drivers after their trips using different modes of transportation, such as cars, trucks, buses and trains. The research indicates that offering feedback through methods like alerts, game-like features, guidance and introducing consequences or incentives could improve driver performance and reduce the likelihood of road crashes. (Koppel et al., 2019) reviewed studies that have investigated mindfulness interventions, as potential behavioral prevention strategies among risky drivers, while other studies reviewed the role of feedback in promoting safe and eco-friendly driving behaviors (Rios-Torres & Malikopoulos, 2016; H. Singh & Kathuria, 2021b).

Similarly, (Feng & Donmez, 2013) and (D. Tselentis et al., 2020) reviewed driver feedback contextually using research from both naturalistic studies and driving simulators, with the first study proposing a cognitive model for driver-feedback interaction, offering a framework for future feedback design and empirical research, although lacking a comprehensive review on driver feedback and its impact on driver behavior and safety.

Building on this background, the current review focuses specifically on the effect of driver feedback. Addressing the research gap identified in a recent bibliometric analysis by (Yang et al., 2023), which highlighted the need for more robust evaluation of interventions in real driving conditions, this review aims to contribute to a deeper understanding of driver behavior.

Despite considerable contributions in the existing literature, no review to date has thoroughly examined the impact of feedback within naturalistic driving studies, making this review both innovative and essential. By merging insights from various research domains, the present research aims to enrich academic understanding while offering practical applications for enhancing road safety. The findings from this review have the potential to inform the development of more effective driver feedback systems that could significantly improve safety outcomes and enhance the overall road safety and insurance domains.

2.1.3 Review objective

Due to the widespread inclusion of driver feedback devices in naturalistic driving studies, a need remains in the literature for empirical and systematic investigation regarding their effects in driver safety. This review aims to:

- Synthesize the different types of feedback utilized in naturalistic driving studies and the technologies employed to deliver driver feedback.
- Examine the methodologies used by researchers to design and evaluate the effectiveness of driver feedback (i.e. experimental design and statistical analysis).
- Present evidence-based findings on the impact of feedback on driver behavior and safety.

Following the Introduction section, this section is organized as follows. Section 2.2 presents the methodology used for the synthesis of this systematic review employing the PRISMA approach. Section 2.3 showcases the results from the reviewed studies in terms of feedback types, experimental framework, modelling approaches, and ultimately the impact of driver feedback on behavior and safety. In section 2.4 the main findings are elaborated and the implications for road safety and current practices are pinpointed. Furthermore, limitations of current research alongside the directions for future research are discussed. Finally, in section 2.5 key conclusions of the current section are drawn.

2.2 Review Methodology

To gain a deeper understanding of the impact of driver feedback on road safety and considering the extensive range of studies in this area, the method of systematic review was chosen to effectively consolidate the existing literature. The present review was conducted in alignment with the Preferred Reporting Items for Systematic Reviews And Meta-analyses (PRISMA) statement guidelines (Moher et al., 2015; Page et al., 2021).

2.2.1 Search strategy and selection criteria

The sources of information for the current study were ScienceDirect, TandFonline, IEEExplore. IEEE journals, Google Scholar, and TRID. These databases were searched using keyword variation of {"(driving OR driver) AND (behavior OR safety) AND (feedback OR intervention)"} and "{(naturalistic OR telematics OR UBI OR "real conditions")}." Only peer-review academic journal and conference articles were selected for the authors to maintain control over quality of the inputs. From all the selected studies identified in the final search (n =29), backward referencing was applied to check for any other available literature that was missed during the initial search, leading to 34 studies in conclusion.

The selection of studies for inclusion in the systematic review is illustrated in Figure 2.1. A total of 597 records were identified from all the database sources using the combined keyword searches. 42 duplicate records were screened out, and the remaining articles (n = 555) were screened based on the titles and abstract, to determine whether each study fulfils the criteria to further be examined for the review. After the exclusion of 402 records that were not in the criteria list (which is shown below), 153 full text articles were assessed for eligibility. A full-text review of each included study was conducted and the following data items were extracted into a prepared data extraction sheet: reference, country of the study, sample size of the experiment, driver demographics (gender and age), experimental design including experimental groups, feedback phases and experiment duration, data collection method, means of feedback provided, timing of feedback, feedback content, monitored risk indicators, analysis method, feedback effects and post-feedback effects.

2.2.2 Inclusion and exclusion criteria

The main reasons for exclusion from the systematic review were the following:

- Driving simulator studies
- Qualitative studies / self -reported studies
- Studies focused on the technical aspects of a feedback or intervention driver system
- Advanced Driver Assistance Systems (ADAS) studies
- Autonomous driving or automated/autonomous vehicles

The scope of the present review is cautiously restricted to studies examining driver feedback under naturalistic driving conditions, in other words, real-world driving scenarios. As a result, driving simulator studies were excluded, with the review focusing exclusively on naturalistic driving studies. To elucidate, naturalistic driving studies are specifically designed to capture driver behavior under real-world conditions, eschewing any form of experimental manipulation. These studies are invaluable in providing insights into the interaction between drivers, their vehicles, and the surrounding environment during routine driving (Ziakopoulos et al., 2020). In simulated environments, drivers often behave differently than they would in real-world settings (Caird et al., 2018; Zöller et al., 2019), and it was critical to ensure that possible discrepancies would not compromise the accuracy of present findings.

There are many reasons why drivers behave differently in a driving simulator experiment; reduced risk awareness, making participants less cautious, and the unfamiliarity with simulator controls, which can feel unnatural. Psychological factors like lower arousal and the artificial environment also contribute, leading to behavior that does not fully reflect real-world driving, resulting in different results between the two study methods (Wijayaratna et al., 2019). For similar reasons, qualitative studies and self-reported data were also excluded from this review, as they lack the objective, real-world behavioral data necessary for a comprehensive assessment of the impact of driver feedback on road safety.

Additionally, the current review emphasizes a critical distinction: Since the primary objective presently is to understand the impact of driver feedback on human factors, several studies were excluded as they predominantly centered on the technological aspects of driver assistance applications or car systems. Studies focusing on Advanced Driver Assistance Systems (ADAS) tend to be more oriented towards technological details, primarily aiming to evaluate the efficacy and interaction of specific driver assistance technologies in enhancing safety. Often, they use naturalistic driving experiments to assess the impact of the driving context parameters on the design, development and evaluation of ADAS (Fleming et al., 2019; Orlovska et al., 2020). As a result, these two approaches are frequently found together in literature. By distinguishing between these two types of studies, this review fosters a more nuanced comprehension of driver interactions and responses to feedback in various contexts. Nonetheless, the reader can refer to (Furlan et al., 2020; Moujahid et al., 2018) for studying the mechanisms and effects of ADAS on road safety.

Lastly, a group of studies that was also excluded were studies regarding autonomous driving or self-driving cars. The primary reason for this exclusion is the fundamental difference between human-driven and autonomous vehicles in terms of overall behavior, driving adjustments and safety interventions. Feedback systems designed for human drivers focus on modifying behavior, while autonomous driving studies primarily investigate the technological capabilities of vehicles to operate independently of human input (Fleming et al., 2019), therefore the two groups are not directly comparable.

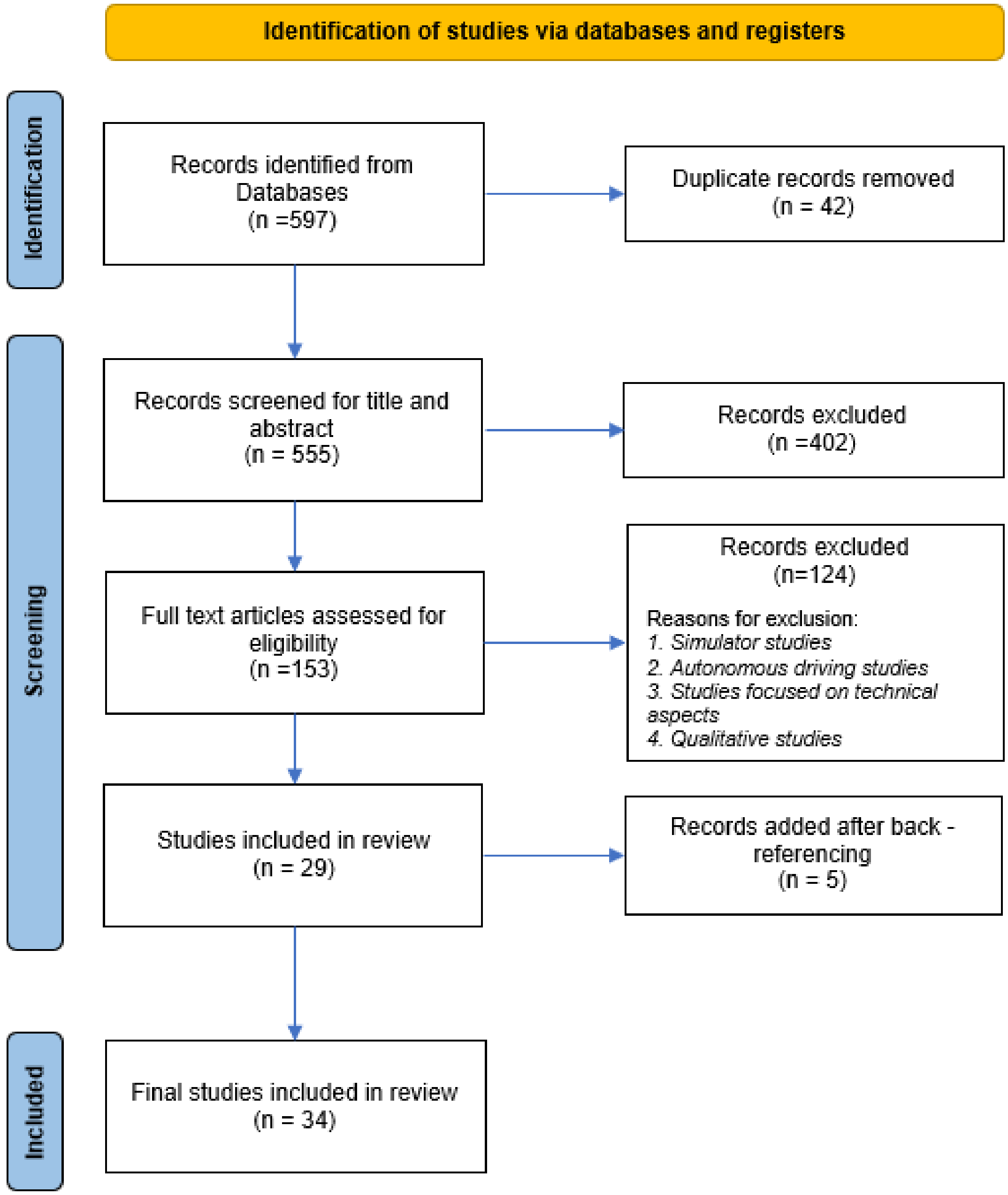


Figure 2.1: Flowchart for articles selection for systematic review

2.3 Review Findings

The findings from the systematic literature review reveal several interesting aspects of driver feedback, including the different types of driver feedback systems, the experimental frameworks used in the studies, the various analysis methods applied, and finally and most significantly, the impact of feedback on driving behavior and safety. The details of the examined studies are summarized in Table 2.1 (ordered from newest to oldest).

2.3.1 Types of driver feedback systems

The reviewed studies utilized a variety of feedback systems to examine the effect on driver behavior and road safety, each differing in how the feedback was delivered and when it was provided to drivers in relation to their trips. As illustrated in Table 1, the most widespread systems used in the experiments include (i) smartphone applications, (ii) in-vehicle systems, (iii) text messages, and (iv) personalized websites and platforms (Figure 2.2). Each type of feedback system is designed to encompass different contexts, ranging from real-time alerts to post-trip summaries, often tailored to the specific driving behavior parameter(s) being monitored during the naturalistic driving experiment.

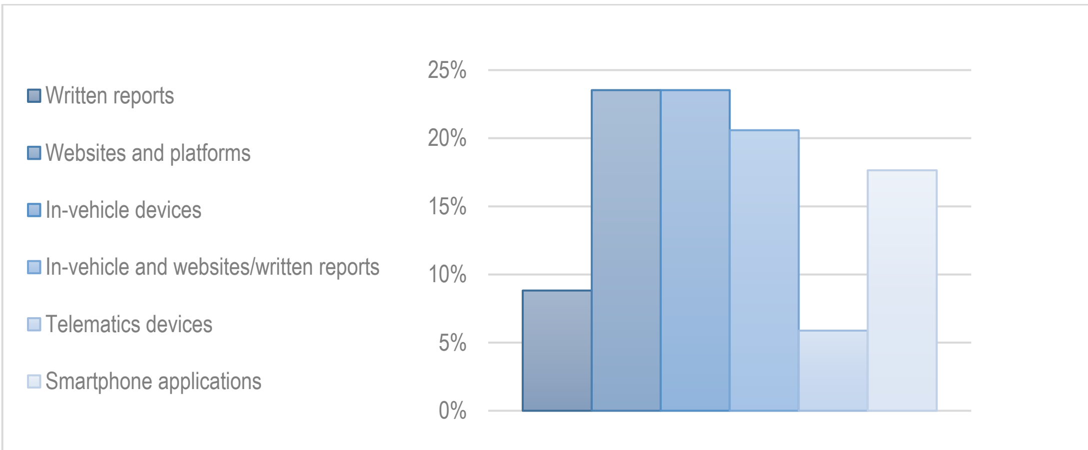


Figure 2.2: Types of driver feedback systems utilized in the reviewed studies

Furthermore, it can be observed that the types of driver feedback systems appear to evolve over time, largely driven by progress in technology, access to new platforms, and the increasing sophistication of data collection methods. In the early 2000s, feedback systems were relatively simple, focusing on in-vehicle devices and written reports, reflecting the technological restrictions of that period. Subsequently, as internet connectivity became more advanced, web-based applications started to lead at the driver feedback concept, allowing drivers and fleet managers to access detailed driving performance reports online, while more recent studies have adopted even more sophisticated feedback mechanisms utilizing smartphone applications, telematics, and machine learning algorithms, providing personalized, data driven insights. In the following paragraphs, the various types of feedback used in this review will be discussed.

Table 2.1. Systematic review table of reviewed studies

Reference	Sample size	Driver characteristics		Experimental design			Data collection	Means of feedback	Timing of feedback	Feedback content	Monitored risk indicators	Analysis method	Feedback effects	Post-feedback effects
		Gender	Age	Experimental groups	Feedback phases	Duration								
Ziakopoulos et al. 2023	176 car drivers	55% female drivers	all ages	one group / within-subjects design	• baseline • scorecard • maps • peer comparison • competitions • no feedback	21 months	smartphone application	smartphone application	post trip	personalized scorecard and gamification	speeding, harsh acceleration, harsh braking, use of mobile phone	supervised ML XGBoost algorithms and SHAP values	positive impact in reducing mobile phone use	NA
Meuleners et al., 2023	57 car drivers	44% female drivers	17-20 years old	• control • intervention	• baseline • personalized feedback	11 weeks	smartphone application	text message	post trip	personalized feedback about the participant's risky driving behavior	speeding, harsh acceleration, harsh braking	four separate Generalized Estimating Equations (GEE) linear regression models	no statistical difference in the driving outcomes examined; however, the treatment effects for feedback were consistently in the expected positive direction	NA
Molloy et al., 2023	70 car drivers	46% female drivers	18-25 years old (M=20,SD=1.68)	• control • feedback *1 • feedback*2 • feedback*3 • feedback*4	• feedback to 4 groups • feedback to 3 groups • feedback to 2 groups • feedback to 1 group • no feedback	6 months	Advanced Driver Awareness System	graphical information from the researcher.	post trip	driver's speeding performance	speeding, maximum speed	5 × 6 mixed repeated measures ANOVA and Tukey HSD (honest significant difference) post hoc test	control group showed the worst speed compliance, while feedback provided on only one or two occasions was the most effective	positive for low-speed zone (50km/h)
Ghamari et al., 2022	1,289 bus and 104 taxi drivers	all male drivers	> 20 years old	• control • intervention	• baseline • text message • no feedback	17 weeks	telematics device	text message	weekly feedback	score ranking and peer comparison	speeding, harsh acceleration, harsh braking, harsh turning	Mann-Whitney U-Test and Generalized Estimating Equations (GEE)	bus drivers: significant positive effect in stages 1 and 2. taxi drivers: significant positive effect in stage 1	positive; the reformed behavior persisted in bus drivers even after the intervention
Kontaxi et al., 2021	65 car drivers	54% female drivers	all ages	one group / within-subjects design	• baseline • scorecard	22 weeks	smartphone application	smartphone application	post trip	personalized scorecard	speeding, harsh acceleration, harsh braking, use of mobile phone	t-test analyses and Generalized Linear Mixed Models	significant decrease in instances of harsh accelerations (12%), brakings (10%), and speeding (40%) in phase 2	NA

Reference	Sample size	Driver characteristics		Experimental design			Data collection	Means of feedback	Timing of feedback	Feedback content	Monitored risk indicators	Analysis method	Feedback effect	Post feedback effect
		Gender	Age	Experimental groups	Feedback phases	Duration								
Chen & Donmez, 2021	58 car drivers	45% female drivers	all ages	• real-time only • real-time and financial incentives • real-time and post-drive feedback	• baseline • feedback • no feedback	16 weeks	in-vehicle device	in-vehicle visual feedback device and website	real-time and post-drive	speed limit compliance via the in-vehicle display and post-drive summary	speed limit compliance	linear mixed-effects model and mixed-effects beta regression	significant positive effect on real-time feedback alone and in conjunction with financial incentives, but not on post-drive feedback	negative: the effects did not sustain when feedback and incentives were removed
Stevenson et al., 2021	174 car drivers	67% female drivers	18-50 years old	• control group • driver feedback • driver feedback plus incentives	one phase / between-subjects design	28 weeks	OBD / smartphone application	text message, online dashboard or smartphone application	post trip and weekly summary	personalized scorecard	speeding, harsh acceleration, harsh braking	Generalised Estimating Equation (GEE) models and two-sided tests	composite measure of risky driving was statistically significant improved, but not speeding, harsh acceleration and harsh braking	NA
Peer et al., 2020	130 car drivers	45% female drivers	<25 years old	• app-based feedback only • app-based feedback and financial incentives	• baseline • feedback • no feedback	18 weeks	OBD / smartphone application	smartphone application	post trip	scorecard and map-based overview	speeding, harsh acceleration, harsh braking, harsh cornering, use of mobile phone	χ2 or Fisher-exact tests and OLS regressions	positive effect on safety-relevant driving behavior; higher if feedback was combined with an incentive for safe driving	due to limited data, no conclusive results
Camden et al., 2019	92 car drivers	NA	NA	one group / within-subjects design	• baseline • driver awareness • feedback	66 weeks	OBD / Geotab telematics device	online web-based instruction program	post trip	personalized web-based instructions	speeding, harsh acceleration, harsh braking, harsh cornering	Wilcoxon signed-rank tests	feedback resulted in statistically significant reductions in the rate of harsh braking by 52.17%, harsh cornering by 51.35%, and speeding by 73.93%	NA
Soleymanian et al., 2019	40,527 car drivers	48% female drivers	M: 39.3	NA	NA	26 weeks	telematics device	telematics device	immediate and daily feedback	UBI score	harsh breaking, driven distance	Fixed-Effects Regression Analysis	decrease of daily average harsh braking frequency by an average of 21%	suggestive but not conclusive positive post feedback effect evidence
Sullman, 2019	34 car drivers	80% female drivers	NA	• control • intervention	• baseline • feedback	34 weeks	in-vehicle data monitors (IVDM)	IVDM and email	real-time and weekly feedback	in-vehicle alerts and personalized feedback	speeding, harsh acceleration, harsh braking, seatbelt use	ANOVA, F-tests	significant reduction in the number of risky driving behaviors/100 km and increase in seatbelt use	NA

Reference	Sample size	Driver characteristics		Experimental design			Data collection	Means of feedback	Timing of feedback	Feedback content	Monitored risk indicators	Analysis method	Feedback effect	Post feedback effect
		Gender	Age	Experimental groups	Feedback phases	Duration								
Pozueco et al., 2017	158 car drivers	36% female drivers	all ages	• feedback based on the engaged gear • feedback based on harsh events	• baseline • feedback	Scenario #1: 3 months Scenario #2: 11 months	OBD-II interface	feedback device	real-time	types of efficiency recommendations	driving speed, rpms, acceleration	Kruskal–Wallis test	feedback during longer periods of time results in improvement in the use of recommended gears of around 9.22%	NA
Bell et al., 2017	315 professional truck drivers	NA	NA	• control • intervention	• baseline • real-time feedback • video watching with supervisors • no feedback	20 months	in-vehicle monitoring system including two camera views	small box-like device and one-on-one coaching between supervisor and driver	real-time and post trip	light indications and view and assessment of videos	aggressive driving, texting on hand-held phone, hands off the wheel, driving the wrong way, belt use	logistic regression and generalized estimating equation (GEE)	intervention group: significantly greater reduction in odds of risky driving behaviors during Coaching + IDF feedback periods, but not in IDF feedback periods separately	significant smaller decline of risky driving behavior
Toledo & Shiftan, 2016	314 car drivers	15% female drivers	18-21 years old	one group / within-subjects design	• baseline • feedback to| the lowest safety scores drivers • feedback to all drivers	12 months	in-vehicle data recorder	verbal feedback and written reports	post trip and bi-weekly	personalised score and comparison with peers	braking, lateral acceleration, speeding	Poisson regression model and two-paired t-test analysis	reduction of of 8% in safety incidents and 3–10% in fuel consumption, with a larger reduction obtained for large vehicles	NA
Aidman et al., 2015	15 car drivers	1% female drivers	21-59 years old (*M* = 41.3, *SD* = 11.1)	• feedback off / feedback on • feedback on / feedback off	• baseline • feedback	4-8 weeks	in-vehicle Optalert Alertness Monitoring System	in-vehicle dashboard display with auditory and visual warnings	real-time	drowsiness score	levels of drowsiness, drivers' own ratings of lane keeping, safe distance keeping and responsiveness	univariate linear mixed models	reduced peak drowsiness (lower maximum JDS scores), improved self-reported alertness (reduced sleepiness in KSS) and drivers' own ratings of headway	NA
Rolim et al., 2016	40 light-duty vehicles drivers	45% female drivers	M: 43.4, S.D.: 12.2	• control • intervention	• baseline • feedback	6 months	on-board device	written report	weekly feedback	personalized scorecard and fuel consumption information	speeding, engine speed, braking and acceleration	Paired t-tests	decreases between 5% and 28% in average speed, fuel cut off, excess rpm and excess speeding; negative feedback higher rates compared to positive feedback	NA

Reference	**Sample size**	**Driver characteristics**		**Experimental design**			**Data collection**	**Means of feedback**	**Timing of feedback**	**Feedback content**	**Monitored risk indicators**	**Analysis method**	**Feedback effect**	**Post feedback effect**
		Gender	**Age**	**Experimental groups**	**Feedback phases**	**Duration**								
Ellison et al., 2015	106 car drivers	58% female drivers	all ages	one group / within-subjects design	• baseline • feedback and financial incentives	10 weeks	in-vehicle GPS device	website and financial incentives	post trip	driving behavior indicators and remaining financial incentive	speeding, acceleration, braking	ANOVA analyses and multilevel regression risk model	intervention was successful in reducing speeding risk scores, and by extension, the risks of involvement in a casualty crash	NA
Merrikhpour et al., 2014	37 car drivers	46% female drivers	all ages	one group / within-subjects design	• baseline • feedback • no feedback	16 weeks	in-vehicle device	in-vehicle display and website	real-time and weekly feedback	speed limit compliance, safe headway maintenance and reward points	speeding and headway	Cluster analysis and Mixed linear models	both cluster groups improved in speeding during intervention and post-intervention phase, while only cluster B improved in headways	behavior indicators decreased slightly in the post-intervention phase, but still better than the baseline
Husnjak et al., 2014	22 car drivers	NA	NA	NA	NA	NA	telematics device	dashboard online portal	post trip	NA	speeding, harsh braking, accelerating and cornering events	summary statistics	positive impact with an average reduction of parameters related to accident risk of 38%	NA
Birrell et al., 2014	40 car drivers	25% female drivers	>21 years old	one group / within-subjects design	• baseline • feedback	50 min route	OBD II port / adapted LDW camera / smartphone	in-vehicle smart driving system	real-time	in-vehicle smart driving feedback	headway, lane departure, gear speed, acceleration and braking	ANOVA analyses	increase in mean headway to 2.3 sec and reduction in time spent traveling closer than 1.5 sec to the vehicle in front	NA
Farah et al., 2014	217 car drivers	all male drivers	17-22 years old	• control group • individual feedback • family feedback • parental training	one phase / between-subjects design	12 months	in-vehicle data recorder	web-based application and in-vehicle display	real-time and post trip	driving aggressiveness level and summary report on trip information and events	events rates	One-way ANOVA analysis	significant improvement of the driving behavior by means of events rates	NA
Newnam et al., 2014	16 work-related car drivers	44% female drivers	M= 45.0	one group / within-subjects design	• baseline • feedback • no feedback	5 weeks	OBD II	written report	weekly feedback	personalized feedback on speeding, comparison with peers and goal setting	speeding	A Wilcoxon Signed Ranks Test	the majority of drivers reduced their overall number of over-speed violations from pre to post intervention	suggestive but not conclusive positive post feedback effect evidence

Reference	Sample size	Driver characteristics		Experimental design			Data collection	Means of feedback	Timing of feedback	Feedback content	Monitored risk indicators	Analysis method	Feedback effect	Post feedback effect
		Gender	Age	Experimental groups	Feedback phases	Duration								
Reagan et al., 2013	50 car drivers	48% female drivers	M=27.8	• control group • driver feedback • driver feedback plus incentives	• baseline • feedback • no feedback	4 weeks	in-vehicle device	alerting system	real-time	auditory and visual advisory signals when speeding	speeding	A series of 3 × 4 mixed factorial ANCOVAs	incentive system resulted in significant reductions in speeding, while the feedback system led to modest changes	NA
Strömberg et al., 2013	54 bus drivers	NA	NA	• access to system & training • access to system • control group	• baseline • feedback	6 weeks	in-vehicle device	in-vehicle system and training	real-time and post trip	information histograms on driving behavior indicators and fuel consumption	speeding, idle, rollout, harsh deceleration	A standard F-test for within-subject design and a Kolmogorov–Smirnov two-sample one-tailed test	feedback led to reduced frequency of harsh decelerations, speeding and 6.8% reduction in fuel consumption and	NA
Hari et al., 2012	15 light commercial vehicles	NA	NA	one group / within-subjects design	• baseline • feedback	4 weeks	OBD / CAN	device instrument panel display	real-time	two sets of light emitting diodes (LEDs) showing short term and long-term driver performance	engine speed, acceleration and inertial power surrogate	summary statistics and cumulative probability distribution	reduction in harsh accelerations and early gear shifting into higher gears	NA
Takeda et al., 2012	33 car drivers	NA	NA	• control • intervention	• baseline • feedback	two 1.5 h driving sessions	continuous data recorders / front-viewed camera	web application	post trip	illustration of hazardous situations on an actual driving map and summary statistics	speeding, harsh deceleration, harsh acceleration, risky steering, following distance, ignoring a traffic light	summary statistics	the number of detected hazardous scenes decreased by approximately 50% for the non-expert drivers	positive effect but not reliable results due to very small sample
Hickman & Hanowski, 2011	100 truck drivers	all male drivers	M=47.0	• control • intervention	• baseline • feedback	17 weeks	On-board monitoring devices (two video cameras and three accelerometers)	in-vehicle system and training	real-time and post trip	feedback light on the OBSM device when an event was detected and interactions with fleet safety managers	safety–critical events	paired sample t-tests	significant reduction in the mean rate of safety-related events/10,000 miles traveled from the Baseline to Intervention phases; group A: 39.9%, group B: 52.2%	NA

Reference	Sample size	Driver characteristics		Experimental design			Data collection	Means of feedback	Timing of feedback	Feedback content	Monitored risk indicators	Analysis method	Feedback effect	Post feedback effect
		Gender	Age	Experimental groups	Feedback phases	Durati on								
Bolderdijk et al., 2011	141 car drivers	40% female drivers	M=24.4, SD=2.2	• control • intervention	• baseline • feedback and reward • feedback and reward • no feedback	8 weeks	in-vehicle GPS device	personalized website	post trip	speed violations, mileage, nighttime driving, and financial inventives in a form of discount	speeding	mixed design through Huynh–Feldt tests	PAYD is estimated to have reduced volitional speeding by 14%	the incentive group increased speeding when the financial incentive was removed
Farmer et al., 2010	84 car drivers	55% female drivers	16-17 years old	• alert and web feedback • alert then web feedback • web only • control	• baseline • feedback • no feedback	24 weeks	in-vehicle device	in-vehicle alert and website	real-time and post trip	audible alerts immediately following each event / website: mapping of the location and nature of each event	speeding, harsh braking, harsh acceleration, seat belt nonuse	Poisson regressions	improvements in seat belt use when parents were informed, and even greater when in-vehicle alerts were activated; reductions in speeding only in-vehicle alerts were activated	performance decreased during the post-feedback phase but still better than baseline; but no statistically significant changes
Toledo et al., 2008	191 car drivers	2% female drivers	M=41.0	one group / within-subjects design	• baseline • feedback	9 months	in-vehicle data recorder	website	post trip	summary reports on driving behavior	speeding, lane changing, harsh acceleration, harsh braking and harsh turning	ANOVA analyses	significant reduction of 38% in crash rates, but not in fault crash rates	Positive effect but not reliable results due to no sufficient data
McGehee et al., 2007	26 car drivers	54% female drivers	16-17 years old	one group / within-subjects design	• baseline • feedback • no feedback	12 months	event-triggered video recording system	two LED lights on the face of the recorder and parent/ teen mentoring sessions	real-time and post trip	real-time alert when threshold value was exceeded / weekly report card on unsafe behaviors and seatbelt use	speeding, harsh acceleration, harsh braking, following distance, signs, belt use, near crash and crash events	t-tests	significant decrease in events for the more at-risk teen drivers	high-frequency drivers maintained a low number of safety-relevant events
Mazureck & Hattem, 2006	62 car drivers	2% female drivers	M=47.0	one group / within-subjects design	• baseline • feedback	24 weeks	in-car technology system	in-vehicle dashboard and website	real-time	driver points based on driving behavior	speeding and headway	summary statistics	significant increase of the percentage of kilometers traveled within the speed limit (26%) driven a safe distance from the car in front (33%)	most drivers returned to their old habits after the feedback phases

Reference	Sample size	Driver characteristics		Experimental design			Data collection	Means of feedback	Timing of feedback	Feedback content	Monitored risk indicators	Analysis method	Feedback effect	Post feedback effect
		Gender	Age	Experimental groups	Feedback phases	Duration								
Toledo & Lotan, 2006	33 car drivers	NA	NA	one group / within-subjects design	• baseline • feedback	5-6 months	in-vehicle data recorder	personal web pages	post trip	records on personalized behavior and comparison with peers	speeding, lane changing, harsh acceleration, harsh braking and harsh turning	summary statistics and linear regression	the average driving risk indices reduced by 38% in the first month that feedback was provided	by the 5th month, driving risk indices were back to the initial values and even slightly higher
Wouters & Bos, 2000	840 different types of vehicles	NA	NA	• control • intervention	• baseline • feedback	24 months	accident data recorder / OBD	provided by the fleet owners	post trip	NA	accident rates	summary statistics	average estimated accident reduction of some 20%	NA

2.3.1.1 Smartphone application

Most of the reviewed recent studies utilized smartphone applications to deliver feedback to drivers (Kontaxi et al., 2021b; Meuleners et al., 2023; Peer et al., 2020; Stevenson et al., 2021; Ziakopoulos et al., 2023). Smartphones are equipped with a wide range of sensors like accelerometers, gyroscopes, GNSS receivers, and cameras, and thus provide an effective way to collect driving data for analysis and feedback. Once driving is detected, sensors data are collected, processed using advanced machine learning and big data techniques, and meaningful driving indicators are generated and subsequently analyzed. After this procedure, smartphone applications typically provide personalized scorecards or driving performance information after trips regarding safety indicators such as speeding, harsh events and mobile phone usage while driving.

The design of these applications may vary; for example, (Meuleners et al., 2023) and (Stevenson et al., 2021) used an existing telematics application with a 0-5 scoring system, featuring a color-coded scale, namely green code for safe (no-risk driving), amber for some at-risk driving and red for at-risk driving. From a different perspective, (Kontaxi et al., 2021b) implemented a customized app with a score range of 0-100, alongside a 5-star system.

One of the key advantages of using smartphone-based feedback in driving behavior experiments is the accessibility and convenience it offers. This method is low-cost once the back-end is established, and it enables the easier engagement of participants in the study. While smartphone applications do have limitations – such as noisy camera-based indices or high battery consumption – ongoing advancements in technology (e.g. real-time edge computations) can significantly enhance the use of smartphone-based naturalistic driving studies in research (Grimberg et al., 2020). Context-wise, smartphone applications used in the examined studies included a variety of feedback features; from personalized scorecards and map-based overviews to gamification and comparison with peers. In particular, (Ziakopoulos et al., 2023) found that the variety of feedback context across different experimental phases significantly influenced mobile phone use, with each different feature of feedback having a distinct impact on driver behavior. The present finding suggests that further investigation into the impact of different feedback features is necessary.

2.3.1.2 In-vehicle devices

Feedback delivered through in-vehicle systems was another widely used method. These devices have become an integral part of modern driver behavior monitoring systems, also offering real-time and post-trip feedback aimed at enhancing road safety. In-vehicle feedback systems typically use in-vehicle data monitors/recorders or OBD (I or II) for driver data collection (Amarasinghe et al., 2016; Pérez et al., 2010). Equipped with sensors and advanced communication technologies, they track various aspects of driving behavior, such as speed, braking, acceleration, and seat belt use. Over the years, in-vehicle systems have evolved from simple alert mechanisms to more sophisticated tools that offer detailed visual, auditory, and sometimes even haptic feedback. The devices vary in the manners in which they provide drivers with feedback in terms of both timing and content. For example, (Chen & Donmez, 2021) used an in-vehicle visual feedback device for driver speed compliance, while other studies employed auditory feedback; either through verbal prompts via the in-vehicle device (G. Toledo & Shiftan, 2016) or alerts when drivers engaged in risky behaviors such as speeding, harsh events, seat belt use, etc. (Sullman, 2020). Some in-vehicle devices combined auditory and visual signals to alert drivers to unsafe behaviors (Aidman et al., 2015; Reagan et al., 2013).

It is noteworthy that in-vehicle feedback devices primarily focus on delivering real-time feedback, which is one of the main advantages of in-vehicle data monitors, particularly when compared to smartphone applications (Grimberg et al., 2020). However, in many studies, feedback was also provided post-trip in the form of written reports or websites, offering summary reports on driving behavior indicators in various formats, i.e. histograms (Strömberg & Karlsson, 2013), via mapping of the location and nature of each event (Farmer et al., 2010), personalized score/reward points (Farah et al., 2014; Mazureck & Van Hattem, 2006; Merrikhpour et al., 2014) or even comparison with peers (Toledo & Shiftan, 2016).

2.3.1.3 Websites and platforms

Other studies have utilized a variety of platforms to deliver personalized feedback to drivers, including websites (Bolderdijk et al., 2011; Ellison et al., 2015; Takeda et al., 2012; Toledo et al., 2008; Toledo & Lotan, 2006), a dashboard online portal (Husnjak et al., 2015), and written reports (Newnam et al., 2014; Rolim et al., 2016). The website used in (Ellison et al., 2015) and the dashboard online portal in (Husnjak et al., 2015) were similar in terms of logging, feedback display, and providing participants with information about their driving behavior, as well as tracking their remaining incentives, as both studies also examined the effects of monetary incentives on driving behavior alongside driver feedback. An interesting aspect of these websites was that researchers could use the number of logins as a proxy for participants' exposure to the provided information (such as frequency of risky behaviors and remaining incentives), allowing for safer and more accurate analysis results.

In some cases, feedback was provided in graphical form, directly from researchers, allowing drivers to visually assess their driving behavior (Molloy et al., 2023). Additionally, certain feedback systems offered recommendations or instructions on how to improve driving behavior (Camden et al., 2019). More precisely, in (Camden et al., 2019), apart from providing feedback to the participant, a Geotab device wirelessly transmitted data on risky driving behaviors to the Predictive Coach program, which automatically assigned educational courses targeting specific behaviors once a driver exceeded a predefined threshold. For example, if a driver exceeded the speed limit three times in a single day, they were assigned an online course designed to reduce speeding, with alerts sent to both the driver and their manager to ensure course completion within seven days.

2.3.1.4 Telematics devices

Furthermore, telematics devices have also been used for delivering driver feedback and examining its impact on driver safety. Both (Soleymanian et al., 2019) and (Ghamari et al., 2022) leveraged telematics for looking deeper to understanding the role of feedback through score ranking and Usage Based Insurance (UBI) score, aiming to understand how feedback influences driver behavior. As noted in the introduction, driver telematics have been extensively employed over the past two decades, yet only a small number of studies have quantitatively examined the impact of feedback on driver safety through telematics systems (Boylan et al., 2024). Data protection issues are a frequent reason why insurance companies are reluctant to give researchers access to large datasets collected from their customers (McDonnell et al., 2021). The lack of access makes it difficult to examine feedback mechanisms in-depth or perform more comprehensive patterns analyses on the broader demand upon driver safety. Consequently, many telematics studies require collaboration with the insurance industry or are limited to small custom-built systems and findings can be more difficult to generalize across larger populations.

2.3.2 Experimental framework

The importance of the experimental framework of each feedback-oriented study has been emphasized in scientific research, as it serves the foundation for producing valid, reliable and interpretable results (Cash et al., 2016; Leik, 1997). In the context of naturalistic studies, a well-constructed experimental framework allows researchers to isolate the specific effects of various feedback mechanisms on driving behavior (van Schagen & Sagberg, 2012). In the present review, the experimental framework of considered studies is examined across the following key aspects: sample characteristics (size, gender balance and age), experimental design, risk indicators, feedback phases (i.e., number and content of phases), and the experimental duration.

2.3.2.1 Sample characteristics

The sample size is crucial in an experiment for ensuring reliability and generalization of the results. The reviewed studies varied widely in terms of the sample size, including samples from small groups of 15 drivers (Aidman et al., 2015) to large-scale studies with over 40,000 drivers (Soleymanian et al., 2019). This variation seems logical for two main reasons: first, the recruitment in such experiments can be challenging and second, because each study has its own objectives, research questions and often target groups. Nevertheless, larger sample sizes, such as (Pozueco et al., 2017) with 158 drivers, offer a higher statistical power and potentially more reliable conclusions.

As for driver characteristics, the reviewed studies demonstrate a wide range of gender representation and age groups. More precisely, the gender composition varies, with some studies having a balanced representation of both males and females, such as (Ziakopoulos et al., 2023) with 55% of the participants being female, and (Ellison et al., 2015) with 58% female participants, while others focus entirely on male drivers (Farah et al., 2014; Ghamari et al., 2022). This variability could affect result generalizability, as male and female drivers may respond differently to feedback interventions. It is worth noting that the representation of female drivers seems to be inconsistent among the studies, ranging from as low as 1% in (Aidman et al., 2015) to 67% in (Stevenson et al., 2021).

In terms of age, many studies are related to a specific age group, mostly young drivers who remain a group that is globally linked with a higher probability of crash risk (A. F. Williams, 2003). For example, (Farmer et al., 2010; Meuleners et al., 2023) investigated the impact of feedback on young drivers aged 13-22, while other studies, like (Hickman & Hanowski, 2011) examined older drivers (mean age of 47-year-old). The variety in both gender and age makes it possible to draw conclusions regarding the way different demographics respond to feedback, but further analyses with more inclusion of sample characteristics would improve understanding of how generalizable these findings are across a diverse population of drivers.

2.3.2.2 Experimental design

Two primary experimental designs emerge from the reviewed studies: within-subjects design and between-subjects design. Within-subject design is frequently used in studies where all participants experience multiple phases of feedback and act as their own control group (Birrell & Fowkes, 2014; Hari et al., 2012; Kontaxi et al., 2021b; Merrikhpour et al., 2014; Newnam et al., 2014; Toledo & Shiftan, 2016; Ziakopoulos et al., 2023). The key advantage of this design is the reduced variability and increased statistical power due to each subject acting as their own control. In other words, since each driver is exposed to both baseline (pre-feedback) and feedback phases, researchers can directly detect the effects of the feedback mechanisms on driving behavior, reducing the variability related to individual differences. However, these studies are not without concerns that must be addressed. For instance, carryover effects and fatigue effects pose significant risks. In longer studies, behavior in one phase might influence behavior in subsequent phases (carryover), while fatigue can cause participants to become tired over time, impacting their performance in later phases (Nichols & Maner, 2008).

Certain studies (Farah et al., 2014; Reagan et al., 2013; Rolim et al., 2016; Stevenson et al., 2021) adopted the between-subjects design, where different groups of drivers received different types of feedback or no feedback at all (control group) through one experimental phase, allowing researchers to examine and analyze driving behavior across different conditions of feedback influence. The main advantage of this design is the reduced risk of carryover effects, as each participant is exposed to only one condition, allowing for clearer comparisons across feedback conditions without the cofounding influence of multiple phases. However, this design comes with some concerns as well, the most critical being the fact that the design requires a larger sample size to ensure that the differences across the groups are not due to chance, as variability across individuals can musk true effects (Charness et al., 2012).

However, most of the studies in the reviewed literature utilized a combination of both within-subjects and between-subjects designs to maximize the strengths of each approach (Bell et al., 2017; Bolderdijk et al., 2011; Chen & Donmez, 2021; Farmer et al., 2010; Ghamari et al., 2022; Hickman & Hanowski, 2011; Molloy et al., 2023; Pozueco et al., 2017; Strömberg & Karlsson, 2013). The combination can range from simple control/intervention group or baseline/feedback phase distinctions to more complex designs. For example, Chen & Donmez (2021) compare three feedback types: (i) real-time only, (ii) real-time with financial incentives, and (iii) real-time combined with post-drive feedback, across three phases: (i) baseline, (ii) feedback, and (iii) no feedback, enabling the evaluation of the effects of various feedback mechanisms on driver behavior. Such a design allows researchers to control for individual differences (as in within-subjects designs) while at the same time avoiding risks like carryover effects and fatigue through the use of between-subjects design.

In addition, this mixed type of approach allows for more sophisticated analyses, enabling researchers to capture both group-level differences and individual changes over time, providing a more comprehensive understanding of feedback mechanisms. There are additional nuances as well. Indicatively, (Molloy et al., 2023) took a quite different approach, aiming to examine the effect of feedback frequency. The experimental design utilized a 5 × 6 mixed repeated measures structure. The feedback variable was the only between-groups factor, consisting of five levels: (i) a control group with no feedback, (ii) feedback provided once, (iii) feedback provided twice, (iv) feedback provided three times, (v) feedback provided four times. Results revealed an interesting finding: that the most effective feedback frequency was when it was provided once or twice.

2.3.2.3 Monitored risk indicators

According to the reviewed studies, the most common monitored indicators include speeding, harsh braking, harsh acceleration, and use of mobile phone while driving. These driving parameters are frequently selected in road safety studies, as they constitute some of the most critical risk factors contributing to road crashes that can be measured in real-time and serve as basis for feedback provision. For example, speeding (both considering the time spent driving above speed limits and the degree to which those limits were exceeded) has been highly correlated with not only the frequency of road crashes, but their severity as well (Aarts & Van Schagen, 2006). Harsh events are also linked with aggressive and stressful behavior (Stephens & Groeger, 2009), while they have also been used as surrogate safety measures in historical crash investigations (Nikolaou et al., 2023). The literature has also pinpointed mobile phone use as a major cause of distracted driving, which can lead to a high likelihood of involvement in road crashes (Caird et al., 2008). Apart from being critical risk factors, these behaviors can be also effectively monitored via naturalistic driving experiments and potentially improved through driver feedback.

A few studies also track more specific behaviors, such as seatbelt use (Farmer et al., 2010; Sullman, 2020) and drowsiness levels (Aidman et al., 2015). This diversity in monitored indicators ensures that a broad spectrum of risky driving behaviors has the potential to be addressed, from aggressive driving to minor traffic violations. However, core behaviors like speeding and harsh events (e.g., acceleration, braking) remain a priority across most studies due to their strong correlation with crash risks.

2.3.2.4 Feedback phases

As shown in Table 1, feedback is often administered across multiple phases, which can range from baseline only to more complex designs with personalized scorecards, competitions or financial incentives. As per the aforementioned, between-subjects design studies used only one phase and examined feedback between different groups of participants. Apart from those studies, other studies utilized two phases: baseline and a single phase of feedback (Farmer et al., 2010; Rolim et al., 2016). These studies typically focused on immediate behavior changes post-feedback without considering how prolonged feedback or multiple interventions affect long-term behavior.

In that context, (Meuleners et al., 2023) created two phases in their experiment: (a) a three-week baseline period with no intervention and (b) an eight-week intervention period for the intervention group, while the control group did not receive any feedback on their driving behavior throughout the study. However, the majority of reviewed studies used multiple feedback phases from three to six phases. For instance, (Ziakopoulos et al., 2023) included phases such as scorecards, maps and highlights, peer comparisons, and competitions to measure cumulative effects on driving behavior. These phases allow for a more nuanced understanding of how different types of feedback affect drivers' behaviors over time and under various contexts.

Furthermore, it should be noted that about a third of the total studies introduced a final phase with no feedback, returning drivers to the baseline phase, as they started. The inclusion of this phase makes it possible to examine the post-feedback effect on drivers, i.e. whether there is a sustained positive impact (if any) on their behavior after receiving feedback for one or more phases in the experiment. Did the participants maintain their shift to safer behavior with reduced risks, or did drivers regress back to risky behaviors after an initial improvement phase? Studies with no feedback phase at the end of the experiment are crucial for determining the long-lasting effects of feedback mechanisms and whether they need to be periodically refreshed to maintain their effectiveness.

2.3.2.5 Experimental duration

Regarding the duration of each experiment, it is observed that the length of the experiment varies widely across studies, ranging from just a few weeks to several months or even years. For instance, (Aidman et al., 2015; Hari et al., 2012; Hickman & Hanowski, 2011) only provided feedback over a few weeks. It is important to underline that all these studies examined real-time feedback and did not examine post-feedback effects. This means that feedback duration appears to be correlated closely with the structure of the experiment design, and thus influenced by researcher limitations. Short-term studies are ideal for examining immediate reactions to feedback. A significant number of the studies, such as (Toledo & Lotan, 2006) and (Rolim et al., 2016), used feedback periods that spanned several months. These studies are better suited to understanding behavior changes that evolve over a longer time span, as they provide insights into how drivers adapt to feedback beyond the initial phase. On the contrary, studies like (Ziakopoulos et al., 2023) (21 months) and (Wouters & Bos, 2000) (24 months) showed the most extensive feedback periods, being able to determine the long-lasting effects of feedback mechanisms.

Nevertheless, the length of a naturalistic driving experiment surely depends on a variety of factors, such as study objectives, intervention style and evidently practical constraints that are not in the researchers' direct control. On that note, an interesting finding from the reviewed studies is that the baseline period is usually shorter than the feedback phase(s), giving the participants enough time to adjust their behavior and observe any differences between the phases.

2.3.3 Modelling approaches

The reviewed studies utilized various modelling approaches to assess the impact of driver feedback, as shown analytically in Table 1. In summary, it can be observed that most analytic methods focus on assessing the correlations between driving-related factors before and after the feedback is provided, in order to draw some conclusions on the effect of driver feedback, following methods established in traffic psychology. The employed methods include both simpler approaches, such as correlation tests and more complex statistical methods that delve deeper into the driver feedback mechanisms.

2.3.3.1 Choice of analysis method

It is important to note that the choice of analysis method used in each study depends on the experimental design (within-subjects design vs. between-subjects design). For example, within-subjects design studies (Birrell & Fowkes, 2014; Hari et al., 2012; Kontaxi et al., 2021b; Merrikhpour et al., 2014; Newnam et al., 2014; Toledo & Shiftan, 2016; Ziakopoulos et al., 2023) frequently use repeated measures and mixed-effects models to account for the multiple measurements taken from the same subjects, employing techniques such as ANOVA, mixed ANOVA, Generalized Estimating Equations (GEE), Wilcoxon signed-rank tests, and paired t-tests.

In contrast, between-subjects design studies (Farah et al., 2014; Reagan et al., 2013; Rolim et al., 2016; Stevenson et al., 2021) often use methods that compare means or distributions between two independent groups, such as T-tests, Mann-Whitney U-Tests, ANOVA, ANCOVA, logistic regression, and Poisson regression (for binary and count data, respectively).

Certainly, most of the studies (Bell et al., 2017; Bolderdijk et al., 2011; Chen & Donmez, 2021; Farmer et al., 2010; Ghamari et al., 2022; Hickman & Hanowski, 2011; Molloy et al., 2023; Pozueco et al., 2017; Strömberg & Karlsson, 2013) utilize a mix of both within-subjects and between-subjects designs depending on the specific setup of feedback phases using linear and logistic regression models, mixed-effects models, and cluster analysis.

2.3.3.2 Basic and exploratory analysis methods

Having established the experimental designs, the methodologies used in the reviewed studies can be explored. In terms of analysis, it is evident that several studies focus solely on basic correlation tests, displaying the critical summary statistics between the feedback and non-feedback phase or groups to assess the impact of driver feedback. Although these methods are relatively simple, several earlier studies relied on descriptive statistics for a clear understanding of the effects of feedback on driver behavioral indicators or safety (Hari et al., 2012; Husnjak et al., 2015; Mazureck & Van Hattem, 2006; McGehee et al., 2007; Rolim et al., 2016; Takeda et al., 2012; Wouters & Bos, 2000). For instance, (Mazureck & Van Hattem, 2006) examined speeding and headway during a baseline and a feedback phase comparing the two phases to observe any change due to driver feedback, concluding to quantified portions of the percentage of kilometers travelled within the speed limit and the number of kilometers driven a safe distance from the car in front.

In contrast, in other studies, these methods are often used as an initial data exploration phase preceding the creation of more sophisticated statistical models (Kontaxi et al., 2021b; Toledo & Lotan, 2006). Simple statistical summaries between feedback and non-feedback phases or groups can highlight trends, to be later examined more rigorously using advanced statistical techniques.

2.3.3.3 ANOVA and test hypothesis

The Analysis of Variance (ANOVA) and hypothesis testing are well-known techniques involving the comparison of means and variances between groups to test hypotheses. Within studies concerning driver feedback, ANOVA tests are used to compare multiple groups (Ellison et al., 2015; Molloy et al., 2023; Sullman, 2020; Toledo et al., 2008), while t-tests compare two groups (Hickman & Hanowski, 2011; Kontaxi et al., 2021b; McGehee et al., 2007; Rolim et al., 2016; Toledo & Shiftan, 2016). Advancing one step further, Post-hoc tests like Tukey HSD and Huynh–Feldt tests are performed following ANOVA to adjust for multiple comparisons and sphericity, respectively (Bolderdijk et al., 2011; Molloy et al., 2023). In that context, (Reagan et al., 2013) implemented a series of 3 × 4 mixed factorial ANCOVA, a type of ANOVA with controlling linear effect of covariate variables by using regression analysis in the aim of examining the different types of feedback provided. The covariates were used to control for driving exposure, included measurements of miles driven over the 4-week trials, considering the respective speed limit zones in the study analysis.

On the other hand, non-parametric tests are used when data does not meet the assumptions of parametric tests, such as normal distribution. Tests like the Mann-Whitney U test were used by (Ghamari et al., 2022) to compare feedback phases in terms of score ranking, while Wilcoxon signed-rank were employed to examine the differences of speeding (Newnam et al., 2014) during the different experimental phases, or other monitored risk indicators, e.g. harsh braking and cornering (Camden et al., 2019). In addition, (Pozueco et al., 2017) applied the Kruskal-Wallis test to compare performance between sub-groups in terms of gender and age, providing interesting findings regarding the effect of demographics on driver feedback. (Strömberg & Karlsson, 2013) used a Kolmogorov– Smirnov two-sample one-tailed test to examine comparisons of the effects of feedback in terms of idle time (%), over speed (%), harsh decelerations/100 km, time using bus momentum to roll out (%) and fuel consumption (liter/100 km). In this case, the researchers used the non-parametric test, because of the small number of data points.

2.3.3.4 Regression and Machine Learning approaches

Regression analysis methods are often used to understand relationships between variables, to accommodate different data structures, and to handle repeated measures or correlated data. Many of the reviewed studies have implemented regression models varying of course in the form of regression depending on the data distribution or even the response variable and the research questions. From simple regression models, such as ordinary least squares (OLS) in order to determine the factors affecting the score differences among the 4 different experiment phases (Peer et al., 2020); linear regression to calculate the individual risk index (Toledo & Lotan, 2006), to more complex regressions like linear mixed effects models to examine the effect of feedback as well as demographic characteristics on driving indicators (Chen & Donmez, 2021; Merrikhpour et al., 2014).

Within a regression framework, (Soleymanian et al., 2019) deployed fixed-effects regression analysis to explore the number of daily hard brakes and capture the effect of negative signal on driving performance, while (Aidman et al., 2015) developed univariate linear mixed models to investigate drowsiness and performance metrics across the two different experimental conditions and identify whether there were any improvements during the feedback phase. Additionally, Poisson regression, often used for count data, has been selected as an appropriate analysis method when the dependent variable has been treated as rare unsafe driving behaviors or incidents in terms of rates, like in (Farmer et al., 2010). While (Kontaxi et al., 2021b) deployed Generalized Linear Mixed-Effects Models to model the frequencies of harsh events during baseline and feedback phase and examine the differences. It is critical to note that regression methods often face caveats such as endogeneity, multicollinearity and the assumption of normalcy, which researchers must anticipate and mitigate, to the extent that is possible within their study designs.

Generalized Estimating Equations (GEE) are typically used when analyzing repeated measures with non-normal response variables. Linking that asset with the determined studies, studies that examined risk indicators in terms of driver performance scores/levels employed GEE models (Bell et al., 2017; Ghamari et al., 2022; Meuleners et al., 2023; Stevenson et al., 2021). For example, (Bell et al., 2017) used GEE to identify the differences between groups accounting for the same vehicles over time, while (Stevenson et al., 2021) used GEE models to account for repeated measures at the participant level. In the same context, (Meuleners et al., 2023) employed GEE to control for driving exposure. After conducting the GEE models, many studies also used binary logistic models to investigate the overall effect/score of feedback in a yes/no form of context.

Finally, a recent study by (Ziakopoulos et al., 2023) applied supervised Machine Learning methods, namely XGBoost algorithms, to create classifiers for predicting drivers´ decisions related to the use of mobile phone while driving as function of naturalistic driving parameters, questionnaire features and the variable of feedback phase. Additionally, SHAP values were computed to highlight for the most relevant features, increasing that way the model explainability.

2.3.4 Impacts of driver feedback on driving behavior and road safety

The reviewed studies reveal that feedback has a generally positive impact on driving behavior and safety, particularly when targeting specific risky driver behaviors like speeding, harsh events, and mobile phone use or when calculating driver crash risk based on the monitored driving behavior parameters.

2.3.4.1 Impact on speeding behavior

Speeding was the most monitored and analyzed risk indicator in the reviewed studies. Feedback resulted in reductions in speeding behavior, specifically in terms of the overall duration of time spent speeding during the trip, with values ranging from 5% as much as 74%. For example, (Mazureck & Van Hattem, 2006) reported a 26% reduction, (Kontaxi et al., 2021b) found a 40% decrease in the feedback phase where participants received a scorecard regarding their behavior, and (Camden et al., 2019) observed a remarkable 74% reduction in speeding instances. Some studies, such as (Ellison et al., 2015), did not quantify the reduction in percentages but did ascertain via statistical modeling that feedback reduced the speeding risk scores that participants received via personalized websites.

An interesting finding from the reviewed studies is that feedback combined with financial incentives further enhances these positive effects. For example, (Chen & Donmez, 2021) reported that real-time feedback with financial incentives significantly improved speed limit compliance, particularly in the highway where typically higher speeds are observed, and consequently higher speeds above the speed limit. Similarly, (Bolderdijk et al., 2011) demonstrated that when participants received financial incentives in the form of discounts, along with post-trip feedback on their behavior, speeding was reduced by 14%. Moreover, (Reagan et al., 2013) found that incentive system resulted in significant reductions in speeding in comparison with the feedback system alone, which led to more modest changes.

2.3.4.2 Impact on (harsh) acceleration and braking behavior

Harsh braking and harsh acceleration are also frequently monitored behaviors, especially given the recent rise of surrogate safety measures as a prominent venue in road safety research. Within the studies examined, the impact of feedback on harsh braking ranged from a 10% reduction to as much as a 52% reduction. Camden et al. (2019) reported a 52.17% reduction in harsh braking and a 51.35% reduction in harsh cornering. (Kontaxi et al., 2021b) and (Soleymanian et al., 2019) also reported reductions in average harsh braking frequency, ranging between 10% and 52%. Similarly, (Strömberg & Karlsson, 2013) also investigated whether real-time and post trip feedback would have a positive impact on eco and safe driving behavior, resulting in reduced frequency of harsh decelerations.

Some studies, such as (Stevenson et al., 2021), reported improvements in overall risky driving behaviors, but no reductions were shown to reported measures such as harsh acceleration or braking. These mixed overall results of the studies highlight that feedback positively induced braking and acceleration, but factors including the type of feedback provided to participants, means of delivery and timing might also influence the effectiveness of such interventions.

Fewer studies have focused directly on harsh accelerations as an individual risk factor and mostly these were nested within analyses focusing on the overall crash risk. Both (Hari et al., 2012) and (Kontaxi et al., 2021b) assessed the direct impact of feedback on harsh accelerations, reporting positive results, with the latter study finding a 12% reduction during the feedback phase compared to the baseline. Other studies however found no statistically significant improvements in harsh acceleration when compared to other risky behaviors such as speeding and harsh braking (Camden et al., 2019; Farmer et al., 2010; Pozueco et al., 2017). A plausible explanation of this discrepancy is that drivers may consider harsh acceleration to be weakly linked to risky driving behavior, meaning that their response to feedback on this particular behavior may lack a connection to safe driving and the related motivation.

2.3.4.3 Impact on distracted driving behavior

Mobile phone use while driving was another monitored risk indicator impacted by driver feedback. Although it was not examined as much as speeding and braking, both (Ziakopoulos et al., 2023) and (Camden et al., 2019) found a notable reduction in mobile phone use following feedback interventions, highlighting the importance of well-structured feedback systems in addressing distracted driving behaviors, which are closely linked to crash risk. However, some studies, such as (Stevenson et al., 2021), reported improvements in composite measures of risky driving but did not find significant effects on distracted driving or mobile phone use specifically.

Nevertheless, mobile phone use while driving has traditionally been studied through driving simulators, where participants were instructed to text, talk, or browse on their phones while receiving real-time feedback (Backer-Grøndahl & Sagberg, 2011; Papantoniou et al., 2015). Again, this allowed for investigating the effects of distractions on driving behavior in safe but controlled environments, as well as the effects of immediate feedback on distracted driving. However, a series of recent advances in technology and use of Internet-of-Things (IoT) have opened new opportunities, i.e. sensors that can detect mobile phone usage while driving nearly continuously and without direct intervention (Camden et al., 2019; Stevenson et al., 2021). These developments allow researchers to measure distracted driving in the real world and identify how feedback affects mobile phone use behind the wheel.

2.3.4.4 Impact on Crash Risk/ Safety incidents

Several studies focused on the impact of feedback on crash risk in the form of a calculated rate of safety incidents. (Hickman & Hanowski, 2011; Husnjak et al., 2015; Takeda et al., 2011; Toledo & Shiftan, 2016), found reductions in safety incidents ranging from 8% to 52% following feedback interventions. Additionally, an earlier study by (Wouters & Bos, 2000) conducted during a two-year span reported an estimated 20% reduction in accident rates after professional drivers received feedback on their driving behavior.

In the same context, (Ghamari et al., 2022), recruited 1,289 bus drivers and 104 taxi drivers to examine the effect of peer comparison and performance feedback on their behavior. The primary outcome of the research was a driver score changing pattern throughout the study period, calculated by a neuro-fuzzy scoring system composed of four factors: speed violation, harsh acceleration, harsh braking, and harsh turning. The researchers used the score as an overall performance safety score and found a significant positive effect for both groups during the feedback phase.

(Peer et al., 2020) extends previous work by examining whether intrinsic motivation to change risky driving behavior is sufficient for feedback-driven improvements, or if higher effects depend on a further boost of extrinsic motivation through incentives that reinforce behavioral recurrence. This study conducted using a smartphone application collecting multiple safety-related driving behaviors (speeding, phone use, cornering and brake/acceleration events) aimed to evaluate the transferability of earlier findings from (Reagan et al., 2013) and (Mullen et al., 2015) – the latter using a driving simulator – to novice drivers and a broader range of driving behaviors. The study reported that feedback led to significantly enhanced driving behavior in terms of safety, but the effect was at its maximum when in addition to providing people with awareness regarding their violations and errors (e.g., informing them about numbers of exceedance through provision only), some type of incentive for safe driving were provided.

2.3.4.5 Post-feedback effects

The post-feedback effects of interventions vary significantly across the studies. In some cases, the behavioral changes resulting from feedback were sustained after the intervention period. For example, (Ghamari et al., 2022) found that bus drivers maintained their reformed behavior even after feedback was removed, indicating a lasting effect. Similarly, (Merrikhpour et al., 2014) reported that, although behavior indicators slightly decreased post-intervention, they remained better than the baseline, reflecting some persistence of the positive effects. (Molloy et al., 2023) also reported a sustainable improvement in speed compliance, but only for low-speed zones (i.e. 50 km/h).

However, some studies show that once feedback is withdrawn, drivers tend to revert to their previous behaviors. (Bolderdijk et al., 2011) observed that the incentive group increased their speeding once the financial incentives were removed, suggesting that the impact of feedback combined with incentives might not be sustained without continuous reinforcement. Likewise, (Toledo & Lotan, 2006) found that the initial reduction in driving risk indices disappeared after five months, with indices returning to or even exceeding their original levels. (Mazureck & Van Hattem, 2006) also noted that most drivers reverted to old habits after the feedback phase ended.

Conversely, other studies (McGehee et al., 2007) that included formal follow-up in the form of coaching or ongoing monitoring, reported maintainable reductions in the number of safety-critical events per mile; high-frequency drivers receiving feedback through continuous reinforcement mechanisms such as training and performance improvement plans demonstrated continued ability to sustain behavioral change post-feedback.

Certain studies indicate that feedback interventions yielded durable benefits (Soleymanian et al., 2019; Takeda et al., 2011; Toledo et al., 2008), but these findings were deemed to be of low certainty due to limited data and/or attrition bias in the data. This illustrates one of the major challenges concerning experiments over long periods, as attrition due to study dropout before a formal end point is quite common. With repeated interventions or a drop-off in motivation stemming from long participation, drivers can lose engagement over time, potentially skewing the results and undermining the statistical power of the findings. This drop-off may not only affect the reliability of the data but could also mask the true long-term effects of feedback, leaving open the question of whether behavior changes could have been sustained with continued participation.

2.4 Discussion

2.4.1 Elaboration of main findings

The results of this systematic review demonstrate a positive overall effect of driver feedback on modifying driving behavior and enhancing road safety. Measuring the quantifiable impact of driver feedback, it appears that when drivers receive feedback during naturalistic driving studies, they may decrease speeding from 5% up to 74%, harsh events from 10% up to 52%, safety incidents from 8% to 52% and also reduce road crashes up to 20%.

However, it is worth noting that the impact of driver feedback depends on several important contextual factors. Specifically: how is feedback provided, when it is delivered (timing), and what is the content of the feedback. Studies consistently indicate that real-time feedback results in the most immediate change at a behavior level; especially when it comes to reducing higher risk driving behaviors such as speeding or harsh braking (Camden et al., 2019; Chen & Donmez, 2021). Post-trip feedback also demonstrates effectiveness, particularly when combined with personalized reports or peer benchmarks (Peer et al., 2020). More precisely, only two of the reviewed studies compared the impact of post-trip and real-time delivered feedback, and the results are inconclusive with (Chen & Donmez, 2021) finding that only real-time feedback had a significant impact on speeding, while (Bell et al., 2017) found that risky driving behaviors reduced significantly less when drivers received only in-vehicle real-time feedback.

The timing of feedback is undoubtedly a crucial factor to consider when designing an effective feedback mechanism and certainly warrants further research. An interesting study by (Dijksterhuis et al., 2015) used a driving simulator to examine the issue, concluding that both systems improved driving performance, though the initial advantage of the real-time feedback group diminished substantially over time. Nevertheless, the answer lies in the type of feedback that is provided; specific and actionable insights about safe driving tend to lead to more positive changes than general, non-specific assessments of unsafe behavior (Merrikhpour et al., 2014).

In addition to how, when and what feedback to provide, another notable observation is the varying effectiveness of feedback depending on the frequency of its delivery. For instance, studies like (Molloy et al., 2023) found that driver feedback was more effective when provided only once or twice compared to multiple times. It is highly likely that the feedback mechanism suffers a reliability loss by reporting too frequently to drivers. This outcome underscores the need to find the optimal frequency of feedback to sustain the acquired improvement in driving behavior, but without triggering drivers with overload or desensitization.

Another interesting finding from the review process is the high correlation of incentives to the effectiveness of feedback. Studies report that the provision of financial incentives can substantially increase behavioral gains when used in conjunction with real-time or post-trip feedback (Bolderdijk et al., 2011; Stevenson et al., 2021). Systems based on motivation incentives, such as price reductions in insurance premiums for safe driving or monetary rewards for absence of hazardous driving phenomena persuade drivers to continue their improved behavior in the long term. Nevertheless, it is also observed that when incentives are withdrawn, behavior can relapse and some participants will revert to their old habits, such as in the case of (Bolderdijk et al., 2011). This highlights the need to create feedback and incentive mechanisms that incentivize persistence and sustainable behavioral change over time.

In the context of experimental frameworks, findings suggest that most of the reviewed studies combine within-subjects with between-subjects design by examining feedback effects both on individual drivers and broader groups. This combination allows researchers to alleviate the concerns of carryover and fatigue effects, while guaranteeing valid and generalizable study results. Many studies explore the post-feedback effects as well (Bell et al., 2017b; Farmer et al., 2010; Ghamari et al., 2022b; Mazureck & Van Hattem, 2006; Molloy et al., 2023), revealing that in some cases, improvements in terms of driving behavior carry on even after the finishing point of the feedback phase. Yet, the permanence of these effects is inconsistent as some drivers return to unsafe behaviors when feedback is removed (Toledo & Lotan, 2006). The long-term effects of driver feedback are a common topic discussed in most of the reviewed studies, often emphasized as a challenging issue. The need for repeated experiments that are conducted sometime after the initial study is underlined in order to effectively examine the sustainability of the impact of feedback over time.

As for the analyses used in the reviewed studies, most of the reviewed studies used conventional statistical models such as Generalized Estimating Equations (GEE), linear and logistic regression, and Analysis of Variance (ANOVA) to evaluate the impact of driver feedback on behavior and safety. Although these methods are useful and provide valuable insights, especially for hypothesis testing and group-level analysis, they are limited in capturing complex non-linear behaviors or patterns when dealing with larger datasets or several feedback phases. Particularly, there is a limited application of Machine Learning methods that may provide more advanced predictive modeling and a comprehensive understanding of specific driver behavior patterns. Such methods as random forests or gradient boosting may facilitate real-time personalization and feedback mechanism optimization. This gap suggests a future research opportunity to employ ML methods for greater adaptability and individuality in feedback systems.

2.4.2 Implications for road safety and current practices

The previous passages reveal the benefits of driver feedback on enhancing road safety, which in turn provides potential guidance for both policy and technological development. Feedback systems serve as a powerful instrument to impact driving behavior, especially when focusing on critical risk factors like speeding. Driven by new technological advancements, feedback systems are becoming more broadly viable, offering tools for deployment across different settings, from urban networks to rural roads. Integrating driver safety programs based on real-time and post-trip feedback creates a shortcut for transportation authorities and fleet managers to enhance driving performance. In addition to reducing risky driving behavior, these systems promote safer road use and facilitate advanced sustainable development goals.

Given that smartphone applications and telematics devices provide a cost-effective way for continuous data acquisition, it is an appropriate method to monitor and make necessary modifications in various driving behaviors by insurance companies, particularly under the Usage-Based Insurance (UBI) models (Tselentis et al., 2017). The systems enable insurers to better understand the risks associated with each driver and to price policies according to an individual's actual driving behavior. However, these research efforts are challenged by issues like privacy and the availability of large datasets for analysis (Reimers & Shiller, 2019). The timing of feedback, its content, and incorporating financial incentives to support safety outcomes are important insights for insurers, stemming from this review. Focusing on these factors, car insurance companies can create a feedback mechanism that not only encourages sustained behavioral change and solves the engagement problem for the driver, but also promotes safe driving.

Another promising emerging area for improving feedback processes is the integration of connected vehicles (CVs). Communications between vehicles (V2V) and infrastructure (V2I), collectively known as connected vehicle technology, can transform an individual vehicle into a node within a real-time information network. This connectivity has the potential to revolutionize driver feedback by exploiting near real-time updates on road conditions, historical and predicted hazards, as well as surrounding traffic behavior (Abdelkader et al., 2021). For example, feedback systems can be enabled in connected vehicles with predictive analytics to provide proactive alerts and customized instructions for drivers that leverage the knowledge of an entire network, thus not limited in their own driving behavior. Indeed, the input of feedback might play a far more extensive role in such systems and is expected to be increasingly simplified as connected vehicles will require more advanced feedback mechanisms that adapt to new forms of shared data and road interactions.

Regardless of automation, the adoption and higher market penetration of feedback schemes will lead to the acquisition of more voluminous driver behavior data, which will lead to the refinement of feedback in turn. This course will initiate a virtuous circle that is well-placed to enhance the acceptance and feasibility of driver feedback as part of everyday driving. Within the research field, feedback and the related data have the potential to provide a broader and more accurate overall image of safety data within the transport network. Automated feedback data will be completely automated and as such will not suffer from delays, misclassifications, compatibility requirements and other issues such as underreporting.

2.4.3 Limitations of current research and directions for future research

The investigation of the impacts of driver feedback has revealed some valuable insights; however, this review additionally highlights several limitations in the existing literature that should be addressed in future research. An important gap in the reviewed studies is the lack of consideration of traffic conditions and how they may affect the effectiveness of driver feedback. There are a few studies that consider the type of the road, i.e. urban, rural or highway in their analyses (Birrell & Fowkes, 2014; McGehee et al., 2007; Strömberg & Karlsson, 2013). However, traffic data such as volume, speed, density etc. should also be incorporated into future naturalistic driving experiments for a more comprehensive and contextual understanding of how driving behavior and the effectiveness of driver feedback may be affected by external factors. In that context, weather conditions should also be accounted for in future studies when exploring the effectiveness of driver feedback.

Another significant limitation in many of the reviewed studies is the reliance on self-selected participants who voluntarily chose to take part in such research experiments, except for a few studies that offered monetary incentives to participants (Chen & Donmez, 2021; Ghamari et al., 2022). This approach may introduce selection bias, as these individuals might already possess higher levels of safety awareness or be more technologically apt compared to the general driving population. As a result, the outcomes of such studies may not fully represent the broader population, leading to potential overestimation of feedback effectiveness. Exploring this aspect further in future studies could improve the design and interpretation of experiments by offering a more comprehensive view of how feedback impacts a wider, more diverse driving population.

Furthermore, in one line of current research, a variety of intervention systems (e.g., context-sensitive distraction warning systems (Kujala et al., 2024) require further development and comparison. The rapid advancements in artificial intelligence, mobile internet, and big data, have introduced an innovative approach to active driving behavior intervention through warning systems and workload generation tools (Yang et al., 2023). However, questions remain about their efficacy, including the extent to which drivers might rely too heavily on these systems and how they may affect driver’s workload and situational awareness. Future research should investigate these impacts to determine the proper division of labor between technology and drivers.

A particularly overlooked area is the feedback provision for two-wheeler drivers, who are considered vulnerable road users. While some studies have explored feedback for motorcyclists under naturalistic conditions (Kontaxi et al., 2021a; Will et al., 2020; V. Williams et al., 2016), there remains a gap in understanding how feedback could better support safety among these drivers. Two-wheeler drivers face unique risks, such as reduced visibility to other road users and greater vulnerability in crashes, which require tailored feedback mechanisms. Future research should focus on developing and testing feedback systems specifically designed for two-wheelers, potentially incorporating real-time alerts and adaptive safety interventions to help mitigate their elevated risk. Comparable considerations can also be given to the provision of feedback for professional drivers as well, who face their own challenges and particularities. Given that the operation of in-vehicle information systems has very low safety impacts for professional drivers (Ziakopoulos et al., 2019) the development of related feedback systems is a promising, yet unexplored venue.

Lastly, although data-driven methods such as machine learning algorithms have been widely used to identify dangerous driving behaviors, few innovative approaches have investigated how feedback itself can be beneficial for improving safety. Research often focuses on recognizing patterns without considering how feedback can alter driving behavior (Elamrani Abou Elassad et al., 2020; Lattanzi & Freschi, 2021). A critical avenue for investigation involves applying machine learning to better understand the mechanisms by which feedback works, leading to more individualized, data-driven approaches for driving behavior change. Future work could combine machine learning with driver feedback systems to produce dynamic, personalized feedback tailored to specific driving styles, enhancing long-term road safety.

On that note, it is important to anticipate, or at least stay on track with, technological developments regarding new driver interaction venues that are being opened. These venues can be novel in terms of both the manner of feedback and the device itself. Indicative examples include smartwatches and similar wearables, which provide a very direct interaction manner with their wearer, which could open the possibility of tactile real-time driver feedback nudges. Moreover, and in parallel with automation, vehicle connectivity is ever increasing, and offers integration capabilities for many devices which can calculate and provide feedback. With the aforementioned developments in edge-computing applications, driving feedback can be provided on the spot, under real-time or almost real-time conditions, and can also be paired with related services such as e-call or rapid crash reporting as well.

2.5 Research Questions

Based on the findings of the comprehensive literature review conducted within the framework of this Dissertation, several gaps and challenges in the existing body of knowledge have been identified. These gaps have informed the development of the following research questions, which aim to address critical aspects of the study and contribute to advancing understanding in this domain:

1. How does feedback influence driver speeding and distracted behavior in terms of the percentage of trip time during which the speed limit was exceeded and mobile phone was used while driving?

2. How does feedback influence harsh driving events, in terms of the number of harsh accelerations and harsh brakings?

3. Do different feedback features (e.g., scorecards, maps, peer comparisons, motivations, gamification, rewards) have different effects on driver behavior? Which feature demonstrates the most significant impact?

4. How does the post-feedback effect influence long-term driver behavior, and to what extent are the changes sustained after the feedback is removed?

5. How can advanced statistical techniques be applied to understand the mechanisms of driver feedback and develop more individualized, data-driven approaches for driving behavior change?

3 Methodological Approach

3.1 Overall Methodological Framework

This section outlines the methodology employed to investigate the impact of driver feedback on driver behavior through a naturalistic driving experiment. The overall framework is visually represented in Figure 3.1, providing a structured approach to achieving the objectives of this dissertation. Subsequent sections delve deeper into the theoretical and experimental methods utilized.

The methodological framework began with an extensive literature review, which explored key aspects of driver feedback systems under naturalistic driving conditions. This review focused on experimental design, feedback types, modeling approaches, and key indicators for evaluation. This phase guided the formulation of research questions, including the impact of feedback on behavior and safety, the effects of feedback features, and the post-feedback influence on long-term driver behavior.

Building on the findings of the literature review, a comprehensive methodological background was developed, combining theoretical approaches and experimental design principles. This included the application of advanced modeling techniques, such as Generalized Linear Mixed Effects Models (GLMMs), Structural Equation Models (SEMs), and Survival Analysis Models, alongside the design of a naturalistic driving experiment (ND Experiment Design). The experimental setup involved a cohort of 130 drivers, encompassing car drivers, van professionals, and motorcyclists, evaluated over six feedback phases using a within-subjects design.

The research utilized data from the BeSmart Research Project, focusing on key driving indicators such as speeding and mobile phone use as measures of risky behavior, and harsh accelerations and harsh brakings as proxies for safety-critical events. The analysis unfolded in three key pillars:

1. Impact of feedback: This phase assessed the immediate effects of feedback on driving behavior (e.g., speeding, mobile phone use, harsh events frequency) across the three driver groups.

2. Effects of different feedback features: A Structural Equation Model (SEM) was developed to explore the relationship between feedback features and driving behavior factors, revealing the differential impact of feedback elements.

3. Post-feedback effects: A survival analysis was conducted to examine the long-term influence of feedback mechanisms. Models such as Cox Proportional Hazards, Accelerated Failure Time (AFT), and Random Survival Forests (RSFs) were applied and compared to identify the best-fitting model.

By employing this comprehensive methodology, the driver behavior telematics feedback mechanism is effectively achieved. The integration of an extensive literature review, advanced modeling techniques, and a well-structured naturalistic driving experiment provided a robust framework to investigate the impact of feedback on driver behavior. The research focused on critical driving indicators such as speeding, mobile phone use, harsh accelerations, and harsh braking events, ensuring a holistic assessment of driver risk factors. The systematic analysis of immediate feedback effects, the influence of different feedback features, and the long-term impact of feedback mechanisms culminated in actionable insights for designing and implementing telematics feedback systems. These findings lay the foundation for a scalable and data-driven feedback mechanism aimed at improving driving behavior, reducing risky behaviors, and enhancing road safety.

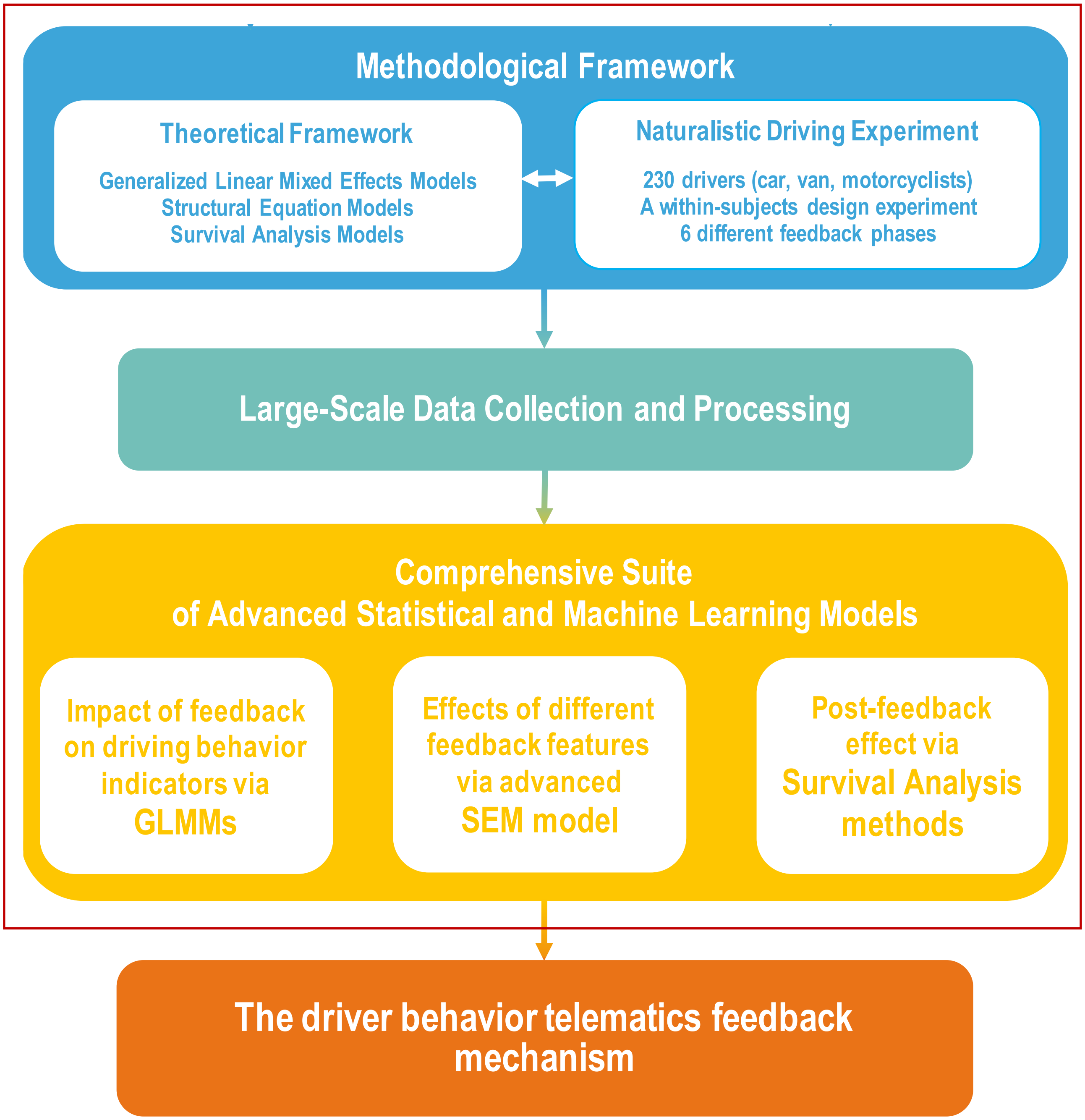


Figure 3.1: Overall methodological framework of the doctoral dissertation

3.2 Theoretical Framework

To address the objectives outlined in this PhD thesis, a robust theoretical framework was developed, leveraging both foundational and advanced statistical methodologies. This theoretical background builds upon existing statistical tools while incorporating innovative approaches tailored to the complexities of analyzing driver behavior and performance. The methodologies were selected and refined based on a comprehensive review of the relevant literature, as discussed in earlier chapters. This ensures that the selected approaches address the limitations and gaps identified in previous studies.

A key finding from the literature review was that most studies rely on traditional statistical techniques, such as repeated measures ANOVA, descriptive statistics, or linear regression models. While these methods provide valuable insights, they are often limited in addressing the intricate dependencies and latent relationships inherent in driver performance data. The review also revealed that advanced techniques, such as Structural Equation Modeling (SEM) and Survival Analysis, remain underexplored in the field, despite their potential to address complex relationships and time-dependent variables.

To build a comprehensive and innovative methodological foundation, the theoretical background of this thesis incorporates both conventional and advanced techniques. Descriptive statistics form the basis of the analysis, offering fundamental insights into the data distribution and relationships. Inferential tests, such as paired samples t-tests and Wilcoxon signed-rank tests, provide robust comparisons between experimental conditions, especially in cases where parametric assumptions are violated.

The statistical modeling framework further expands into regression models, including General Linear Models (GLMs) and Generalized Linear Mixed Models (GLMMs), which accommodate the hierarchical structure of the data and enable the analysis of fixed and random effects. These models are enhanced by variations such as random intercepts and random slopes, which capture the heterogeneity of driver behavior across individuals and conditions. Structural Equation Models (SEMs) are introduced to explore latent relationships between exposure indicators and different features of driver feedback, providing a deeper understanding of their direct and indirect impacts on driving performance and safety.

The methodological approach also incorporates advanced Survival Analysis techniques to examine time-to-event data, such as relapse of driving behavior after the end of feedback period. Kaplan-Meier curves and Cox Proportional Hazards models with frailty terms assess survival probabilities under varying conditions, while Weibull Accelerated Failure Time (AFT) models and Random Survival Forests allow for more flexible modeling of time-dependent risks. A systematic model comparison ensures that the most appropriate techniques are applied for the respective research questions, contributing to the reliability and validity of the findings.

This theoretical background not only integrates a wide range of statistical tools but also emphasizes their sequential and complementary application, ensuring that the analysis is comprehensive, robust, and capable of addressing the multifaceted nature of the feedback mechanism and its effect to driving behavior and safety. The following sections provide a detailed explanation of each methodological approach and its application within the context of this thesis.

3.2.1 Descriptive Analysis

Descriptive analysis is a fundamental step in data analysis that provides a solid foundation for advanced statistical modeling. Before diving into complex models, it is crucial to understand the underlying structure and behavior of the data through summary statistics and visualizations. Key components of descriptive analysis include measures like the five-number summary (minimum, first quartile, median, third quartile, and maximum), which helps to capture the range, central tendency, and spread of the data. These metrics give an overview of the data's variability and highlight potential outliers that could skew the results of subsequent analyses. Additionally, calculating the standard deviation quantifies the degree of dispersion around the mean, offering insight into how tightly or loosely data points are distributed.

3.2.2 Inferential statistical tests

Inferential statistical tests are used to draw conclusions about a population based on sample data. These tests help evaluate hypotheses, determine relationships, and assess differences between groups or conditions. Unlike descriptive statistics, inferential tests use probability theory to determine whether observed patterns are statistically significant or could occur by chance. Two commonly used inferential tests for comparing paired data are the Paired Samples t-test and the Wilcoxon Signed-Rank Test, which are both designed to assess differences between two related groups.

These tests are particularly useful in scenarios where the same subjects are measured under different conditions, such as pre-treatment and post-treatment measurements in clinical studies. The choice of test depends on whether the data meet assumptions of normality or require non-parametric methods for analysis.

3.2.2.1 Paired Samples t-test

The Paired Samples t-test (also known as the Dependent Samples t-test) is a parametric test used to compare the means of two related groups. It assumes that the differences between paired observations are normally distributed.

The test calculates the t-statistic as follows:

$$\boldsymbol{t = \frac{\bar{d}}{s_d/\sqrt{n}}} \qquad \textbf{\textit{(3.1)}}$$

Where:

- $\bar{d}$ is the mean difference between the paired observations,
- s_d is the standard deviation of the differences,
- n is the number of pairs.

The resulting t-statistic is compared against a critical value from the t-distribution table based on the degrees of freedom (df=n-1) and the desired significance level (α, typically 0.05). If the p-value is below α, we reject the null hypothesis (H_0), which states that the mean difference is zero.

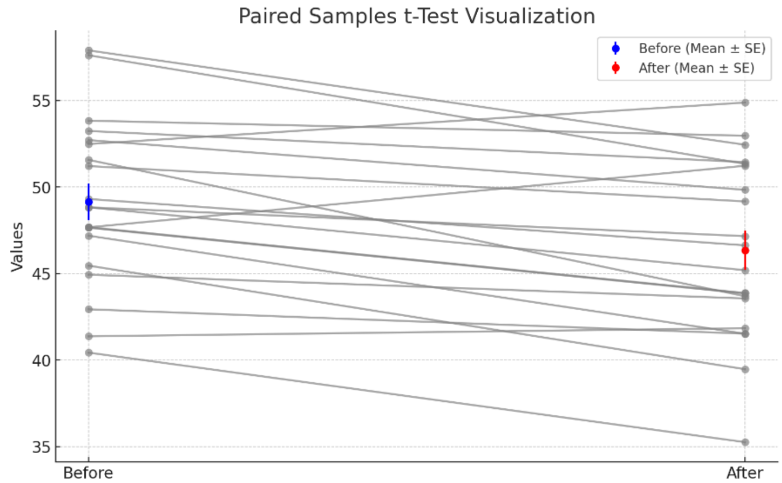


Figure 3.2: Schematic diagram of a paired samples t-test

In the above schematic diagram:

- Each gray line represents a pair (before and after values for the same subject).
- Blue and red markers show the mean for the "Before" and "After" groups, along with error bars indicating standard error.

3.2.2.2 Wilcoxon signed-rank test

The Wilcoxon Signed-Rank Test is a non-parametric alternative to the Paired Samples t-test, used when the data do not meet the normality assumption. This test compares the median of paired differences and is robust to outliers and skewed distributions. For instance, it can be used to assess whether a mindfulness program affects stress levels, where pre- and post-program stress scores are ranked and analyzed.

The procedure involves ranking the absolute differences between paired observations, assigning positive or negative signs based on the direction of the difference, and calculating the test statistic WWW, which is the sum of signed ranks:

$$\boldsymbol{W = \sum R_i^+} \qquad \textit{(3.2)}$$

Where R_i^+ are the ranks with positive differences.

The test statistic is then compared to a critical value from the Wilcoxon Signed-Rank distribution or converted into a z-score for large samples:

$$\boldsymbol{z = \frac{W - \mu_W}{\sigma_W}} \qquad \textit{(3.3)}$$

Where μ_W and σ_W are the mean and standard deviation of the rank sum distribution under the null hypothesis. Like the t-test, if the p-value is below the significance level (α), the null hypothesis (H_0: no difference in medians) is rejected.

Both tests provide valuable insights into paired data, with the Wilcoxon Signed-Rank Test offering a more robust solution when the normality assumption is violated.

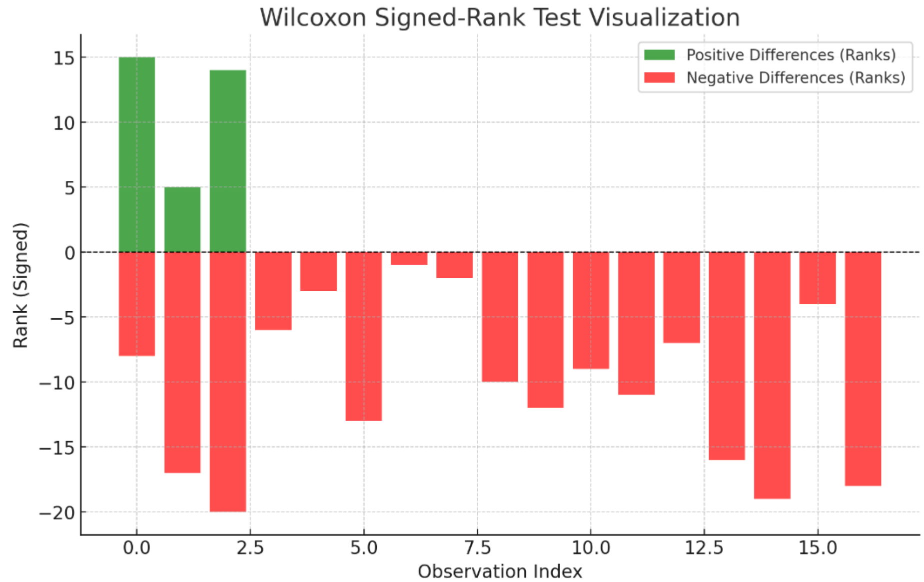


Figure 3.3: Schematic diagram of a Wilcoxon Signed-Rank Test

In the above schematic diagram:

- The bars represent the ranked differences between paired observations.
- Green bars correspond to positive differences (the first variable is greater), while red bars represent negative differences (the second variable is greater).
- The heights of the bars indicate the rank assigned to each absolute difference.
- Positive ranks are displayed above the zero line, while negative ranks are below it.
- The sum of these signed ranks is used to compute the test statistic for assessing whether the median difference is significantly different from zero.

3.2.3 Generalized Linear Models

Generalized Linear Models (GLMs) extend traditional regression methods by allowing the dependent variable to follow distributions from the exponential family, such as Poisson, binomial, or Gaussian. For count data, such as the frequency of harsh events (e.g., harsh braking or harsh acceleration) per trip, a Poisson distribution is typically used.

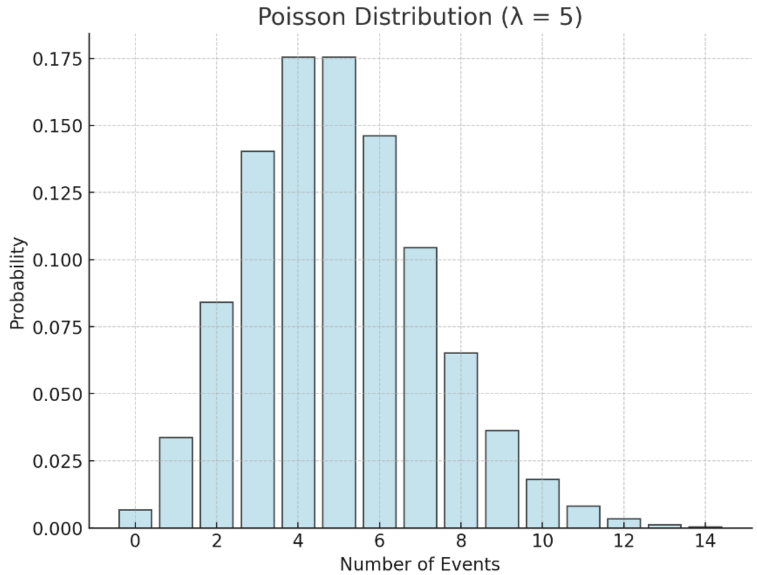


Figure 3.4: Example of a Poisson distribution (λ=5)

Following Breslow and Clayton (1993), if y_i represents the observed frequency of harsh events during trip i and λ_i represents the predicted expected frequency, the GLM is specified as:

$$y_i \sim Poisson\ (\lambda_i) \tag{4.4}$$

The log of the expected frequency (λ_i) is then modeled via a linear predictor:

$$log(\lambda_i) = \beta_0 + \beta_n x_n + \varepsilon \tag{4.5}$$

Where β are the fixed-effect parameters (constant and coefficients) for n independent variables, and ε is the error term.

3.2.3.1 Generalized Linear Mixed-Effects Models

Generalized Linear Mixed-Effects Models (GLMMs) extend GLMs by adding random effects to account for grouping or clustering in the data, such as repeated measures for the same drivers. This approach accounts for the variability in behavior between drivers that could influence the frequency of harsh events. The GLMM builds on the GLM framework by adding a random effect (u_i) to the linear predictor:

$$log\ (\lambda_{ij}) = \beta_0 + \textstyle\sum_n\ \beta_n x_{nij} + u_i + \epsilon \tag{4.6}$$

Here, $\lambda_{ij}$is the expected frequency of harsh events for trip j by driver i, and $u_i \sim N(0, \sigma^2_{s,0})$ represents the random effect for driver i, capturing unobserved heterogeneity. This structure ensures that differences between drivers are modeled explicitly, improving the accuracy of fixed-effect estimates.

For instance, a GLMM could model harsh braking events by accounting for predictors such as trip duration or distance, while also incorporating a random effect for each driver to reflect differences in their driving styles.

3.2.3.2 GLMM with Random Intercepts

A GLMM with random intercepts models variability in the baseline frequency of harsh events across drivers while keeping the effect of predictors constant. The model is expressed as:

$$log\left(\lambda_{ij}\right) = \beta_0 + \sum_n \beta_n x_{nij} + u_i \quad (4.7)$$

Here, u_i represents the random intercept for driver i, distributed as $u_i \sim N(0, \sigma_u^2)$. This allows the model to account for differences in the baseline frequency of harsh events across drivers.

For example, one driver might consistently have a higher baseline frequency of harsh braking, regardless of trip duration or speed. A random intercept captures this variation, allowing the fixed effects (e.g., trip duration) to be estimated without bias.

3.2.3.3 GLMM with Random Slopes

A GLMM with random slopes further extends the model by allowing the effect of a predictor, such as trip duration or distance, to vary across drivers. The model is specified as:

$$log\left(\lambda_{ij}\right) = \beta_0 + (\beta_1 + u_{1i})x_{1ij} + \sum_{n\neq 1} \beta_n x_{nij} + u_{0i} \quad (4.8)$$

Here, u_{0i} is the random intercept for driver i, and u_{1i} is the random slope for predictor x_{1ij} (e.g., trip duration), with $u_{1i} \sim N(0, \sigma_{u1}^2)$.

For instance, the impact of trip duration on the frequency of harsh braking events might vary between drivers. For one driver, longer trips might significantly increase the likelihood of harsh events, while for another, the effect might be weaker. A random slope allows the model to capture this variability, providing a more detailed understanding of driver-specific behaviors.

3.2.3.4 Interpreting Coefficients and Goodness-of-fit metrics

Coefficient result interpretation is more intuitive when using relative risk ratios (sometimes called incidence rate ratios). Relative risk ratios are obtained by transforming the predictor to obtain the frequency. For an increase of one unit in one specific variable, k, with all other parameters remaining equal, the predicted original frequency λ_i is multiplied by: $\lambda_{ki} = \exp(\beta_{ki}) * \lambda_i$

As (McCulloch, 2003) mentions, random effect models may use correlated independent variables as input, circumventing the limitations of traditional GLMs. Furthermore, it should be mentioned that for computational reasons during the GLMM fitting, the trip data underwent z-score scaling, a common standardization process which does not affect the obtained coefficients. Mathematically, for every parameter x with a mean $\bar{x}$ and a standard deviation S a scaled value is obtained:

$$x_{scaled} = (x - \bar{x})/S \quad (4.9)$$

The best-fitting model which contains the more informative variable combination and explains the highest degree of variance per given dataset is selected as the one with the minimum Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

$$AIC = -2ln\,(L) + 2k \qquad (4.10)$$

Where:
- L: Maximized likelihood of the model.
- k: Number of parameters in the model.

$$BIC = -2ln\,(L) + kln\,(n) \qquad (4.11)$$

Where:
- n: Number of observations.

Finally, it is critical to note that the added value of any random effects is assessed by conducting a custom ANOVA (log-likelihood test) between the fixed effects GLM and any formulated GLMMs.

3.2.4 Structural Equation Models

Structural Equation Modeling (SEM) is a technique within the family of latent variable analysis. It is a multivariate method that supports both multiple-input and multiple-output modeling. In this study, SEM is used to formulate several unobserved constructs as latent variables from different types of variables collected through the naturalistic driving experiment.

SEM is a widely recognized methodology with numerous applications. It has been employed in various studies to model complex interrelationships involving unobserved concepts expressed as latent variables. This includes applications in traffic engineering and road safety. For example, SEM has been used to model driving behavior and the probability of crash risk (Papantoniou et al., 2019; Wu & Wang, 2021) Additional examples include the use of SEM to connect task complexity and coping capacity with driving risk (Roussou et al., 2023) or perception of risk and driving tasks on road safety attitudes of drivers (Ram & Chand, 2016; Zhao et al., 2019).

The underlying mathematical structure of SEMs can be defined as follows (Jöreskog & Sörbom, 2001):

$$\eta = \beta\eta + \gamma\xi + \varepsilon \qquad (4.12)$$

where η is a vector expressing the dependent variables, ξ expressing the independent variables, ε expressing the regression error term, β expressing the regression coefficients for the dependent variables, and γ expressing the regression coefficients for the independent variables.

3.2.4.1 Path analysis

Path analysis, a subset of SEM, focuses on modeling the structural relationships between variables. It visualizes these relationships using a path diagram, where arrows represent regression paths. For example, in a study analyzing driving behavior, latent variables such as "risk perception" (ξ) might influence "event occurrence" (η) through observed indicators like average speed, acceleration, or harsh braking events. Path analysis also allows for mediating effects, where one variable influences another indirectly through a third variable.

Path diagrams clarify the roles of explanatory variables, showing whether they act as correlated causes, mediated causes, or independent predictors. For instance, "trip duration" (X_1) and "driver age" (X_2) might have direct effects on "harsh braking" (Y_1) while "trip speed" (X_3) mediates the relationship. This structured representation not only simplifies complex models but also facilitates hypothesis testing, enabling researchers to specify whether certain pathways are significant or not.

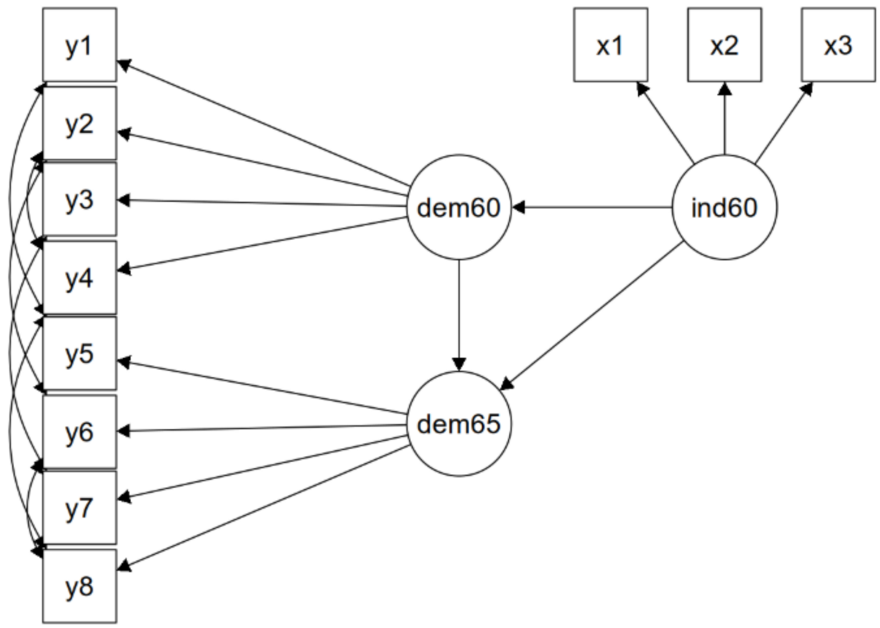


Figure 3.5: Example of SEM path analysis created using the lavaan package in R

The diagram represents a structural equation model (SEM) created using the lavaan package in R. This model combines measurement components (latent variables derived from observed indicators) and structural relationships (regressions and covariances). Below is a breakdown of the elements in the diagram:

Latent Variables

- ind60: A latent variable representing an unobserved construct measured by three observed indicators: x1, x2, and x3. These arrows indicate that the latent variable ind60 is inferred from the shared variance of these observed variables.
- dem60: A latent variable representing another unobserved construct, measured by the observed indicators y1, y2, y3, and y4. These arrows suggest that the latent variable dem60 is inferred from these observed variables.
- dem65: A latent variable inferred from the observed indicators y5, y6, y7, and y8.

Structural Relationships

- Regression Paths:
    - The arrow from ind60 to dem60 indicates that the latent variable dem60 is regressed on ind60, meaning ind60 explains some portion of the variability in dem60.

    - The arrows from ind60 and dem60 to dem65 represent a structural relationship where both ind60 and dem60 jointly explain the variability in dem65.

Residual Covariances

- Correlated Residuals:
    - The bidirectional arrows connecting y1 and y5, y2 and y4 and other pairs indicate residual covariances between the corresponding observed variables.

Observed Variables

- The observed variables (x1, x2, x3 and y1, y2, y3, y4, y5, y6, y7, y8) are directly measured in the dataset. These are used to infer the latent variables or are part of the structural relationships. The "x" variables (x1, x2, x3) are indicators for ind60, while the "y" variables (y1 to y8) serve as indicators for dem60 and dem65.

3.2.4.2 Goodness-of-fit metrics

In the realm of model configuration, Goodness-of-Fit measures play a crucial role in any statistical model assessment. The goodness-of-fit metrics utilized in the current analysis are listed below.

The Comparative Fit Index (CFI) compares the fit of a hypothesized model with an independence model. Values range from 0 to 1, with over 0.90 generally accepted as a good fit. The formula is represented as follows:

$$CFI = 1 - \frac{max(x_H^2 - df_H, 0)}{max(x_H^2 - df_H, x_I^2 - df_l)} \tag{4.13}$$

where x_H^2 and $df_H$are the chi-square value and degrees of freedom of the hypothesized model, and x_I^2 and df_l are those of the independence model.

The Tucker–Lewis Index (TLI) evaluates model parsimony, with values above 0.95 indicating a good fit. The formula is represented as follows:

$$TLI = \frac{\frac{x_I^2}{df_I} - \frac{x_H^2}{df_H}}{\frac{x_I^2}{df_l} - 1} \tag{4.14}$$

The Root Mean Square Error Approximation (RMSEA) measures the unstandardized discrepancy between the population and the fitted model, with values below 0.08 typically considered a good fit and values below 0.05 indicating an excellent fit. The formula is represented as follows:

$$RMSEA = \sqrt{\frac{x_H^2 - df_H}{df_H(n-1)}} \tag{4.15}$$

where x_H^2 is the chi-square value, df_H is the degrees of freedom, and nnn is the sample size.

The Standardized Root Mean Square Residual (SRMR) is the square root of the difference between the residuals of the sample covariance matrix and the hypothesized covariance model, with values below 0.05 indicating an excellent fit. The formula is represented as follows:

$$SRMR = \sqrt{\frac{\sum_{i=1j}^{n} \sum_{j=1}^{n} \left(s_{ij}-\sigma_{ij}\right)^2}{n(n+1)/2}} \tag{4.16}$$

where s_{ij} is the observed covariance between variables i and j, σ_{ij} is the predicted covariance between variables i and j based on the model and n is the number of observed variables.

3.2.5 Survival Analysis

Survival analysis is a branch of statistics designed to analyze time-to-event data, where the primary objective is to model the time until a specific event occurs, such as a failure, death, or, in driving contexts, the time until an adverse event like a harsh braking occurs. Unlike other statistical methods, survival analysis accounts for censored data, which arises when the event of interest has not occurred for some subjects during the study period. Survival analysis is widely used across disciplines, including medical research, engineering, and transportation safety (Klein & Moeschberger, 2003; Clark et al., 2003; Machin et al., 2006).

Over the years, survival analysis has developed to be one useful technique of driving behavior investigation through the modeling of time to critical events-such as harsh braking, near-crashes, or collisions (Parmet et al., 2014; Shangguan et al., 2020; Samani & Mishra, 2024). This method explicitly allows researchers to consider censored data, such as trips with no adverse events occurring, while considering the impact of covariates like driver behavior and trip characteristics on event occurrence.

Basic terminologies for the survival analysis for present study are defined as following (Washington et al., 2020):

Event: In this study, an "event" is defined as a "relapse" in driving behavior, specifically when the driver's harsh accelerations per 100 km exceed a predefined threshold. This threshold is calculated as the mean harsh acceleration rate observed during the feedback phase, a period of active intervention. Exceeding this threshold in the post-feedback phase signals a decline in driving behavior, which is considered an "event" for survival analysis purposes.

Duration Variable (Time to Event): The duration variable in this analysis is represented by the number of trips taken until a relapse event occurs (i.e., harsh acceleration rate exceeds the feedback phase threshold). In survival analysis terms, the duration is a continuous random variable T with a cumulative distribution function F(t) and probability density function f(t). The survival analysis tracks the probability of a driver maintaining improved driving behavior over successive trips in the post-feedback phase.

$$F(t) = P(T < t) = \int_0^t f(t)dt \tag{4.17}$$

Survival Rate (S(t)): The survival rate S(t) gives the probability that a driver will maintain driving behavior below the harsh acceleration threshold for a given number of trips, denoted by t. This can be interpreted as the probability of no relapse occurring within that period. Mathematically,

$$\boldsymbol{S(t) = P(T \geqslant t) = 1 - F(t)} \tag{4.18}$$

where F(t) is the cumulative probability of a relapse occurring by trip t.

<u>*Hazard Rate (h(t)):*</u> The hazard rate h(t) represents the conditional probability of a relapse occurring at a particular trip t, given that no relapse has occurred up until that trip. It provides an instantaneous risk of relapse at each point in time (number of trips). In this study, as the number of trips increases in the post-feedback phase, the probability of relapse also tends to increase, indicating a rising hazard rate over time. The hazard function h(t) can be defined as:

$$h(t) = \frac{f(t)}{[1-F(t)]} \tag{4.19}$$

where f(t) is the probability density function of relapse events.

The key models discussed in this section—Kaplan-Meier curves, Cox proportional hazards models, and Weibull accelerated failure time models—offer different approaches to understanding survival data, with extensions to handle heterogeneity and clustered data.

3.2.5.1 The Kaplan-Meier curves

The Kaplan-Meier estimator, also known as the product-limit estimator, is a non-parametric statistical method used to estimate the survival function from time-to-event data (Kaplan & Meier, 1958). It is commonly applied in fields such as medicine, engineering, and social sciences to understand the likelihood of an event (e.g., failure, relapse) occurring over time, especially when data include censored observations (i.e., when the event has not occurred for some subjects by the end of the observation period).

The Kaplan-Meier survival function S(t) is defined as the probability that the event of interest has not occurred by a certain time t:

$$S(t) = P(T \geqslant t) \tag{4.20}$$

where T represents the time to event. The Kaplan-Meier estimator calculates the survival probability at each time point where an event occurs, updating the cumulative survival probability accordingly. The survival probability at each event time tj is calculated by:

$$\hat{S}(t) = \prod_{t_j \leq t} \left(1 - \frac{d_j}{n_j}\right) \tag{4.21}$$

where:

- d_j is the number of events (e.g., relapses, failures) occurring at time t_j
- n_j is the number of subjects at risk just prior to time t_j

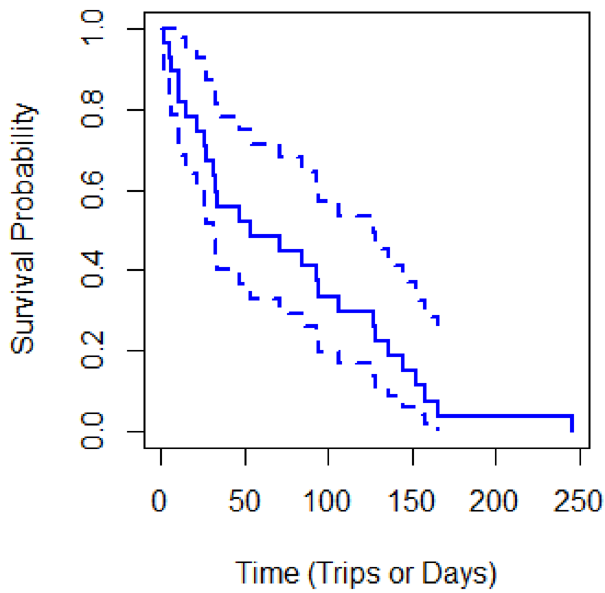


Figure 3.6: Example of Kaplan-Meier survival curve

The Kaplan-Meier survival curve in the above example visualizes the probability of survival over time, accounting for censored data. Some key points:

- Stepwise Survival Function: The curve shows the estimated survival probability S(t)S(t)S(t) as a step function, which decreases only at observed event times. At each event time, the survival probability is updated based on the proportion of individuals still at risk who experience the event.

- Censoring: The Kaplan-Meier method accounts for censored observations (instances where the event has not occurred during the study period). Censored individuals contribute to the survival probability up to the time they are last observed but do not lower the survival probability

- Confidence Intervals: The dashed lines represent the 95% confidence intervals around the survival estimate, providing a range within which the true survival probability is likely to fall. This helps assess the reliability of the survival estimates at different time points.

3.2.5.2 Cox-PH Model with Frailty

The Cox proportional hazards (Cox-PH) model is a semi-parametric regression method used to examine the relationship between survival time and explanatory variables (Cox, 1972). The hazard function $h(t \mid x)$ is modeled as:

$$h(t \mid x) = h_0(t) exp\,(\beta^{\top} x) \qquad (4.22)$$

where $h_0(t)$ is the baseline hazard function, x is a vector of covariates, and β is a vector of regression coefficients. The Cox model assumes the proportional hazards assumption, meaning the effect of covariates on the hazard is constant over time.

To handle heterogeneity in grouped data (e.g., trips by the same driver), the Cox-PH model can incorporate frailty terms:

$$h(t \mid x, u) = h_0(t) exp\,(\beta^{\top} x + u) \tag{4.23}$$

where $u \sim N(0, \sigma^2)$ is a random effect capturing unobserved heterogeneity (Wienke, 2011). Frailty models are particularly useful for survival data with correlated observations or clustered structures, such as repeated trips by the same driver.

3.2.5.3 Weibull AFT Model with Clustered Heterogeneity

The Weibull Accelerated Failure Time (AFT) model is a parametric approach to survival analysis that directly models the survival time T as a function of covariates:

$$log\,(T) = \beta^{\top} x + \varepsilon \tag{4.24}$$

where ε is a random error term and β represents the regression coefficients. The Weibull distribution is commonly used due to its flexibility in modeling hazard rates (e.g., increasing or decreasing over time), with the survival function given as:

$$S(t) = exp\,(-\lambda t^{\gamma}) \tag{4.25}$$

where λ and γ are scale and shape parameters, respectively.

To account for clustering, the AFT model incorporates random effects (Wang, 2006):

$$log\,(T_{ij}) = \beta^{\top} x_{ij} + u_i + \varepsilon_{ij} \tag{4.26}$$

where u_i represents cluster-level heterogeneity (e.g., driver-specific effects) and ε_{ij} captures individual-level variability. This extension is particularly suitable for datasets with clustered survival times, such as trips nested within drivers (Hougaard, 2000).

3.2.5.4 Random Survival Forest

Random Survival Forest (RSF) is a machine learning approach to survival analysis that extends random forests to time-to-event data. RSF constructs multiple decision trees using bootstrap samples of the data, and the survival function is estimated as the ensemble average of the survival probabilities across all trees. Each tree splits the data based on covariates to maximize survival differences between groups, using a splitting rule such as the log-rank split statistic (Ishwaran et al., 2008).

RSF can handle non-linear relationships, high-dimensional covariates, and complex interactions without requiring proportional hazards or parametric assumptions. It is particularly effective in datasets with heterogeneous risk factors and is robust to censoring. The cumulative hazard function for an individual i is estimated as:

$$H_i(t) = \frac{1}{B}\sum_{b=1}^{B} H_i^{(b)}(t) \tag{4.27}$$

where $H_i^{(b)}(t)$ is the cumulative hazard function from the b-th tree, and B is the total number of trees.

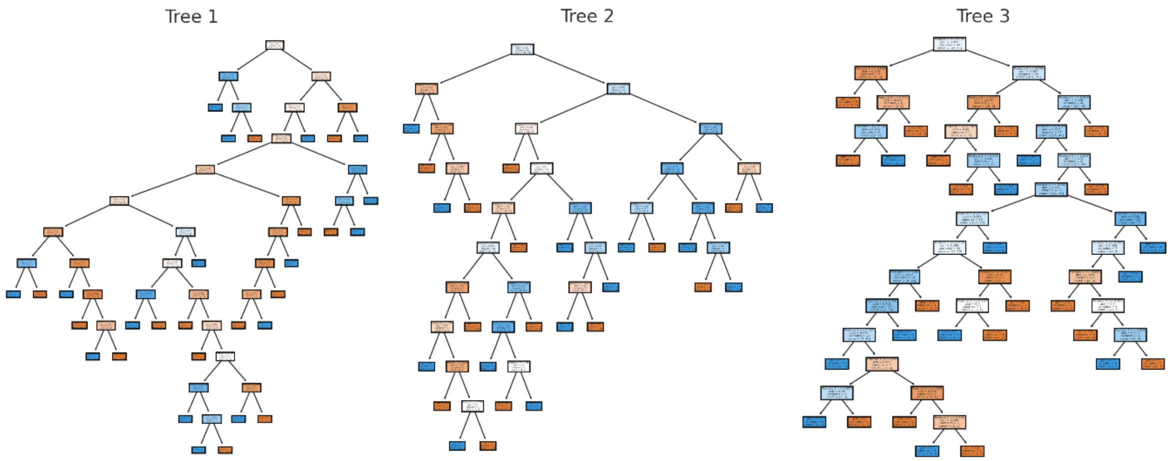


Figure 3.7: Example of tree structure in Random Survival Forest

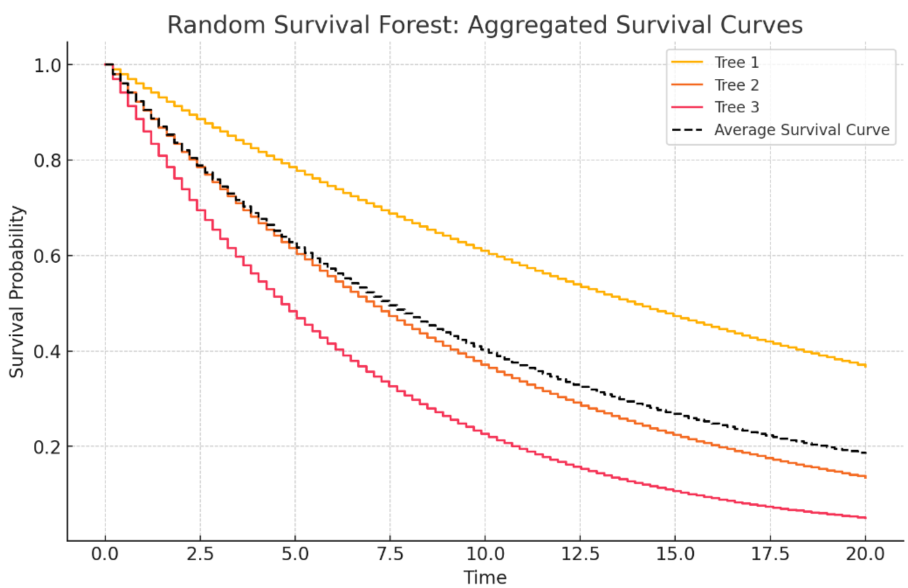


Figure 3.8: Example of survival curves in Random Survival Forest

Some key points from the examples above:

- Tree Structures: The first figure shows the decision paths of three individual trees within the RSF.

- Survival Curves: The second figure illustrates survival probabilities estimated by each tree and their aggregated average curve (dashed line).

RSF is increasingly popular in survival analysis applications, such as identifying high-risk drivers based on multiple behavioral and environmental factors, offering interpretability through variable importance measures and partial dependence plots.

3.2.5.5 Model evaluation

Survival models are typically assessed based on their ability to accurately capture the underlying relationships between covariates and survival times, while also accounting for censored observations. Model evaluation typically focuses on two key aspects: discrimination and calibration**.** Discrimination refers to a model's ability to correctly differentiate between individuals at higher or lower risk of experiencing the event, while calibration assesses how well the predicted survival probabilities align with the observed outcomes over time (Harrell, 2015).

For discrimination, commonly used metrics include the concordance index (C-index) and time-dependent receiver operating characteristic (ROC) curves. The C-index measures the proportion of concordant pairs, where an individual with a shorter survival time has a higher predicted risk, and typically ranges from 0.5 (no better than random guessing) to 1.0 (perfect discrimination):

$$C = \frac{\sum_{i,j} 1(x_i > x_j) \cdot 1(T_i < T_j)}{\sum_{i,j} 1(T_i < T_j)} \tag{4.28}$$

Where:

- x_i and x_j: Predicted risk scores for individuals i and j,
- T_i and T_j: Observed survival times,
- 1(·): Indicator function returning 1 if the condition is true, 0 otherwise.

Time-dependent ROC curves provide a more dynamic evaluation of the model's discriminatory power at different time points, offering insights into its performance across the study period. Calibration, on the other hand, is evaluated by plotting predicted survival probabilities against observed survival probabilities, often using a calibration curve. A well-calibrated model will show a close match between predicted and observed values along the diagonal line of the plot.

$$\textit{Calibration Curve: } S_{obs}(t) = f\left(S_{pred}(t)\right) \tag{4.29}$$

Where:

- $S_{pred}(t)$: Predicted survival probability at time t,
- $\boldsymbol{S_{obs}(t)}$**:** Observed survival probability at time t.

Beyond discrimination and calibration, model fit is another important criterion. For parametric models, such as the Weibull AFT model, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) (described in GLMM section) are widely used to compare competing models. Lower AIC or BIC values indicate a better balance between goodness-of-fit and model complexity. For semi-parametric models like the Cox-PH model, log-likelihood ratio tests are used to assess the significance of added covariates. Additionally, residual diagnostics, such as

Martingale residuals or Schoenfeld residuals, are essential for verifying assumptions like proportional hazards in the Cox model.

Martingale Residuals for testing functional form of covariates:

$$M_i = \delta_i - \hat{\Lambda}(T_i) exp\left(X_i^T \beta\right) \tag{4.30}$$

Where:

- δ_i: Event indicator (1 if event occurs, 0 if censored),
- $\hat{\Lambda}(T_i)$: Estimated cumulative hazard,
- $X_i^T \beta$: Linear predictor for individual i.

Schoenfeld Residuals for testing the proportional hazards assumption:

$$R_{ij} = X_{ij} - \frac{\sum_{l \in R(T_j)} X_{lj} exp\left(X_l^T \beta\right)}{\sum_{l \in R(T_j)} exp\left(X_l^T \beta\right)} \tag{4.31}$$

Where:

- R_{ij}: Risk set at time T_j,
- X_{ij}: Covariate value for individual i at time j.

For machine learning approaches like random survival forests, model evaluation often involves out-of-bag (OOB) error estimation, which provides a measure of prediction accuracy by testing each tree on data not used in its construction (Ishwaran et al., 2008).

$$OOB\ Error = \frac{1}{n} \sum_{i=1}^{n} L\left(S_{pred,i}, S_{true,i}\right) \tag{4.32}$$

Where:

- L: Loss function, often based on a survival-specific metric like integrated Brier score,
- $S_{pred,i}$: Predicted survival probability for individual i,
- $S_{true,i}$: True survival outcome for individual i.

4 Naturalistic Driving Experiment

4.1 Experimental Framework

As already mentioned, the primary objective of this dissertation is to examine the driver behavior telematics feedback mechanism and its effect on driving behavior. The dissertation adopts a comprehensive approach to explore feedback's role in modifying driving behavior, with a focus on three key pillars:

1. Assessing the impact of feedback on driving behavior of different road user groups (car drivers, professional van drivers, and motorcyclists) in various road environments.
2. Investigating the effects of different feedback features across the experimental phases.
3. Analyzing the long-term, post-feedback effects on driving behavior.

To achieve these objectives, a naturalistic driving experiment was thoroughly designed and implemented, building on the results of the literature review regarding the experimental framework of existing naturalistic studies. The importance of the experimental framework of feedback-oriented study has been emphasized in scientific research, as it serves the foundation for producing valid, reliable and interpretable results (Cash et al., 2016; Leik, 1997). In the context of naturalistic studies, a well-constructed experimental framework allows researchers to isolate the specific effects of various feedback mechanisms on driving behavior (van Schagen & Sagberg, 2012). In the present Dissertation, the experimental framework is discussed across the following key aspects: experimental design, recruitment process, sample characteristics (size, type of vehicle), and feedback phases (i.e., number and content of phases).

Table 4.1: Framework of the naturalistic driving experiment

Experimental design	Feedback phases	Duration
one group / within-subjects design	• baseline • scorecard • maps • peer comparison • competitions • no feedback	21 months

4.1.1 Experimental design

Naturalistic driving studies primarily employ two experimental designs: within-subjects and between-subjects designs. In a within-subjects design, the same participants are exposed to multiple conditions or interventions, allowing researchers to assess changes within individuals over time. All participants experience multiple phases of feedback and act as their own control group (Birrell & Fowkes, 2014; Hari et al., 2012). Conversely, a between-subjects design involves assigning different participants to separate groups, each exposed to distinct conditions. Studies where different groups of drivers received different types of feedback or no feedback at all (control group) through one experimental phase (Farah et al., 2014; Reagan et al., 2013; Rolim et al., 2016; Stevenson et al., 2021), allowing researchers to examine and analyze driving behavior across different conditions of feedback influence. This design eliminates carryover effects and simplifies the study structure but requires larger sample sizes to account for variability between participants.

The present naturalistic driving experiment constitutes a within-subject experimental design, which was deliberately chosen to align with the primary objectives of this dissertation. This design enables the examination of changes in individual driving behavior outcomes as drivers are exposed to different feedback mechanisms across multiple experimental phases. By using a within-subject design, each participant serves as their own control, allowing for a more precise assessment of the direct effects of telematics feedback on driving behavior. This approach is particularly well-suited for isolating the impact of feedback, as it minimizes the influence of individual differences, such as baseline driving skills, attitudes, or risk tolerance, which could confound results in a between-subjects design (Newnam et al., 2014; Toledo & Shiftan, 2016). Furthermore, it ensures that the observed behavioral changes can be attributed more reliably to the feedback interventions rather than external factors.

The within-subject experimental design also supports the comprehensive goals of the study, which include evaluating feedback's effects across various road user groups, phases, and long-term contexts. This design facilitates a nuanced exploration of how feedback modifies behavior over time and under different conditions, such as varying road environments and vehicle types. It is particularly effective in capturing the longitudinal effects of feedback, enabling the analysis of both immediate and post-feedback impacts.

4.1.2 Recruitment process

The official launch of the naturalistic driving experiment took place on July 1, 2019. The participant onboarding process was conducted via an automated email system sent to prospective participants and included the following steps:

- 1st Email to Prospective Participant: An invitation to participate in the BeSmart research project, along with a brief description of the experiment. The Participant Information Sheet was attached.
- 2nd Email: After the prospective participant expressed interest in participating in the BeSmart experiment, they were sent the Participant Consent Form, which they were required to complete in order to join the experiment.
- 3rd Email: Once the prospective participant completed the Participant Consent Form, they received a unique username and a download link for the app (available in Android and iOS versions). The BeSmart App User Guide was also attached.

The form of Consent for the participants can be found in Annex II. Additionally, OSeven's team created an application usage guide specifically for the research team, aimed at providing a more comprehensive understanding of the application's functionality to address potential participant technical questions. It is further emphasized that all principles, requirements, and recommendations of the General Data Protection Regulation (GDPR) were followed during the recruitment of prospective participants.

The majority of participants were recruited through the project team's contact networks, which facilitated effective collaboration and communication throughout the experiment. Regarding the recruitment of a sufficient number of professional drivers, the team collaborated with "Nea Odos," which agreed to participate by providing its professional drivers as participants, believing that proper driver education and awareness could play a vital role in road safety.

4.1.3 Sample distribution and characteristics

Data regarding invited prospective participants per mode of transportation, gathered by the research team at NTUA responsible for sending the invitations, is presented in the following table.

Table 4.2: Invited Prospective Participants by Mode of Transportation

Mode of Transportation	Number of Invited Candidates
Passenger Vehicles	260
Professionals (Car/Van)	80
Motorcycles	55
Bicycles	10
Total	**405**

Data on application users, provided by OSeven's backend team, is shown in the table Table 4.3.

Table 4.3: BeSmart Application Users

BeSmart App Installations		
Ios	Android	
13	238	
BeSmart Environment		
Total Users	At least one Signin	At least one trip
225	225	208
BeSmart Moto Environment		
Total Users	At least one Signin	At least one trip
26	26	22

Finally, the sample of active participants per mode of transportation is shown in the following table Table 4.4.

Table 4.4: Sample Distribution by Mode of Transportation

Mode of Transportation	Number of Drivers	Percentage of Total Drivers
Passenger Vehicles	176	76.5%
Professionals (Car/Van)	27	11.7%
Motorcycles	22	9.6%
Bicycles	5	2.2%
Total	**230**	**100%**

4.1.4 Experimental phases

The experiment was divided into six phases, each providing different types of feedback to drivers. Figure 4.1 displays the different feedback features provided at each phase. As it is shown, Phase 1 served as the baseline phase where drivers were recorded through the smartphone app and only a trip list was available with no other information about the driver behavior. In phases 2,3,4 and 5 different feedback features were added, as shown below, while drivers returned to no feedback in phase 6 for researchers to examine the post feedback effect.

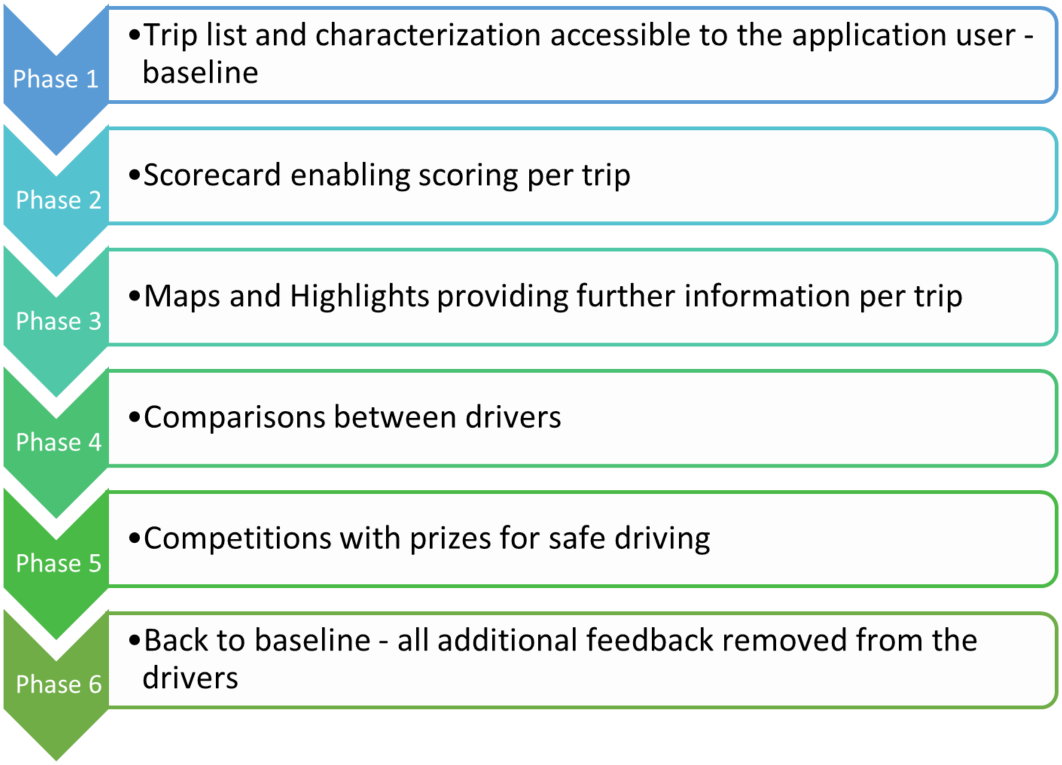


Figure 4.1: Description of the experiment phases

The initial plan was to conduct each phase over two months, culminating in a total study duration of one year. However, the onset of the COVID-19 pandemic required the research team to adapt to the unprecedented circumstances. Due to government-imposed lockdowns and restrictions on movement, particularly between March 23, 2020, and May 4, 2020, the number and frequency of trips made by participants dropped significantly, with up to a 70% reduction in some cases. As a result, it became necessary to extend the experiment beyond the initially planned twelve months. The research team decided to prolong Phase 4, allowing participants to resume regular driving patterns as restrictions eased. Despite the easing of lockdown measures in May 2020, the travel activity remained lower than anticipated. Consequently, the start of the next phase (Phase 5 - Competitions and Challenges) was delayed until October 2020. The impact of COVID-19 both on mobility and road safety has been discussed in recent studies with very interesting insights (Du et al., 2024; Katrakazas et al., 2021).

Table 4.5: Naturalistic driving experiment timetable

Date	**Experiment Phase**	**Feedback feature No1**	**Feedback feature No2**	**Feedback feature No3**	**Feedback feature No4**	**Feedback feature No5**
Jul-19	Phase 1 – No feedback					
Aug-19	Phase 1 – No feedback					
Sep-19	Phase 1 – No feedback					
Oct-19	Phase 2 – Feature 1	Scorecard				
Nov-19	Phase 2 – Feature 1	Scorecard				
Dec-19	Phase 3 – Feature 1+2	Scorecard	Maps and Highlights			
Jan-20	Phase 3 – Feature 1+2	Scorecard	Maps and Highlights			
Feb-20	Phase 4 – Feature 1+2+3	Scorecard	Maps and Highlights	Comparisons		
Mar-20	Phase 4 – Feature 1+2+3	Scorecard	Maps and Highlights	Comparisons		
Apr-20	Phase 4 – Feature 1+2+3	Scorecard	Maps and Highlights	Comparisons		
May-20	Phase 4 – Feature 1+2+3	Scorecard	Maps and Highlights	Comparisons		
Jun-20	Phase 4 – Feature 1+2+3	Scorecard	Maps and Highlights	Comparisons		
Jul-20	Phase 4 – Feature 1+2+3	Scorecard	Maps and Highlights	Comparisons		
Aug-20	Phase 4 – Feature 1+2+3	Scorecard	Maps and Highlights	Comparisons		
Sep-20	Phase 4 – Feature 1+2+3	Scorecard	Maps and Highlights	Comparisons		
Oct-20	Phase 5 – Feature 1+2+3+4	Scorecard	Maps and Highlights	Comparisons	Competitions	
Nov-20	Phase 5 – Feature 1+2+3+4	Scorecard	Maps and Highlights	Comparisons	Competitions	
Dec-20	Phase 5 – Feature 1+2+3+5	Scorecard	Maps and Highlights	Comparisons		Challenges
Jan-21	Phase 5 – Feature 1+2+3+5	Scorecard	Maps and Highlights	Comparisons		Challenges
Feb-21	Phase 6 – No feedback					
Mar-21	Phase 6 – No feedback					
Apr-21	End of the experiment					

Furthermore, Figure 4.2 presents example screenshots from the application features in all experiment phases, illustrating the progression of feedback mechanisms provided to users. Phase 1 and Phase 6 represent control phases with no feedback, showcasing only basic trip details. Phase 2 introduces the Scorecard feature, offering users a performance score and visual feedback on driving behavior metrics. Phase 3 highlights the Maps feature, where users can view trip routes alongside key driving event markers, enabling a spatial understanding of their behaviors. Phase 4 includes the Comparison feature, allowing users to evaluate their performance against past trips or peer averages. Phase 5 showcases the Competitions and Challenges feature, incorporating gamified elements like ranking systems, points, and personalized challenges to motivate safer driving practices.

Figure 4.2. Example screenshots from the application features in all experiment phases

4.2 Smartphone Application

4.2.1 Data collection system

For the purpose of the experiment, an innovative smartphone application, developed by OSeven Telematics (www.oseven.io), was utilized to assess and improve driver behavior and safety. This application records driver behavior using the smartphone's hardware sensors and various APIs to collect sensor data.

Smartphones are equipped with a variety of sensors (Bluetooth, GPS receiver, GSM/GPRS connectivity, 3D accelerometer, gyroscope, magnetometer, proximity sensor, compass, barometer, etc.). The BeSmart application is designed using the latest software advancements, providing users with the best possible experience. The app is activated periodically by the Android and iOS operating systems, collecting position and movement data for a few seconds to determine if the user is in a vehicle. This process is called "Driving Detection." If the app verifies that the user is in a vehicle, it begins recording primary data. Otherwise, data collection stops, and the app deactivates until the system reactivates it next time.

The frequency of data recording varies depending on the type of sensor, with a minimum rate of 1 Hz. It is worth noting that this method enables the collection of a large amount of driving characteristic data solely through the mobile phone application. The journey recording continues even if the vehicle is idle for five minutes to account for scenarios where the driver may resume a trip after a brief stop.

The primary data collected include the following:

- Date and time
- GPS data (longitude, latitude, altitude, speed, horizontal accuracy, altitude accuracy)
- Accelerometer data
- Gyroscope data (Yaw, Pitch, Roll)
- Smartphone orientation data
- Activity data (e.g., walking, driving, stationary)
- Screen status (without access to screen content)

A variety of APIs (Application Programming Interfaces) are used to read the recorded sensor data and temporarily store it in the smartphone's database before transferring it to the central database. All additional information collected after driving ends is deleted. Figure 4.3 summarizes the stages of data flow.

Figure 4.3: The OSeven data flow system

4.2.2 Data processing and feature engineering

To process the primary data for extracting driver behavior and vehicle usage information (e.g., speeding, mobile phone use, sudden accelerations and braking, user scores), specialized software from OSeven periodically retrieves the Primary Data from the upload area and begins processing it to calculate driving behavior and vehicle usage data. For this purpose, OSeven has servers located in the data centers of the aforementioned cloud service providers within the EU where this software is executed.

Advanced machine learning (ML) and data fusion algorithms are used for processing the recorded data. The following procedures are carried out:

- Data filtering and cleaning: Detecting outlier values (data deemed unreliable is discarded). Advanced signal processing techniques remove noise from the raw data, retaining only information relevant to driving behavior. Data smoothing is also performed for parameters requiring it (e.g., when abnormal deviations are observed).

- Speeding detection: Speed limits are calculated based on map data from providers such as Google and OSM. Speed limit violations and their duration are then computed.

- Recognition of sudden events: This includes detecting sudden accelerations, braking, and turns, accounting for the aggressiveness and intensity of the events.

- Mobile phone use detection: This includes calls, texting, and browsing. A precise detection algorithm has been developed, measuring phone usage based on data fusion and ML algorithms. These algorithms use only smartphone sensor data, without accessing other app activity for data privacy protection.

- Identification of dangerous driving hours: Distance traveled between midnight and 5 a.m. is tracked, as these hours are considered risky due to potential alcohol consumption and driver fatigue.

- Driver vs. passenger identification: During the trip, the system detects whether the user is the driver or a passenger and identifies the mode of transport (car, motorcycle, public transport, bicycle).

- Scoring model: The primary data is processed and converted into metadata for use in the scoring model. Calculated metadata includes risk exposure and driving behavior indicators: total trip distance, driving duration, road type, time of day (peak hours, risky hours), trip purpose (specified by the driver), speed limit violations (duration and extent), number and severity of sudden events, driving aggressiveness (e.g., braking, acceleration), and mobile phone distractions (Figure 4.4).

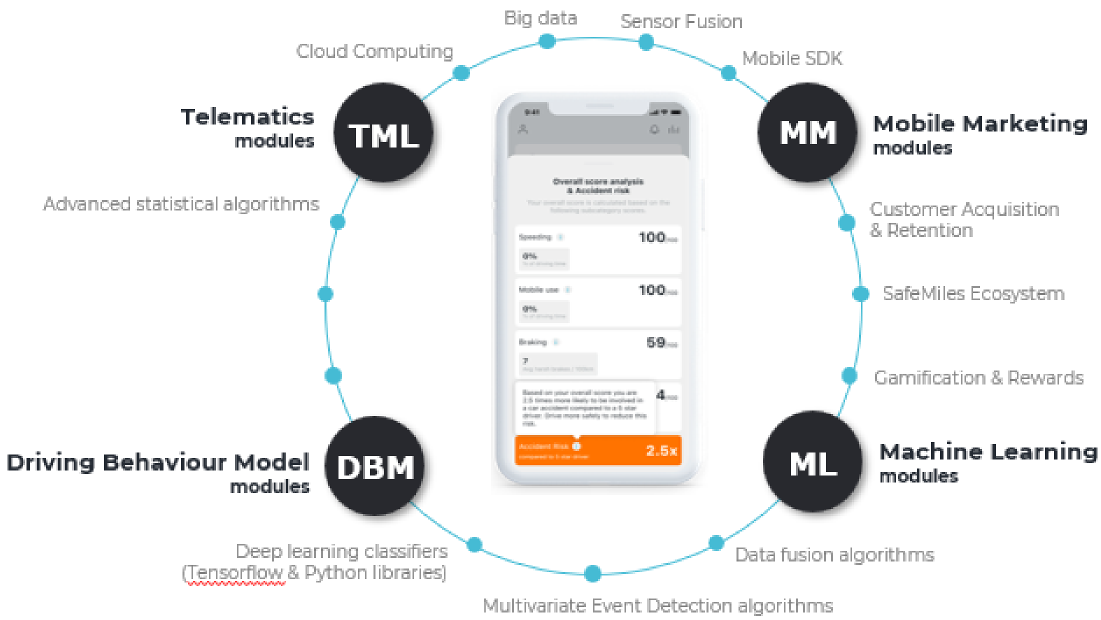


Figure 4.4: OSeven Telematics Platform Overview

Additionally, a separate environment was developed within the app for users who primarily use motorcycles, after identifying sensor limit issues in recording and evaluating motorcycle driving. Key differences in the upgraded version from the original BeSmart version include a different weighting factor for the score concerning the frequency of sudden events and the exclusion of mobile phone use from trip scoring categories. Data is analyzed, and the scoring system is calibrated based on the overall sample.

4.2.3 Metadata and driver behavior indicators

The metadata is then transferred to the server for statistical calculations per user. The results of this entire process are accessible via the BeSmart mobile app, enabling users to view all detected events and their locations on the map. This provides drivers with a user-friendly way to identify trip segments with risky driving behavior and encourages them to avoid similar behaviors in the future.

A variety of different metadata are eventually calculated, including the following exposure indicators:

- Total distance (mileage)
- Driving duration
- Type(s) of the road network used (given by GPS position and integration with map providers e.g. Google, OSM)
- Time of the day driving (rush hours, risky hours)
- Weather conditions (under development, on the basis of integration
- with weather data providers)
- Trip purpose (set by the driver himself by using the smartphone app)

The driving behavior indicators that are also calculated from the data include:

- Speeding (duration of speeding, speed limit exceedance etc.)
- Number and severity of harsh events
- Harsh braking (longitudinal acceleration)
- Harsh acceleration (longitudinal acceleration)
- Distraction from mobile phone use (mobile phone use is considered to be any type of phone use by the driver e.g. talking, texting etc.).

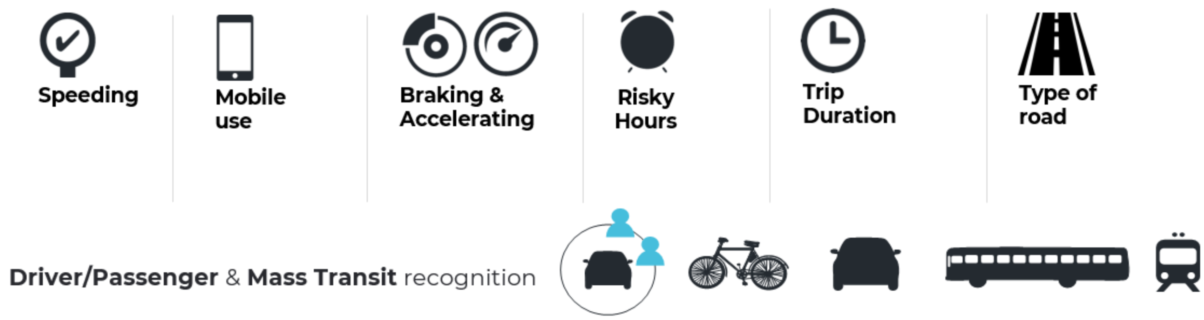


Figure 4.5: OSeven Driving Behaviour Scoring

It should be noted that harsh events are calculated via data fusion and machine learning algorithms and not a rule-based approach using as input the values of the accelerometer as well as values from additional sensors (e.g. orientation, magnetometer, GPS, gyroscope). Therefore, the determination of the harsh events is not based on specific thresholds. Yet, some indicative examples of speed and acceleration data related with specific harsh events from the available dataset are illustrated, so that harsh events can be better comprehended. Indicative harsh accelerations: (i) speed increase from 31km/h to 40km/h within one second and (ii) longitudinal acceleration 0.28g; Indicative harsh brakes: (i) speed decrease from 51km/h to 40km/h within one second and (ii) longitudinal deceleration 0.30g. It is stressed out that these are four indicative cases of harsh events and the respective values cannot be considered as thresholds, as the determination of the events is based on the coevaluation of several time series. In addition, the reliability of the OSeven algorithms has been extensively evaluated against literature data, OBD data, on-road experiments by certified experts on the assessment of driving behavior, and experiments on driving simulators (Tselentis et al., 2019; Petraki et al., 2020). Taking into consideration that the cited studies deal with car drivers, the authors clarify that the mentioned algorithms have been calibrated based on (i) on-road annotated experiments, and (ii) the comparison between the car and motorcycle data that quantify the vehicle dynamics.

These indicators along with other data (e.g. from map providers) can be subsequently exploited to calculate individual driver statistics, on all road networks (urban, rural, highway, etc.) and under various driving conditions, enabling the creation of a large database of individual trip/driver characteristics.

4.3 Self-reported Questionnaire Data

4.3.1 Structure and content

A Driver Behavior Questionnaire was developed simultaneously (see Annex II). Each participant is asked to complete a questionnaire about their driving habits and driving behavior. The questions were carefully selected based on existing literature regarding self-reported driver behavior. The sections of the questionnaire are as follows:

A. Driving Experience – Journeys
B. Vehicle
C. Driving Behavior
D. Demographic Information

The driving experience section includes questions about participants' driving experience and habits, which will be used in the analyses as potential critical factors in evaluating the participants' driving performance. This section also includes questions about participants' driving experience over different time frames, e.g., frequency of trips on a daily, weekly basis, etc., thus providing more detailed information on their driving experience.

The second section of questions concerns the characteristics of the vehicle driven by participants, such as the age and engine size of the vehicle, as well as average fuel consumption during trips. The vehicle is one of the three factors impacting road safety, alongside the road environment and the human factor.

The driving behavior section consists of two parts: the first addresses the participants' accident history over their lifetime as drivers, while the second involves the participants' self-assessment of their driving skills and performance. This section is very important, as the information provided here will be correlated with the drivers' performance in the various phases of the driving experiment.

The final section includes demographic questions that are also relevant for the subsequent analyses to be conducted throughout the project.

Additionally, it is worth noting that four different questionnaires were developed, one for each vehicle type (passenger vehicles, professional drivers, motorcycles, bicycles), with corresponding adjustments in the questions to address the different modes of transportation. After the questionnaires were created in paper format, it was decided to use the online platform SurveyMonkey, with OSeven handling the online questionnaire administration.

4.3.2 Statistical summary

The following tables contain descriptive statistics for the questionnaire variables collected from the exploited BeSmart car driver sample, amounting to 87 drivers in total who completed the questionnaire: Table 4.6 contains the discrete variables &

Table 4.7 contains the continuous variables.

The Table 4.6 present descriptive statistics of questionnaire responses.

Table 4.6: Discrete descriptive statistics of the questionnaire responses

Variable	**Classes: Responses (Percentage)**						
	1:	**2:**	**3:**	**4:**	**5:**	**6:**	**7:**
days_per_week	1 (1.15%)	1 (1.15%)	2 (2.30%)	6 (6.90%)	15 (17.24%)	21 (24.14%)	41 (47.13%)
km_per_week	3 (3.45%)	21 (24.14%)	29 (33.33%)	19 (21.84%)	15 (17.24%)		
daily_trips	3 (3.45%)	36 (41.38%)	25 (28.74%)	13 (14.94%)	10 (11.49%)		
km_daily_avg	17 (19.54%)	1 (1.15%)	27 (31.03%)	17 (19.54%)	8 (9.20%)	17 (19.54%)	
km_year	5 (5.75%)	13 (14.94%)	30 (34.48%)	14 (16.09%)	25 (28.74%)		
vehicle_ownership	65 (74.71%)	19 (21.84%)	0 (0.00%)	2 (2.30%)	1 (1.15%)		
cc	14 (16.09%)	2 (2.30%)	17 (19.54%)	31 (35.63%)	19 (21.84%)	1 (1.15%)	3 (3.45%)
vehicle_age	14 (16.09%)	16 (18.39%)	37 (42.53%)	20 (22.99%)			
fuel_cons_avg	9 (10.34%)	4 (4.60%)	32 (36.78%)	29 (33.33%)	13 (14.94%)	0 (0.00%)	
tickets_3y	63 (72.41%)	11 (12.64%)	5 (5.75%)	3 (3.45%)	5 (5.75%)		
declared_speeding	4 (4.60%)	26 (29.89%)	43 (49.43%)	13 (14.94%)	1 (1.15%)		
declared_mbu	1 (1.15%)	14 (16.09%)	13 (14.94%)	31 (35.63%)	28 (32.18%)		
declared_harsh_brk	7 (8.05%)	43 (49.43%)	30 (34.48%)	7 (8.05%)	0 (0.00%)		
declared_harsh_acc	11 (12.64%)	3 (3.45%)	44 (50.57%)	29 (33.33%)	0 (0.00%)		
declared_harsh_turn	27 (31.03%)	3 (3.45%)	44 (50.57%)	13 (14.94%)	0 (0.00%)		
declared_aggressive	33 (37.93%)	30 (34.48%)	19 (21.84%)	3 (3.45%)	2 (2.30%)		
declared_careful	0	0	14	44	29		

Variable	Classes: Responses (Percentage)					
	(0.00%)	(0.00%)	(16.09%)	(50.57%)	(33.33%)	
education	**1:** 48 (55.17%)	**2:** 26 (29.89%)	**3:** 3 (3.45%)	**4:** 4 (4.60%)	**5:** 6 (6.90%)	
smartphone_familiarity	**1:** 0 (0.00%)	**2:** 3 (3.45%)	**3:** 13 (14.94%)	**4:** 25 (28.74%)	**5:** 46 (52.87%)	
gender	**1:** 48 (55.17%)	**2:** 39 (44.83%)	**3:** 0 (0.00%)			
age	**1:** 1 (1.15%)	**2:** 7 (8.05%)	**3:** 54 (62.07%)	**4:** 14 (16.09%)	**5:** 8 (9.20%)	**6:** 3 (3.45%)
marital_status	**1:** 3 (3.45%)	**2:** 24 (27.59%)	**3:** 59 (67.82%)	**4:** 1 (1.15%)		

From Table 4.6, it is evident that the considered driver sample is quite active, with the majority using their vehicles 5 to 7 days per week. The majority of driving trips appear to have a non-trivial length, i.e. more than 6 kilometers daily and more than 20 kilometers weekly, while there is also a modest representation of other daily and weekly trip lengths. This indicates a well-rounded sample of both shorter and longer trips for drivers, indicative of moving predominantly within a large urban environment.

The driver sample has good representation of men and women, while most drivers belong to the 25-34 and 35-44 years old age categories. In combination with modest to full familiarity with smartphones, this outlines a more technology-oriented driver sample, rather than traditional drivers. Another interesting insight refers to self-assessment of drivers: While the majority of drivers self-report that they sometimes or often engage in speeding, harsh accelerations and harsh turning/cornering, only a minority self-characterize as aggressive (the majority (strongly) disagrees with the characterization). At the same time, they are mostly inclined to self-characterize as careful drivers, despite self-reporting a very frequent use of their mobile phones when driving, though this can also include more passive functions such as GPS-issued directions.

Table 4.7: Continuous variable descriptive statistics of the questionnaire responses

Variable	Min	Median	Max	St. Dev.
licence_year	1978	2009	2019	7.758
driving_exp	1	10	42	8.628
crashes_until_now	0	1	5	1.344
crashes_3y	0	0	3	0.776
crashes_victims_until_now	0	0	3	0.471
crashes_victims_3y	0	0	1	0.255
crashes_pdo_until_now	0	0	6	1.270
crashes_ pdo_3y	0	0	5	1.074
family_members	0	3	5	1.368

In addition to the previous, it can be observed from

Table 4.7 that driving experience was more or less compatible with license year, indicating practiced drivers that were active after obtaining their licenses. In 2019, when the questionnaires were completed, the median value for driving experience was 10 years, denoting a driving sample that is moderately experienced. Most drivers were not involved in crashes of any type, as hinted by the fact that the median drops to zero when separating property damage only (PDO) crashes from crashes with victims (i.e. fatalities or injuries).

4.4 Big Data Processing

4.4.1 Data integration and organization

After recording, collection, storage, and processing of data by OSeven's integrated system, as detailed in the previous chapter, the next step was to organize the large-scale data to be suitably utilized as the database for statistical analyses conducted by the NTUA research team.

The resulting database was provided in Microsoft Excel files to enable data processing by the project's research team. Specifically, there were two different types of databases, each referring to distinct driving characteristics, to be used in various analyses aimed at investigating complex relationships among different groups of variables, with the ultimate goal of assessing driving behavior and relevant interventions.

Alongside the integrated data transfer system by OSeven, the Postman platform was utilized by the NTUA research team to extract a small amount of data for investigation and analysis purposes.

Postman is an API platform for creating and using APIs. The Postman platform simplified and improved collaboration between the two teams, enabling better and faster data retrieval. In summary, the NTUA team was able to send requests to the OSeven backend system to retrieve data from drivers' trips for specific time periods. Additionally, two environments were created: one for the BeSmart application (car drivers) and one for the BeSmart MotorBike application (motorcyclists). A snapshot from the Postman platform is shown in Figure 4.6: Snapshot from the Postman Platform.

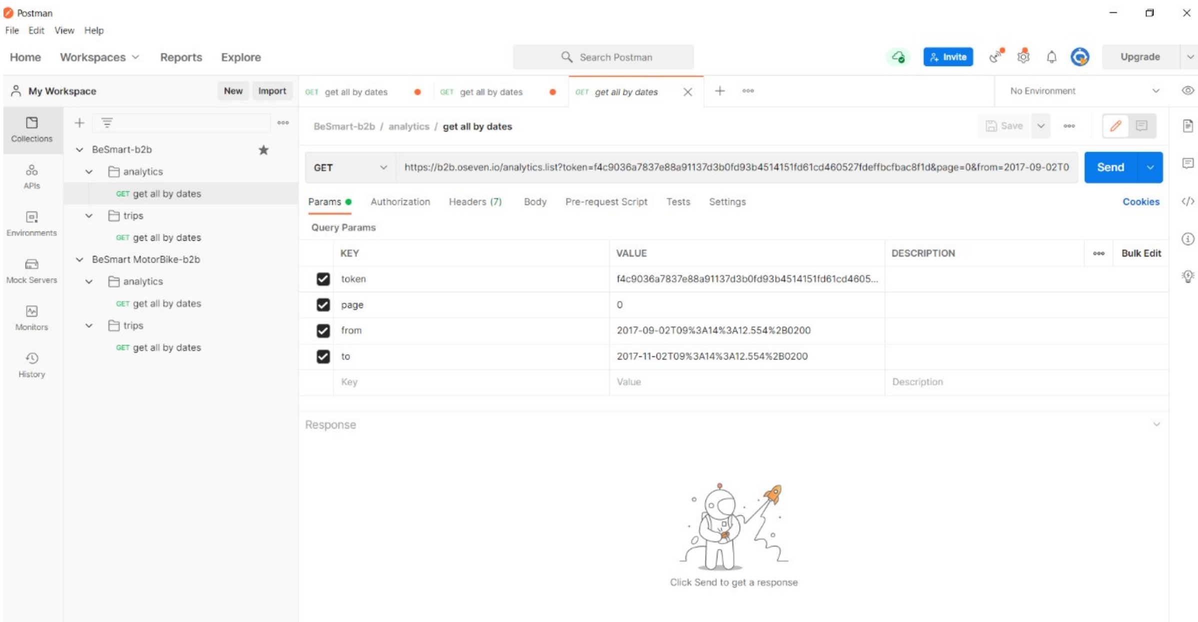


Figure 4.6: Snapshot from the Postman Platform

4.4.2 Database structure

One database focused on drivers' travel characteristics. Each row represented a driver's journey, while each column contained a variable for each journey. Some of the variables included in this database were as follows:

- **tripid**: trip code
- **userid**: driver code
- **duration**: total trip duration in seconds (s)
- **driving_duration**: net driving time excluding stops, in seconds (s)
- **totaldist**: total trip distance in kilometers (km)
- **ha**: number of harsh accelerations (absolute count)
- **hb**: number of harsh decelerations (absolute count)
- **hc**: number of sharp turns (absolute count)
- **ha_intensity_high**: high-intensity harsh acceleration
- **ha_intensity_low**: low-intensity harsh acceleration
- **ha_intensity_medium**: medium-intensity harsh acceleration
- **avaccel**: average acceleration (km/h/s)
- **avdecel**: average deceleration (km/h/s)
- **smooth_corner**: average turning speed (degrees/s)
- **mobileusage**: percentage of driving_duration where the driver used a mobile phone
- **perc_speeding**: percentage of driving_duration spent above the speed limit
- **av_speeding**: average percentage of speed limit excess
- **riskyhourdistance**: driving distance between 22:00 and 05:00 (km)

Additionally, for better management of large-scale data, supplementary databases were created for each experimental phase and vehicle type to enhance their effective use in statistical analyses.

Finally, the collection and entry of personal data into an MS Excel file was carried out by members of the research team, with the file being password-protected. Once personal data collection and entry were completed, only the Project Coordinator and Data Management Officer had access to this file. Each participant was assigned a unique code with no association to their personal data. The research team analyzed driving behavior data using these anonymous codes and selected demographic data that did not allow for identification of the individual. This ensured that driving behavior data could not be linked to individuals by research team members who no longer had access to the Personal Data File.

4.5 Descriptive Statistics

Before proceeding with the model development, it is essential to examine the descriptive statistics of the six experimental phases for the three driver groups: car drivers, professional drivers, and motorcyclists. These statistics provide a preliminary understanding of the effects of the different feedback interventions on key driving behavior indicators, including speeding percentage, mobile phone use, harsh accelerations, and harsh braking. (Table 4.8).

Table 4.8: Descriptive statistics of the per driver values of the recorded driving behavioral indicators (mean value and the respective standard deviation in parenthesis)

Speeding percentage (%)	**Phase 1**	**Phase 2**	**Phase 3**	**Phase 4**	**Phase 5**	**Phase 6**
Car drivers	5.98 (0.05)	3.66 (0.03)	3.63 (0.04)	4.38 (0.04)	3.25 (0.04)	3.88 (0.04)
Professional drivers	-	1.61 (0.02)	0.78 (0.01)	0.96 (0.01)	1.05 (0.01)	2.26 (0.01)
Motorcyclists	10.42 (0.09)	7.91 (0.09)	9.58 (0.12)	9.23 (0.09)	7.86 (0.08)	9.09 (0.07)
Mobile use percentage	**Phase 1**	**Phase 2**	**Phase 3**	**Phase 4**	**Phase 5**	**Phase 6**
Car drivers	4.20 (0.06)	3.58 (0.05)	3.89 (0.06)	3.66 (0.06)	3.03 (0.07)	2.83 (0.04)
Professional drivers	-	0.80 (0.01)	0.76 (0.01)	0.82 (0.01)	1.00 (0.02)	1.44 (0.01)
Motorcyclists	-	-	-	-	-	-
Harsh accelerations per 100km	**Phase 1**	**Phase 2**	**Phase 3**	**Phase 4**	**Phase 5**	**Phase 6**
Car drivers	9.31 (9.76)	9.01 (8.21)	10.62 (9.89)	10.51 (11.33)	8.81 (9.49)	8.33 (5.61)
Professional drivers	-	0.58 (1.07)	0.59 (0.98)	1.04 (1.51)	0.42 (0.65)	2.09 (1.21)
Motorcyclists	48.38 (29.77)	19.76 (19.66)	28.53 (28.83)	19.26 (14.74)	8.27 (10.30)	12.24 (13.83)
Harsh braking per 100km	**Phase 1**	**Phase 2**	**Phase 3**	**Phase 4**	**Phase 5**	**Phase 6**

Car drivers	18.04 (14.14)	17.35 (11.94)	18.31 (10.70)	17.35 (11.90)	12.08 (9.94)	13.71 (8.31)
Professional drivers	-	2.57 (3.39)	2.24 (2.89)	2.90 (3.04)	1.37 (2.13)	4.79 (1.58)
Motorcyclists	31.36 (24.36)	26.38 (22.27)	39.95 (41.94)	33.81 (33.81)	13.68 (9.77)	18.18 (9.80)

Speeding Behavior Across Phases

Speeding is a critical indicator of risky driving behavior. Car drivers exhibit a reduction in speeding percentage during the initial feedback phases, decreasing from 5.98% in Phase 1 (baseline) to 3.63% in Phase 3, where map-based feedback was introduced. However, a slight increase is observed in later phases, with 4.38% in Phase 4 and 3.88% in Phase 6, suggesting a potential relapse effect after the feedback interventions were removed. Professional drivers, who entered the study from Phase 2 onward, demonstrated significantly lower speeding percentages than car drivers, starting at 1.61% in Phase 2 and gradually increasing to 2.26% in Phase 6. Motorcyclists initially exhibited a much higher speeding percentage (10.42% in Phase 1), which decreased in subsequent phases but fluctuated between 7.86% and 9.23% in later phases.

Mobile Phone Use Trends

Mobile phone use while driving is a major distraction and safety risk. Car drivers exhibited a steady decrease in mobile use, from 4.20% in Phase 1 to 2.83% in Phase 6, indicating a potential positive impact of feedback interventions in reducing this risky behavior. Professional drivers consistently showed very low percentages of mobile phone use, starting from 0.80% in Phase 2 and reaching 1.44% in Phase 6. This might be attributed to the nature of their driving behavior, as they may be more aware of company policies or safety regulations. Motorcyclists had no recorded mobile phone use, which is likely due to the limitations in measuring phone use while riding a motorcycle.

Harsh Acceleration Patterns

Harsh accelerations per 100km serve as an indicator of aggressive driving tendencies. Car drivers started with 9.31 harsh accelerations per 100km in Phase 1, which slightly fluctuated but ultimately decreased to 8.33 in Phase 6. Professional drivers showed much lower values than car drivers, averaging below 2 harsh accelerations per 100km in all phases, indicating that this driver group tends to engage in less aggressive acceleration. Motorcyclists displayed significantly higher initial values (48.38 in Phase 1), which decreased over time but remained relatively high compared to other driver groups.

Harsh Braking Trends

Harsh braking per 100km is another critical risk factor in evaluating driving safety. Car drivers demonstrated a relatively stable pattern, beginning at 18.04 in Phase 1 and declining to 13.71 in Phase 6. Professional drivers had much lower harsh braking values, starting at 2.57 in Phase 2 and peaking at 4.79 in Phase 6. Motorcyclists initially displayed high values (31.36 in Phase 1), with significant fluctuations across phases, reaching their highest at 39.95 in Phase 3 before gradually decreasing.

These descriptive statistics highlight the changes in driving behavior across phases and provide a foundation for further analysis using statistical modeling.

5 Feedback Impact on Driver Speeding and Distracted Behavior

5.1 Feedback Impact on Speeding Behavior of Motorcyclists of Phases 1 & 2

5.1.1 Introduction

Motorcyclists constitute a vulnerable road user group with up to 30 times higher fatality rates compared to passenger cars (Johnson et al., 2008). In 2017, motorcyclists accounted for 18% of the total number of road deaths in the EU countries; specifically, about 3,850 users (riders and passengers) of motorcycles and about 600 riders of mopeds were killed in EU countries in road traffic accidents (European Commission, 2018). During 2017, Greece had the highest rate of motorcycle fatalities per million population in EU-28 Countries, 20.1 deaths, while the EU average was 7.5 fatalities per 1 million population (European Commission, 2018). Although motorcyclists represent approximately the 15% of total vehicles in Greece, they were involved in 36% of total road accidents in 2018 (ELSTAT, 2020).

Specific factors affecting the accident injury severity of motorcyclists have been determined in the literature: Albalate & Fernandez-Villadangos (2010) identified gender, excess speed, road width, and alcohol consumption as factors affecting powered two-wheeler (PTW) injury severity. Theofilatos & Ziakopoulos (2018) determined that traffic and speed variations increase PTW injury severity, while increased truck proportions in the traffic mix were found to relatively reduce injury severity, possibly due to behavioral adaptations on behalf of PTW riders.

Behavioral issues are major moderating factors to both frequency and severity of motorcycle accidents. Speeding, sensation seeking, aggressiveness, perceived risk, errors, violations and attitudes towards road safety are considered to be crucial behavioral risk factors (Vlahogianni et al., 2012; Theofilatos & Yannis, 2015). This is corroborated by a related in-depth accident investigation as well. Using data collected from 500 accidents involving PTWs and bicycles, Ziakopoulos et al. (2018) found speeding to be a contributing factor to both accident frequency and severity. Another in-depth accident study using the framework of GIDAS (German In-Depth Accident Study) investigated the injury protection and accident causation parameters for motorcyclists, among groups of vulnerable road users, alongside pedestrians and bicyclists. It was found that in the majority of the cases, motorcycle riders failed to correctly evaluate the information received from the traffic environment or the situation of their own vehicle (Otte et al., 2012).

The accurate monitoring of motorcyclist riding behavior is therefore of high importance. Although motorcycle accidents have been widely investigated by researchers, the lack of detailed naturalistic riding data remains a persistent obstacle for the scientific community. Several existing studies have shown promising results on the analysis of motorcyclist riding behavior by means of naturalistic experiments (Espié et al., 2013). Williams et al. (2016) examined the contributing factors that influence motorcyclist accident risk, exploiting data from the MSF 100 Motorcyclists Naturalistic Study (31,000 trips from 100 riders). Their results can be utilized for further investigation of road safety levels during daily riding. Other recent studies have attempted to develop methodological techniques to allow either for rider profile detection (Will et al., 2020) or for measuring the lean angles of motorcycles on a large scale (Stanglmayr et al., 2020). Furthermore, the validity of approaches based on smartphone applications has been demonstrated in past studies. Using applications solely, stopping and dangerous riding events have been detected with a reported accuracy of 90.1% for scooters (Hsieh et al., 2014) and 86.8% for bicycles (Gu et al., 2017).

However, to the best of the authors' knowledge, this is the first attempt to understand behaviors and risks related to rider speeding on the basis of data collected from smartphone sensors. Additionally, since several studies have correlated speeding with rider aggressiveness (Lin & Kraus, 2009; Vlahogianni et al., 2012) and with environmental/exposure metrics, as expressed, for instance, by night-time riding (de Rome et al., 2016), the authors' aim is to investigate the respective influence factors and come to conclusions regarding their impact on speeding. More precisely, rider speeding is expected to increase due to:

- rider aggressiveness
- travelled distance
- nighttime riding

In light of the aforementioned, the objective of the present study is twofold: (i) to explore the riding behavior of motorcyclists while speeding, based on detailed riding analytics collected by smartphone sensors, and (ii) to investigate whether personalized feedback can improve rider behavior.

5.1.2 Rider panel and descriptive statistics

Originally, 20 motorcyclist riders volunteered to participate in the experiment and to allow for monitoring their riding behavior through the respective smartphone application. However, for the present analysis it was decided that the final sample should consist only of riders who have participated equally in both phases on terms of trips. An additional criterion was set; all riders selected for the analysis were required to have ridden for at least 40 trips. This number approximately equals the typical monthly number of working trips, assuming that each rider drives 2 trips a day for 5 working days a week. This number is reasonable to filter out riders for which there are not enough observations, and it is also the 'industrial' criterion set by OSeven to start providing rider evaluation to clients. As a result, from the 20 motorcyclists, 13 riders (4 female, 9 male) were ultimately selected. All participants possessed a valid driving license while the 11 out of 13 rode their own motorcycle (the other two rode the motorcycle of a family member). The participants were aged between 25-34 (n=9) and between 35-45 (n=4). Detailed sample information is presented in Table 5.1.

Table 5.1: Participant panel description regarding riding and vehicle data (N=13)

Riding parameter	**Distribution of participants**			
Riding experience (number of years)	<5	5-10	11-20	>20
	7.7%	30.8%	53.8%	7.7%
Driven distance per year (km)	<5000	5001-10000	10001-15000	>15000
	15.4%	7.7%	61.5%	15.4%
Motorcycle engine size (cc)	<251	251-500	501-1000	>1000
	46.1%	7.7%	23.1%	23.1%

Overall, during the two phases of the experiment a large dataset of 3,537 trips from a sample of 13 motorcyclists were recorded. Before presenting the model development, exploratory descriptive analysis of the data is implemented, allowing for an overview of the percentage of speeding while riding, as well as the other three riding indicators that are presented via the application in the feedback phase. Particularly, the descriptive statistics of values of the respective variables are shown in Table 5.2 and they reveal some interesting first findings. It is obvious that all risk factors show a significant reduction when riders receive feedback about their riding behavior, which constitutes an incentive for modelling the impact of feedback. Additionally, this finding is observed in all different road types, indicating a smoother performance of the respective riding metrics when riders are receiving feedback. However, in the case of highways, it is remarkable that the reduction of the examined riding indicators is much slighter compared to the urban and rural road network. This seems logical, taking into consideration the riding pattern in highways, as longer distances are covered and high speeds are developed.

Table 5.2: Descriptive statistics of the per trip values of the variables and the respective standard deviations (in parenthesis) recorded for Phase 1 (Baseline) and Phase 2 (Feedback)

	Road type					
Variable	**Urban**		**Rural**		**Highway**	
	Baseline	Feedback	Baseline	Feedback	Baseline	Feedback
Average speed [km/h]	35.82 (0.30)	33.66 (0.29)	50.08 (0.66)	38.28 (0.48)	97.01 (0.58)	77.62 (0.66)
Speeding percentage [%]	13.41 (0.41)	9.84 (0.33)	10.33 (0.62)	3.07 (0.34)	5.60 (0.96)	5.35 (0.90)
Harsh accel. [count]	2.54 (0.08)	1.70 (0.06)	2.02 (0.09)	1.38 (0.06)	0.81 (0.11)	0.22 (0.04)
Harsh brakings [count]	1.59 (0.05)	1.14 (0.04)	1.27 (0.07)	0.81 (0.04)	0.45 (0.08)	0.16 (0.03)

5.1.3 Generalized Linear Mixed-Effects Models for speeding

In order to model the expected percentage of speeding per trip for the participant riders, mixed-effect models in a GLM framework (GLMMs) were calibrated. Specifically, GLMMs were fitted in R-studio (with the lme4 package) via maximum likelihood and using z-factor scaling. A number of models were tested with different configurations in the collected parameters in both fixed effects and random effects. Furthermore, it is important to note that the sample consists of driver trips, as opposed to drivers directly, thus allowing the training of statistically robust models.

The selected variables were chosen after taking into account the following: lowest Akaike Information Criterion (AIC) for dealing with the trade-off between the goodness of fit of the model and the simplicity of the model, high statistical significance of variables, low multicollinearity, and finally rational interpretation of their impact on the dependent variable. After conducting log-likelihood test ANOVA comparisons, the most informative configuration of random effects was included both random intercepts and random slopes in the GLMMs to capture unique rider traits. Table 5.3 provides a description of the parameters of the models.

Table 5.3: Log-likelihood comparison of mixed effect configuration for overall speeding model

Model Family	**Model Configuration**	**D.f.**	AIC	BIC	logLik	χ^2	**P(>χ^2)**

GLM	Fixed effects only [baseline]	7	52468	52511	-26227	–	–
GLMM	Fixed effects & Random Intercepts	8	38795	38844	-19390	13674.6	< 2e-16
GLMM	Fixed effects, Random Intercepts & Random Slopes	10	37114	37176	-18547	1684.9	< 2e-16

The final models are presented in Table 5.4, Table 5.5, and Table 5.6. Modelling results reveal some interesting findings: The parameters of trip duration, the distance driven during risky hours, morning peak hours and the number of harsh accelerations have all been determined as statistically significant and positively correlated with the percentage of speeding. In the same context, riding during the feedback phase of the experiment, as well as afternoon peak hours are statistically significant and negatively correlated with speeding percentage.

Table 5.4: GLMM overall model for speeding behavior in all road network environments

Random Effects				
Group	**Variable**	**SD**	**Variance**	
Identifier	Intercept	0.9935	0.9870	
	duration	0.3397	0.1154	
Fixed Effects				
Variable	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**
Intercept	1.898	0.276	6.874	<0.001
Rider Feedback	-0.144	0.013	-10.911	<0.001
Trip duration	0.194	0.095	2.030	0.042
Harsh accelerations	0.246	0.005	53.127	<0.001
Risky hours	0.018	0.003	5.161	<0.001
Morning Rush	0.066	0.015	4.356	<0.001
Afternoon Rush	-0.287	0.015	-18.826	<0.001
AIC	37114.1			
BIC	37175.9			
logLik	-18547.1 s			

Table 5.5: GLMM urban model for speeding behavior in urban environment

Random Effects				
Group	**Variable**	**SD**	**Variance**	
Identifier	Intercept	3.081	<0.001	
	duration	0.004	<0.001	
Fixed Effects				
Variable	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**
Intercept	1.810	0.351	5.152	<0.001
Rider Feedback	-0.031	0.011	-2.801	0.005
Trip duration	0.001	0.000	3.563	<0.001
Harsh accelerations	-	-	-	-
Risky hours	0.006	0.013	7.279	0.001
Morning Rush	0.093	0.013	-22.969	<0.001
Afternoon Rush	-0.303	0.351	5.152	<0.001

AIC	54460.9			
BIC	54516.4			
logLik	-27221.4			

Table 5.6: GLMM rural model for speeding behavior in rural environment

Random Effects				
Group	**Variable**	**SD**	**Variance**	
Identifier	Intercept	3.081	<0.001	
	duration	0.004	<0.001	
Fixed Effects				
Variable	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**
Intercept	-	0.870	-0.689	-
Rider Feedback	-0.420	0.019	-22.014	<0.001
Trip duration	0.003	0.001	2.819	0.004
Harsh accelerations	0.056	0.002	28.986	<0.001
Risky hours	0.019	0.002	8.611	<0.001
Morning Rush	0.130	0.020	6.551	<0.001
Afternoon Rush	-0.436	0.023	-19.270	<0.001
AIC	34576.3			
BIC	34638.0			
logLik	-17278.2			

The aforementioned results could be further interpreted, calculating the relative risk ratio of every variable and thus measuring the increase in probability of speeding while riding. The exposure metrics of trip duration, trip distance during risky hours and morning peak hours seem to increase speeding percentage by a factor of exp(B=0.194) = 1.214, exp(B=0.018) = 1.018 and exp(B=0.067) = 1.069 respectively for the overall model. The variables are found to have similar significant effects both in urban and in rural areas. In other words, motorcyclists seem prone to speeding while riding under circumstances that increase their impatience and/or stress such as long trip durations, riding during hours of increased traffic conflicts, lane splitting, hurrying while commuting, etc.

Additionally, the riding behavioral parameter of harsh accelerations increases speeding percentage by a factor of 1.281 in the overall model, indicating the pattern of a stressful riding style. The variable is not found statistically significant for the urban road model, but regarding rural riding, the number of harsh accelerations seem to increase the speeding percentage by a factor of 1.060.

Providing motorcyclists with feedback about their riding performance during experiment Phase 2 led to a remarkable decrease of speeding percentage by 14.5%. Particularly, in the developed models rider feedback seems to decrease speeding percentage, having a risk ratio of exp(B=-0.145) = 0.865 for the overall model, and exp(B=-0.031) = 0.970 and exp(B=-0.420) = 0.657 for urban and rural road types respectively. As explained above, during the feedback phase, riders received personalized feedback regarding their weak points, namely speeding and aggressive riding (harsh accelerations and harsh breakings) by means of a scorecard through the smartphone application. Therefore, the quantification of the positive effect of rider feedback on riding performance indicates new ways of improving road safety.

Furthermore, the chart below depicts the random effects estimates for individual participants (besmartmotousers) in the overall GLMM model analyzing speeding behavior, with "Trip Duration" as a fixed effect and an intercept for each user. The random intercepts capture the variability in baseline speeding tendencies among users, while the slopes associated with "Trip Duration" indicate how the effect of trip duration varies across individuals. The spread and overlap of the confidence intervals around the estimates suggest heterogeneity in individual driving behaviors. Participants like besmartmotouser21 show a significant deviation in the intercept and trip duration effect compared to others, highlighting potential individual-specific factors influencing speeding behavior. This variability emphasizes the importance of including random effects in the GLMM to account for individual differences in driver response.

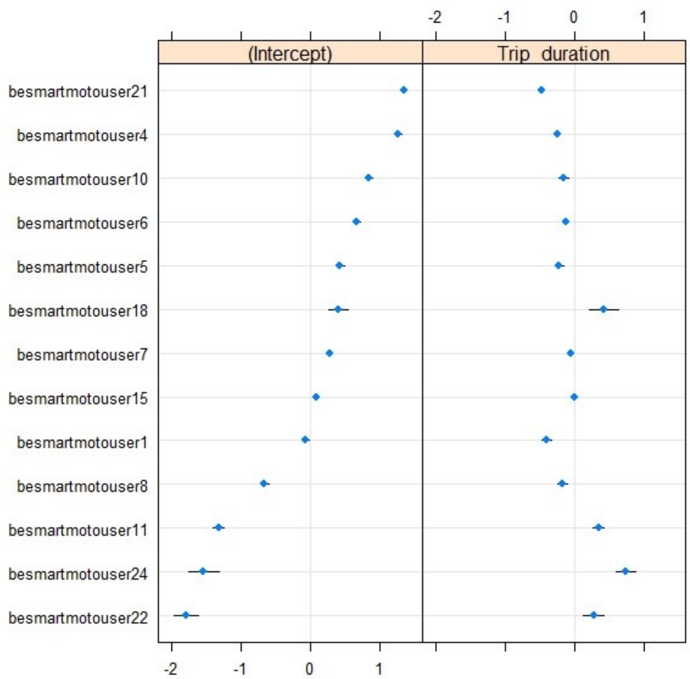


Figure 5.1: Random Intercepts and Random Slopes for total trip duration

5.1.4 Discussion of results

This section aimed: (i) to explore the riding behavior of motorcyclists while speeding based on detailed riding analytics collected by smartphone sensors, and (ii) to investigate whether personalized feedback can improve riding behavior. For that purpose, high-resolution smartphone data collected from a naturalistic riding experiment with a sample of 13 motorcyclists were utilized. Using risk exposure and riding behavior indicators calculated from smartphone sensor data, a statistical analysis was carried out for correlating the percentage of riding time over the speed limit with other riding behavior indicators, namely by means of Generalized Linear Mixed-Effects Models. In particular, an overall model was developed for all trips, and additional separate models were developed for riding on urban and rural roads.

The results from the interpretation of the estimated parameters of the models can be summarized as follows: Trip length and riding during the morning rush and night-time risky hours are exposure metrics significantly associated with the odds of speeding while riding. Harsh accelerations are also associated with the odds of someone exceeding the speed limits, outlining a pattern of an overall unsafe riding behavior.

Furthermore, the outcomes of this study entail both scientific and social impacts. The present research contributes a preliminary example of the quantitative documentation of the impact of personalized rider feedback on one of the most important human risk factors; speeding. The ultimate objective when providing feedback to riders is to: (i) trigger their learning and self-assessment process, thus enabling them to gradually improve their performance and (ii) monitor the shift of riding behavior as the application provides feedback. The present results capture and quantify the positive effects of rider feedback, thus providing needed impetus for larger-scale applications as well as relevant policy interventions. State-of-the-art interventions can include approaches for driver or rider training and support through innovative rider behavior monitoring and feedback tools for different types of riders, such as cyclists or motorcyclists.

The current lack of motorcycle instrumentation certainly limits the potential of detailed multifaceted analysis. Nevertheless, the proposed methodology reaches noteworthy findings, even though the riding exposure and behavior metrics are solely based on smartphone sensors and no other methods, e.g. optical sensors. It is important to highlight that there is little or no previous experience on analyzing and predicting rider speeding behavior through microscopic riding behavior metrics collected from such a portable and low-cost device, and therefore the results of the present research cannot be directly compared to those of the existing literature.

Furthermore, as already discussed, the calculation of the harsh events for motorcycles is based on ML and data fusion algorithms, utilizing data from several smartphone sensors, after the application of the appropriate filtering and data cleaning algorithms on the raw signals. These algorithms evaluate the processed time series from the smartphone sensors of the complete trip. Although the applied algorithms increase the accuracy of harsh events detection; by nature, they do not include specific threshold values or explicit explanations that could be presented in the existing section for verification purposes – nonetheless, some practical examples were illustrated. Harsh events were treated as a given input of rider behavior for the investigation of their speeding behavior; the examination of their definition or detection falls outside the scope of this research.

5.2 Feedback Impact on Distraction of Car Drivers of Phases 1 & 2

5.2.1 Introduction

Mobile phone use while driving has a significant negative effect on driving behavior in terms of reaction time (Haque and Washington, 2014), vehicle lateral control (Niu et al., 2019), headway (Saifuzzaman et al., 2015), speed variation (Wijayaratna et al., 2019). When engaging with a mobile phone (hand-held or hands-free) drivers' work load is highly increased resulting in impaired driving performance. A plethora of studies have examined the impact of mobile phone while driving on driving performance and some risk compensation behaviors have been observed, namely lower driving speeds (Choudhary & Velaga, 2017) and increased car following headways (Li et al., 2019). Nevertheless, the revealed compensatory behaviors do not necessarily result in lower crash risk (O'Connor et al., 2017), highlighting the high importance of investigating the distracted behavior due to mobile phone use while driving.

However, it is also significant to underline that the context and circumstances in which drivers decide to engage in mobile phone use (i.e. road geometry, presence of lead or following car, line crossings, etc.) matter considerably. Tivesten & Dozza (2015) used video recordings of 1,432 trips in order to investigate in which way the driving context influences driver's decision of engagement in visual-manual phone tasks. Researchers found that drivers were more likely to initiate a visual-manual task while standing still (i.e. red traffic lights), and less likely when driving at high speeds. Similar results occurred by the presence of a passenger. In addition, results showed that drivers did adjust their task timing to specific driving contexts, such as when the lead vehicle increased speed or after making sharp turns and lane change maneuvers.

In that environment, a recent study also investigated the relation between driving demands and distraction using a naturalistic driving study dataset (Risteska et al., 2021). Results implicate the role of the driving demands and their impact on secondary task engagements. Similarly, Khan et al. (2021) analyzed driver distraction due to mobile phone use exploiting a multimodal dataset consisting of vehicle data, drivers' smartphone activities, and environmental and contextual data. Going one step further, the researchers assessed existing driver distraction smartphone-based solutions via maxed mode survey, i.e., questionnaires and interviews from 98 participant drivers. Results showcase that future smartphone solutions should focus on the identification of driving context and thus generate context-aware adaptive interfaces enhancing driver safety.

Monitoring distraction by means of smartphone devices is a quite new possibility which is gaining more and more attraction due to the continuous development of new technologies in the field of telecommunications industry. So far, many proposed systems have been introduced with the aim of accurately monitoring driver distraction by exploiting various mobile phone sensors, such as accelerometer and gyroscope, magnetometer, barometers, photometers, and thermometers, microphone and cameras, etc. (Paruchuri & Kumar, 2015; Nambi et al., 2018; Papadimitriou et al., 2019).

As an example, Bo et al. (2013) used personal smartphones by leveraging inertial sensors in order to detect the engagement of mobile phone use while driving. Particularly, the researchers aimed at distinguishing drivers and passengers and detecting when the driver is texting through a Hidden Markov Model (HMM). The results of their study revealed that their proposed system titled "TEXIVE" had an 87.18% accuracy and a 96.67% precision. Similarly, Streiffer et al. (2017) proposes a unified data collection and analysis framework, titled "DarNet" which allows the detection and classification of distracted driving behavior. The proposed system collects two kinds of data; image data from a facing camera and Inertial Measurement Unit (IMU) data from smartphone sensors. Implementing deep learning techniques, they achieved 87.02% accuracy in detecting distracted driving. The smartphone camera has also been used in prior work (Wesley et al., 2010; Shabeer & Wahidabanu, 2012). Nambi et al. (2018) used both smartphone cameras in order to monitor distracted driving; the front camera for driver monitoring and the back camera for driver behavior monitoring.

In that context, the purpose of this study is to examine the impact of feedback on the use of mobile phone use in different environments.

5.2.2 Descriptive statistics and preliminary analysis

Overall, during the first two phases of the experiment, 26,619 trips from a sample of 147 car drivers have been recorded. However, for the present analysis it was decided that the final sample should consist of drivers who have participated equally in both phases only. An additional criterion was set; all drivers chosen to be included in the analysis were required to have driven at least for 40 trips. This number approximately equals the typical monthly number of working trips for a driver assuming that each driver drives 2 trips per day for 5 working days per week. This number is reasonable to filter out drivers for which there are not enough observations, and it is also the 'industrial' criterion set by OSeven to start providing driver evaluation. As a result, from the 147 car drivers, 65 were ultimately selected creating a large dataset of 21,167 trips. Demographic information regarding the drivers' gender and age are shown in Table 6.1.

Table 5.7: Overview of the selected sample

	Age groups			
	<25	25-55	>55	Total%
Male	0	27	3	46%
Female	3	31	1	54%
Total %	5%	89%	6%	100%

The key indicator, and response variable for the purpose of this research is the use of mobile phone while driving during the two first experiment phases. Additional basic road safety indicators (i.e. possibly suggesting risky or reckless behavior) are the following: exceeding the speed limit (percentage of time driving over the speed limits per trip driving duration), and mobile usage (percentage of driving time using the mobile phone per trip driving duration). On the basis of the literature review results, it is assumed that the occurrence of harsh events is partly correlated with drivers' speeding behavior and mobile phone use, and thus some of its variance may be explained by changes in those variables.

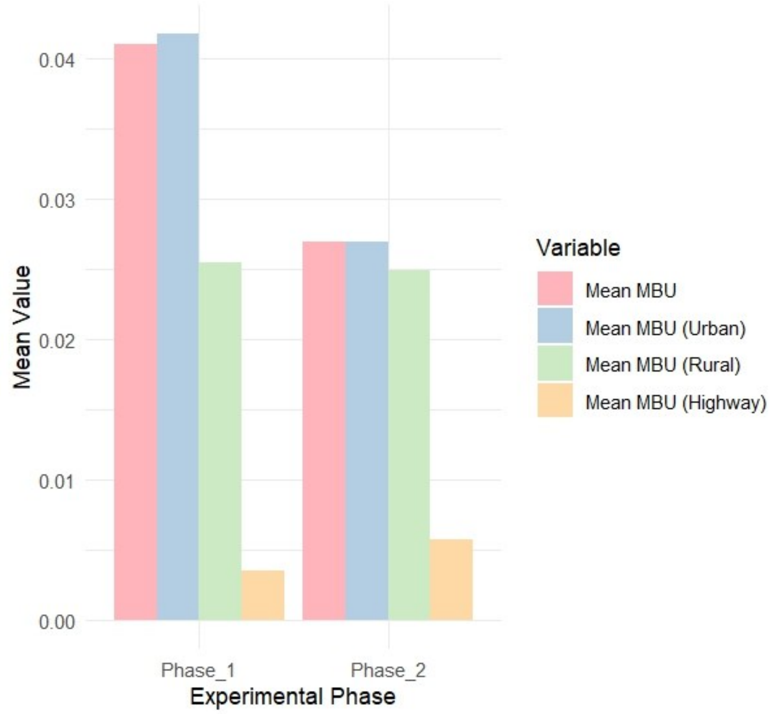


Figure 5.2: Average percentage of mobile phone use per road type

Figure 5.2 compares the mean values of mobile phone use across two experimental phases (Phase 1 and Phase 2) and further disaggregates the data by road types: urban, rural, and highway. In both phases, the overall mean mobile phone use percentage and mean values for urban and rural roads are substantially higher compared to highway conditions. A notable trend is the reduction in mean values from Phase 1 to Phase 2 across all categories, suggesting an improvement in distracted driving behavior over time or due to intervention effects. Urban areas consistently exhibit higher mean values than rural and highway, indicating more frequent mobile phone use in these environments, likely due to more stops due to driving. Highway conditions show the lowest mean value in both phases, which seems logical, taking into consideration the driving pattern in highways; longer distances covered and high speeds developed.

5.2.3 Generalized Linear Mixed Effects Models for mobile phone use

In order to model the expected percentage of mobile phone use while driving per trip for the participant drivers, mixed-effect models in a GLM framework (GLMMs) were calibrated. Specifically, GLMMs were fitted in R-studio (with the lme4 package) via maximum likelihood and using z-factor scaling. A number of models were tested with different configurations in the collected parameters in both fixed effects and random effects. Furthermore, it is important to note that the sample consists of driver trips, as opposed to drivers directly, thus allowing the training of statistically robust models.

The selected variables were chosen after taking into account the following: lowest Akaike Information Criterion (AIC) for dealing with the trade-off between the goodness of fit of the model and the simplicity of the model, high statistical significance of variables, low multicollinearity, and finally rational interpretation of their impact on the dependent variable. After conducting log-likelihood test ANOVA comparisons, the most informative configuration of random effects was included both random intercepts and random slopes in the GLMMs to capture unique rider traits. Table 5.8 provides a description of the parameters of the models.

Table 5.8: Log-likelihood comparison of mixed effect configuration for mobile phone use while driving in all road network environments

Model Family	Model Configuration	D.f.	AIC	BIC	logLik	χ^2	P(>χ^2)
GLM	Fixed effects only [baseline]	7	267329	267385	-133658	–	–
GLMM	Fixed effects & Random Intercepts	8	193060	193123	-96522	74271.3	< 2e-16
GLMM	Fixed effects, Random Intercepts & Random Slopes	10	191573	191652	-95777	1490.9	< 2e-16

The analysis of mobile phone use while driving reveals critical insights into the effectiveness of feedback and other predictors across various driving environments. The comparison of models in Table 5.8 highlights the importance of accounting for individual variability. The baseline fixed-effects model showed a poor fit (AIC = 267,329), while the inclusion of random intercepts significantly improved the model (AIC = 193,060). Introducing random slopes for trip duration further enhanced the model's performance (AIC = 191,573), indicating that the influence of trip duration varies significantly among drivers. The Table 5.9, Table 5.10, Table 5.11, Table 5.12 present the results of the final models, namely the overall model, urban, rural and highway.

Table 5.9: GLMM overall model for mobile phone use while driving in all road network environments

Random Effects				
Group	**Variable**	**SD**	**Variance**	
Identifier	Intercept	1.4024	1.9667	
	duration	0.2827	0.0799	
Fixed Effects				
Variable	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**
Intercept	0.6528	0.1754	3.7220	0.0002
Driver Feedback	-0.4276	0.0081	-52.5660	< 2e-16
Trip duration	0.1514	0.0374	4.0440	5.25e-05
Harsh accelerations	0.0424	0.0034	12.4380	< 2e-16
Risky hours	-0.0543	0.0046	-11.6940	< 2e-16
Morning Rush	-0.3390	0.0124	-27.4020	< 2e-16
Afternoon Rush	0.1586	0.0089	17.9030	< 2e-16
AIC	191573.2			
BIC	191652.3			
logLik	-95776.6			

In the overall model (Table 5.9), feedback emerges as a strong and significant factor in reducing mobile phone use while driving (Estimate =−0.4276, p<2e−16), demonstrating its efficacy in mitigating phone use during driving. Longer trip durations and harsh accelerations were associated with increased MBU, while risky hours and morning rush periods reduced mobile phone use while driving, possibly reflecting heightened caution during these times. The substantial variability in random intercepts (SD=1.4024) suggests notable differences in baseline mobile phone use while driving levels among drivers, while moderate variability in trip duration (SD=0.2827) highlights differences in how trip length affects mobile phone use while driving behavior.

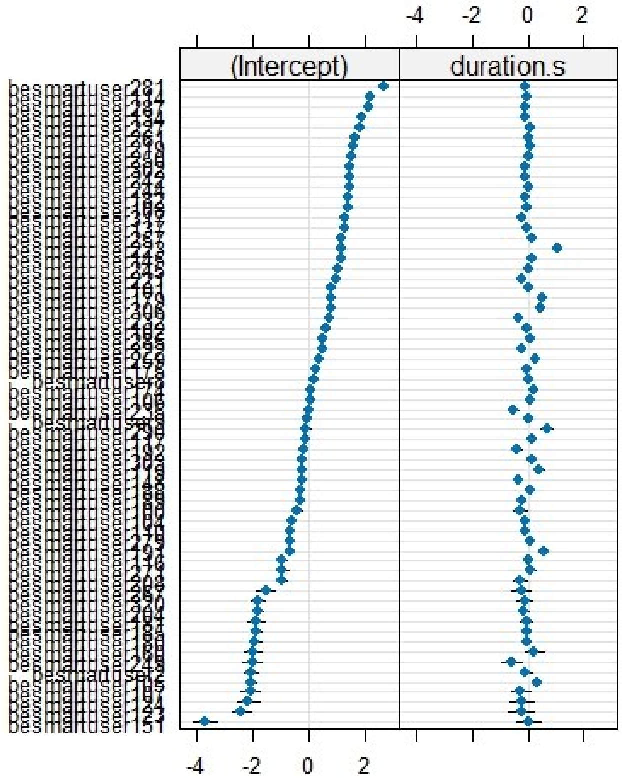


Figure 5.3: Random intercepts and random slopes for total trip duration for the overall model

Figure 5.4 presents the random effects estimates from the overall model - analyzing mobile phone use while driving across all road network environments. The left panel displays the variability in baseline mobile phone use (intercepts) among participants, indicating substantial heterogeneity, with some drivers showing consistently higher or lower likelihoods of using their phones while driving. The right panel illustrates the impact of trip duration (duration.s) on mobile phone use, with most drivers showing little effect (estimates clustered near zero), though a few participants exhibit notable increases or decreases in phone use with longer trips. The spread of estimates and confidence intervals highlights significant individual differences in behavior, underscoring the importance of modeling random effects to capture variability in mobile phone usage influenced by personal habits, trip characteristics, or external factors.

In urban environments (Table 5.10), feedback remains a significant deterrent to mobile phone use while driving (Estimate=−0.3687, p<2e−16), though its effect is slightly reduced compared to the overall model. Urban-specific dynamics, such as frequent stop-and-go driving, may explain this reduced impact. Risky hours (Estimate=−0.3298, p<2e−16) showed a stronger protective effect in urban settings, indicating that drivers might be more vigilant in complex traffic conditions. Interestingly, harsh accelerations had a slight negative association with mobile phone use while driving, contrary to other contexts, potentially reflecting heightened attention during abrupt driving maneuvers.

Table 5.10: GLMM model for mobile phone use while driving in urban environment

Random Effects				
Group	**Variable**	**SD**	**Variance**	
Identifier	Intercept	1.3711	<0.001	
	duration	0.0008	<0.001	
Fixed Effects				
Variable	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**
Intercept	0.4750	0.1721	2.7590	0.0058
Rider Feedback	-0.3687	0.0083	-44.2790	<2e-16
Trip duration	0.0004	0.0001	3.9160	9e-05
Harsh accelerations	-0.0131	0.0013	-9.8620	<2e-16
Risky hours	-0.3298	0.0124	-26.6670	<2e-16
Morning Rush	0.1533	0.0088	17.3810	<2e-16
Afternoon Rush	0.4750	0.1721	2.7590	0.0058
AIC	214914.1			
BIC	214985.3			
logLik	-107448.0			

In rural environments (Table 5.11), the effect of feedback is weaker than in urban areas (Estimate=−0.1180, $p<2e-16$) indicating that rural drivers may be less responsive to interventions. Trip duration and harsh accelerations both showed positive associations with mobile phone use while driving, consistent with overall findings. However, risky hours and morning rush periods continued to reduce MBU, reinforcing the protective effects of these time-specific factors. The greater variability in random intercepts (SD=1.6843) suggests that rural drivers exhibit more diverse baseline behaviors compared to urban counterparts.

Table 5.11: GLMM model for mobile phone use while driving in rural environment

Random Effects				
Group	**Variable**	**SD**	**Variance**	
Identifier	Intercept	1.6843	<0.001	
	duration	0.0007	<0.001	
Fixed Effects				
Variable	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**
Intercept	-0.2269	0.2102	-1.0790	0.2810
Rider Feedback	-0.1180	0.0095	-12.4290	<2e-16
Trip duration	0.0008	0.0001	8.5510	<2e-16
Harsh accelerations	0.0511	0.0044	11.5500	<2e-16
Risky hours	-0.0114	0.0012	-9.5000	<2e-16
Morning Rush	-0.2634	0.0136	-19.3220	<2e-16
Afternoon Rush	0.1624	0.0104	15.6600	<2e-16
AIC	191416.0			
BIC	191495.1			
logLik	-95698.0			

The highway model (Table 5.12) presented an unexpected outcome: feedback was positively associated with MBU (Estimate=0.5490, $p<2e-16$). This counterintuitive result suggests that feedback mechanisms may inadvertently encourage compensatory behaviors, or that drivers perceive lower risks on highways. Trip duration had the strongest positive effect on mobile phone use while driving in this model (Estimate=0.6892, p=0.0049), aligning with the prolonged exposure to driving conditions. Risky and rush hour periods continued to have protective effects, consistent with the notion of increased driver caution in these contexts.

Table 5.12: GLMM model for mobile phone use while driving in highways

Random Effects				
Group	**Variable**	**SD**	**Variance**	
Identifier	Intercept	3.536	12.506	
	duration	1.931	3.731	
Fixed Effects				
Variable	**Estimate**	**Std. Error**	**z value**	**Pr(>|z|)**
Intercept	-4.1676	0.4586	-9.0880	< 2e-16
Rider Feedback	0.5490	0.0235	23.4120	< 2e-16
Trip duration	0.6892	0.2451	2.8120	0.0049
Harsh accelerations	0.0883	0.0055	16.0250	< 2e-16
Risky hours	-0.0653	0.0097	-6.7170	0.0000
Morning Rush	-0.4071	0.0281	-14.4940	< 2e-16
Afternoon Rush	-0.4439	0.0285	-15.5990	< 2e-16
AIC	59690.4			
BIC	59769.5			
logLik	-29835.2			

Overall, the findings emphasize the critical role of feedback in reducing mobile phone use while driving but also highlight the need for context-specific interventions. While feedback is broadly effective in urban and rural environments, its limitations on highways underscore the need for alternative strategies, such as dynamic feedback or stricter enforcement. These insights provide a foundation for designing tailored policies to enhance road safety and mitigate distracted driving.

5.2.4 Discussion of results

This present research aimed to investigate the impact of mobile phone use on driving behavior and road safety through the investigation of driving analytics collected by smartphone sensors. In order to achieve that objective, a naturalistic driving experiment was carried out in order to examine distracted driving as expressed by the use of mobile phone while driving. Results substantiate correlations of mobile phone use with specific exposure and driving behavior indicators.

The examination of mobile phone use while driving across different driving environments and models underscores the importance of tailored interventions to address distracted driving. Feedback mechanisms have demonstrated consistent effectiveness in reducing MBU in overall, urban, and rural contexts, but their impact varies by setting, with reduced effectiveness in rural areas and an unexpected positive association on highways. These findings highlight the complexity of driver behavior and the influence of environmental factors on feedback efficacy. To maximize the benefits of such interventions, future policies should focus on refining feedback systems to account for contextual differences, leveraging dynamic adjustments, and integrating complementary strategies such as education, enforcement, and technology-based solutions. By addressing the unique characteristics of each driving environment, these measures can contribute to significant reductions in MBU and improvements in overall road safety.

6 Feedback Impact on Harsh Events

6.1 Feedback Impact on Harsh Events of Car Drivers of Phases 1 & 2

6.1.1 Introduction

Driving behavior is influenced by a range of factors, including individual tendencies, environmental conditions, and external interventions. Among these, feedback mechanisms have emerged as a valuable tool for modifying driver behavior and reducing risky driving practices. Harsh events, such as sudden accelerations, harsh braking, and abrupt maneuvers, are critical indicators of driving safety (Nikolaou et al., 2023). Understanding the impact of feedback on these events is essential for designing effective interventions aimed at promoting safer driving habits. By examining the relationship between feedback and harsh events, this section provides insights into how targeted interventions can enhance driver safety.

Although both harsh accelerations and harsh brakings are associated with unsafe driving behavior, they constitute two different types of events and should therefore examined as such. More specifically, harsh acceleration events may reveal high levels of anxiety and anger while driving (Stephens et al., 2009; Roidl et al., 2014) leading to a risky driving behavior characterized by drivers' involvement in situations of high risk. Harsh brakings may indicate driver struggle to anticipate the occurrence of a critical situation, which most of the times would not have occurred at the first place if it were not for driver's inattention, high speed development, inadequate distances from adjacent vehicles and other unsafe behavior indicators. As a result, harsh breaking events harsh are often used to locate safety critical events in Naturalistic Driving (ND) data (Hanowski et al., 2005; Jansen & Wesseling, 2018).

Additionally, given the strong correlation between harsh events and driving risk, it is not surprising why harsh accelerations and harsh brakings have been investigated by insurance industry in the context of usage-based motor insurance (UBI) schemes (Boquete et al., 2010; Paefgen et al., 2013), allowing for more behavioral parameters being used in UBI models. Harsh events, in combination with other driving behavioral indicators such as speeding and distracted driving are being increasingly used by Pay How You Drive (PHUD) Usage Based Insurance schemes as the critical risk factor indicators in terms of driving behavior (Tselentis et al., 2017).

6.1.2 Descriptive statistics and preliminary analysis

Overall, during the first two phases of the experiment, 26,619 trips from a sample of 147 car drivers have been recorded. However, for the present analysis it was decided that the final sample should consist of drivers who have participated equally in both phases only. An additional criterion was set; all drivers chosen to be included in the analysis were required to have driven at least for 40 trips. This number approximately equals the typical monthly number of working trips for a driver assuming that each driver drives 2 trips per day for 5 working days per week. This number is reasonable to filter out drivers for which there are not enough observations, and it is also the 'industrial' criterion set by OSeven to start providing driver evaluation. As a result, from the 147 car drivers, 65 were ultimately selected creating a large dataset of 21,167 trips. Demographic information regarding the drivers' gender and age are shown in Table 6.1.

Table 6.1: Overview of the selected sample

	Age groups			
	<25	25-55	>55	Total%
Male	0	27	3	46%
Female	3	31	1	54%
Total %	5%	89%	6%	100%

The key indicator, and response variable for the purpose of this research is the frequency of harsh acceleration and harsh braking events during the two first experiment phases. Additional basic road safety indicators (i.e. possibly suggesting risky or reckless behavior) are the following: exceeding the speed limit (percentage of time driving over the speed limits per trip driving duration), and mobile usage (percentage of driving time using the mobile phone per trip driving duration). On the basis of the literature review results, it is assumed that the occurrence of harsh events is partly correlated with drivers' speeding behavior and mobile phone use, and thus some of its variance may be explained by changes in those variables.

Table 6.2 provides a description of the variables selected. Regarding the harsh event frequencies, it is noted that both during Phase 1 and Phase 2 drivers seem to incur more harsh brakings than harsh accelerations during their trip.

Table 6.2: Description of the variables used in the analysis

Variable	**Description**
Total Trip Duration [s]	Total trip duration [sec]
Total Trip Distance [km]	Total trip distance [km]
Average Speed [km/h]	Mean driving speed per trip [km/h]
Maximum Speed [km/h]	Maximum of driving speed per trip [km/h]
Percentage of Mobile Use Duration [%]	Share of mobile use per trip [%]
Percentage of Speeding Duration [%]	Share of time over the speed limit per trip [%]
Harsh accelerations [count]	Harsh acceleration events per trip [count]
Harsh brakings [count]	Harsh braking events per trip [count]

The descriptive statistics of the parameters that were recorded per trip for both experiment phases are shown in Table 6.3 for Phase 1 and Table 6.4 for Phase 2.

Table 6.3: Descriptive statistics of the per trip values of the variables recorded for Phase 1

Variable	**Mean**	**Minimum**	**Median**	**Maximum**	**St. Dev.**
Total Trip Duration [s]	992.55	61.00	696.00	17642	1051.86
Total Trip Distance [km]	10.17	0.50	4.8.00	387.20	20.14
Average Speed [km/h]	37.27	14.00	33.00	59.00	15.35
Maximum Speed [km/h]	68.64	19.00	64.00	206.00	26.68
Percentage of Mobile Use Duration [%]	4.00	0.00	0.00	93.80	10.00
Percentage of Speeding Duration [%]	5.00	0.00	1.00	70.00	9.00
Harsh accelerations [count]	0.76	0.00	0.00	19.00	1.52
Harsh brakings [count]	1.51	0.00	1.00	33.00	2.34

Table 6.4: Descriptive statistics of the per trip values of the variables recorded for Phase 2

Variable	Mean	Minimum	Median	Maximum	St. Dev.
Total Trip Duration [s]	950.00	61.00	687.00	11349.00	904.81
Total Trip Distance [km]	8.25	0.50	4.30	325.50	14.09
Average Speed [km/h]	34.55	14.00	30.00	56.35	14.07
Maximum Speed [km/h]	64.90	18.00	59.00	228.00	26.42
Percentage of Mobile Use Duration [%]	3.00	0.00	0.00	93.20	10.00
Percentage of Speeding Duration [%]	3.00	0.00	0.00	65.00	7.00
Harsh accelerations [count]	0.67	0.00	0.00	21.00	1.36
Harsh brakings [count]	1.36	0.00	1.00	27.00	2.20

For further investigation, the differences between the two Phases were also explored with t-tests based on the drivers' trips during Phase 1 and Phase 2, respectively. A t-test analysis, based on the 65 selected drivers' trips, rejected the null hypothesis that both distributions of harsh events have an equal mean, demonstrating that the decrease in harsh events to be significant; both harsh accelerations (t-stat = 4.90, $p < 0.001$) and harsh brakings (t-stat = 4.82, $p < 0.001$). Additionally, results reveal that both the average speed (t-stat = 13.417, $p < 0.001$) and the maximum mean speed (t-stat = 10.25, $p < 0.001$) seem to be statistically significantly reduced during Phase 2 of the experiment. This also applies to the percentage of time driving over the speed limits per trip driving duration, taking into consideration the significant reduction of the particular variable, namely from 5% to 3% (t-stat = 17.63, $p < 0.001$). Finally, although the percentage of time using the mobile phone per trip driving duration is also found to be reduced in Phase 2 of the experiment, from 4% to 3%, the reduction is not statistically significant (t-stat = 1.15, $p = 0.248$).

6.1.3 Generalized Linear Mixed-Effects Models for harsh events

In order to model the expected frequency of events per trip for the participant drivers, models in a GLM framework were calibrated, as previously explained. Since the BeSmart application allows for a high resolution, big-data oriented collection scheme, it was attempted to include random effects in order to capture the unique driving behavior traits for each driver. This entails having a critical minimum sample of trips for each driver to achieve a meaningful outcome. Therefore, a screening was made among participant drivers, as described above, and drivers that had over 40 trips each were selected for the GLMM analysis.

GLMMs were fitted in R-studio (with the lme4 package) via maximum likelihood and using z-factor scaling, following Bates et al. (2014). A number of models were tested with different configurations in the collected parameters in both fixed effects and random effects. The Poisson function with the log-odds link function was implemented. After conducting log-likelihood test (ANOVA) comparisons, the most informative configuration of random effects was the inclusion of both random intercepts and random slopes in the GLMMs to capture unique driver traits (lowest LogLikelihood and highest χ^2). It should be noted that conceptually, for harsh event frequencies, random slope models without random intercepts make little sense and are thus avoided.

Results of mixed effect configuration are shown on Table 6.5 for harsh accelerations and on Table 6.5for harsh brakings:

Table 6.5: Log-likelihood comparison of mixed effect configuration for harsh acceleration models for both phases

Experiment Phase	Model Family	Model Configuration	D.f.	LogLikelihood	χ^2	P(>χ^2)	Sig.
Phase 1	GLM	Fixed effects only [baseline]	6	-11078.6	–	–	–
	GLMM	Fixed effects & Random Intercepts	7	-9929.5	2298.34	<2e-16	***
	GLMM	Fixed effects, Random Intercepts & Random Slopes	9	-9860.8	137.35	<2e-16	***
Phase 2	GLM	Fixed effects only [baseline]	6	-10639.5	–	–	–
	GLMM	Fixed effects & Random Intercepts	7	-9787.4	1704.32	<2e-16	***
	GLMM	Fixed effects, Random Intercepts & Random Slopes	9	-9741.7	91.34	<2e-16	***

Significance codes: ‘***’: 0.000 | ‘**’: 0.001 | ‘*’: 0.01 | ‘.’: 0.05 | ‘ ’: ≥ 0.1

Table 6.6: Log-likelihood comparison of mixed effect configuration for harsh braking models

Experiment Phase	Model Family	Model Configuration	D.f.	LogLikelihood	χ^2	P(>χ^2)	Sig.
Phase 1	GLM	Fixed effects only [baseline]	5	-15634	–	–	–
	GLMM	Fixed effects & Random Intercepts	6	-14352	2565.04	< 2e-16	***
	GLMM	Fixed effects, Random Intercepts & Random Slopes	8	-14158	388.04	< 2e-16	***
Phase 2	GLM	Fixed effects only [baseline]	5	-15588	–	–	–
	GLMM	Fixed effects & Random Intercepts	6	-14045	3086.73	<2e-16	***
	GLMM	Fixed effects, Random Intercepts & Random Slopes	8	-13965	159.73	<2e-16	***

Significance codes: ‘***’: 0.000 | ‘**’: 0.001 | ‘*’: 0.01 | ‘.’: 0.05 | ‘ ’: ≥ 0.1

The final models were selected as the ones with the lowest AIC values. Fixed effect results appear on Table 6.7 for harsh acceleration frequencies and on Table 6.8 for harsh braking frequencies. A dash sign ('–') on the tables indicates that the specific variable was not used in the particular model. Furthermore, taking into account the value range of the examined data, the models do not yield any negative predictions, thus leading to no concerns for the negative value of the intercept which describes residual unexplained variance.

Table 6.7: GLMMs for harsh acceleration frequencies of 65 drivers (fixed effects)

Trip characteristic	GLMM for Phase 1 Estimate	s.e.	p-value	Sig.	Relative Risk Ratio	GLMM for Phase 2 Estimate	s.e.	p-value	Sig.	Relative Risk Ratio
Intercept	-0.927	0.091	0.000	***	0.395	-1.127	0.085	0.000	***	0.324
Maximum Speed	0.321	0.022	0.000	***	1.378	0.412	0.021	0.000	***	1.509
Percentage of Speeding Duration	0.074	0.013	0.000	***	1.076	0.035	0.012	0.003	**	1.035
Percentage of Mobile Use Duration	0.042	0.011	0.000	***	1.042	-	-	-	-	-
Log(Total Trip Duration)	0.848	0.051	0.000	***	2.334	0.729	0.050	0.000	***	2.073
Log(Total Trip Distance)	-0.231	0.050	0.000	***	0.793	-0.087	0.046	0.047	*	0.916

Significance codes: '***': 0.000 | '**': 0.001 | '*': 0.01 | '.': 0.05 | ' ': ≥ 0.1

Table 6.8: GLMMs for harsh braking frequencies of 65 drivers (fixed effects)

Trip characteristic	GLMM for Phase 1 Estimate	s.e.	p-value	Sig.	Relative Risk Ratio	GLMM for Phase 2 Estimate	s.e.	p-value	Sig.	Relative Risk Ratio
Intercept	-0.182	0.067	0.006	**	0.833	-0.313	0.075	0.000	***	0.731
Maximum Speed	0.327	0.016	0.000	***	1.387	0.331	0.015	0.000	***	1.395
Percentage of Speeding Duration	0.097	0.010	0.000	***	1.102	0.081	0.009	0.000	***	1.084
Log(Total Trip Duration)	0.885	0.045	0.000	***	2.423	0.723	0.038	0.000	***	2.061
Log(Total Trip Distance)	-0.298	0.036	0.000	*	0.742	-0.082	0.033	0.015	*	0.921

Significance codes: '***': 0.000 | '**': 0.001 | '*': 0.01 | '.': 0.05 | ' ': ≥ 0.1

Modelling results regarding the harsh acceleration frequencies reveal some interesting findings; the parameters of maximum speed, percentage of speeding duration and total trip duration have all been determined as statistically significant and positively correlated with harsh acceleration frequencies for both experiment phases. In the same context, total trip distance is statistically significant and negatively correlated with harsh acceleration frequencies for both experiment phases as well. Mobile use duration was found statistically significant only for Phase 1 with a small positive correlation.

More specifically, the aforementioned results could be further interpreted by calculating the relative risk ratio of every variable and thus measuring the increase in log-odds of the harsh acceleration frequencies. Maximum driver speed increases the frequencies of harsh accelerations; the effect appears to be higher in Phase 2 than Phase 1, as the respective estimates are 0.412 and

0.321, corresponding to risk ratios (in other words, incidence rate ratios for frequency frequencies) of 1.509 and 1.378. Exceeding the speed limit seems to increase the odds of harsh acceleration frequencies. A possible explanation is that certain drivers develop high driving speeds, reaching higher than the speed limit, they accelerate in a more abrupt way, allowing the application to detect a harsh acceleration event. In other words, overly aggressive drivers do not only exceed speed limits, they do so by accelerating harshly. When comparing the two phases, it is found that the effect of exceeding the speed limit is higher in Phase 1 than in Phase 2, with respective estimates of 0.074 and 0.035, corresponding to risk ratios of 1.076 and 1.035, respectively.

Regarding the effect of mobile use while driving, it appears that 1% of mobile use duration per trip increases harsh acceleration event frequencies by 4% which makes it the least impactful variable among the examined ones, based on the respective estimate of 0.042. The fact that mobile phone use was not found significant for Phase 2 may be interpreted by the aforementioned reduction of the specific parameter during the two phases. Overall, it is noteworthy that all examined behavioral parameters (namely speed, percentage of speeding and mobile phone duration) are positively correlated with harsh acceleration frequencies, confirming the strong correlation between harsh events and unsafe driving behavior.

As for the exposure parameters (trip distance and duration), there is a remarkable finding similar for both experiment phases. Trip duration seems to increase the odds of harsh acceleration frequencies; 1 sec of driving time increases acceleration frequencies by 2.3 and 2.1 times in Phase 1 and 2, respectively; risk ratios originating from parameter estimates of 0.848 and 0.729, respectively. However, the exposure parameter of total trip distance was found to be negatively associated with the odds of higher harsh acceleration counts, possibly because drivers may be prepared more effectively, physically and psychologically, when they are aware of the fact that they will cover a long distance. More precisely, a single unit of total trip distance travelled appears to reduce harsh acceleration counts by a lower degree in Phase 2 compared to Phase 1, as the respective estimates are -0.087 and -0.231, corresponding to risk ratios of 0.916 and 0.793.

With respect to harsh braking events, apart from the mobile use variable, all the other variables, both driving behavioral and exposure ones, seem to have effects that are similar to the ones they have on the harsh acceleration events. The occurred finding is more obvious when it comes to both exposure parameters; trip distance and duration, where the effects appear to be identical. Specifically, trip duration appears to increase the odds of harsh braking frequencies; 1 sec of driving time increases braking frequencies by 2.4 and 2.1 times in Phase 1 and 2, respectively; risk ratios originating from parameter estimates of 0.885 and 0.723, respectively. On the other hand, similarly to the harsh acceleration models, the exposure parameter of total trip distance was found to be negatively associated with the odds of higher harsh braking counts; the effect seems higher in Phase 1 compare to Phase 2, as the respective estimates are -0.298 and -0.082, corresponding to risk ratios of 0.742 and 0.921.

The examined behavioral parameters, namely the parameters of maximum speed and percentage of speeding duration, have all been positively correlated with harsh braking frequencies for both experiment phases. Maximum driving speed increases the frequencies of harsh brakings; the effect appears to be similar for both Phases, as the respective estimates are 0.327 for Phase 1 and 0.331 for Phase 2, corresponding to risk ratios of 1.387 and 1.395, respectively. In the same context, exceeding the speed limit appears to increase the odds of harsh braking frequencies; the effect appears to be higher in Phase 1 than Phase 2, as the respective estimates are 0.097 and 0.081, corresponding to risk ratios of 1.102 and 1.084. As already mentioned in the harsh accelerations models, aggressive driving, expressed by higher driving speeds and higher percentage of speeding while driving, increases harsh braking frequencies, confirming once again the close relationship between harsh events and unsafe driving behavior. The absence of mobile phone use in the statistical model could be interpreted by the fact that drivers reduce speed while distracted, and therefore are less prone to harsh brakings. In other words, drivers decelerate in order to compensate for the distraction from mobile use, so they have the ability to brake normally and avoid harsh braking.

The visual representations of values of random intercepts and random slopes for the log of total trip duration per driver for both Phase 1 and Phase 2 are shown in the constructed caterpillar plots below for harsh acceleration frequencies (Figure 6.1) and for harsh braking frequencies (Figure 6.2), respectively. Personal differences per driver from the fixed effect intercept and slope are thus included in the linear predictor.

To visualise these results, driver-level deviations are shown relative to the mean (0.00; vertical gray line) with 95% confidence intervals around each intercept (blue circle). With respect to harsh acceleration frequencies, shown on Figure 6.1, the visual comparison of the intercept estimates illustrates that overall prediction averages in Phase 1 are similar to the ones of Phase 2. This is an indication that the unexplained variance for harsh acceleration occurrence is somewhat constant for each driver between the two Phases. Estimates for random slope effects of the log of total trip duration show a more erratic effect during Phase 1 comparted to Phase 2. In Phase 1, the impact of total duration on harsh acceleration counts appears to be mathematically higher for certain drivers. In practice, this indicates that there are certain individuals for whom total trip duration plays very different roles for harsh acceleration occurrence when they are not provided with feedback, a trend that greatly diminishes when feedback is provided.

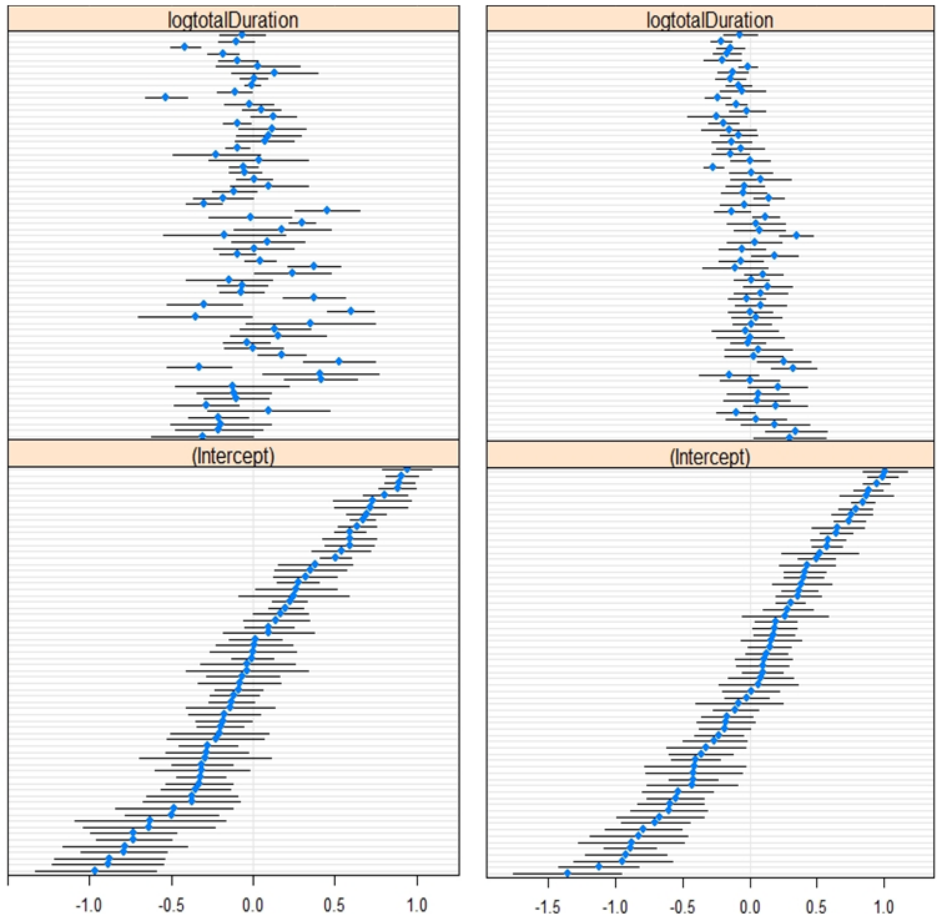


Figure 6.1: Random Intercepts and Random Slopes for log(total trip duration)
[Both in the GLMM for harsh acceleration frequencies during Phase 1 (left) and Phase 2 (right)]

However, different observations are drawn from the caterpillar plots when examining harsh braking frequencies, shown on Figure 6.2. The visual comparison of intercept estimates illustrates that prediction averages in Phase 2 overlap by a larger degree than in Phase 1. Furthermore, the gradient when connecting the blue data points of intercepts is steeper for Phase 2 if one takes the reduced range of the x-axis into account. Estimates for random slope effects of the log of total trip duration show little difference across the two Phases. It seems that the provision of feedback in Phase 2 led to overall reductions of harsh braking events per trip duration unit if the driver sample is examined on an aggregate level. Nonetheless, these reductions seem to have a certain amount of divergence across individual drivers.

As previously stated, the personal differences are captured in the random effects of the GLMMs. These differences may be parameters unobserved in the present models such as driver age, experience, aggressiveness, alertness and performance levels and other similar human factors. Since the dependent variables are harsh event frequencies, which also depend on the road environment, additional parameters which may affect harsh event occurrence can be thought to be integrated therein as well. Examples include temporal and spatial headways and more unforeseen events such as traffic conflicts or the presence of obstacles.

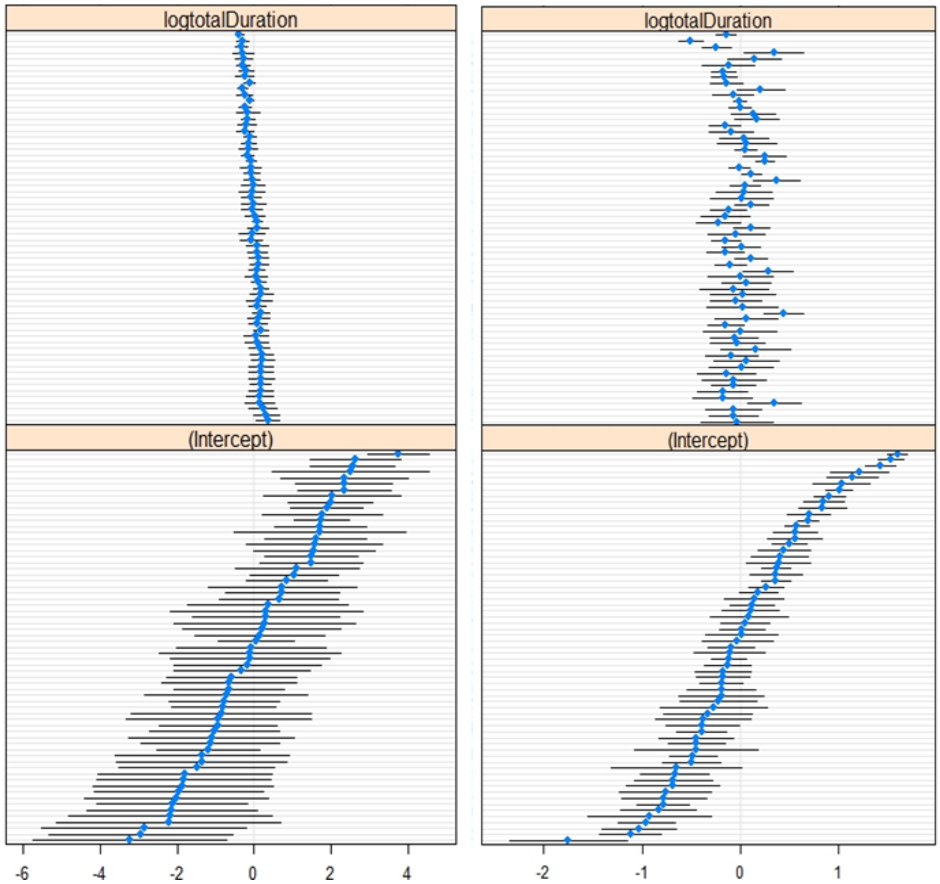


Figure 6.2: Random Intercepts and Random Slopes for log(total trip duration)
[Both in the GLMM for harsh braking frequencies during Phase 1 (left) and Phase 2 (right)]

6.1.4 Discussion of results

This section aimed to investigate the impact of detailed trip characteristics on the frequency of harsh acceleration and harsh braking events recorded by smartphone sensors. In order to achieve that objective, an ongoing naturalistic driving experiment was carried out within the framework of the BeSmart project in order to examine driving behavior as expressed by the frequencies of harsh accelerations and harsh brakings. In the present research, the first two phases were considered; Phase 1, where participants were asked to drive in the way they usually did, without receiving any feedback on their driving behavior and Phase 2, where participants were provided with personalized feedback, a trip list and a scorecard regarding their driving behavior, allowing them to identify their critical deficits or unsafe behaviors.

Generalized Linear Mixed-Effects Models were fitted to the trips of 65 car drivers who made frequent trips during both experiment phases in order to model the frequencies of harsh acceleration and harsh braking events. Results reveal correlations of harsh event frequencies with specific driving behavior and exposure metrics, at a more detailed level than existing studies (e.g. exceeding speed limit, trip duration). More specifically, results indicate that maximum speed, the percentage of speeding duration and total trip duration are positively correlated with both harsh acceleration and harsh braking frequencies. On the other hand, the exposure metric of total trip distance was found to be negatively correlated with both harsh event types.

Results for Phase 1 and Phase 2 indicate that both types of harsh events are influenced by the same explanatory variables, with the exception of mobile use while driving, which is found statistically significant only for harsh acceleration models for Phase 1 of the experiment. Additionally, for the majority of the variables, coefficient values seem to change between the two experiment phases in a similar direction for both harsh acceleration and harsh braking events. Specifically, the effect of speeding was found to be higher for the frequency of harsh events in Phase 1 than Phase 2. On the contrary, the effect of maximum drive speed is higher in Phase 2 than Phase 1 for both acceleration and braking events. A similar pattern is noticed for the exposure parameters, as it seems that trip duration effect is higher in Phase 1 than Phase 2, while trip distance effect was found higher in Phase 2 than in Phase 1, for both types of events.

Although initial findings from descriptive statistics suggest that drivers improved their driving behavior with regards to all the recorded driving behavior metrics, namely maximum driver speed, percentage of speeding duration and using the mobile phone duration, the application of suitable statistical methodologies for before-after evaluation leads to more reliable and accurate results. As the experiment progresses, different types of personalized feedback will be communicated to all drivers allowing them to identify their critical deficits or unsafe behaviors, while incentives within a social gamification scheme, with personalized target setting, benchmarking and comparison with peers will also be developed and provided through the smartphone application.

Alternative approaches to frequency modelling, such as the modelling of harsh event rates (harsh events per km), were considered. However, due to the nature of the data, they were ultimately discarded. Linear modelling led to negative predictions and very poor independent variable and overall model fits. Additionally, trips with zero harsh events in the dataset did not allow for log-normal transformations, and their subsequent removal would bias results. This section does not explicitly focus on the evaluation of the effectiveness of driver feedback on improving driving behavior and increasing road safety levels from the smartphone application; this will be the dedicated focus of future research. It should be mentioned that this analysis is macroscopic overall, and should be treated as a high-level behavioral investigation. Within the present approach, there is no option to statistically examine if mobile phone use was exactly simultaneous with harsh events, but only to verify that they both occurred within the same trip. In other words, the temporal coincidence of data was not considered. Addressing these limitations will require the development of additional dedicated methodologies in the future, and the examination of in-depth datasets analyzing each trip per trip-second.

6.2 Feedback Impact on Harsh Events of Professional Drivers on Highways of Phases 4 & 5

6.2.1 Introduction

Road accidents are the leading cause of death from work-related accidents in industrialized countries. For truck, coach, and company car drivers, fatigue and speeding are the most common causes of accidents. (Professional Drivers | Mobility and Transport). Yuan et al. (2021) analyzed the risk factors associated with truck-involved fatal crashes on various group of truck drivers. The findings revealed that extreme adverse weather, risky driving behavior (fatigue, driving under the influence of alcohol), the use of one or more trailing units, and trucks with heavy weights, were all linked to an increased risk of serious accidents. Uddin & Huynh (2017) examined injury severity in crashes with trucks, concluding that lighting conditions, age and gender of occupant, truck types, speed, and weather condition were found to be factors that have impact on injury severity. Brouwer et al. (2015) conducted a simulator experiment with 26 professional truck drivers to investigate the effectiveness of personalized feedback. Han & Zhao (2020) investigated driving behavior of professional urban bus drivers in China.

An experiment can be carried out mainly in two ways: in a naturalistic experiment, where actual conditions are used or in a simulator experiment, in a more secure and contained environment. Driver monitoring through naturalistic driving is one of the most recent developments in road safety. Approaches of that area include the use of high-end technological solutions, exploitation of On-Board Diagnostics (OBD) and smartphone data collection. The last approach has many proven advantages, including uninterrupted and rapid data collection and broad application capabilities, as well as lower costs per examined driver. Dahlinger et al. (2018) performed naturalistic experiments and collected data via smartphone. Elvik (2014) conducted naturalistic trials in several countries to evaluate the impact of rewards on drivers.

There is a variety of studies that used driving simulators for experiments (Dijksterhuis et al., 2015; Molloy et al., 2018; Zhao & Wu, 2012). Donmez et al. (2007) conducted a simulator study to investigate real time feedback on drivers. Mullen et al. (2015) used a driving simulator as a cost efficient and effective solution. Brouwer et al. (2015) and Yuan et al. (2021) both used a driving simulator experiment in order to evaluate truck drivers' behavior.

Gamification is the application of game-based design techniques and game-inspired mechanics (e.g. scoring and achievement measurement methods) to non-game contexts. It is a powerful tool that can be used to enable drivers to adopt improved driving behavior (behavior persuasion) (Rossetti et al., 2013). According to Toledo and Lotan (2006) safety-related scores calculated based on in-vehicle monitoring and given to drivers through personal web pages had a major positive impact on driver results. Elvik (2014) carried out a comprehensive analysis of experiments to reward safe and eco-driving and found that they were all successful in encouraging rewarded behaviors. Hamari et al. (2014) found that the effects of gamification (e.g. scores, competition, social pressure, incentives and rewards, tips and recommendations) are positive, although controlled by several factors such as the context in which it is applied as well as the profile of targeted users. Mantouka et al. (2019) found that economic rewards tend to have a major effect on users' willingness to use a mobile application for airports.

The objective of the present study is: (i) to explore the speeding and aggressive behavior of professional drivers based on detailed driving analytics collected by smartphone sensors, and (ii) to investigate whether incentives in a social gamification scheme can improve driving behavior. For that purpose, high-resolution smartphone data collected from a naturalistic driving experiment with a sample of 25 professional drivers is utilized. Generalized Linear Mixed-Effects Models (GLMMs) with the Poisson function are estimated using high-level trip data of professional drivers, to estimate the percentage of driving time over the speed limit and the frequency (counts) of harsh-acceleration and harsh-braking events.

6.2.2 Professional drivers sample and descriptive statistics

Originally, 27 professional drivers volunteered to participate in the experiment and allow for monitoring their driving behavior through the respective smartphone application. However, for the present analysis it was decided that the final sample should consist only of drivers who have participated equally in both phases on terms of trips. An additional criterion was set; all drivers selected for the analysis were required to have driven for at least 20 trips. As a result, from the 27 professional drivers, 19 drivers (all male) were ultimately selected. The participants were aged between 25-34 (n=9), between 35-45 (n=9) and between 45-54 (n=1). Detailed sample information is presented in Table 6.9.

Table 6.9: Participant panel description regarding driving and vehicle data (N=19)

Driving experience (n. of years)	**Percentage %**	**Driven distance per year (km)**	**Percentage %**	**Engine size (cc)**	**Percentage %**
<5	5.6	< 10.000	0.0	<1500cc	0.0
5 - 10	16.7	10.000 - 20.000	22.2	1500 - 1700cc	0.0
11 - 20	38.9	20.000 - 30.000	16.7	1701 - 1900cc	0.0
21 - 30	38.9	30.000 - 40.000	0.0	1901 - 2200cc	55.6
>30	0.0	> 40.000	61.1	> 2200cc	44.4
Total	100.0	**Total**	100.0	**Total**	100.0

Overall, during the two phases of the experiment a large dataset of 5,345 trips from a sample of 19 professional drivers were recorded. Before presenting the model development, it should be highlighted that the majority of professional drivers' trip distance was travelled on highways; namely 84% of the total travelled distance, while 12% and 3% were travelled in the rural and the urban environment, respectively (Figure 6.3). Taking that into consideration, in combination with the specific driving patterns noticed on highways, the authors decided to analyze explicitly the total of trips travelled on highways.

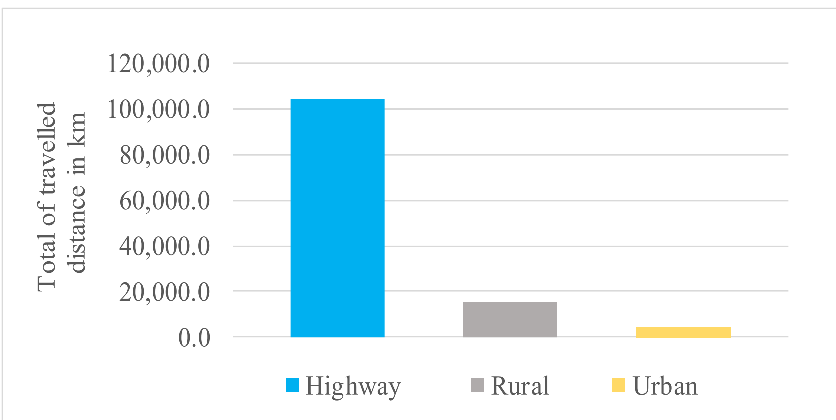


Figure 6.3: Total of travelled distance in km per road type

6.2.3 Generalized linear mixed models for harsh events

In order to model the expected speeding percentage as well as the frequency of events per trip for the participant drivers, models in a GLM framework were calibrated, as previously explained. Since the BeSmart application allows for a high resolution, big-data oriented collection scheme, it was attempted to include random effects in order to capture the unique driving behavior traits for each driver. This entails having a critical minimum sample of trips for each driver to achieve a meaningful outcome. Therefore, a screening was made among participant drivers, as described above, and drivers that had over 20 trips each were selected for the GLMM analysis.

GLMMs were fitted in R-studio (with the lme4 package) via maximum likelihood and using z-factor scaling. A number of models were tested with different configurations in the collected parameters in both fixed effects and random effects. The selected variables were chosen after taking into account the following: lowest Akaike Information Criterion (AIC) for dealing with the trade-off between the goodness of fit of the model and the simplicity of the model, high statistical significance of variables, low multicollinearity, and finally rational interpretation of their impact on the dependent variable. Table 6.10 provides a description of the selected variables.

Table 6.10: Description of the variables used in the analyses

Variable	Description
Competition (binary dummy variable)	Competition phase (yes/no)
Trip Duration (continuous numerical variable)	Total trip duration (sec)
Harsh Accelerations (discrete numerical variable)	Number of harsh accelerations per trip
Weekend (binary dummy variable)	Trip realized during the weekend (yes/no)
Speeding Percentage (numerical variable)	Share of time over the speed limit per trip (%)
Harsh accelerations (discrete numerical variable)	Harsh acceleration events per trip (count)
Harsh brakings (discrete numerical variable)	Harsh braking events per trip (count)

To model the frequency of harsh accelerations events, log-likelihood test ANOVA comparisons were conducted. As is shown in Table 6.11, the most informative configuration of random effects was the inclusion of random intercepts in the GLMMs to capture unique driver traits (lowest LogLikelihood and highest χ^2). Table 6.11 provides a description of the results of mixed effect configuration.

Table 6.11: Log-likelihood comparison of mixed effect configuration for harsh acceleration model

Model Family	Model Configuration	D.f.	Log Likelihood	χ^2	P(>χ^2)	Sig.
GLM	Fixed effects only [baseline]	4	-1198.8	–	–	–
GLMM	Fixed effects & Random Intercepts	5	-1156.8	83.91	<2e-16	***
GLMM	Fixed effects, Random Intercepts & Random Slopes	7	-1151.2	11.13	0.004	**

Results for the harsh acceleration model indicate that the exposure metrics of trip duration as well as driving during the weekend are statistically significant and correlated with the frequency of harsh events. More precisely, trip duration seems to increase the odds of harsh acceleration frequencies; 1 sec of driving time increases acceleration frequencies by 1.558 times. On the other hand, driving during the weekend compared to the weekdays seems to reduce the probability of a harsh acceleration occurrence while driving by a factor of 0.661. This finding can be explained by the different driving style over the week, indicating a less stressful one on the weekends. With respect to the impact of the competition on driving behavior, similar to the speeding model, it is found that drivers seem prone to reducing the frequency of harsh accelerations events when participating in social gamification scheme with prizes and awards; namely by a factor of 0.348.

Table 6.12: GLMMs for harsh accelerations of 19 professional drivers (fixed effects)

Trip characteristic	Estimate	s.e.	p-value	Sig.	Relative Risk Ratio

Intercept	-3.531	0.341	0.000	***	-
Competition	-1.054	0.219	0.000	***	0.348
Trip Duration	0.444	0.026	0.000	***	1.558
Weekend	-0.414	0.175	0.000	*	0.661

Significance codes: '***': 0.000 | '**': 0.001 | '*': 0.01 | '.': 0.05 | ' ': ≥ 0.1

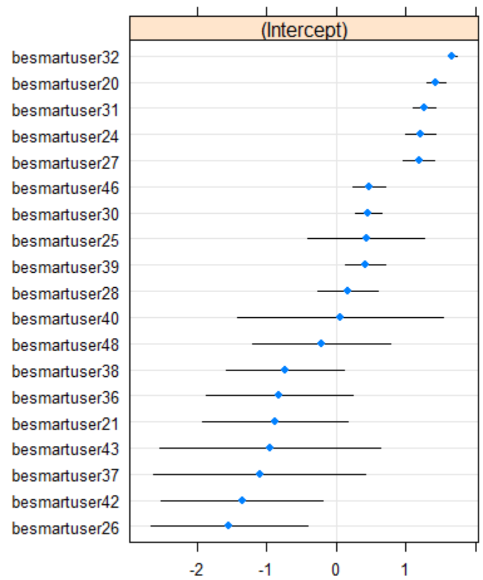


Figure 6.4: Random Intercepts for the harsh acceleration model

Figure 6.4 shows that the inclusion of random intercepts, as identified through log-likelihood test ANOVA comparisons, allows the model to account for individual variability in baseline tendencies for harsh accelerations. Each blue point represents a driver (identifier), and the horizontal lines denote the confidence intervals for their random intercept estimates. The variation in intercepts highlights significant differences in drivers' inherent likelihood of harsh acceleration events, independent of fixed effects. Drivers like "besmartuser32" and "besmartuser20" exhibit higher baseline tendencies (positive intercepts), while others, such as "besmartuser42" and "besmartuser26," show lower tendencies (negative intercepts). This variability underscores the importance of incorporating random effects to better model individual-level differences, capturing unique driver traits that influence harsh acceleration behavior.

Similar to harsh acceleration model, the inclusion of random intercepts was the most informative configuration of random effects in order to capture the unique driver traits. The log-likelihood test ANOVA comparisons are presented in Table 6.13. (lowest LogLikelihood and highest χ^2).

Table 6.13: Log-likelihood comparison of mixed effect configuration for harsh braking model

Model Family	Model Configuration	D.f.	Log Likelihood	χ^2	P(>χ^2)	Sig.
GLM	Fixed effects only [baseline]	4	-3324.6	–	–	–
GLMM	Fixed effects & Random Intercepts	5	-3093.4	462.38	<2e-16	***
GLMM	Fixed effects, Random Intercepts & Random Slopes	7	-3092.3	2.14	0.343	

With respect to harsh braking events model (Table 6.14), all the independent variables, both driving behavioral and exposure ones, seem to have effects that are similar to the ones they have on the harsh acceleration events. The occurring finding is more obvious when examining the relative risk ratio of every variable. Specifically, trip duration appears to increase the odds of harsh braking frequencies; 1 sec of driving time increases braking frequencies by 1.564 times. On the other hand, similarly to the harsh acceleration models, the exposure parameter of the variable “weekend”, was found to be negatively associated with the odds of higher harsh braking counts, corresponding to risk ratio of 0.748. Finally, once more, the competition seems to motivate drivers to improve their performance by adopting a less aggressive style, with lower frequencies of harsh brakings. Notably, driving during the competition phase, the probability of a harsh braking occurrence is reduced by a factor of 0.404.

Table 6.14: GLMMs for harsh brakings of 19 professional drivers (fixed effects)

Trip characteristic	Estimate	s.e.	p-value	Sig.	Relative Risk Ratio
Intercept	-2.384	-8.161	0.000	***	-
Competition	-0.907	-7.738	0.000	***	0.404
Trip Duration	0.447	45.106	0.000	***	1.564
Weekend	-0.290	-3.432	0.001	***	0.748

Significance codes: ‘***’: 0.000 | ‘**’: 0.001 | ‘*’: 0.01 | ‘.’: 0.05 | ‘ ’: ≥ 0.1

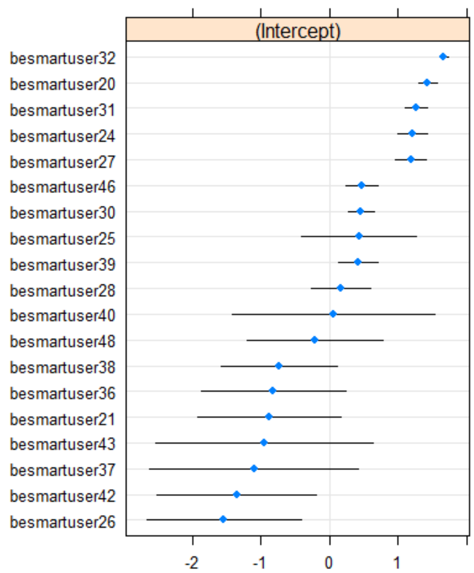


Figure 6.5: Random Intercepts for the harsh braking model

As with the harsh acceleration model, the inclusion of random intercepts was identified as the most informative random effects configuration, based on log-likelihood test ANOVA comparisons (lowest LogLikelihood and highest χ^2). Each blue point corresponds to an individual driver, and the horizontal lines represent the confidence intervals for their random intercept estimates. The variability in intercepts highlights substantial differences in drivers' baseline likelihood of harsh braking events, independent of other fixed effects in the model. For instance, drivers like "besmartuser32" and "besmartuser20" exhibit higher tendencies toward harsh braking (positive intercepts), whereas drivers such as "besmartuser42" and "besmartuser26" display lower tendencies (negative intercepts). This heterogeneity underscores the importance of accounting for unique driver traits in modeling harsh braking behavior.

6.2.4 Discussion of results

This section aimed: (i) to explore the speeding and aggressive behavior of professional drivers on based on detailed driving analytics collected by smartphone sensors, and (ii) to investigate whether incentives in a social gamification scheme can improve driving behavior. For that purpose, high-resolution smartphone data collected from a naturalistic driving experiment with a sample of 19 professional drivers were utilized. Using risk exposure and driving behavior indicators calculated from smartphone sensor data, statistical analyses were carried out for correlating the percentage of driving time over the speed limit, as well as the frequencies of harsh events, with other driving behavior indicators, namely by means of Generalized Linear Mixed-Effects Models.

The results from the interpretation of the estimated parameters of the models can be summarized as follows: Trip duration has a positive correlation with harsh eevents, while driving during the weekends seems to reduce the frequency of harsh events; both accelerations and brakings. In addition, harsh accelerations are associated with the odds of someone exceeding the speed limits, outlining a pattern of an overall unsafe driving behavior.

Furthermore, the present research contributes a preliminary example of quantitative documentation of the impact of encouraging rewarded behaviors on the examined human risk factors; aggressive behavior as expressed by the frequency of harsh accelerations and harsh brakings. Professional drivers constitute a high-risk road user group mainly due to the increased driving time and distance travelled. In that context, rewarding safe driving behavior and providing drivers with motivations and incentives within a social gamification scheme seems to have successful results. State-of-the-art interventions can include approaches for driver training and support through innovative driver behavior monitoring and feedback tools in a variety of ways, personalized feedback with scorecards as well as incentives within a social gamification scheme, with personalized target setting, benchmarking and comparison with peers.

7 Feedback Effects of Different Features on Driver Behavior

7.1 Introduction

Despite considerable progress in road safety over the past decade, road traffic crashes remain a pervasive public health issue globally, resulting in around 1.19 million road traffic deaths in 2021 (Global Status Report on Road Safety 2023. Geneva: World Health Organization; 2023.), corresponding to a rate of 15 road traffic deaths per 100 000 population. The identification of critical risk factors leading to road traffic crashes has been researched by numerous studies over the years. Among these factors, human elements are consistently recognized as the most significant, accounting for the vast majority of road crashes. In fact, human error is cited as the cause of 95% of all road crashes (Singh, 2015). This underscores the importance of understanding and addressing driver behavior as a key component of road safety initiatives. By analyzing driver behavior, targeted interventions can be developed, aiming to mitigate risky actions such as distracted driving, speeding, and impaired driving.

The significance of driver monitoring is becoming more widely acknowledged in the transportation sector (Koesdwiady et al., 2017). Nevertheless, researchers encounter difficulties in collecting accurate real-time driving data with affordable collection and processing techniques. In this context, the widespread use of smartphones and social networks presents new opportunities for monitoring and analyzing driver behavior (Chan et al., 2020). The capabilities of smartphone applications, combined with their low cost and ease of use, facilitate data collection. These advancements facilitate the provision of direct feedback and trip analysis to drivers, potentially reducing road crashes and casualties. Going one step further, the conduction of driving experiments under naturalistic conditions using smartphones allows for the recording of drivers in their normal driving environments without external influences, and thus for the effective assessment of driver behavior (Ziakopoulos et al., 2020). Despite the growing interest from both manufacturing companies and transportation researchers in driver behavior, there is a notable gap in research quantifying the influence of driver feedback on road safety, particularly in terms of comparing data before and after feedback provision.

In this regard, the present study aims to leverage large-scale trip data from smartphone sensors to assess the impact of driver feedback on key performance indicators, such as speeding, harsh braking, and harsh acceleration events. For this purpose, a naturalistic driving experiment has been conducted thousands of trips have been used first to examine the trend of the risk driving indicators and then, Structural Equations Models (SEM) are applied to identify feedback effects to risky driving indicators. The outputs of the two methods are combined to provide some critical insights on whether driver feedback influences driving behavior and in what extent.

7.2 Descriptive Statistics and Preliminary Analysis

Overall, during the 21-months experiment 73,869 trips were recorded from a sample of 175 car drivers (54% female, all ages) who had participated in all experiment phases. This subchapter presents the trend in the most crucial indicators of driving behavior during the experiment, noting the change in the phase of the experiment, in order to draw some initial conclusions about the effect of feedback on driver behavior. Table 7.1 shows the summary statistics of the selected variables, while Figure 7.1 illustrate the trend of the same variables across the different experiment phases.

The trends identified are the following: (i) There was an overall improvement in driving behavior from Phase 1 to Phase 2; (ii) The Covid-19 pandemic and subsequent lockdown measures significantly reduced the travel; (iii) There is a fluctuating improvement in driver behavior in subsequent phases; (iv) There is an improvement in driver behavior during the Competition phase; (v) There is a relapse to worse driving behavior once the Competition and Challenge phase is completed.

Table 7.1: Descriptive statistics of the per trip values of the variables recorded during the experiment

Experiment Phases	**Percentage of mobile use**		**Percentage of speeding time**		**Harsh brakings per 100km**		**Harsh accelerations per 100km**		**Speed above the speed limits**	
	mean	**std**	**mean**	**std**	**mean**	**std**	**mean**	**std**	**mean**	**std**
Phase 1	4.69%	0.13	6.17%	0.10	18.90	31.23	9.20	21.34	4.81	6.37
Phase 2	3.95%	0.12	3.71%	0.07	19.35	32.72	9.64	22.37	4.22	6.30
Phase 3	4.55%	0.13	3.70%	0.08	19.75	31.87	11.34	25.11	3.71	5.76
Phase 4	4.44%	0.13	3.88%	0.08	17.70	31.62	10.53	23.98	3.33	5.28
Phase 5/ Competition	3.03%	0.10	2.69%	0.06	13.21	24.92	8.23	20.44	2.38	4.13
Phase 5/ Challenges	3.04%	0.11	3.37%	0.07	16.21	29.26	8.62	21.86	2.51	4.34
Phase 6	2.35%	0.09	3.72%	0.07	16.89	29.79	8.01	19.72	3.01	4.51

Figure 7.1: Trends of the driving behavior parameters over the different experimental phases

Before moving to the advanced statistical analysis, Wilcoxon signed-rank test was used to compare the examined variable of the different phases and assess whether their population mean ranks differ. Since none of the variables met the assumptions for normality and homogeneity of variances, the Wilcoxon signed-rank test (a non-parametric test) was used for all variables (Table 3). Previous within-subjects design studies in the field of naturalistic driving research have also utilized the test to compare means among different feedback phases (Camden et al., 2019; Newnam et al., 2014).

The Wilcoxon signed-rank test results, showed in Table 7.2, reveal significant changes in driving behaviors across different phases. Notable improvements were observed between Phase 1 and Phase 2, with substantial reductions in mobile use (26.20%), speeding time (41.40%), and harsh braking (12.90%), indicating the effectiveness of feedback interventions. Between Phase 2 and Phase 3, although mobile use and speeding time continued to decrease, harsh braking slightly increased, suggesting mixed outcomes. The comparison between Phase 3 and Phase 4 showed an increase in mobile use (9.60%) but a decrease in harsh accelerations (11.20%), highlighting persistent distractions despite some improvements. Lastly, the shift from Competition and Challenges to the last phase resulted in significant increases in all risky behaviors, emphasizing that drivers may relapse to unsafe behaviors once not receiving anymore feedback.

Table 7.2: Wilcoxon signed-rank test results comparing means of risky driving behavior parameters across each phase

Phases compared	Driving behavior parameters	mean diff.	S statistic	p-value
Phase 1 - Phase 2	Percentage of mobile use	-26.20%	1.50591E+11	**<0.01**
Phase 1 - Phase 2	Percentage of speeding time	-41.40%	4.63629E+11	**<0.01**
Phase 1 - Phase 2	Harsh brakings per 100km	-12.90%	4.05868E+11	**<0.01**
Phase 1 - Phase 2	Harsh accelerations per 100km	-2.50%	2.07426E+11	**<0.01**
Phase 1 - Phase 2	Speed above the speed limits (km/h)	-18.00%	4.14412E+11	**<0.01**
Phase 2 - Phase 3	Percentage of mobile use	-26.80%	2.74877E+11	**<0.01**
Phase 2 - Phase 3	Percentage of speeding time	-16.70%	5.50967E+11	**<0.01**
Phase 2 - Phase 3	Harsh brakings per 100km	1.50%	7.48152E+11	**<0.01**
Phase 2 - Phase 3	Harsh accelerations per 100km	0.00%	4.86203E+11	**<0.01**
Phase 2 - Phase 3	Speed above the speed limits (km/h)	-27.60%	5.1229E+11	**<0.01**
Phase 3 - Phase 4	Percentage of mobile use	9.60%	1.08069E+12	**<0.01**
Phase 3 - Phase 4	Percentage of speeding time	-5.80%	2.46395E+12	**<0.01**
Phase 3 - Phase 4	Harsh brakings per 100km	-10.00%	2.68961E+12	**<0.01**
Phase 3 - Phase 4	Harsh accelerations per 100km	11.20%	1.98769E+12	**<0.01**
Phase 3 - Phase 4	Speed above the speed limits (km/h)	1.30%	2.27401E+12	**<0.01**
Phase 4 - Competition	Percentage of mobile use	-3.90%	27327559116	0.657
Phase 4 - Competition	Percentage of speeding time	-13.10%	72000240177	**<0.01**
Phase 4 - Competition	Harsh brakings per 100km	-3.20%	82995973298	**<0.01**
Phase 4 - Competition	Harsh accelerations per 100km	-10.30%	47511875657	**<0.01**
Phase 4 - Competition	Speed above the speed limits (km/h)	-20.90%	72401356187	**<0.01**
Competition - Challenges	Percentage of mobile use	10.00%	1502836826	**<0.01**
Competition - Challenges	Percentage of speeding time	50.70%	4321598228	**<0.01**
Competition - Challenges	Harsh brakings per 100km	41.50%	5775250817	**<0.01**
Competition - Challenges	Harsh accelerations per 100km	30.00%	2623882294	**<0.01**
Competition - Challenges	Speed above the speed limits (km/h)	24.30%	4081454761	**<0.01**
Challenges - Phase 6	Percentage of mobile use	2.90%	4712410638	**<0.01**
Challenges - Phase 6	Percentage of speeding time	4.00%	18068246934	**<0.01**
Challenges - Phase 6	Harsh brakings per 100km	-4.90%	21576313967	**<0.01**
Challenges - Phase 6	Harsh accelerations per 100km	1.80%	9765516199	**<0.01**
Challenges - Phase 6	Speed above the speed limits (km/h)	13.00%	18002330803	**<0.01**

The results of the Wilcoxon signed-rank test provide a comprehensive understanding of how driver behavior evolves across different feedback phases and contexts. The significant reductions in mobile phone use, speeding time, and harsh braking between Phase 1 and Phase 2 strongly indicate the initial effectiveness of feedback interventions in curbing risky behaviors. These findings align with previous research that highlights feedback as a powerful tool for behavior modification in naturalistic driving settings. The observed improvements suggest that drivers responded positively to feedback, adopting safer driving practices during the early stages of intervention.

However, the subsequent phases reveal more nuanced outcomes, highlighting the challenges of sustaining long-term behavioral changes. Between Phase 2 and Phase 3, while mobile use and speeding continued to decrease, a slight increase in harsh braking suggests potential adaptation effects or situational factors that influenced driver responses. The increase in mobile use observed between Phase 3 and Phase 4, despite reductions in harsh accelerations, underscores the persistent challenge of distractions and the complexity of driver behavior over time. Importantly, the significant increases in risky behaviors during the shift from the Competition phase to the Challenges phase emphasize the potential for relapse when feedback or incentives are removed. This relapse underscores the necessity of continuous engagement strategies or reinforcement mechanisms to maintain safe driving behaviors over extended periods.

7.3 Structural Equations Models

This section presents the results of the SEM analysis, focusing solely on the final models, presented in Table 7.3. In addition to the previously mentioned hard goodness-of-fit measures, the coefficient estimates produced were evaluated to ensure they provided logical and interpretable results. Efforts were made to avoid model misspecification by considering the appropriateness of the theoretical framework and the outcomes produced. During the modeling process, it was evident that certain model structures were significantly better suited to the experimental data based on specific criteria; only these best-fitting models are presented here. Variations within each latent variable structure were explored using the backwards elimination technique.

All statistical analyses were conducted in R-studio (R Core Team, 2013) and SEM analysis in particular utilized the lavaan R package. Ultimately, the proposed SEM structure retained two latent unobserved variables:

- Feedback, expressing the influence of the different features of the smartphone app during the different phases of the experiment, namely Baseline, Scorecard feature, Maps feature, Compare feature, Competition and Challenges feature.

- Exposure, expressing the influence of the exposure metrics, namely Distance (for driving speed 30km/h – 50km/h), Morning peak and Afternoon peak.

It is important to note that mobile phone use while driving was initially included as a variable in the examined models. However, it did not prove to be statistically significant and was subsequently omitted from the final model. The analysis instead focused on other recorded driving behavior indicators, specifically the percentage of speeding time, harsh brakings per 100 km, and harsh accelerations per 100 km.

Table 7.3: SEM model of Percentage of speeding time, Harsh Brakings per 100km & Harsh Accelerations per 100km

SEM Components		**Parameters**	**Estimate**	**S.E.**	**z-value**	**P(>|z|)**
Latent Variables	Feedback	Baseline	1.000	–	–	–
		Scorecard feature	2.076	0.014	148.640	**0.000**
		Maps feature	1.646	0.010	157.864	**0.000**
		Compare feature	1.215	0.029	41.754	**0.000**
		Competition & Challenges feature	2.053	0.038	54.447	**0.000**
	Exposure	Distance (for driving speed 30km/h – 50km/h)	1.000	–	–	–
		Morning peak	2.473	0.350	7.072	**0.000**
		Afternoon peak	-1.360	0.129	-10.579	**0.000**
Regressions	Percentage of speeding time	Intercept	0.409	0.003	138.941	**0.000**
		Exposure	0.326	0.043	7.627	**0.000**
		Feedback	-0.214	0.014	-15.655	**0.000**
	Harsh Accelerations per 100km	Intercept	0.099	0.001	95.037	**0.000**
		Exposure	0.028	0.010	2.769	0.006
		Feedback	0.026	0.004	6.493	**0.000**
		Competition & Challenges feature	-0.001	0.000	-2.748	**0.000**
		Afternoon peak	0.006	0.002	3.095	**0.002**
	Harsh Brakings per 100km	Intercept	0.184	0.001	158.258	**0.000**
		Exposure	0.077	0.014	5.542	**0.000**
		Feedback	-0.027	0.005	-4.976	**0.000**
Covariances	Percentage of speeding time	Harsh Brakings per 100km	0.007	0.001	7.686	**0.000**
	Harsh Accelerations per 100km	Percentage of speeding time	0.006	0.001	9.526	**0.000**
	Harsh Brakings per 100km	Harsh Accelerations per 100km	0.021	0.000	75.739	**0.000**
	Feedback	Exposure	-0.001	0.000	-5.558	**0.000**
Goodness-of-fit measures	CFI		0.940			
	TLI		0.944			
	RMSEA		0.049			0.845
	SRMR		0.025			

The model's goodness-of-fit measures further support the robustness of these findings, with a CFI of 0.940, TLI of 0.944, RMSEA of 0.049, and SRMR of 0.025. All four examined goodness-of-fit measures and the signs of the estimated coefficients indicate an excellent model fit. Additionally, the model's AIC was the lowest among the tested combinations, and no negative variances were observed, which would have suggested model misspecification (variance outputs are omitted for brevity). To enhance model fit, several variables were scaled linearly by factors of 10, which reduced variance discrepancies without affecting the interpretation of the coefficients. Moreover, an iterative process was employed to integrate covariances of the measured variables into the model. This involved comparing observed and fitted covariance correlations and addressing the largest discrepancies by including relevant covariance pairs, provided there were no significant theoretical objections. This approach significantly improved the overall model fit.

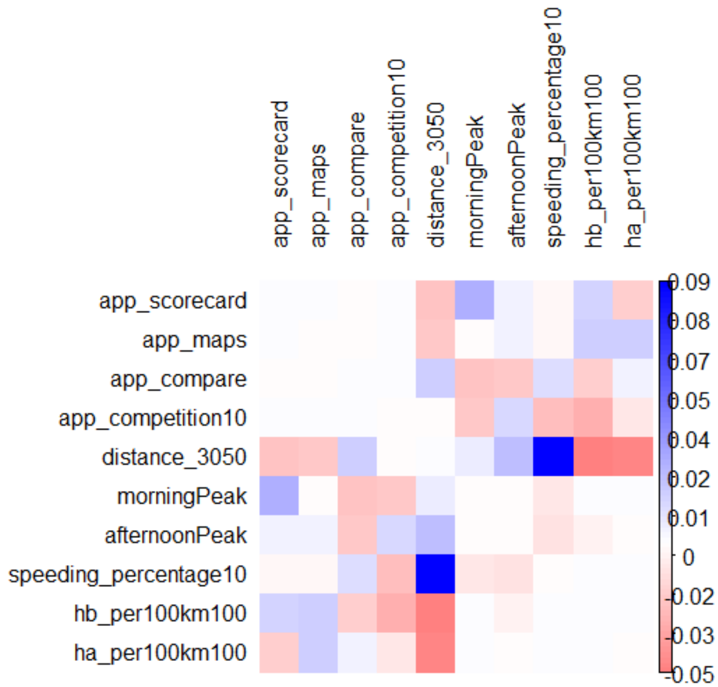


Figure 7.2: Residual correlation heatmap for SEM model

Furthermore, Figure 7.2 provides a heatmap visualization of the residual correlations among the observed variables in the SEM model. The colors indicate the degree and direction of residual discrepancies: blue represents overestimation by the model, while red indicates underestimation. The strongest correlations (blue) are observed between the latent variables app_competition10 and distance_3050, as well as between speeding_percentage10 and hb_per100km100. This suggests that while the model fits well overall, these relationships highlight potential areas for further refinement or covariate inclusion. The relatively low range of residual values (maximum absolute value = 0.09) supports the robustness of the model. However, minor adjustments to better account for specific interactions—such as those involving distance_3050—could enhance the model's explanatory power.

The path diagram of the present model is presented on Figure 7.3; green arrows denote positive correlations, while red arrows denote negative correlations. Several useful insights can be obtained from the produced SEM model results that are discussed in detail in the following subsections.

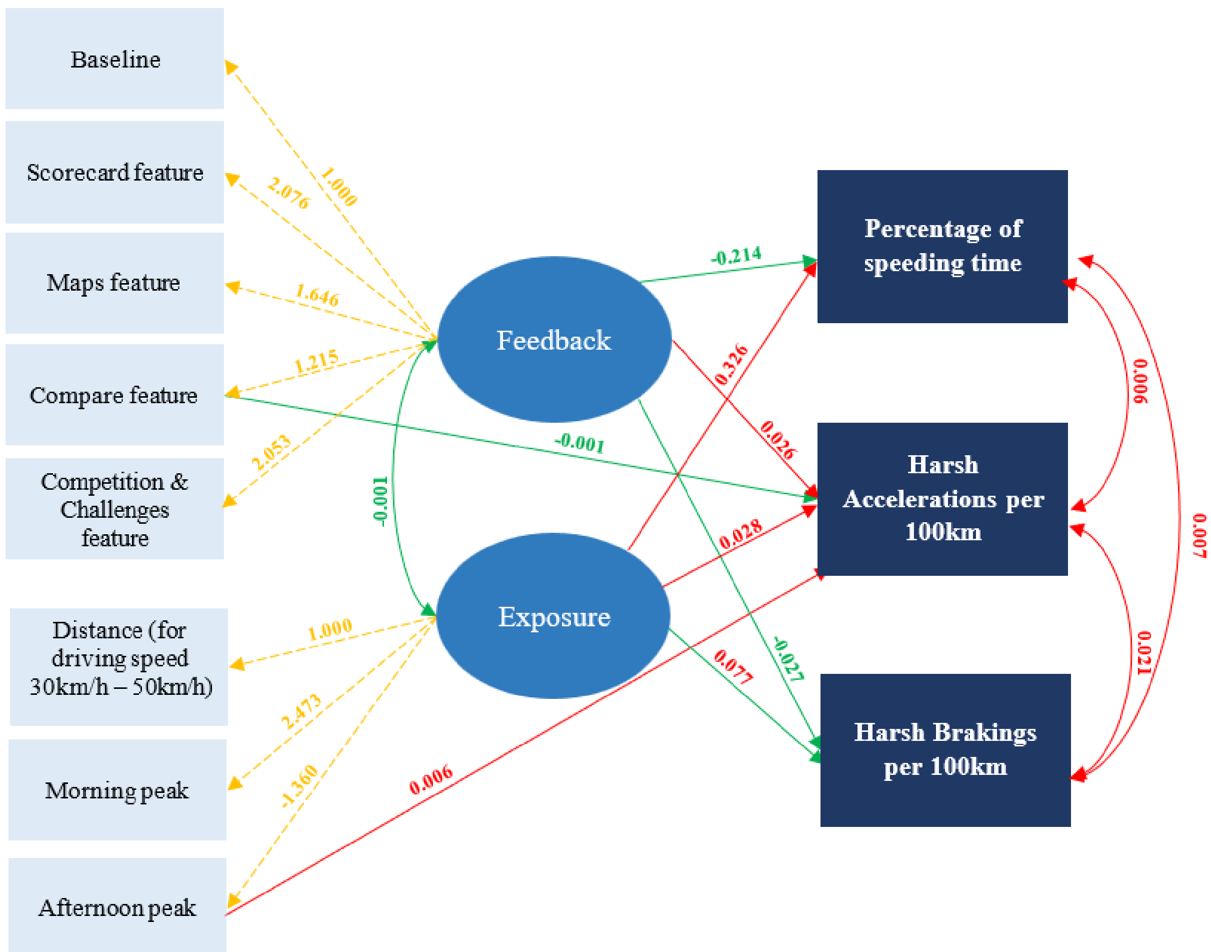


Figure 7.3. Path diagram of SEM model for percentage of speeding time, harsh accelerations per 100km and harsh brakings per 100km

Feedback

The SEM analysis reveals that feedback mechanisms, including the scorecard, maps, compare, and competition & challenges features, significantly impact driver behavior. The scorecard feature has the highest positive estimate at 2.076 ($p < 0.001$), indicating its crucial role in modifying driving habits. This finding is particularly interesting as it demonstrates the powerful impact of feedback, leading to an immediate shift in driving behavior once participants are made aware of their risky actions. Similarly, the maps feature shows a strong influence with an estimate of 1.646 ($p < 0.001$), suggesting that providing drivers with map-based feedback can effectively encourage safer driving practices.

The compare feature, with an estimate of 1.215 ($p < 0.001$), helps drivers assess their performance relative to others, positively influencing behavior. Additionally, the competition & challenges feature, with an estimate of 2.053 ($p < 0.001$), motivates drivers through competitive elements, reinforcing safe driving behaviors. Overall, these feedback mechanisms are effective in reducing the percentage of speeding time (feedback estimate: -0.214, $p < 0.001$) and harsh braking incidents (feedback estimate: -0.027, $p < 0.001$), although there is a slight increase in harsh accelerations (feedback estimate: 0.026, $p < 0.001$), which may require further refinement of the feedback system.

Exposure
Exposure factors, particularly the times of day, play a significant role in driving behaviors. Morning peak exposure is associated with increased driving aggressiveness, as indicated by the significant positive estimate of 2.473 ($p < 0.001$). This suggests that drivers are more likely to engage in risky behaviors, such as speeding and harsh braking, during morning peak hours. This is likely due to their rush to reach work or university, resulting in a sense of urgency.

In contrast, afternoon peak exposure has a negative estimate of -1.360 ($p < 0.001$), indicating that driving behaviors may be less aggressive during this time. The distance driven at speeds between 30km/h and 50km/h serves as a reference parameter with an estimate of 1.000. Understanding these exposure patterns can help in designing targeted interventions to mitigate risky driving behaviors during specific times of the day.

Regressions
The regression analysis provides insights into how exposure and feedback influence specific driving behaviors. The percentage of speeding time is positively associated with exposure, with an estimate of 0.326 ($p < 0.001$), indicating that increased driving time leads to more speeding. However, feedback mechanisms significantly reduce speeding, with an estimate of -0.214 ($p < 0.001$), highlighting their effectiveness. Harsh accelerations per 100km show a slight positive association with exposure (estimate: 0.028, $p = 0.006$) and feedback (estimate: 0.026, $p < 0.001$), suggesting that while feedback reduces some risky behaviors, it may inadvertently increase others.

The competition & challenges feature slightly reduces harsh accelerations (estimate: -0.001, $p < 0.001$), demonstrating its potential in moderating aggressive driving. Afternoon peak exposure increases harsh accelerations (estimate: 0.006, $p = 0.002$), further emphasizing the need for targeted interventions. Harsh brakings per 100km are positively influenced by exposure (estimate: 0.077, $p < 0.001$) but significantly reduced by feedback (estimate: -0.027, $p < 0.001$), reinforcing the importance of feedback in promoting safer driving.

Covariances
The covariance analysis reveals strong interrelationships between various driving behaviors. There is a positive correlation between the percentage of speeding time and harsh brakings per 100km (estimate: 0.007, $p < 0.001$), indicating that drivers who speed are also more likely to brake harshly. Similarly, there is a positive correlation between speeding and harsh accelerations (estimate: 0.006, $p < 0.001$), suggesting that these behaviors often co-occur in aggressive driving patterns. The strong positive correlation between harsh brakings and harsh accelerations (estimate: 0.021, $p < 0.001$) further supports this finding.

Additionally, a negative correlation between feedback and exposure (estimate: -0.001, $p < 0.001$) indicates that increased feedback is associated with decreased exposure to risky driving conditions. These covariance relationships highlight the interconnected nature of different driving behaviors and the importance of comprehensive intervention strategies to address multiple aspects of driver behavior simultaneously.

7.4 Discussion of Results

Rapid technological advances, especially in telematics and Big Data analytics, as well as the increasing penetration and use of information technology by drivers (e.g. smartphones), provide new capabilities for monitoring and analyzing driving behavior. In this section, the effect of driver feedback via a smartphone application on driving behavior risk indicators within a multiphase naturalistic driving experiment was examined. First, a preliminary analysis highlighted the beneficial effects of upgraded feedback on key risk indicators across experiment phases. Subsequently, SEM analysis on a 73,869-trip dataset provided significant insights into how feedback mechanisms and exposure factors influence driving behaviors.

Feedback features, namely scorecard, maps, compare, and competition & challenges seem to be effective in reducing risky driving behaviors like speeding and harsh braking, although they may slightly increase harsh accelerations, at least some of the feedback features. Morning peak exposure is associated with more aggressive driving, while afternoon peak exposure tends to be less risky. Additionally, the strong positive correlations between speeding, harsh braking, and harsh accelerations highlight the interconnected nature of aggressive driving behaviors, confirming previous studies (Su et al., 2023) and showcasing the importance of driver behavior analysis.

The present findings come with some practical implications, as well. First, further in depth examination of driver feedback is necessary to quantify the complex relationships involved in various driving tasks, taking also into account the two other road safety pillars; environment and vehicle. On that note, modifications to vehicle and mobile phone interfaces could be beneficial, particularly with the expected rise in connectivity and automation in the future. Additionally, incorporating eco-driving feedback is crucial to understanding and enhancing the effectiveness of feedback mechanisms.

As the transportation landscape evolves, particularly with the rise of micromobility within multimodal transport systems and the emergence of new travel patterns in the wake of COVID-19, understanding these shifts in decision-making processes is crucial. In that context, the ultimate goal of providing feedback to drivers is to activate the process of learning and self-assessment, enabling them to gradually improve their performance and monitor their progress. This process involves establishing detailed cause-and-effect relationships between aggressive driving and associated risks, offering valuable insights for improving road safety. Such information is beneficial for insurance companies, fleet management applications, and identifying hazardous geographical locations on the road network. Additionally, feedback can serve as a tool for objectively proving driving behavior, allowing users to gain benefits from their insurance companies or to regain their driver's license after revocation.

While the study provides valuable insights into the impact of driver feedback mechanisms on driving behavior, it is not without limitations. One significant limitation is the exclusion of mobile phone use as a variable in the final model due to its lack of statistical significance. This omission may have overlooked potential nuances in driver distraction and its interaction with other risky behaviors. Additionally, the naturalistic driving experiment was conducted within specific geographic and temporal contexts, which may limit the generalizability of the findings to different regions or periods. Moreover, the reliance on self-selected participants who voluntarily used the feedback application could introduce selection bias, as these individuals may inherently be more safety-conscious or tech-savvy compared to the general driving population. This aspect of the experiment design is particularly intriguing and warrants further investigation; i.e. a follow-up questionnaire directed at participants could provide additional insights, allowing for a deeper understanding of their motivations and behaviors.

Undoubtedly, much more work needs to be done in this promising area and many challenges need to be addressed. Many different techniques have been used and there is a need for comparison integration to optimize the reliability and level of detail of research results. The advent of new technologies (e.g. in sensors) is progressively simplifying the application of naturalistic driving experiments, and new analysis techniques, such as artificial intelligence and machine learning, allow for a more efficient and in-depth analysis of results. Finally, considering other vulnerable road users (cyclists, pedestrians) in naturalistic driving research allows for the identification and understanding of problems specific to these groups of road users and the identification of possible solutions in ways that were not previously applicable.

Future research should also focus on complex spatio-temporal analysis of data to develop advanced statistical models for identifying dangerous locations and road segments, through comparison with road accident locations, and reduced fuel consumption. The richness of the collected data can be exploited to develop easy-to-use reliable road maps by examining changes in driving risk exposure under changing traffic conditions and in different regions. The transfer and scalability potential of the methodology and research derivatives will be a legacy for the development of maps at any region and scale.

8 Post-Feedback Effect on Long-Term Driver Behavior

8.1 Introduction

Feedback-based interventions have been widely explored as a means to improve driver behavior and enhance road safety. However, the long-term effects of these systems remain inconsistent, with post-feedback behaviors varying significantly across studies. In some cases, feedback has demonstrated enduring positive impacts on driver behavior, even after the feedback period ended. For instance, Ghamari et al. (2022) observed that bus drivers sustained their improved behavior beyond the intervention phase, highlighting the potential for lasting change. Similarly, Merrikhpour et al. (2014) found that although behavioral improvements slightly diminished post-intervention, they remained better than baseline levels, reflecting a degree of persistence. Molloy et al. (2023) also reported sustained improvements in speed compliance, albeit limited to low-speed zones (e.g., 50 km/h).

On the other hand, several studies have shown that once feedback is discontinued, drivers often revert to pre-feedback behaviors. For example, Bolderdijk et al. (2011) observed an increase in speeding among drivers once financial incentives tied to the feedback were removed, suggesting that such effects may not persist without ongoing reinforcement. Similarly, Toledo and Lotan (2006) found that initial reductions in driving risk indices dissipated within five months, with metrics returning to or even surpassing baseline levels. Mazureck and Van Hattem (2006) also noted that most drivers reverted to their prior habits following the feedback phase.

Interestingly, some studies have highlighted the importance of continuous reinforcement mechanisms, such as follow-up coaching or ongoing monitoring, in sustaining behavioral changes. McGehee et al. (2007) reported a durable reduction in safety-critical events among high-frequency drivers when feedback was coupled with consistent reinforcement, such as training or performance improvement plans. This underscores the role of structured follow-ups in prolonging the benefits of feedback interventions.

Certain studies also suggest that feedback interventions can deliver long-term benefits under specific conditions. For example, Soleymanian et al. (2019), Takeda et al. (2011), and Toledo et al. (2008) reported sustained improvements in driver behavior. However, these findings were often limited by low certainty due to issues such as data attrition or study dropout. Attrition bias is a critical challenge in long-term experiments, as drivers may lose engagement over time, leading to reduced statistical power and potential skewing of results. This loss of participation raises the question of whether sustained behavior change could have been achieved with ongoing interventions and consistent participant engagement.

These mixed findings highlight the complexity of achieving sustained behavioral change through feedback interventions and underline the need for further research to understand the factors that influence long-term outcomes. This section seeks to address this gap by examining the post-feedback effects on driver behavior over extended periods, focusing on the mechanisms that contribute to or hinder sustained improvements through survival analysis.

8.2 Participants and Preliminary Analysis

In this study, a cohort of 31 car drivers who participated across all experimental phases—baseline, feedback, and post-feedback—was analyzed. Over the course of the 21-month experiment, these drivers completed a total of 24,904 trips, with each driver contributing a minimum of 20 trips in the post-feedback phase. Overall, the broader experiment involved a larger sample of 175 drivers and documented 73,869 trips, and feedback features different effect on driver behavior was investigated in the previous section of this thesis. This subset of 31 drivers was selected to closely examine the long-term effects of feedback interventions on driving behavior. Summary statistics for the selected sample are shown in Table 8.1, while critical driving indicators are shown in Table 8.2 and Table 8.3.

Table 8.1: Overview of the selected sample

	Age groups			
	18-34	35-54	>55	Total%
Male	6	5	3	45%
Female	8	5	4	55%
Total %	45%	32%	23%	100%

Table 8.2: Summary statistics for critical driving indicators during the different phases

Indicators	Baseline	Feedback	Post-Feedback
Mean of the percentage of mobile phone use while driving (sd)	3.34% (0.11)	2.17% (0.09)	2.33% (0.09)
Mean of the percentage of speeding while driving (sd)	5.42% (0.09)	2.81% (0.06)	3.74% (0.07)
Mean of harsh accelerations per 100 km (sd)	6.68 (17.27)	6.88 (18.72)	7.96 (19.67)
Mean of harsh braking per 100km (sd)	16.86 (27.15)	14.479 (26.50)	16.734 (29.55)

To rigorously examine the impact of feedback on driving behavior across the Baseline, Feedback, and Post-Feedback phases, a Wilcoxon signed-rank test was conducted for each key behavioral indicator. Given that the data did not meet assumptions for normality and homogeneity of variances, the non-parametric Wilcoxon signed-rank test was chosen to compare median ranks across the three phases. This approach is consistent with methodologies used in prior naturalistic driving studies that explore changes in driver behavior under feedback interventions (Camden et al., 2019; Newnam et al., 2014).

Table 8.3: Results of the non-parametric Wilcoxon signed-rank test for the examined indicators

Variable	Comparison	Statistic	p value	Mean difference
mbu_percentage	Baseline vs. Feedback	2.05E+11	<0.01	-0.35%
mbu_percentage	Baseline vs. Post-Feedback	2.42E+09	<0.01	-0.28%
mbu_percentage	Feedback vs. Post-Feedback	9.26E+10	<0.01	0.16%
speeding_percentage	Baseline vs. Feedback	1.41E+12	<0.01	-2.74%
speeding_percentage	Baseline vs. Post-Feedback	1.79E+10	<0.01	-2.48%
speeding_percentage	Feedback vs. Post-Feedback	4.17E+11	<0.01	0.41%
hb_per100km	Baseline vs. Feedback	1.20E+12	<0.01	-0.93

hb_per100km	Baseline vs. Post-Feedback	1.50E+10	<0.01	0.45
hb_per100km	Feedback vs. Post-Feedback	5.15E+11	<0.01	1.98
ha_per100km	Baseline vs. Feedback	4.21E+11	<0.01	0.47
ha_per100km	Baseline vs. Post-Feedback	5.34E+09	<0.01	0.96
ha_per100km	Feedback vs. Post-Feedback	2.32E+11	<0.01	0.69

As shown in Table 8.2, the comparison between the Baseline and Feedback phases showed a significant improvement in driving behavior. Indicators of risky driving, such as mobile phone use, speeding percentage, and frequency of harsh braking, decreased notably during the Feedback phase. This trend aligns with previous studies indicating that post trip feedback can enhance safe driving behaviors by making drivers more aware of their risky actions (Camden et al., 2019; Newnam et al., 2014). Feedback mechanisms have been shown to promote self-regulation and encourage drivers to adjust their behavior to safer standards, supporting the hypothesis that feedback can be an effective tool for improving road safety in the short term.

Despite the overall improvement in safe driving during the Feedback phase, some fluctuations in risky behavior were observed. Certain indicators, i.e. harsh accelerations per 100 km showed intermittent increases, suggesting that driver adaptation to feedback may vary across individuals and situations. This finding highlights the complexity of driver behavior under feedback conditions and aligns with studies that report varied responses to feedback interventions over time (Johnson et al., 2020). The fluctuation may reflect an adjustment period as drivers acclimate to receiving real-time feedback, underscoring the potential need for prolonged or reinforced feedback to achieve steady behavioral improvements.

A notable relapse in risky driving behavior occurred once feedback was withdrawn in the Post-Feedback phase. Indicators such as mobile phone use, speeding, and harsh braking increased compared to the Feedback phase, suggesting that the absence of feedback may lead drivers to revert to previous, riskier behaviors. This relapse effect is consistent with findings from Newnam et al. (2014) and Johnson et al. (2020), who observed similar tendencies in naturalistic driving studies. The observed relapse highlights a critical challenge in sustaining safe driving habits over the long term, as drivers may rely on continuous feedback to maintain safer behavior.

The results of the Wilcoxon signed-rank test suggest that further research is necessary to better understand the impact of post-feedback effects on long-term driving behavior. To address this need, the present study will employ survival analysis to explore the mechanisms underlying these post-feedback effects, providing insights into the sustainability of behavioral changes and the potential for relapses in the absence of feedback.

8.3 Survival Analysis of Driving Behavior Relapse After Feedback Withdrawal

The integration of survival analysis into driving studies has allowed researchers to understand the temporal patterns in risky driving and the factors that increase or decrease the likelihood of adverse events. For instance, Afghari et al. (2020) applied survival analysis to examine the impacts of distractions on developing time of a critical driving event, further showing its utilitarian value in understanding real-time driving risk. Similarly, the Kaplan-Meier curve and the parametric models, including the Weibull and Cox proportional hazards models, have been one of the widely applied methods to quantify the impact that environmental and behavioral factors exert on the timing of risky events (Wang et al., 2021). This is a methodological approach that has considerably enriched the literature on driver safety and risk mitigation strategies with valuable practical implications for designing interventions and driver monitoring systems.

Survival analysis was employed because of its ability to analyze time-to-event data, making it particularly suitable for understanding driver behavior over time. In the context of post-feedback effect, survival analysis allows for the examination of how long, drivers maintain improved behaviors before relapsing into their original patterns. This method is essential for capturing not only whether behavioral changes occur but also the duration for which these changes persist, providing a dynamic view of the intervention's long-term impact. By focusing on time-to-relapse as the primary outcome, survival analysis enables the evaluation of feedback effectiveness in fostering sustained behavioral improvements. While most driving behavior studies have analyzed time-to-event data within a microscopic framework—such as deceleration adjustment times in car-following scenarios (Adavikottu et al., 2023) or the probability of detecting a pedestrian during mobile phone use (Choudhary & Velaga, 2017)—this study focuses on a macroscopic analysis of time-to-event data, examining driving behavior relapse after feedback withdrawal during the final phase of the experiment.

Another key advantage of survival analysis is its ability to handle censored data, which is common in longitudinal studies. In this research, not all drivers may have experienced a relapse or completed the entire study period, leading to incomplete observations. Survival analysis addresses this issue by incorporating censored data into the model, ensuring that the findings remain robust and reflective of the entire study population. Moreover, the inclusion of covariates such as age, driving experience, and vehicle characteristics in survival models allows for the identification of factors influencing the likelihood and timing of behavioral relapse, providing deeper insights into the mechanisms underlying the post-feedback effects.

Furthermore, survival analysis facilitates the comparison of different intervention strategies by estimating and visualizing survival curves, such as Kaplan-Meier plots, and by fitting advanced models like Accelerated Failure Time (AFT) and Cox proportional hazards models, or even machine learning techniques like Random Survival Forest. These tools help quantify the influence of specific variables on the duration of behavior change while accounting for time-dependent effects and individual heterogeneity through random effects or frailty models. The detailed description of the survival models used in this thesis are presented in Section 3. In summary, basic terminologies for the survival analysis in the present section are defined as following:

- Event: In this analysis, an "event" is defined as a "relapse" in driving behavior, specifically when the driver's examined behavior (i.e. harsh accelerations, harsh brakings, speeding and mobile phone use) exceed a predefined threshold. This threshold is calculated as the mean behavior indicator rate observed during the feedback phase, a period of active intervention. Exceeding this threshold in the post-feedback phase signals a decline in driving behavior, which is considered an "event" for survival analysis purposes.

- Duration Variable (Time to Event): The duration variable in this analysis is represented by the number of trips taken until a relapse event occurs. In survival analysis terms, the duration is a continuous random variable T with a cumulative distribution function F(t) and probability density function f(t). The survival analysis tracks the probability of a driver maintaining improved driving behavior over successive trips in the post-feedback phase.

- Survival Rate (S(t)): The survival rate S(t) gives the probability that a driver will maintain driving behavior below the harsh acceleration threshold for a given number of trips, denoted by t. This can be interpreted as the probability of no relapse occurring within that period.

- Hazard Rate (h(t)): The hazard rate h(t) represents the conditional probability of a relapse occurring at a particular trip t, given that no relapse has occurred up until that trip. It provides an instantaneous risk of relapse at each point in time (number of trips). In this study, as the number of trips increases in the post-feedback phase, the probability of relapse also tends to increase, indicating a rising hazard rate over time.

All survival analyses were performed in R Studio. The survfit function was used to generate Kaplan-Meier curves for survival probability visualization. The coxme function fitted a Cox Proportional Hazards model with frailty effects, while cox.zph validated the proportional hazards assumption. A Weibull Accelerated Failure Time (AFT) model was implemented using survreg, incorporating clustered heterogeneity with the cluster argument. The Random Survival Forest model was built using the ranger package to explore non-linear relationships and identify key predictors. Model performance and comparisons were assessed using metrics like AIC, BIC, and concordance indices computed through predict and concordance functions. These tools facilitated a comprehensive exploration of survival patterns.

8.3.1 Survival analysis of relapse in harsh accelerations

8.3.1.1 The Kaplan-Meier curves

The Kaplan-Meier curve is a stepwise function that visually represents the survival probability over time. Its primary advantage lies in its ability to handle censored data, which is particularly useful in studies where subjects may not experience the event by the end of the observation period. The Kaplan-Meier survival curve shows the proportion of drivers who continue to maintain improved driving behavior without relapse across successive trips. The survival probability is recalculated at each relapse event, giving a stepwise depiction of the declining survival rate as drivers accumulate trips post-feedback. The Kaplan-Meier survival curve for harsh accelerations is shown in Figure 8.1.

The Kaplan-Meier survival curve for harsh accelerations per 100 km in the post-feedback phase provides insights into the long-term impact of feedback on driving behavior. These results demonstrate that while feedback led to temporary behavior improvement, the effect diminishes over time without continued feedback. The Kaplan-Meier analysis highlights the necessity for sustained or recurring interventions to reinforce safer driving behaviors in the long term.

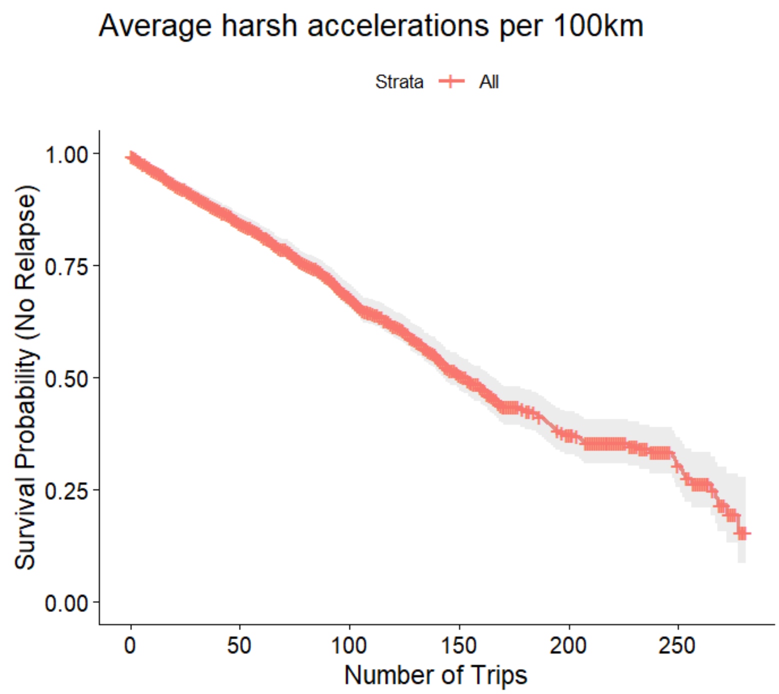


Figure 8.1: Kaplan-Meier survival curve for harsh accelerations per 100 km

Initial Period: The survival probability remains relatively high initially, indicating that most drivers maintain improved behavior within the first several trips.
Declining Survival Probability: As the number of trips increases, the survival probability declines, with notable drops at specific points. For instance:

- 50 trips: approximately 84.8% of drivers still maintain improved behavior (no relapse).
- 100 trips: the survival probability decreases to about 68.7%, showing a progressive decline in adherence to safer driving behavior.
- 150 trips: the probability further drops to around 49.2%, suggesting that nearly half of the drivers have relapsed to pre-feedback levels of harsh acceleration.

8.3.1.2 Cox-PH Model with Frailty

In this study, the Cox Proportional Hazards (Cox-PH) model with frailty was applied to capture the nuanced relationship between driver behavior and relapse after feedback phase, using both naturalistic driving data and questionnaire data, as well. This semi-parametric approach is particularly advantageous because it models the hazard rate of relapse without imposing strict assumptions about the underlying distribution of survival times. By focusing on the hazard rate, the Cox-PH model allows for a dynamic analysis of how various factors, such as age, driving experience, and driving conditions, influence the likelihood of relapse over time. This flexibility is critical in studying complex behavioral patterns where survival times may not conform to traditional distributions.

The inclusion of frailty in the Cox-PH model further enhances its utility by accounting for unobserved heterogeneity across drivers. Frailty acts as a random effect, capturing individual-specific differences that might not be fully explained by observed covariates. This is especially important in driving behavior studies, where factors like personality, driving style, or situational influences can introduce variability in relapse rates. Additionally, the model's ability to provide

interpretable hazard ratios enables a clear understanding of the relative impact of different variables on relapse risk, offering actionable insights for tailoring feedback interventions to specific driver groups. By leveraging these strengths, the Cox-PH model with frailty provides a robust framework for analyzing time-to-relapse data in the context of driver safety feedback.
The results of the Cox Proportional Hazards model with frailty highlight key factors influencing the likelihood of harsh accelerations. Table 8.4 shows the results and the metrics of the model.

Table 8.4: Cox -PH model with frailty results for harsh accelerations per 100km

Random Effects					
Group	**Variable**	**SD**	**Variance**		
Identifier	Intercept	1.189	1.415		
Metrics	**chisq**	**df**	**p**	**AIC**	**BIC**
Integrated loglik	489	9.00	0.00	471	431.4
Penalized Loglik	584	26.66	0.00	530.7	413.2
Fixed Effects					
Variable	**Coef**	**Exp(Coef)**	**SE(Coef)**	**z**	**p**
Participant’s age					
Age [18-34]	Ref.				
Age [35-54]	-1.851	0.156	0.652	-2.84	**0.004**
Age [55+]	-0.930	0.394	1.020	-0.91	0.362
Participant’s gender					
Female	Ref.				
Male	-0.653	0.520	0.520	-1.25	0.209
Self-reported aggressiveness					
Low	Ref.				
High	1.176	3.243	0.651	1.81	**0.070**
Participant’s vehicle cc					
<1400cc	Ref.				
>1400cc	0.500	1.649	0.677	0.74	0.459
Peak hour					
Off peak	Ref.				
Morning peak	-0.244	0.783	0.110	-2.21	**0.026**
Afternoon peak	-0.368	0.691	0.108	-3.39	**<0.001**
Trip duration	0.011	1.011	0.002	4.75	**<0.001**
Concordance Index (C-index):	0.675				
AIC	7588.86				
BIC	7740.94				

Results show that random effects indicate significant unobserved heterogeneity across drivers, as indicated by an intercept standard deviation of 1.189 and variance of 1.415. Among fixed effects, significant predictors include age group [35-54] ($p = 0.0046$), which is associated with a reduction in the hazard of relapse (Exp(Coef) = 0.156), and peak hour99 ($p = 0.0007$), which shows a protective effect against relapse (Exp(Coef) = 0.692). Duration is strongly associated with an increased hazard ($p < 0.0001$), suggesting that longer trips slightly increase the risk of relapse (Exp(Coef) = 1.012). Non-significant predictors include gender and vehicle engine capacity, while aggressive driving tendencies approach significance ($p = 0.070$).

As explained above, one of the main assumptions of the Cox Proportional Hazards model is the proportionality of the covariates. For each of the variables included in the model, a function that correlates the corresponding set of scaled Schoenfeld residuals with time was implemented to test independence between residuals and time. The same test applies to the model as a whole. The results indicate that the proportionality assumption is violated, as p<0.05.

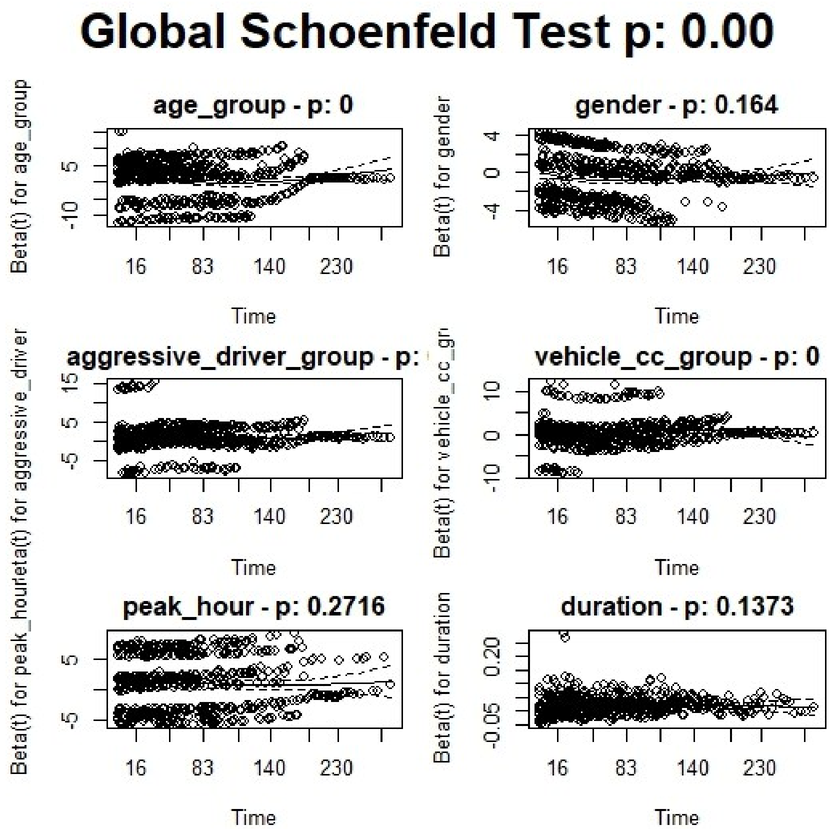


Figure 8.2 Schoenfeld residuals versus time for each of the variables of the Cox model

8.3.1.3 Weibull AFT Model with Clustered Heterogeneity

The Weibull Accelerated Failure Time (AFT) model with clustered heterogeneity provides a powerful alternative for analyzing the time to relapse in driving behavior indicators following the feedback phase. Unlike the Cox-PH model, which focuses on hazard rates, the Weibull AFT model directly models survival time, offering a distinct perspective on how various covariates accelerate or decelerate relapse. This approach is particularly valuable in understanding the temporal dynamics of driver behavior, as it allows for the direct estimation of time-related effects of predictors, such as trip duration or vehicle characteristics. For instance, the model's coefficients can provide actionable insights into how specific factors extend or shorten the time until a relapse occurs, making it a useful tool for tailoring interventions.

Another key strength of the Weibull AFT model is its ability to incorporate random effects to account for unobserved heterogeneity among drivers. By clustering drivers and considering individual-specific differences, the model effectively captures variability introduced by latent factors, such as driving habits or psychological tendencies, that are not explicitly measured. This feature ensures that the analysis reflects the diversity of driver behavior, thereby increasing the reliability of the results. Moreover, the interpretability of the model's coefficients facilitates a straightforward understanding of how covariates like age, driving conditions, or feedback timing influence survival time, making the Weibull AFT model a robust choice for analyzing complex, clustered time-to-event data in driver behavior studies.

The results presented in Table 8.5 and the accompanying plots provide critical insights into the relapse of harsh acceleration behaviors among drivers after feedback withdraw in the experiment.

Table 8.5: Weibull AFT model with clustered heterogeneity results for harsh accelerations per 100km

Variable	Value	Std. Err	(Naive SE)	z	p
(Intercept)	5.011	0.362	0.091	13.82	**<0.001**
Participant's age					
Age [18-34]	Ref.				
Age [35-54]	0.245	0.21892	0.078	1.12	**0.042**
Age [55+]	0.387	0.2108	0.169	1.84	0.085
Participant's gender					
Female	Ref.				
Male	0.382	0.27019	0.073	1.41	0.157
Self-reported aggressiveness					
Low	Ref.				
High	-0.643	0.25925	0.115	-2.48	**0.013**
Participant's vehicle cc					
<1400cc	Ref.				
>1400cc	-0.008	0.28301	0.110	-0.03	**0.046**
Peak hour					
Off peak	Ref.				
Morning peak	0.189	0.10928	0.084	1.73	**0.083**
Afternoon peak	0.411	0.13258	0.080	3.11	**0.001**
Trip duration	-0.015	0.00332	0.001	-4.58	**<0.001**
Log(scale)	-0.253	0.083	0.032	-3.05	**0.002**
Scale	0.776				
Loglik(model)	-3842.7				
Loglik(intercept only)	-3944.9				
Chisq	204.41				**<0.001**
Number of Newton-Raphson Iterations	8				
Concordance Index	0.677				
AIC	7705.44				
BIC	7762.49				

The Weibull AFT model identifies significant predictors of time to relapse, accounting for unobserved heterogeneity through clustering. Aggressive drivers in the high self-reported category had a shorter time to relapse ($p = 0.013$), underscoring the importance of targeted interventions for this group. Afternoon peak hours were associated with a significantly longer time to relapse ($p = 0.001$), while longer trip durations reduced the time to relapse ($p < 0.001$), highlighting the influence of contextual factors such as driving conditions and trip characteristics on behavior. The model's concordance index (0.677) reflects moderate predictive accuracy, indicating that it provides reasonably reliable predictions.

The plots provide a visual representation of the frailty effects and survival probabilities across strata. The frailty plot shows variability in random intercepts across drivers, confirming the importance of considering unobserved individual differences in the analysis. Meanwhile, the survival probability plot reveals that afternoon peak hours (green) are associated with higher survival probabilities compared to morning and off-peak hours, further substantiating the impact of contextual driving conditions. Together, the table and plots underscore the role of both individual and situational factors in shaping relapse behaviors, highlighting actionable areas for improving feedback strategies and ensuring sustained behavioral change. More specifically, The left plot illustrates survival probabilities stratified by peak hour categories (off-peak, morning peak, afternoon peak) over successive trips, highlighting differences in relapse risk. The right plot depicts the random intercepts (frailty effects) by driver identifier, showcasing individual-level variability in relapse behavior.

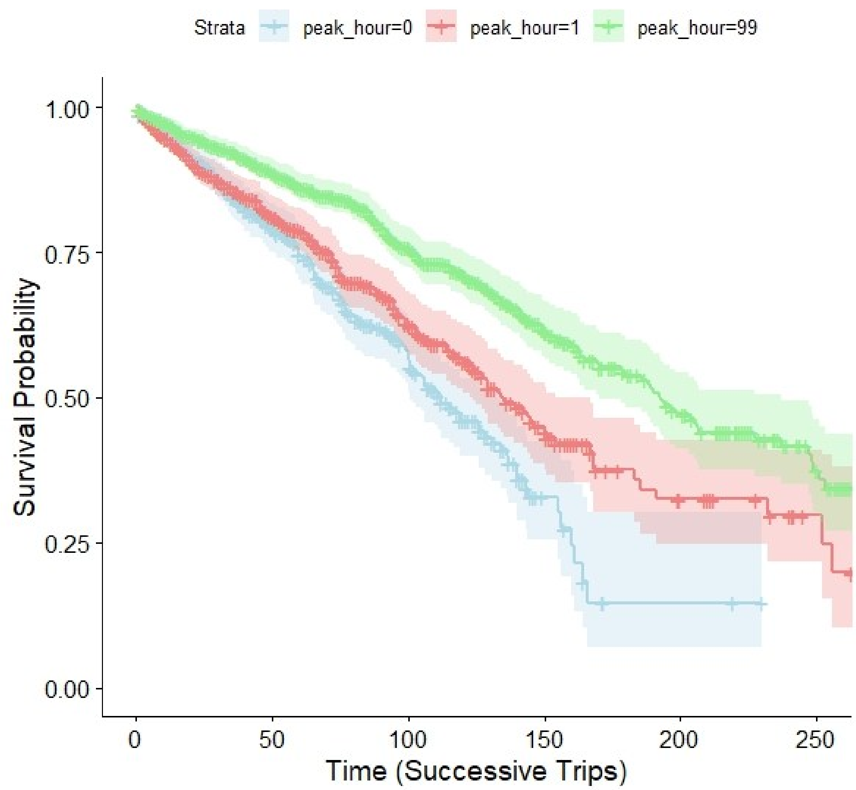


Figure 8.3: Survival probability for harsh accelerations relapse

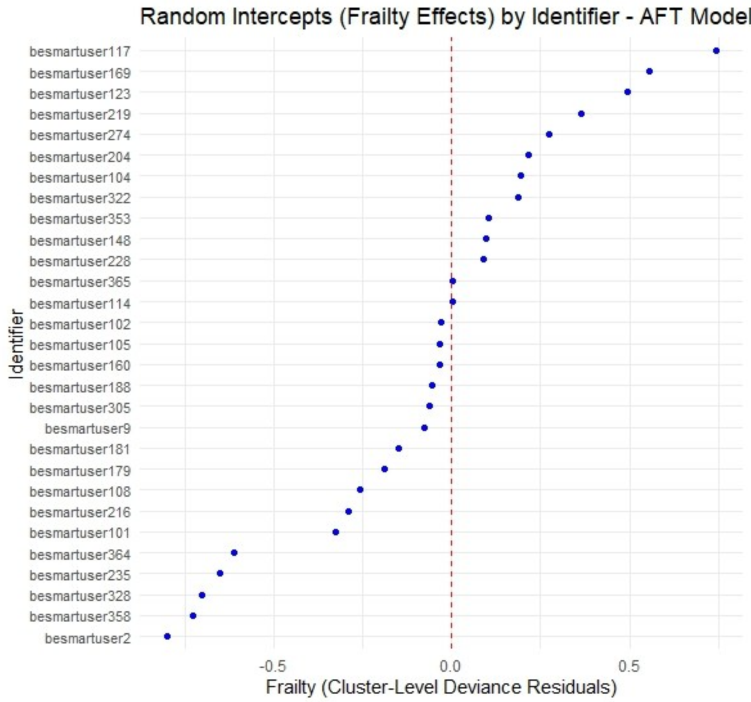


Figure 8.4:Random effects for harsh accelerations relapse

The diagnostic plots for the Weibull AFT model with clustered heterogeneity indicate an overall good model fit for predicting relapse in harsh accelerations per 100km. The histogram and Q-Q plot of deviance residuals show that the residuals are approximately symmetric and centered around zero, with minor deviations at the tails, suggesting the model captures most of the data well but may have difficulty with extreme values. The Cox-Snell residuals display a random scatter, supporting the assumption of independence and appropriate model specification. The observed vs. predicted times plot reveals a reasonable alignment with the diagonal line, indicating that the model predictions are broadly accurate, though variability increases for longer observed times.

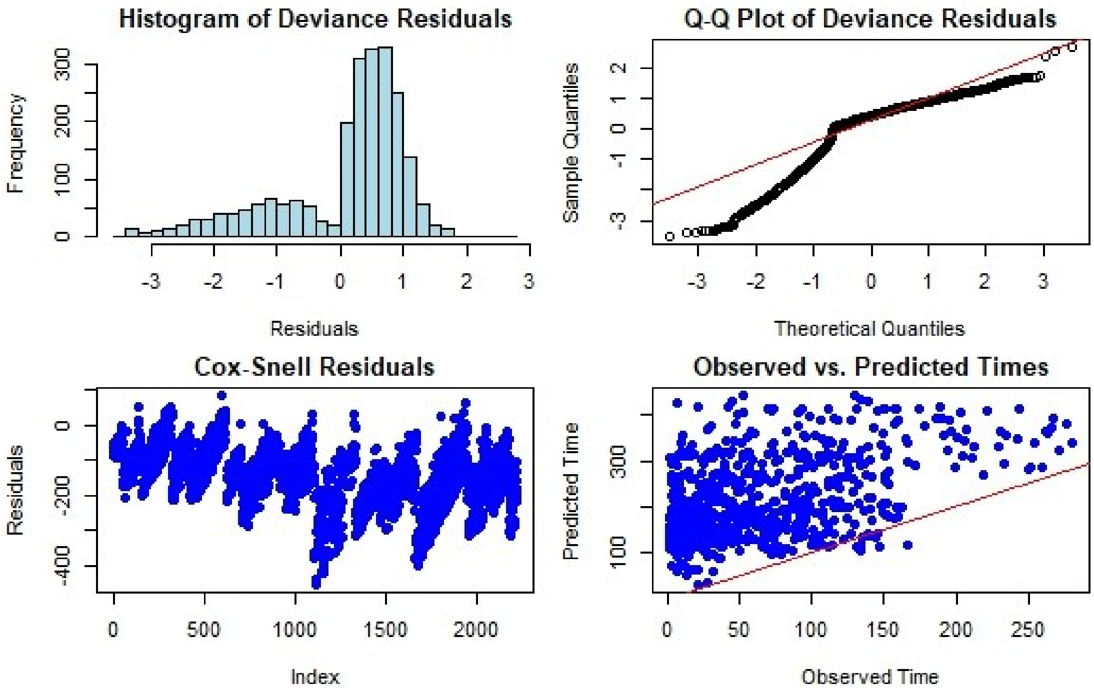


Figure 8.5: Model diagnostics for the Weibull AFT model on harsh accelerations

8.3.1.4 Random Survival Forest

The results of the Random Survival Forest (RSF) model for harsh accelerations emphasize the dominance of trip duration as the most significant predictor of relapse, as illustrated in the variable importance plot. The importance score for trip duration significantly exceeds that of other predictors, emphasizing its critical role in influencing survival outcomes. This finding aligns with results from other survival models, which consistently show that longer trips are associated with a quicker relapse into previous driving behaviors, effectively reducing the likelihood of maintaining improved driving behavior over time. Gender and vehicle engine capacity (vehicle_cc_group) also show moderate importance, suggesting that demographic factors and vehicle characteristics play roles in relapse prediction. Age group, aggressive driving behavior, and peak hour display relatively lower importance, indicating their less direct impact on survival times in the context of harsh accelerations.

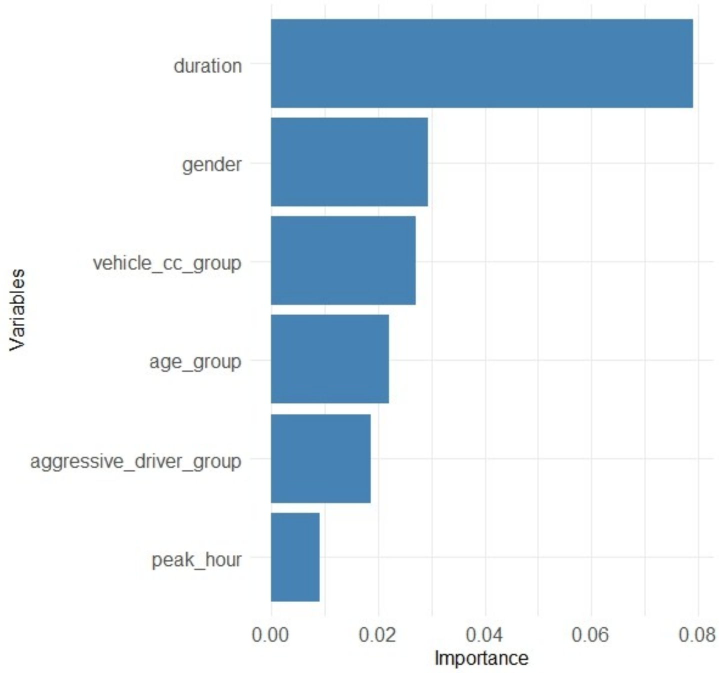


Figure 8.6: Feature importance of Random Forest for harsh accelerations relapse

The predictive performance of the RSF model was evaluated using Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), with values of 91.36 and 69.63, respectively. These errors highlight the model's ability to estimate survival times with moderate accuracy. Additionally, the out-of-bag prediction error (1-C index) of 0.33 indicates reasonable discrimination performance, though slightly lower compared to the Weibull AFT parametric survival model. Overall, the RSF model complements traditional survival analysis methods by capturing complex, non-linear interactions among predictors, which is evident in the nuanced relationships identified between predictors and relapse behavior.

Table 8.6: RSF model results for harsh accelerations

Type	**Survival**
Number of trees	30
Sample size	2220
Number of independent variables	6
Mtry	2
Target node size	5
Variable importance mode	permutation
Splitrule	logrank
Number of unique death times	196
OOB prediction error (1-C)	0.32969

8.3.1.5 Model comparison for harsh accelerations

Table 8.6 provides a comparative analysis of survival models applied to driving behavior relapse, focusing on harsh accelerations per 100km. Among these, the Weibull Accelerated Failure Time (AFT) model demonstrated its strengths, offering a balance of interpretability and predictive accuracy with a C-index of 0.677. Its ability to model survival time directly and account for clustered heterogeneity through a parametric framework makes it particularly effective when interpretability and explicit handling of frailty effects are critical. Key predictors such as trip duration and aggressive driver group were identified, and the model exhibited lower prediction error compared to the Cox model, although it was slightly outperformed by the Random Survival Forest (RSF) in terms of RMSE and MAE.

Table 8.7 Survival models comparison for harsh accelerations per 100km

Aspect	Weibull AFT Model	Cox Model with Frailty	Random Survival Forest (RSF)
Purpose	Models survival time directly	Models hazard rate	Captures non-linear effects
C-index	0.677	0.675	0.670 (OOB)
AIC	7705.44	7588.86	N/A
BIC	7762.49	7740.94	N/A
Key Predictors	Age, aggressive driver group, duration	Aggressive driver group, duration	Duration, gender, vehicle_cc_group
Frailty Effects	Accounted (Clustered Heterogeneity)	Accounted (Shared Frailty)	Implicitly handled (Non-parametric)
Prediction Error (RMSE/MAE)	RMSE: 92.81, MAE: 71.30	RMSE: 173.08, MAE: 152.21	RMSE: 91.36, MAE: 69.63
Strengths	Interpretable, adjusts for clustering	Handles heterogeneity flexibly	Captures complex interactions
Weaknesses	Assumes Weibull distribution	Assumes proportional hazards	Less interpretable

The Cox model with frailty achieved a comparable C-index of 0.675 and handled heterogeneity flexibly through shared frailty terms. However, its reliance on the proportional hazards assumption, which was violated in this dataset, along with higher prediction errors (RMSE: 173.08, MAE: 152.21), limits its applicability in contexts where hazard ratios vary over time or predictive accuracy is essential.

The RSF model, while less interpretable, excelled in capturing complex non-linear relationships and interactions among predictors, making it particularly useful when predictive accuracy and modeling flexibility are paramount. It achieved the best predictive performance, with an RMSE of 91.36 and an MAE of 69.63. However, its lack of transparency and inability to provide clear hazard ratios or survival time estimates present challenges for understanding the underlying dynamics of relapse.

Overall, the choice of the model depends on the priorities of the analysis. If interpretability and the explicit handling of frailty are critical, the Weibull AFT model stands out as the preferred option. However, when predictive accuracy and the ability to model complex relationships are paramount, the Random Survival Forest emerges as a strong candidate, despite its interpretability challenges.

8.3.2 Survival analysis of relapse in harsh braking

8.3.2.1 The Kaplan-Meier curves

The Kaplan-Meier survival curve for harsh braking events per 100 km shows the rate at which drivers revert to higher levels of harsh braking after the feedback phase.

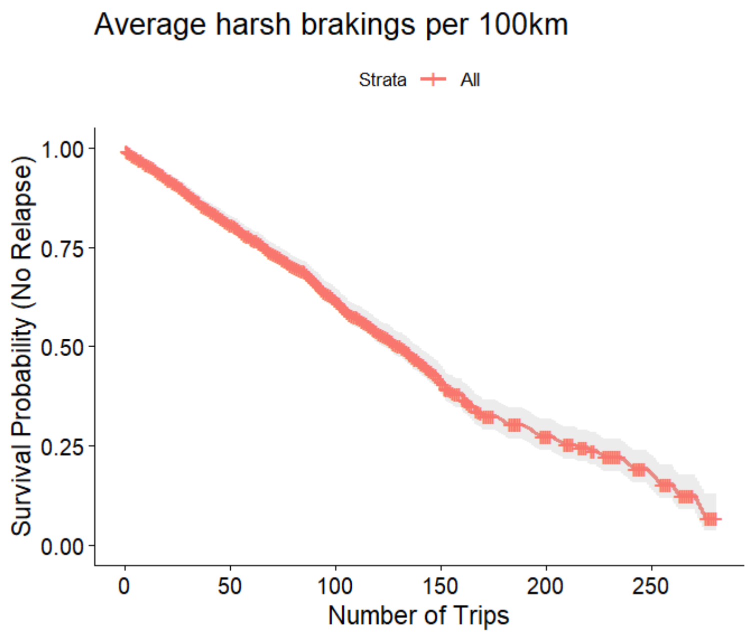


Figure 8.7: Kaplan-Meier survival curve for harsh braking per 100 km

Initial Period: The survival probability remains relatively high in the early trips of the post-feedback phase, indicating that most drivers maintain reduced levels of harsh braking initially. Declining Survival Probability: Over time, there is a steady decline in the survival probability, indicating an increase in harsh braking behaviors. Specific points of note include:

- 50 trips: Approximately 81.5% of drivers maintain their improved behavior, with a notable 18.5% relapsing.
- 100 trips: The survival probability drops to around 61.4%, suggesting that close to 40% of drivers have reverted to higher levels of harsh braking.
- 150 trips: The survival probability falls further to 40.3%, indicating that the majority of drivers have relapsed by this stage.

These results suggest that while feedback leads to temporary improvements in harsh braking behavior, the effect diminishes considerably over time. Continued feedback may be necessary to sustain these improvements long-term.

8.3.2.2 Cox-PH Model with Frailty

The results of the Cox Proportional Hazards model with frailty highlight key factors influencing the likelihood of relapse in harsh brakings.

Table ***8.8*** shows the results and the metrics of the model.

Table 8.8: Cox-PH model with frailty results for harsh braking per 100km

Random Effects					
Group	**Variable**	**SD**	**Variance**		
Identifier	Intercept	0.941	0.885		
Metrics	**chisq**	**Df**	**p**	**AIC**	**BIC**
Integrated loglik	446.9	9.00	0.00	428.9	387.4
Penalized Loglik	537	26.15	0.00	484.8	363
Fixed Effects					
Variable	**Coef**	**Exp(Coef)**	**SE(Coef)**	**z**	**p**
Participant's age					
Age [18-34]	Ref.				
Age [35-54]	-1.659	0.190	0.506	-3.28	**0.002**
Age [55+]	-1.517	0.219	0.812	-1.87	**0.061**
Participant's gender					
Female	Ref.				
Male	-0.260	0.770	0.409	-0.64	0.524
Self-reported aggressiveness					
Low	Ref.				
High	0.962	2.617	0.518	1.85	**0.063**
Participant's vehicle cc					
<1400cc	Ref.				
>1400cc	1.064	2.898	0.516	2.06	**0.039**
Peak hour					
Off peak	Ref.				
Morning peak	-0.311	0.732	0.103	-3.00	**0.002**
Afternoon peak	-0.185	0.831	0.096	-1.92	**0.054**
Trip duration	0.005	1.005	0.002	2.24	**0.024**
Concordance Index (C-index):	0.653				
AIC	9796.77				
BIC	9945.937				

Results reveal that random effects show significant variability between participants, as indicated by the intercept standard deviation (SD = 0.941) and variance (0.885), highlighting the importance of accounting for unobserved heterogeneity. Among the fixed effects, age group [35-54] (Exp(Coef) = 0.190, p = 0.002) significantly reduces the hazard of relapse, while participants with vehicles >1400cc (Exp(Coef) = 2.898, p = 0.039) show an increased risk of relapsing. Morning peak hours (Exp(Coef) = 0.732, p = 0.002) also reduce relapse risk compared to off-peak hours. Trip duration slightly increases the hazard of relapse (Exp(Coef) = 1.005, p = 0.024). However, other factors such as gender and self-reported aggressiveness approach significance but do not achieve statistical significance. The model's concordance index (C-index = 0.275) suggests limited predictive accuracy, while the AIC (9796.77) and BIC (9945.94) indicate adequate model fit. These results emphasize the influence of vehicle characteristics, trip context, and demographic factors in relapse behavior but suggest room for improvement in predictive power.

Simirarly to the harsh accelerations model, the results of the Schoenfeld residuals test indicate that the proportionality assumption is violated.

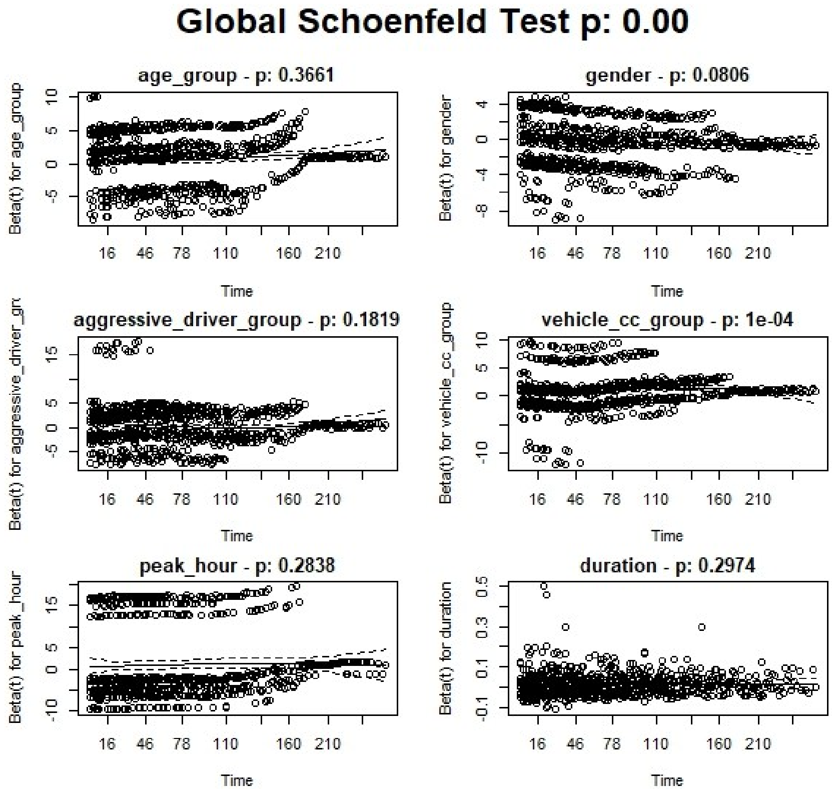


Figure 8.8 Schoenfeld residuals versus time for each of the variables of the Cox model

8.3.2.3 Weibull AFT Model with Clustered Heterogeneity

The Weibull AFT model results for harsh braking per 100km, as presented in Table 8.4, provide key insights into the factors influencing relapse time. Significant predictors include age, where both the [35-54] and [55+] groups exhibit longer survival times compared to the [18-34] reference group, with p-values of 0.010 and 0.005, respectively. Vehicle engine capacity (>1400cc) also plays a role, as drivers with larger vehicles have shorter survival times (p = 0.012). Additionally, trip duration negatively impacts relapse likelihood (p = 0.027), indicating that longer trips reduce the chances of relapse. While morning peak hours approach significance (p = 0.051), other variables, such as gender and aggressiveness, were not significant. The model achieves a concordance index (C-index) of 0.653, suggesting moderate predictive accuracy.

Table 8.9: Weibull AFT model with clustered heterogeneity results for harsh braking per 100km

Variable	**Value**	**Std. Err**	**(Naive SE)**	**z**	**p**
(Intercept)	4.792	0.214	0.077	22.350	**<0.001**
Participant's age					
Age [18-34]	Ref.				
Age [35-54]	0.360	0.139	0.065	2.590	**0.010**
Age [55+]	0.624	0.220	0.159	2.830	**0.005**
Participant's gender					
Female	Ref.				
Male	0.352	0.185	0.064	1.900	**0.057**
Self-reported aggressiveness					
Low	Ref.				
High	-0.271	0.217	0.099	-1.250	0.213
Participant's vehicle cc					
<1400cc	Ref.				
>1400cc	-0.508	0.202	0.085	-2.520	**0.012**
Peak hour					
Off peak	Ref.				
Morning peak	0.191	0.098	0.074	1.960	**0.051**
Afternoon peak	0.151	0.085	0.066	1.780	**0.075**
Trip duration	-0.008	0.003	0.002	-2.220	**0.027**
Log(scale)	-0.325	0.056	0.029	-5.750	**<0.001**
Scale	0.723				
Loglik(model)	-4740.7				
Loglik(intercept only)	-4849.1				
Chisq	216.72				**<0.001**
Number of Newton-Raphson Iterations	8				
Concordance Index	0.724				
AIC	9501.39				
BIC	9558.44				

The survival plot stratified by aggressiveness group highlights clear differences in survival probabilities. Drivers in the "High Aggressiveness" group show a steeper decline in survival probability compared to the "Low Aggressiveness" group, confirming the significant impact of behavioral traits on relapse likelihood. This visualization complements the model results, emphasizing the role of personality factors in determining survival times. Furthermore, the random intercepts plot reveals individual variability in frailty effects, showcasing the importance of accounting for unobserved heterogeneity across drivers. Drivers with positive frailty effects exhibit shorter survival times, while those with negative frailty effects survive longer.

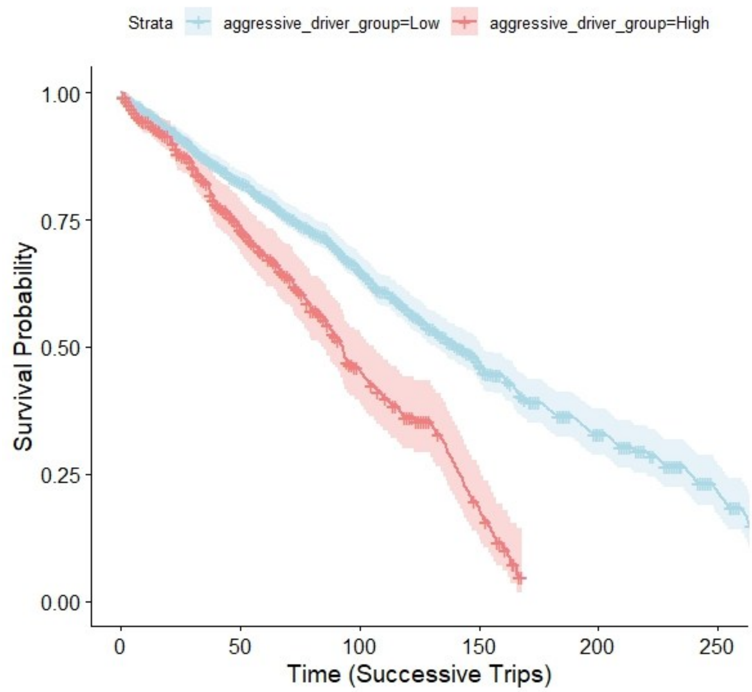


Figure 8.9: Survival probability for harsh braking relapse

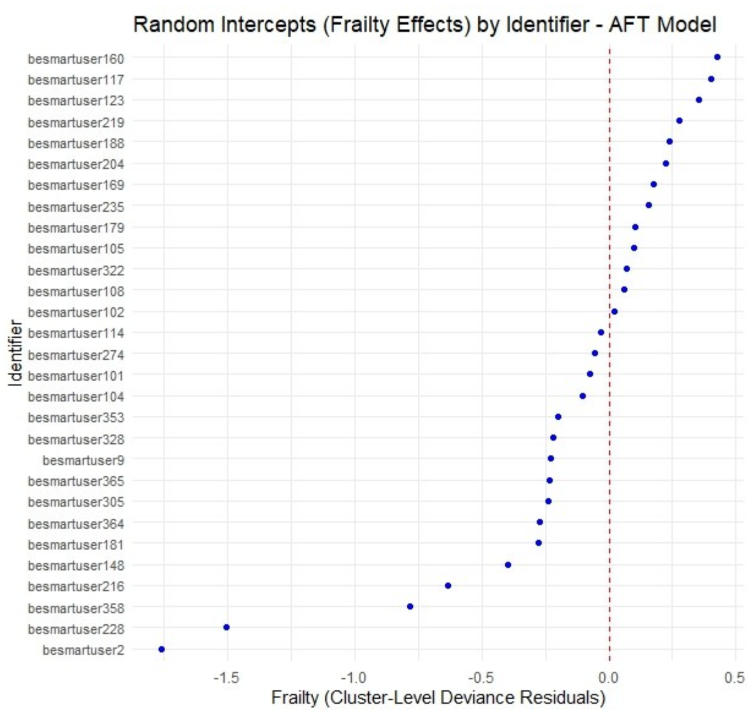


Figure 8.10: Random effects for harsh braking relapse

Lastly, the residual diagnostics provide additional support for the model's adequacy. The histogram and Q-Q plots of deviance residuals show that residuals are roughly symmetric and align well with theoretical expectations, although slight deviations exist at the tails. The Cox-Snell residuals plot confirms the independence of residuals, as no clear patterns emerge, while the observed vs. predicted times scatterplot indicates reasonable alignment around the diagonal, despite variability at higher observed times. Together, the table and plots confirm the robustness of the Weibull AFT model while highlighting areas for refinement, particularly in predicting extreme cases or tailoring interventions based on individual frailty.

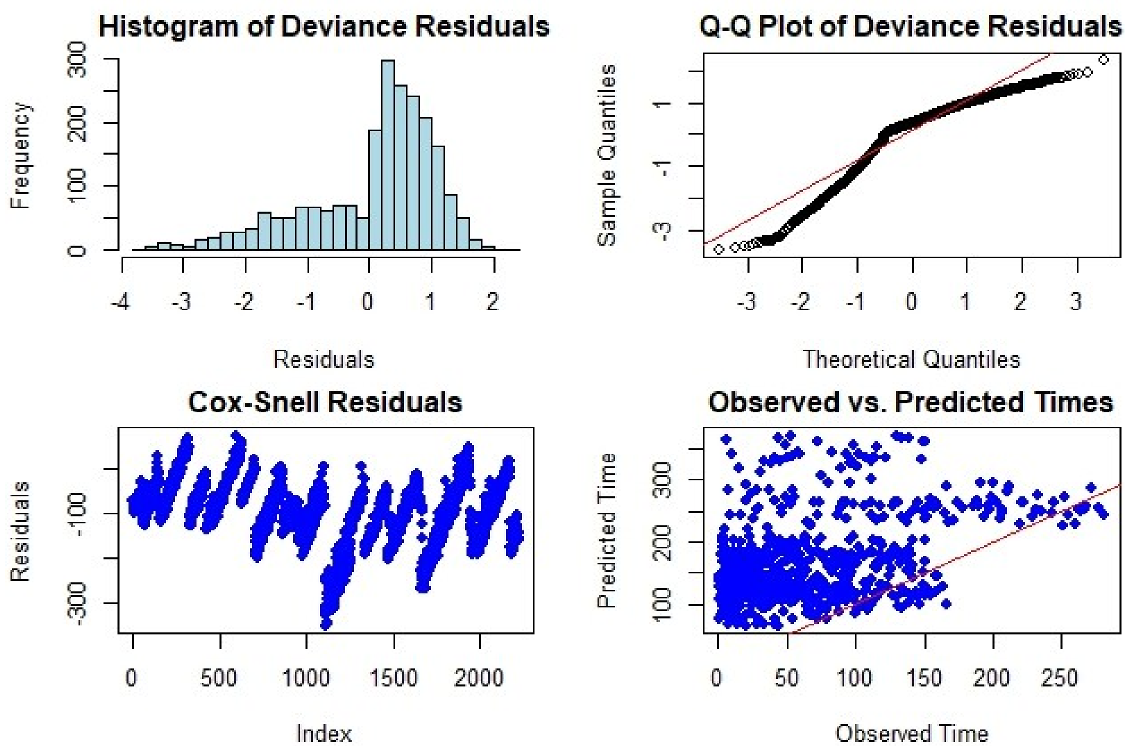


Figure 8.11: Model diagnostics for the Weibull AFT model on harsh braking

Lastly, the residual diagnostics provide additional support for the model's adequacy. The histogram and Q-Q plots of deviance residuals show that residuals are roughly symmetric and align well with theoretical expectations, although slight deviations exist at the tails. The Cox-Snell residuals plot confirms the independence of residuals, as no clear patterns emerge, while the observed vs. predicted times scatterplot indicates reasonable alignment around the diagonal, despite variability at higher observed times. Together, the table and plots confirm the robustness of the Weibull AFT model while highlighting areas for refinement, particularly in predicting extreme cases or tailoring interventions based on individual frailty.

8.3.2.4 Random Survival Forest

The results of the Random Survival Forest (RSF) model for harsh braking reveal key insights into the predictors influencing the relapse event. The variable importance plot indicates that vehicle_cc_group is the most critical predictor, followed by age_group and gender, while factors like duration and aggressive_driver_group have a moderate influence. Interestingly, peak_hour, while included, appears to have the least importance, suggesting it has a minor role in predicting harsh braking relapse. These findings align with the broader understanding that vehicle characteristics and driver demographics are more central to relapse behavior than contextual factors like time of travel.

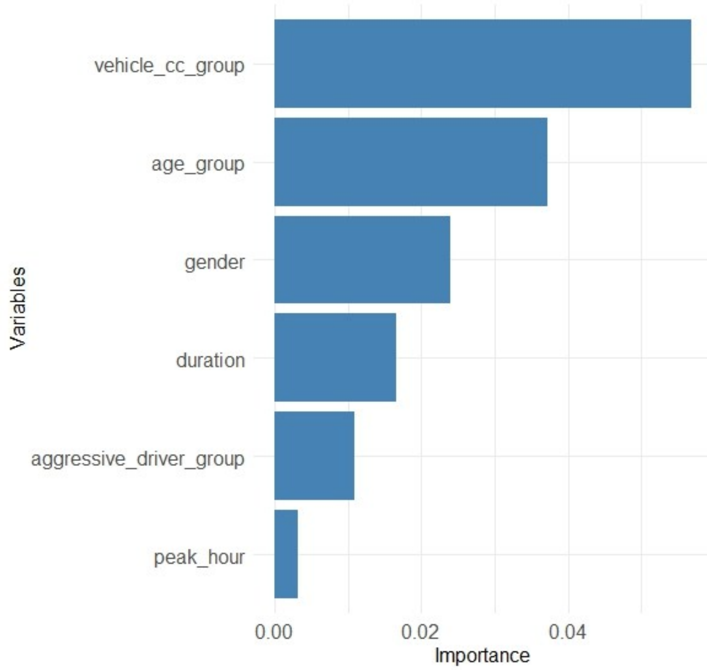


Figure 8.12: Feature importance of Random Forest for harsh braking

The RSF model, trained with 30 trees, demonstrated an out-of-bag (OOB) prediction error of 0.357, highlighting its moderate predictive performance. The model's flexibility in capturing non-linear relationships and interactions adds value, but its interpretability is limited compared to parametric models. The error metrics, with an RMSE of 91.92, and an MAE of 70.67, indicate moderate alignment between predicted and observed survival times. These results suggest the RSF model's utility for identifying critical predictors, which can inform targeted interventions, while underscoring the importance of integrating other models for more nuanced interpretations of hazard dynamics.

Table 8.10: RSF model results for harsh braking

Type	**Survival**
Number of trees	30
Sample size	2220
Number of independent variables	6
Mtry	2
Target node size	5
Variable importance mode	permutation
Splitrule	logrank
Number of unique death times	220
OOB prediction error (1-C)	0.357947

8.3.2.5 Model comparison for harsh braking

Table 8.11 Survival models comparison for harsh braking per 100km

Aspect	Weibull AFT Model	Cox Model with Frailty	Random Survival Forest (RSF)
Purpose	Models survival time directly	Models hazard rate	Captures non-linear effects
C-index	0.724	0.653	0.636 (OOB)
AIC	9501.4	9796.8	N/A
BIC	9558.4	9945.9	N/A
Key Predictors	Age group, vehicle CC group, trip duration	Vehicle CC group, peak hour, trip duration	Vehicle CC group, age group, gender, trip duration
Frailty Effects	Accounted (Clustered Heterogeneity)	Accounted (Shared Frailty)	Implicitly handled (Non-parametric)
Prediction Error (RMSE/MAE)	RMSE: 91.73, MAE: 70.25	RMSE: 121.11, MAE: 102.42	RMSE: 91.92, MAE: 70.67
Strengths	Interpretable, adjusts for clustering	Handles heterogeneity flexibly	Captures complex interactions
	Accounts for driver-specific effects	Provides interpretable hazard ratios	Robust to outliers, identifies non-linear effects
Weaknesses	Assumes Weibull distribution	Assumes proportional hazards	Less interpretable
	Sensitive to outliers	Lower predictive accuracy	Requires larger datasets

The comparison of survival models for harsh braking per 100km highlights the strengths and limitations of the Weibull AFT model, the Cox model with frailty, and the Random Survival Forest (RSF). The Weibull AFT model emerged as the best performer, with a C-index of 0.724 and competitive prediction error metrics, making it both interpretable and effective at accounting for frailty effects. It is particularly well-suited for datasets requiring explicit adjustment for clustered heterogeneity, and it identified key predictors such as age group, vehicle CC group, and trip duration. This combination of strong discriminative ability, predictive accuracy, and interpretability positions the AFT model as the preferred choice for understanding the factors influencing harsh braking behavior.

The Cox model with frailty, while flexible and interpretable due to its shared frailty terms, demonstrated a lower C-index of 0.653 and higher prediction errors (RMSE: 121.11, MAE: 102.42), limiting its utility in this analysis. Its reliance on the proportional hazards assumption, which may not hold in this context, further reduces its applicability. On the other hand, the RSF model excelled in capturing complex non-linear interactions and demonstrated comparable prediction error metrics to the Weibull AFT model. However, its lower C-index (0.636) and lack of interpretability make it less suitable for understanding the dynamics of harsh braking but valuable for predictive tasks where modeling flexibility and accuracy are paramount. Together, these models reflect the trade-offs between interpretability, predictive performance, and methodological assumptions in survival analysis.

8.3.3 Survival analysis of relapse in speeding percentage

8.3.3.1 The Kaplan-Meier curves

The Kaplan-Meier curve for speeding percentage reflects how quickly drivers relapse to pre-feedback levels of speeding.

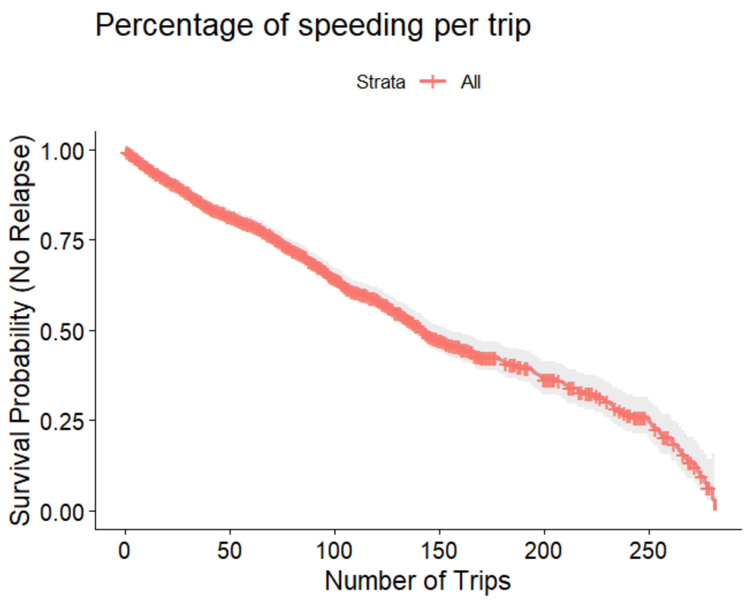


Figure 8.13 Kaplan-Meier survival curve for speeding percentage

Initial Period: At the beginning of the post-feedback phase, the survival probability is high, with most drivers adhering to improved speeding behavior in the initial trips.
Progressive Decline: The survival probability decreases steadily as the number of trips increases. Key milestones include:

- 50 trips: Approximately 82.3% of drivers maintain lower speeding levels.
- 100 trips: The survival probability reduces to 65.2%, suggesting a gradual relapse.
- 150 trips: Around 46.8% of drivers maintain improved behavior, showing a significant relapse among the remaining drivers.

These findings highlight that while drivers initially adhere to safer speeds, many revert to previous behaviors in the absence of ongoing feedback, emphasizing the potential benefit of sustained interventions to address speeding.

8.3.3.2 Cox-PH Model with Frailty

The results of the Cox Proportional Hazards model with frailty highlight key factors influencing the likelihood of speeding behavior Table 8.11 shows the results and the metrics of the model.

Table 8.12: Cox -PH model with frailty results for speeding percentage

Random Effects			
Group	**Variable**	**SD**	**Variance**
Identifier	Intercept	1.486	2.208

Metrics	**chisq**	**Df**	**p**	**AIC**	**BIC**
Integrated loglik	697.8	9.00	0.00	679.8	638.9
Penalized Loglik	810	27.91	0.00	754.1	627.1
Fixed Effects					
Variable	**Coef**	**Exp(Coef)**	**SE(Coef)**	**z**	**p**
Participant's age					
Age [18-34]	Ref.				
Age [35-54]	-1.066	0.344	0.750	-1.42	0.154
Age [55+]	-0.128	0.879	1.224	-0.11	**0.016**
Participant's gender					
Female	Ref.				
Male	-0.583	0.557	0.632	-0.92	0.355
Self-reported aggressiveness					
Low	Ref.				
High	0.632	1.881	0.819	0.77	**0.040**
Participant's vehicle cc					
<1400cc	Ref.				
>1400cc	0.457	1.580	0.826	0.55	0.579
Peak hour					
Off peak	Ref.				
Morning peak	0.191	1.210	0.110	1.73	**0.082**
Afternoon peak	-0.009	0.990	0.110	-0.09	0.927
Trip duration	0.022	1.023	0.002	11.32	**<0.001**
Concordance Index (C-index):	0.696				
AIC	8549.05				
BIC	8708.30				

The Cox Proportional Hazards model with frailty for speeding percentage provides valuable insights into the factors influencing relapse while highlighting the limitations of the model's predictive power. The random effects indicate substantial unobserved heterogeneity, with an intercept standard deviation (SD = 1.486) and variance (2.208), suggesting significant variability across individuals. Among the fixed effects, trip duration emerges as a strong and highly significant predictor of relapse (Exp(Coef) = 1.023, $p < 0.001$), indicating that longer trips are associated with a higher hazard of speeding relapse. Morning peak hours (Exp(Coef) = 1.210, $p = 0.082$) approach statistical significance, suggesting a potential increase in relapse risk during this time. Other variables, such as age, gender, aggressiveness, and vehicle engine size, show limited or no significant associations, though high self-reported aggressiveness (Exp(Coef) = 1.881, $p = 0.040$) is associated with increased relapse risk, and age group [55+] (Exp(Coef) = 0.879, $p = 0.016$) demonstrates a protective trend.

Similarly to the previous models, the results of the Schoenfeld residuals test indicate that the proportionality assumption is violated, with the global p-value being 0.00.

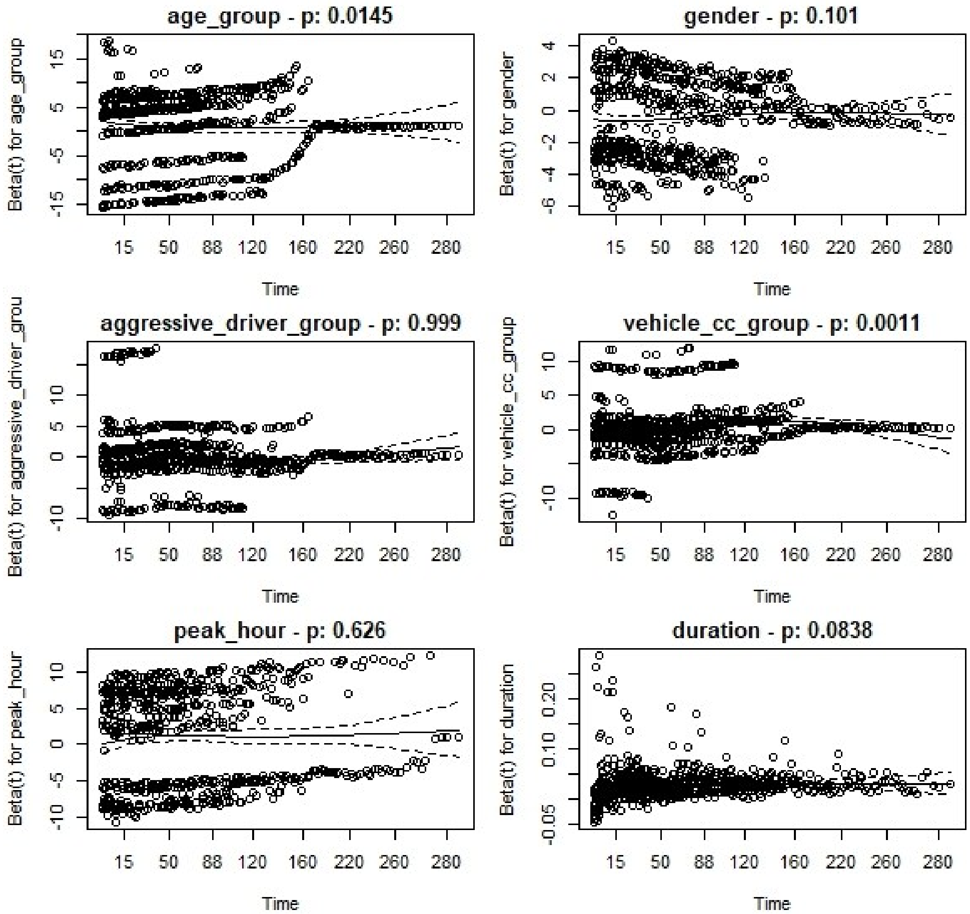


Figure 8.14: Schoenfeld residuals versus time for each of the variables of the Cox model

8.3.3.3 Weibull AFT Model with Clustered Heterogeneity

The Weibull Accelerated Failure Time (AFT) model with clustered heterogeneity provides a detailed examination of factors influencing speeding relapse. The model effectively accounts for survival time and clustered variability, with a concordance index (C-index = 0.700) indicating moderate predictive accuracy. Among the predictors, trip duration emerges as a significant factor, with a negative coefficient (Coef = -0.022, $p < 0.001$), suggesting that longer trips are associated with a shorter time to relapse. This aligns with the expected behavior of trip duration being a key driver of speeding relapse. Additionally, morning peak hours show a significant negative association (Coef = -0.096, p = 0.004), indicating a slightly reduced survival time for relapse compared to off-peak hours. Age group [55+] demonstrates a protective effect (Coef = 0.084, p = 0.022), although other age and gender categories, as well as vehicle engine size, do not show significant contributions.

Table 8.13: Weibull AFT model with clustered heterogeneity results for speeding percentage

Variable	Value	Std. Err	(Naive SE)	z	p
(Intercept)	5.294	0.453	0.095	11.69	**<0.001**
Participant's age					
Age [18-34]	Ref.				
Age [35-54]	0.165	0.215	0.073	0.77	0.441
Age [55+]	0.084	0.237	0.140	0.36	**0.022**
Participant's gender					
Female	Ref.				
Male	0.318	0.305	0.071	1.04	0.298
Self-reported aggressiveness					
Low	Ref.				
High	0.191	0.477	0.121	-0.40	**0.089**
Participant's vehicle cc					
<1400cc	Ref.				
>1400cc	-0.115	0.386	0.108	-0.30	0.766
Peak hour					
Off peak	Ref.				
Morning peak	-0.096	0.186	0.086	-0.52	**0.004**
Afternoon peak	0.146	0.130	0.083	1.13	0.259
Trip duration	-0.022	0.003	0.001	-7.38	**<0.001**
Log(scale)	-0.229	0.083	0.030	-2.76	0.006
Scale	0.796				
Loglik(model)	-4306.10				
Loglik(intercept only)	-4446.00				
Chisq	279.820				<0.001
Number of Newton-Raphson Iterations	8				
Concordance Index	0.700				
AIC	8632.255				
BIC	8689.308				

The survival probability plot stratified by peak hour categories reveals substantial differences in survival probabilities over successive trips. Drivers operating during afternoon peak hours show higher survival probabilities, suggesting that driving conditions during this period may contribute to safer behavior or delayed relapse. By contrast, morning peak hours indicate a slightly higher relapse risk, aligning with contextual factors such as traffic density or stress. Additionally, the frailty effects plot shows variability among individual drivers, with a range of frailty scores highlighting unobserved heterogeneity. This variability emphasizes the importance of accounting for driver-specific differences in the analysis.

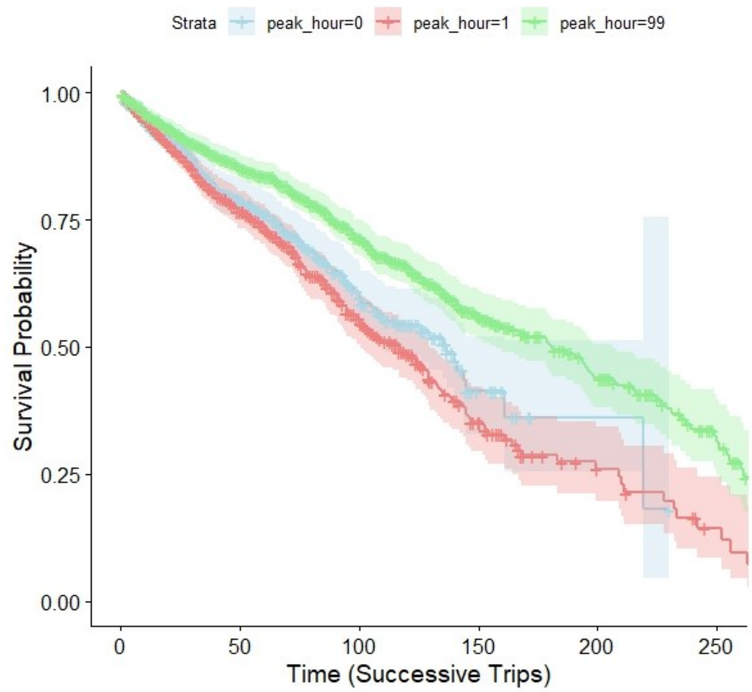


Figure 8.15: Survival probability for speeding relapse

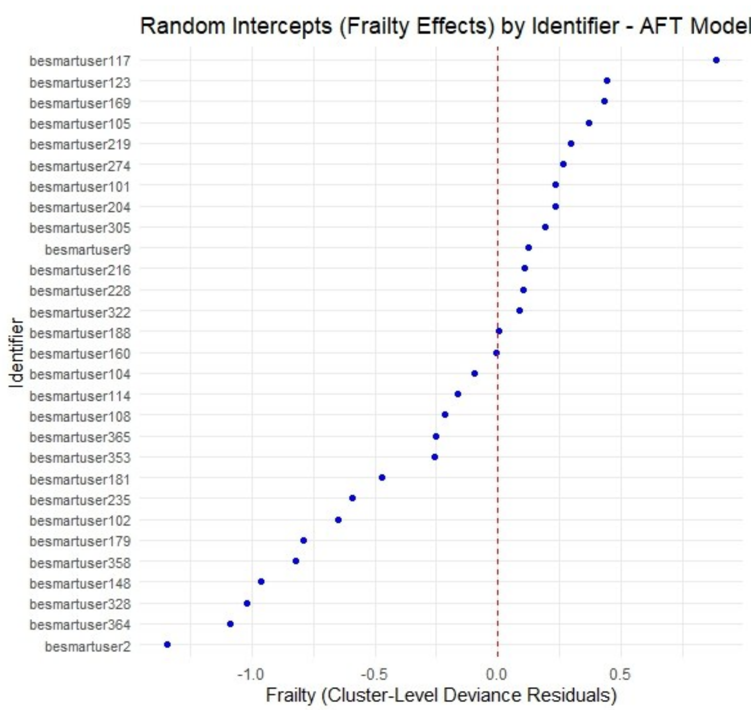


Figure 8.16: Random effects for speeding relapse

The diagnostic plots provide additional evidence of model adequacy. The histogram and Q-Q plots of deviance residuals show a roughly symmetric distribution with slight deviations, while the Cox-Snell residuals confirm the appropriateness of the model specification with no evident patterns. The observed versus predicted times scatterplot shows reasonable alignment with the diagonal, though some variability at higher observed times suggests areas for improvement. Together, these results underscore the robustness of the Weibull AFT model while emphasizing the role of trip-specific and individual-level factors in understanding relapse in speeding behavior.

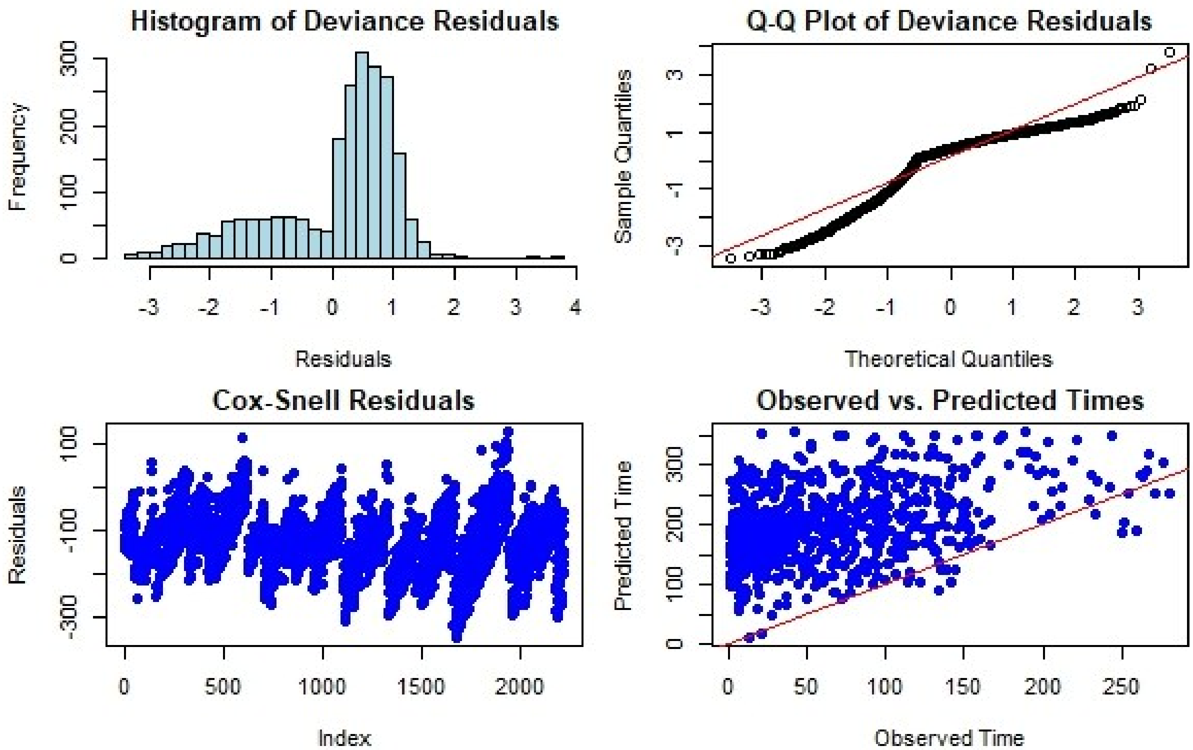


Figure 8.17: Model Diagnostics for the Weibull AFT Model on speeding

8.3.3.4 Random Survival Forest

The variable importance results from the Random Survival Forest (RSF) model for predicting harsh braking events underscore the significance of trip duration as the most critical predictor, showing the strongest association with relapse events. This indicates that the length of a trip directly influences the likelihood of a harsh braking occurrence. The age group and gender of drivers emerged as the next most influential factors, suggesting that demographic characteristics play a key role in driving behavior patterns. Peak hour and vehicle engine size group also showed moderate importance, indicating that external factors like traffic density and vehicle characteristics contribute to harsh braking events. Lastly, the aggressive driver group had a lower relative importance but still highlights that self-reported behavioral tendencies are relevant to predicting such events.

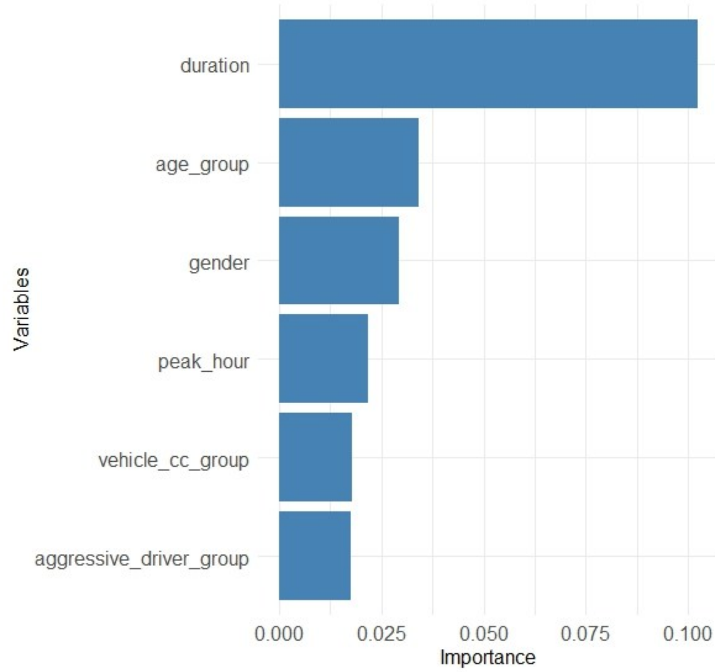


Figure 8.18: Feature importance of Random Forest for speeding

The error metrics provide further insight into the model's performance. The RMSE of 91.87 and MAE: of 70.17 indicate moderate prediction accuracy, reflecting some degree of variance between predicted and observed survival times. These results suggest that while the RSF captures key patterns in the data, the non-linear and interaction effects it models may introduce noise, especially in survival scenarios with a high number of unique death times (196 in this case).

Table 8.14 RSF model results for speeding behavior

Type	**Survival**
Number of trees	30
Sample size	2220
Number of independent variables	6
Mtry	2
Target node size	5
Variable importance mode	permutation
Splitrule	logrank
Number of unique death times	196
OOB prediction error (1-C)	0.32969

8.3.3.5 Models comparison for speeding percentage

The comparison of survival models for relapse in speeding behavior highlights the trade-offs between interpretability, predictive performance, and modeling flexibility. The Weibull AFT model offers a well-rounded approach, balancing interpretability and predictive accuracy with a concordance index (C-index) of 0.70 and reasonable prediction error metrics (RMSE: 92.47, MAE: 70.91). Its explicit handling of clustered heterogeneity through frailty terms makes it particularly effective for datasets with nested structures, such as repeated measures for drivers. Trip duration emerged as the key predictor, underscoring its critical role in understanding speeding relapse. However, its reliance on the Weibull distribution and sensitivity to outliers may limit its applicability in more complex datasets.

Table 8.15: Models comparison summary of relapse in speeding behavior

Aspect	**Weibull AFT Model**	**Cox Model with Frailty**	**Random Survival Forest (RSF)**
Purpose	Models survival time directly	Models hazard rate	Captures non-linear effects
C-index	0.70	0.696	0.704 (OOB)
AIC	8632.26	8549.06	N/A
BIC	8689.31	8708.31	N/A
Key Predictors	Trip duration	Gender, trip duration, vehicle CC group	Trip duration, age group, aggressive driving
Frailty Effects	Accounted (Clustered Heterogeneity)	Accounted (Shared Frailty)	Implicitly handled (Non-parametric)
Prediction Error (RMSE/MAE)	RMSE: 92.47, MAE: 70.91	RMSE: 146.59, MAE: 130.41	RMSE: 91.87, MAE: 70.17
Strengths	Interpretable, adjusts for clustering; Accounts for driver-specific effects	Handles heterogeneity flexibly; Provides interpretable hazard ratios	Captures complex interactions; Robust to outliers, identifies non-linear effects
Weaknesses	Assumes Weibull distribution; Sensitive to outliers	Assumes proportional hazards; Lower predictive accuracy	Less interpretable; Requires larger datasets

The Cox model with frailty demonstrated comparable discriminative ability (C-index = 0.696) but higher prediction errors (RMSE: 146.59, MAE: 130.41), suggesting reduced predictive accuracy. While it flexibly handles heterogeneity through shared frailty terms and provides interpretable hazard ratios, its reliance on the proportional hazards assumption can be restrictive. In contrast, the Random Survival Forest (RSF) excelled in predictive performance, achieving the lowest prediction errors (RMSE: 91.87, MAE: 70.17) and the highest C-index (0.704). Its ability to capture non-linear relationships and complex interactions makes it a strong choice for predictive tasks, though its lack of interpretability and reliance on larger datasets present challenges. Together, these models illustrate the balance between understanding speeding relapse dynamics and achieving high predictive accuracy, with the Weibull AFT model standing out for interpretability and RSF for flexibility and predictive performance.

8.3.4 Survival analysis of relapse in mobile phone use while driving

8.3.4.1 The Kaplan-Meier curves

The survival curve suggests that mobile phone use relapses at a slower rate compared to harsh braking and speeding. However, without continued feedback, drivers gradually return to higher levels of phone use while driving.

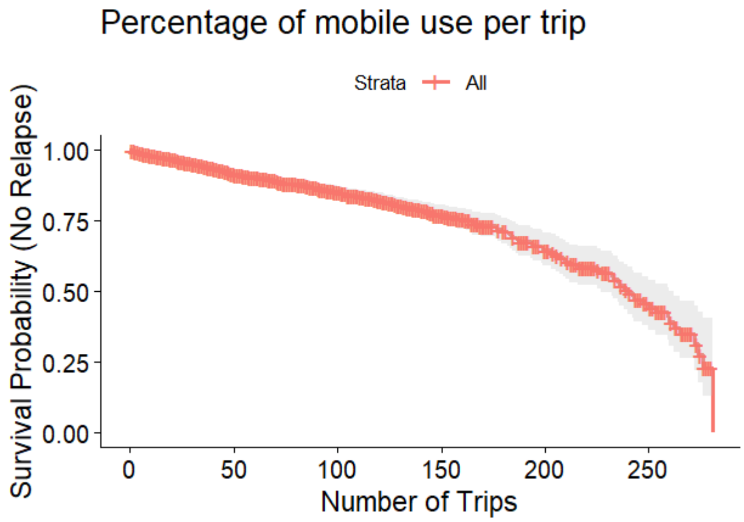


Figure 8.19 Kaplan-Meier survival curve for mobile phone use percentage

Early Maintenance: In the early stages, the survival probability remains high, indicating that drivers initially maintain reduced phone use during driving.

Gradual Decline: Over successive trips, the survival probability gradually decreases, showing a rise in mobile phone use among drivers. Key milestones include:

- 50 trips: About 91.7% of drivers still show restraint in phone use, indicating a slower relapse pattern compared to other indicators.
- 100 trips: The survival probability decreases to approximately 84.8%, showing a steady increase in mobile phone use.

8.3.4.2 Cox-PH Model with Frailty

The results of the Cox Proportional Hazards model with frailty highlight key factors influencing the likelihood of distracted behavior due to mobile phone use while driving. Table 8.16 shows the results and the metrics of the model.

Table 8.16: Cox -PH model with frailty results for mobile phone use

Random Effects					
Group	**Variable**	**SD**	**Variance**		
Identifier	Intercept	1.419	2.014		
Metrics	**chisq**	**Df**	**p**	**AIC**	**BIC**
Integrated loglik	376.2	9.00	0.00	358.2	325.2
Penalized Loglik	454.8	24.98	0.00	404.9	313.3
Fixed Effects					
Variable	**Coef**	**Exp(Coef)**	**SE(Coef)**	**z**	**p**
Participant’s age					
Age [18-34]	Ref.				
Age [35-54]	-2.463	0.085	0.853	-2.89	**0.003**
Age [55+]	-1.324	0.266	1.313	-1.01	0.313
Participant’s gender					
Female	Ref.				

Male	-0.258	0.772	0.643	-0.40	0.687
Self-reported aggressiveness					
Low	Ref.				
High	1.525	4.597	0.825	1.85	**0.064**
Participant's vehicle cc					
<1400cc	Ref.				
>1400cc	-0.910	0.402	0.953	-0.96	0.339
Peak hour					
Off peak	Ref.				
Morning peak	0.167	1.182	0.218	0.77	0.443
Afternoon peak	0.213	1.237	0.217	0.98	0.327
Trip duration	0.026	1.026	0.003	8.57	**<0.001**
Concordance Index (C-index):	0.737				
AIC	3371.42				
BIC	3513.93				

The Cox Proportional Hazards model with frailty for mobile phone use highlights key factors influencing relapse, with a strong predictive accuracy as indicated by a concordance index (C-index) of 0.737. Trip duration emerges as the most significant predictor (Coef = 0.026, Exp(Coef) = 1.026, $p < 0.001$), indicating that longer trips are associated with an increased hazard of mobile phone use relapse. Among demographic factors, age group [35-54] shows a highly significant protective effect (Coef = -2.463, Exp(Coef) = 0.085, p = 0.003) compared to the reference group [18-34], while other age and gender categories do not reach statistical significance. Self-reported aggressiveness trends toward significance at the 0.1 level (Coef = 1.525, Exp(Coef) = 4.597, p = 0.064), suggesting a potential association with increased relapse risk. The random effects indicate substantial variability across individuals, with an intercept standard deviation (SD = 1.419) and variance (2.014), underscoring unobserved heterogeneity. The model's fit is supported by an AIC of 3371.42 and BIC of 3513.93, reflecting a reasonable balance between complexity and predictive power. These results emphasize the critical role of trip duration and individual differences in influencing mobile phone use relapse.

Similar to the previous models, the results of the Schoenfeld residuals test indicate that the proportionality assumption is violated with the global p-value being 0.04.

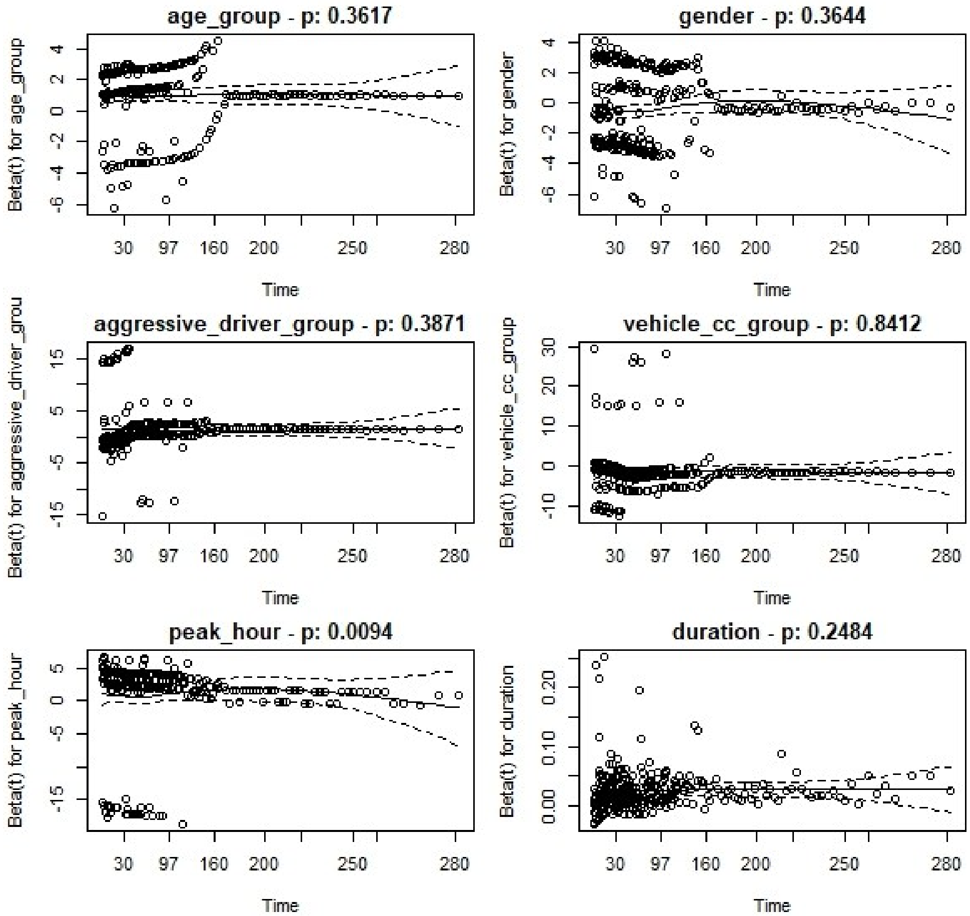


Figure 8.20: Schoenfeld residuals versus time for each of the variables of the Cox model

8.3.4.3 Weibull AFT Model with Clustered Heterogeneity

The Weibull AFT model with clustered heterogeneity for speeding percentage highlights significant predictors while achieving a moderate predictive accuracy, as indicated by a concordance index (C-index) of 0.700. Trip duration emerges as the most significant factor (Coef = -0.022, $p < 0.001$), showing that longer trips are strongly associated with shorter survival times before relapse. Age group [35-54] also shows a statistically significant effect at the 0.05 level (Coef = 0.165, $p = 0.041$), indicating a slight increase in survival time compared to the reference group [18-34], while age group [55+] does not reach significance. Other variables, such as gender, self-reported aggressiveness, and vehicle engine size, trend toward significance at the 0.1 level but do not achieve stronger associations.

The model's fit is supported by significant improvements in log-likelihood values (-4306.10 for the model vs. -4446.00 for the intercept-only model) and a chi-square statistic of 279.820 ($p < 0.001$), with AIC (8632.255) and BIC (8689.308) reflecting reasonable model fit. These findings underscore the importance of trip duration while suggesting limited contributions from other demographic and behavioral factors.

Table 8.17: Weibull AFT model with clustered heterogeneity results for mobile phone use

Variable	**Value**	**Std. Err**	**(Naive SE)**	**z**	**p**
(Intercept)	5.294	0.453	0.095	11.69	**<0.001**
Participant's age					
Age [18-34]	Ref.				
Age [35-54]	0.165	0.215	0.073	0.77	**0.041**
Age [55+]	0.084	0.237	0.140	0.36	0.722
Participant's gender					
Female	Ref.				
Male	0.318	0.305	0.071	1.04	0.298
Self-reported aggressiveness					
Low	Ref.				
High	0.191	0.477	0.121	-0.40	**0.089**
Participant's vehicle cc					
<1400cc	Ref.				
>1400cc	-0.115	0.386	0.108	-0.30	**0.066**
Peak hour					
Off peak	Ref.				
Morning peak	-0.096	0.186	0.086	-0.52	0.604
Afternoon peak	0.146	0.130	0.083	1.13	**0.059**
Trip duration	-0.022	0.003	0.001	-7.38	**<0.001**
Log(scale)	-0.229	0.083	0.030	-2.76	0.006
Scale	0.796				
Loglik(model)	-4306.10				
Loglik(intercept only)	-4446.00				
Chisq	279.820				<0.001
Number of Newton-Raphson Iterations	8				
Concordance Index	0.700				
AIC	8632.255				
BIC	8689.308				

The survival probability plot stratified by peak hour categories reveals substantial differences in survival probabilities over successive trips. Drivers operating during afternoon peak hours show higher survival probabilities, suggesting that driving conditions during this period may contribute to safer behavior or delayed relapse. By contrast, morning peak hours indicate a slightly higher relapse risk, aligning with contextual factors such as traffic density or stress. Additionally, the frailty effects plot shows variability among individual drivers, with a range of frailty scores highlighting unobserved heterogeneity. This variability emphasizes the importance of accounting for driver-specific differences in the analysis.

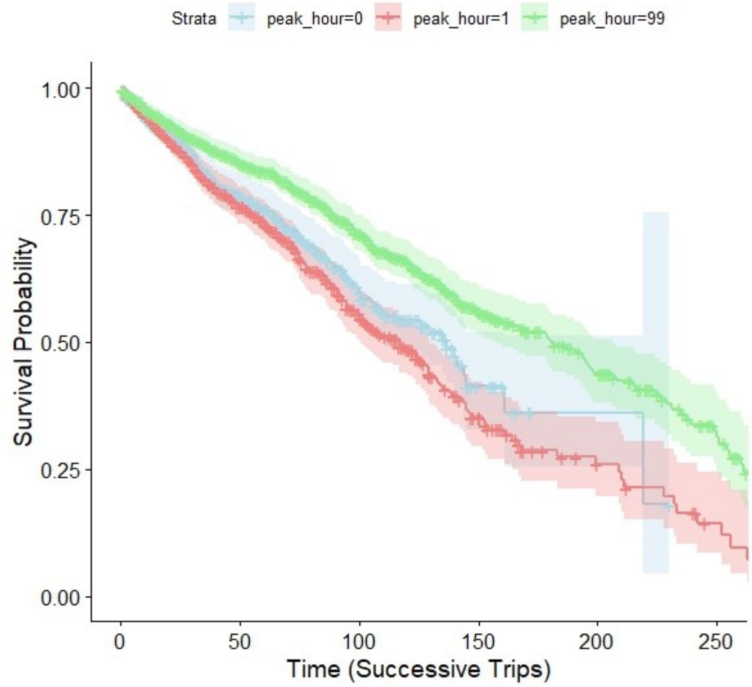


Figure 8.21: Survival probability for mobile phone use relapse

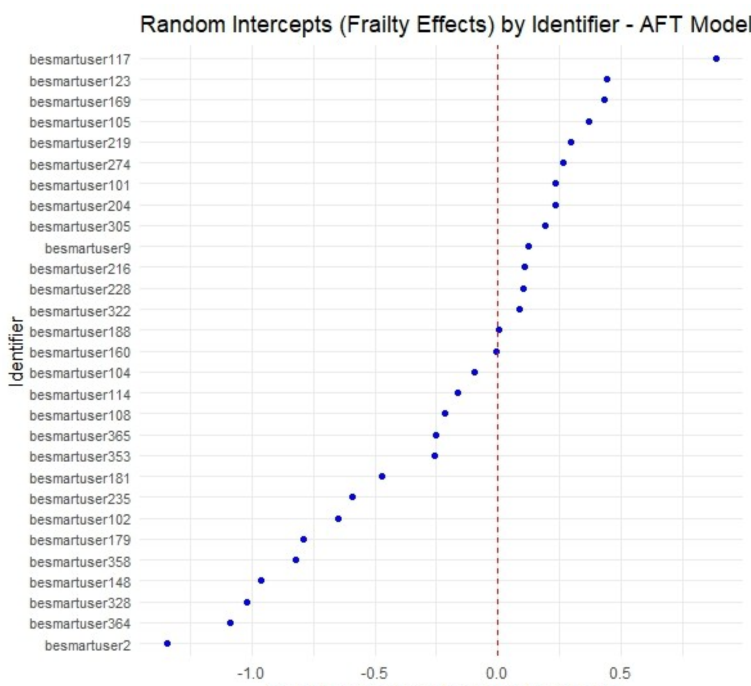


Figure 8.22: Random effects for mobile phone use relapse

The diagnostic plots provide additional evidence of model adequacy. The histogram and Q-Q plots of deviance residuals show a roughly symmetric distribution with slight deviations, while the Cox-Snell residuals confirm the appropriateness of the model specification with no evident patterns. The observed versus predicted times scatterplot shows reasonable alignment with the diagonal, though some variability at higher observed times suggests areas for improvement. Together, these results underscore the robustness of the Weibull AFT model while emphasizing the role of trip-specific and individual-level factors in understanding relapse in speeding behavior.

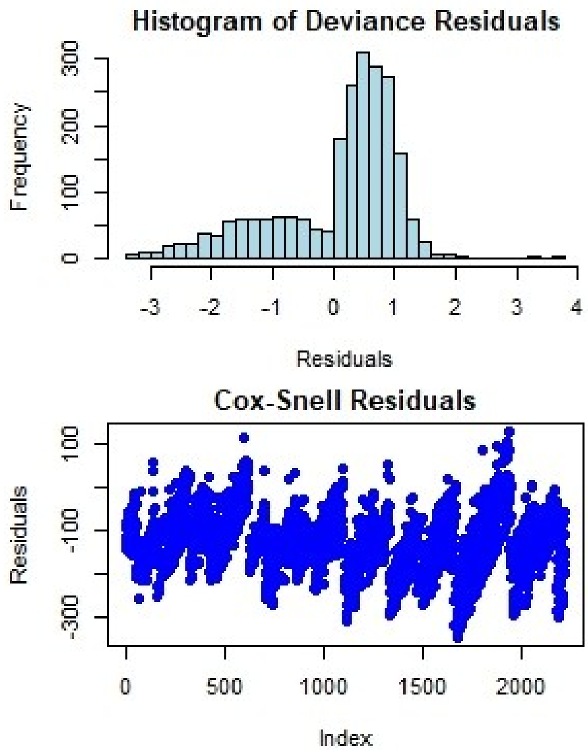


Figure 8.23: Model Diagnostics for the Weibull AFT Model on mobile phone use

8.3.4.4 Random Survival Forest

The results from the Random Survival Forest model highlight the critical predictors influencing speeding relapse among drivers. Among the six independent variables, age group emerged as the most significant predictor, suggesting that younger or older age groups may have distinct behavioral patterns contributing to relapse. Trip duration was the second most influential variable, likely reflecting how prolonged driving may lead to lapses in attention or adherence to speed limits. Vehicle engine size (vehicle_cc_group) also played a key role, indicating that drivers of larger engine vehicles might engage in riskier behaviors or have higher relapse tendencies.

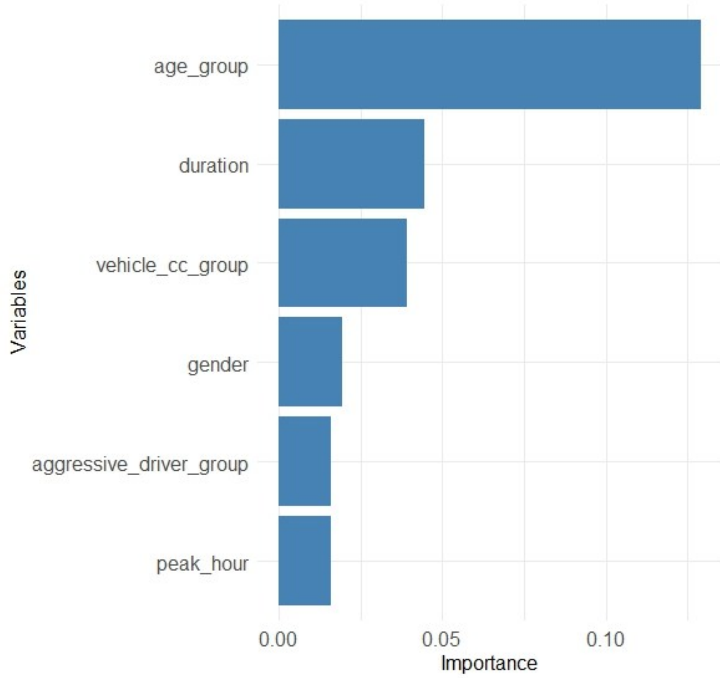


Figure 8.24: Feature importance of Random Forest for mobile phone use

The model's configuration, with 30 trees and a logrank split rule, efficiently handled the survival data, capturing non-linear relationships and interactions. However, while the out-of-bag (OOB) error rate of 0.243 indicates a reasonably good model fit, the predictive performance, as reflected in the Root Mean Squared Error (RMSE) of 89.77 and Mean Absolute Error (MAE) of 68.05, suggests room for improvement in accurately predicting relapse times. Despite this, the model's flexibility in ranking variables provides valuable insights into key behavioral and vehicle-related factors that could inform targeted interventions to reduce speeding relapse among drivers.

Table 8.18 RSF model results for mobile phone use

Type	**Survival**
Number of trees	30
Sample size	2220
Number of independent variables	6
Mtry	2
Target node size	5
Variable importance mode	permutation
Splitrule	logrank
Number of unique death times	154
OOB prediction error (1-C)	0.242985
Root Mean Squared Error (RMSE):	89.77445
Mean Absolute Error (MAE):	68.04955

8.3.4.5 *Model comparison for mobile phone use*

Table 8.19 shows the comparison of models for relapse in mobile phone use. The Weibull AFT model demonstrates the highest discriminative ability with a C-index of 0.773, making it the strongest performer in terms of predictive accuracy. It effectively handles frailty effects while providing interpretable results, identifying key predictors such as age group, aggressive driver group, vehicle CC, and trip duration. These features make it particularly valuable for understanding the dynamics of relapse. However, its AIC (3976.995) and BIC (4034.048) values indicate a poorer fit compared to the Cox model, and its reliance on the Weibull distribution may limit its flexibility in datasets with more complex structures. Despite these drawbacks, its overall interpretability and predictive capability position it as a strong contender for survival analysis.

Table 8.19: Models comparison summary of relapse in mobile phone use

Aspect	Weibull AFT Model	Cox Model with Frailty	Random Survival Forest (RSF)
Purpose	Models survival time directly	Models hazard rate	Captures non-linear effects
C-index	0.773	0.737	0.755
AIC	3976.995	3371.426	N/A
BIC	4034.048	3513.939	N/A
Key Predictors	Age group, aggressive driver group, vehicle CC, duration	Age group, aggressive driver group, duration	Age group, duration, vehicle CC group, aggressive driving
Frailty Effects	Accounted (Clustered Heterogeneity)	Accounted (Shared Frailty)	Implicitly handled (Non-parametric)
Prediction Error (RMSE/MAE)	RMSE: 92.47, MAE: 70.91	RMSE: 105.87, MAE: 85.41	RMSE: 85.87, MAE: 65.41
Strengths	Interpretable, adjusts for clustering; Highlights significant predictors	Adjusts for heterogeneity across clusters; Provides interpretable hazard ratios	Captures complex relationships; Robust to outliers
Weaknesses	Assumes Weibull distribution; Sensitive to deviations and outliers	Lower discrimination ability; Assumes proportional hazards	Less interpretable; Weaker numerical precision compared to parametric models

The Cox model with frailty offers the best model fit with the lowest AIC (3371.426) and a reasonable BIC (3513.939), demonstrating its strength in handling heterogeneity across clusters. However, its lower C-index (0.737) and higher prediction error metrics (RMSE: 105.87, MAE: 85.41) suggest weaker predictive accuracy compared to the Weibull AFT and RSF models. In contrast, the Random Survival Forest (RSF) excels in predictive performance, with a competitive C-index (0.755) and the lowest prediction error metrics (RMSE: 85.87, MAE: 65.41), showcasing its ability to model non-linear relationships and complex interactions. However, the RSF model's lack of interpretability and reliance on larger datasets make it less suitable for explanatory analyses. Overall, the Weibull AFT model is preferable for balancing interpretability and predictive accuracy, while RSF stands out when prioritizing predictive performance and modeling flexibility.

8.4 Discussion of Results

The results of this section provide a comprehensive evaluation of driver telematics feedback on long-term driving behavior using survival analysis techniques, with a particular focus on relapse in mobile phone use, speeding, harsh braking, and harsh accelerations. Across all indicators, driver feedback demonstrated significant short-term improvements in driving behavior during the feedback phase. However, the post-feedback phase revealed varied tendencies toward relapse, underscoring the importance of intervention strategies in shaping driver behavior over time.

The Weibull Accelerated Failure Time (AFT) model consistently emerged as a robust performer across the examined indicators, balancing predictive accuracy and interpretability. The C-index values ranged between 0.677 and 0.773, with the model achieving the highest predictive ability for mobile phone use relapse (C-index = 0.773), indicating strong discriminative capacity in identifying drivers most at risk of relapse. Key predictors such as age group [35-54] ($\beta = 0.165$, $p = 0.041$), trip duration ($\beta = -0.022$, $p < 0.001$), and self-reported aggressiveness (approaching significance at $p = 0.089$) were highlighted, providing actionable insights into relapse behavior. The model also captured heterogeneity across drivers by incorporating random effects, with frailty effects showing significant variability in survival times. For speeding relapse, the Weibull AFT model achieved a C-index = 0.700 with significant predictors including trip duration ($\beta = -0.022$, $p < 0.001$) and morning peak hours ($\beta = -0.096$, $p = 0.004$). Trip duration, in particular, emerged as the dominant predictor, consistently reducing survival time across all relapse indicators, underscoring the role of prolonged driving in behavioral regression. Similarly, in the analysis of harsh braking relapse, the model achieved a moderate predictive accuracy (C-index = 0.724) with significant contributions from variables such as age group [35-54] ($\beta = 0.360$, $p = 0.010$) and vehicle engine capacity (>1400cc) ($\beta = -0.508$, $p = 0.012$). These findings highlight that younger age groups and drivers of larger-engine vehicles are more prone to relapse.

The model's ability to account for clustered heterogeneity through frailty terms further enhanced its robustness, capturing unobserved individual-specific differences that influence relapse times. For instance, random intercept variability was reflected in frailty standard deviations, ranging from 1.189 in harsh accelerations to 1.419 in mobile phone use relapse. Additionally, log-likelihood values and diagnostic metrics such as AIC and BIC demonstrated good model fit, with values like AIC = 7705.44 for harsh accelerations and AIC = 8632.26 for speeding relapse. Despite these strengths, the Weibull AFT model does have limitations. Its reliance on the Weibull distribution assumes a specific survival time shape, which may not fully capture the complexities of behavioral relapse in datasets with non-parametric relationships. Furthermore, the model's sensitivity to outliers can affect predictive performance in cases with extreme observations, particularly in harsh braking and speeding events. Nonetheless, the Weibull AFT model's combination of interpretability and predictive power makes it a valuable tool for understanding relapse dynamics between driving behavior indicators and cofounding parameters.

The Cox Proportional Hazards model with frailty also proved valuable, particularly in its ability to flexibly handle heterogeneity through shared frailty terms. Its strength lay in providing interpretable hazard ratios that elucidated the relative impact of predictors on relapse risks. For example, the model highlighted the protective effects of age (e.g., age group [35–54] reduced relapse hazards for mobile phone use and harsh braking) and the influence of driving conditions such as trip duration and peak hour. However, its C-index values (e.g., 0.737 for mobile phone use and 0.653 for harsh braking) and higher prediction errors indicate lower predictive accuracy compared to the Weibull AFT model and Random Survival Forest (RSF). Furthermore, and most importantly, the proportional hazards assumption, which was violated in several instances, further limited its applicability for these datasets.

The Random Survival Forest (RSF) model excelled in capturing complex, non-linear interactions among predictors, achieving the lowest prediction error metrics across most relapse indicators. With its ability to model non-parametric relationships, the RSF effectively uncovered nuanced dynamics in driver behavior relapse. Specifically, the Root Mean Squared Error (RMSE) values ranged between 91.36 and 85.87, while the Mean Absolute Error (MAE) fell between 70.67 and 65.41, reflecting its strong predictive performance. For mobile phone use relapse, the out-of-bag (OOB) prediction error was the lowest among all models at 24.3%, reinforcing RSF's superior accuracy. Trip duration consistently emerged as the most critical predictor of relapse across all behaviors, while other variables like aggressive driving tendencies and vehicle engine size showed moderate importance. Additionally, the RSF effectively identified the interplay between demographic factors, such as age group and gender, and contextual elements, like peak hours, highlighting the model's flexibility in handling diverse predictors.

Despite its predictive strengths, the RSF model has limitations, primarily its lack of interpretability. Unlike parametric models, RSF operates as a "black box," making it difficult to derive hazard ratios or direct survival time estimates, which are often critical for explanatory analysis. Furthermore, the RSF model's reliance on larger datasets to capture complex interactions makes it less efficient in studies with smaller sample sizes. Nonetheless, the RSF is a powerful tool for analyzing behavioral relapse, offering exceptional predictive performance and the ability to model intricate, non-linear relationships. Future research could enhance its interpretability by incorporating post-hoc techniques, such as SHAP values or partial dependence plots, enabling a deeper understanding of the drivers behind relapse while maintaining the model's predictive accuracy.

Overall, comparing the models, the Weibull AFT model stands out for balancing interpretability and predictive accuracy, making it particularly suited for contexts requiring actionable insights into survival dynamics. The Cox model offers a useful compromise with its interpretability and ability to handle frailty, albeit at the cost of predictive performance. The RSF model is most appropriate for predictive tasks where capturing non-linear relationships and complex interactions is critical, though its lack of transparency limits its utility in understanding the underlying behavioral mechanisms. These findings emphasize the importance of aligning model configuration with research objectives. For studies focused on understanding behavioral dynamics and guiding intervention design, the Weibull AFT model provides robust insights. Conversely, when predictive accuracy is paramount, RSF offers a superior alternative.

The results also underscore the critical need for sustained feedback mechanisms to reinforce safe driving behaviors over the long term, as relapse patterns were consistently observed across all examined indicators once feedback interventions were withdrawn. For harsh accelerations, the Kaplan-Meier survival analysis revealed that approximately 84.8% of drivers maintained improved behavior within the first 50 trips of the post-feedback phase. However, this figure declined to 68.7% at 100 trips and further dropped to 49.2% by 150 trips, indicating that nearly half of the drivers reverted to unsafe acceleration habits within this period. Similarly, for harsh braking, survival probabilities decreased progressively from 81.5% at 50 trips to 61.4% at 100 trips, and finally to 40.3% at 150 trips, demonstrating a steady return to pre-feedback behaviors.

For speeding, the relapse trends were equally pronounced, with survival probabilities showing a gradual decline. At the 50-trip mark, 82.3% of drivers adhered to improved behavior, but by 100 trips, this proportion fell to 65.2%, and by 150 trips, only 46.8% of drivers maintained reduced speeding levels. This highlights that more than half of the participants relapsed into speeding behaviors in the absence of ongoing feedback. Mobile phone use, while showing a slower relapse pattern, followed a similar trajectory: 91.7% of drivers refrained from phone use at 50 trips, but adherence reduced to 84.8% by 100 trips. Although these percentages suggest slightly greater resilience in behavior regarding distracted driving, the decline over time remains evident.

Despite its valuable insights, this study has several limitations that should be acknowledged. The sample size of 31 drivers, while relatively small, is mitigated by the extensive number of trips analyzed over a prolonged period, providing a robust dataset for understanding long-term driving behavior. However, the exclusion of traffic conditions, which are critical contextual factors influencing driver behavior, represents a significant limitation. Incorporating traffic dynamics into future research could provide a more comprehensive understanding of relapse behaviors. Additionally, as this is a macroscopic study focusing on aggregated driving patterns, future research could delve into microscopic elements, such as moment-to-moment decision-making and situational responses. Employing random parameters with heterogeneity-in-means in AFT models in future studies would also allow for a more nuanced analysis (Ali et al., 2022; Sharma et al., 2020) by accommodating greater complexity and variability among drivers and driving behavior indicators, further enhancing the understanding of long-term feedback effects.

9 Conclusions

9.1 Dissertation Overview

Road safety is a critical public health and societal issue, as road traffic crashes claim millions of lives and cause severe injuries globally every year. Beyond the tragic loss of life, these incidents impose immense emotional and economic burdens on families and communities. Research attributes approximately 95% of road crashes to human error, underscoring the critical role of driver behavior in accident prevention. Understanding and addressing risky driving behaviors—such as distracted driving, speeding, and harsh events—are pivotal to enhancing road safety.

Based on the above, the primary aim of this dissertation is to investigate the driver telematics feedback mechanism under the framework of driving behavior and road safety. Despite growing interest from automotive manufacturers and transportation researchers in driver behavior, limited research exists on quantifying the comprehensive impact of driver feedback across its entire lifecycle—encompassing the pre-feedback, feedback, and post-feedback phases. To address this gap, this dissertation adopts a holistic approach to evaluate the effectiveness of feedback on modifying driving behavior and ultimately enhancing road safety.

To achieve these objectives, a series of methodological steps were carefully implemented. The methodological framework provides a structured approach to achieving the objectives of this dissertation.

As a first step, a systematic literature review was conducted to evaluate the effectiveness of driver feedback within naturalistic driving studies. Feedback mechanisms have evolved from in-vehicle devices and paper-based reports to sophisticated, user-friendly smartphone applications. These advancements enable the collection of high-resolution driving data and the delivery of personalized, data-driven feedback. Mechanisms such as real-time alerts, post-trip summaries, and performance reports have demonstrated potential for improving driver behavior.

However, significant gaps remain regarding the long-term sustainability of these effects and the differential impacts of feedback features. While studies highlight the effectiveness of feedback in reducing speeding, harsh braking, and mobile phone use, the influence of feedback type, frequency, and incentives on behavior remains underexplored. Additionally, many studies observe a relapse into risky behaviors once feedback is removed, necessitating further investigation.

Based on the results of the systematic literature review, the following research questions are formulated:

1. How does feedback influence driver speeding and distracted behavior in terms of the percentage of trip time during which the speed limit was exceeded and mobile phone was used while driving?

2. How does feedback influence harsh driving events, in terms of the number of harsh accelerations and harsh brakings?

3. Do different feedback features (e.g., scorecards, maps, peer comparisons, motivations, gamification, rewards) have different effects on driver behavior? Which feature demonstrates the most significant impact?

4. How does the post-feedback effect influence long-term driver behavior, and to what extent are the changes sustained after the feedback is removed?

5. How can advanced statistical techniques be applied to understand the mechanisms of driver feedback and develop more individualized, data-driven approaches for driving behavior change?

To answer the research questions, a robust methodological background was developed, combining theoretical approaches and experimental design principles. This included the application of advanced modeling techniques, such as Generalized Linear Mixed Effects Models (GLMMs), Structural Equation Models (SEMs), and Survival Analysis Models, alongside the design of a naturalistic driving experiment.

A 21-month naturalistic driving experiment involving 230 drivers was conducted. The participants were divided into three groups (car drivers, professional van drivers, and motorcyclists), and their driving behavior was monitored across six distinct feedback phases. These phases were defined as follows:

- Phase 1: Basic trip data and characterization were accessible to drivers.
- Phase 2: Introduction of scorecards with trip-level scoring.
- Phase 3: Addition of maps and highlights for further trip insights.
- Phase 4: Peer comparisons enabled for driver performance benchmarking.
- Phase 5: Competitions and challenges introduced with rewards for safe driving.
- Phase 6: Reversion to Phase 1, removing all additional feedback.

High-resolution data were collected from 106,776 trips, covering a total of 1,317,573 kilometers and 30,532 hours of driving. Behavioral metrics were captured using non-intrusive smartphone sensors, ensuring a seamless and accurate recording of driving behaviors. This sensor data was complemented by self-reported information from participants, providing a holistic understanding of driver perceptions, habits, and behavioral changes. The experimental design adhered to strict ethical standards, having been approved by the Research Ethics and Conduct Committee of NTUA, and ensured full compliance with GDPR guidelines. Continuous communication with participants was maintained throughout the study to address any technical issues, sustain engagement, and monitor the smooth execution of the experiment.

Extensive data processing and cleaning were carried out to ensure the quality and reliability of the dataset. Invalid or incomplete trip data were systematically identified and excluded, while key behavioral metrics such as speeding, mobile phone use while driving, harsh accelerations, and braking events were standardized for analysis. Data preprocessing steps included the conversion of raw sensor outputs into meaningful variables and the integration of self-reported data for cross-validation. This rigorous approach to data management enabled the creation of a robust dataset, facilitating detailed statistical analyses and ensuring the accuracy of the findings presented in this dissertation.

Advanced statistical techniques were employed to analyze feedback effects on critical driving indicators, such as speeding, mobile phone use, harsh accelerations, and harsh brakings. The analysis unfolded in three key pillars:

1. Impact of Feedback: This pillar assessed the immediate effects of feedback on i) driver speeding and distracted behavior, focusing on speeding among motorcyclists and distraction due to mobile phone use while driving in car drivers, and ii) driver harsh events, focusing on harsh braking and harsh accelerations among car drivers, and professional drivers on highways. Generalized Linear Mixed-Effects Models (GLMM) were then employed in all cases to evaluate the effects of feedback while accounting for individual differences and contextual factors.

2. Effects of Different Feedback Features: A Structural Equation Model (SEM) was developed to explore the complex relationships between feedback features (e.g. scorecards, maps, peer comparisons, motivations, gamification, rewards) and driver behavior, exposure metrics and safety outcomes, allowing for the simultaneous analysis of multiple variables and their interactions.

3. Post-Feedback Effects: Effects on long-term driver behavior, with a particular emphasis on understanding the relapse of driving behaviors following the withdrawal of feedback telematics during the last phase of the experiment. Survival Analysis methods were employed to investigate relapse patterns across various indicators, including harsh accelerations, harsh braking, speeding behavior, and mobile phone use while driving. These analyses leverage Kaplan-Meier curves, Cox-PH models with frailty, Weibull AFT models with clustered heterogeneity, and Random Survival Forests to evaluate and compare the predictive power and insights offered by each model.

Ultimately, the synthesis of all the analyses carried out within the framework of this doctoral dissertation resulted in a driver behavior telematics feedback mechanism with numerous original and interesting results, which are discussed in the following concluding subsections.

9.2 Main Findings

9.2.1 Main findings of the feedback impact on driver behavior

The investigation of feedback impacts on driver behavior yielded significant findings across different user groups, driving environments, and behavioral metrics. Overall, during the two phases of the experiment a large dataset of 3,537 trips from a sample of 13 motorcyclists were recorded and analysed. Using Generalized Linear Mixed-Effects Models with random intercepts and random slopes for total trip duration revealed that providing motorcyclists with feedback about their riding performance during experiment Phase 2 led to a remarkable decrease in speeding percentage over a trip. Particularly, in the developed models rider feedback seems to decrease speeding percentage, having a risk ratio of $\exp(\beta=-0.145) = 0.865$ for the overall model (13.5% decrease), and $\exp(\beta=-0.031) = 0.970$ and $\exp(\beta=-0.420) = 0.657$ for urban (3.0% decrease) and rural (34.3%) road types respectively. These results highlight the effectiveness of feedback in targeting high-risk behaviors, offering a foundation for scalable interventions in rider training and policy design.

Similarly, distracted driving, particularly mobile phone use, was examined across urban, rural, and highway contexts via GLMM models for 65 car drivers over 21,167 trips. Feedback emerged as a strong restrictive in using the mobile phone while driving overall ($\beta = -0.4276$, $p < 2e-16$), particularly in urban settings ($\beta = -0.3687$, $p < 2e-16$), while its impact was notably weaker in rural environments ($\beta = -0.1180$, $p < 2e-16$) and unexpectedly positive on highways ($\beta = 0.5490$, $p < 2e-16$), suggesting compensatory behaviors or a perceived lower risk of distraction on high-speed roads. The substantial variability in random intercepts (SD = 1.4024 overall) highlights notable individual differences in baseline behavior, while random slopes for trip duration (SD = 0.2827 overall) show diverse responses to prolonged trips.

In the domain of harsh events such as accelerations and brakings, feedback mechanisms demonstrated significant behavioral improvements, as well. Results from the analysis of 65 car drivers during the first two phases of the experiment, revealed a significant reduction in harsh accelerations (12%) and harsh brakings (10%), both changes being statistically significant ($p < 0.001$). The GLMM models further reinforce these findings, as feedback was consistently associated with reduced frequencies of harsh events, particularly in urban and rural environments. Notably, the relative risk ratios for speeding duration and trip duration indicate strong positive associations with harsh events, though feedback appears to mitigate these effects to a degree in Phase 2 of the experiment. Importantly, driver-specific variability, captured through random intercepts and slopes, underscores the need for tailored feedback mechanisms to address unique behavioral traits.

Professional drivers, due to their prolonged driving hours and distances, were also a key area of exploration. Using GLMMs calibrated on a dataset of 5,345 trips from 19 professional drivers, the analysis revealed that participation in a social gamification scheme with incentives led to notable improvements in harsh events of professional drivers. During the competition phase, the likelihood of harsh accelerations was reduced by a factor of 0.348 ($p < 0.001$), while harsh brakings decreased by a factor of 0.404 ($p < 0.001$), indicating the efficacy of gamification in promoting safer driving practices. Additionally, trip duration showed a positive association with harsh events, with a 1-second increase in driving time raising the odds of harsh accelerations and harsh brakings by factors of 1.558 and 1.564, respectively, highlighting the cumulative effects of extended driving. The inclusion of random intercepts in the models underscored substantial variability in baseline driver behavior, emphasizing the importance of personalized interventions.

9.2.2 Main findings of the different feedback features effects

The Structural Equation Model (SEM) analysis provided significant insights into the impact of different effects of feedback features on driver behavior, specifically speeding, harsh braking, and harsh acceleration events. The dataset, comprising 73,869 trips from 175 car drivers over 21 months, offered a robust basis for modeling. The SEM results identified two latent variables, namely feedback and exposure as critical influences. The model exhibited excellent goodness-of-fit measures, with Comparative Fit Index (CFI) = 0.940, Tucker–Lewis Index (TLI) = 0.944, Root Mean Square Error Approximation (RMSEA) = 0.049, and Standardized Root Mean Square Residual (SRMR) = 0.025, indicating a robust and well-specified structure. The inclusion of covariances among variables, guided by residual correlation analysis, further improved the model fit and highlighted critical relationships, such as those between speeding and harsh braking behaviors.

Among the feedback features analyzed, the scorecard emerged as the most influential feature, with the highest positive estimate ($\beta = 2.076$, $p < 0.001$), demonstrating its powerful role in promoting safer driving habits by immediately altering risky behaviors in comparison with the baseline phase. This result can be attributed to the clear, concise, and actionable nature of scorecards, which provide drivers with straightforward insights into their performance and specific areas for improvement, making it easier to adjust their behavior. Similarly, the maps feature showed a strong impact ($\beta = 1.646$, $p < 0.001$), emphasizing the importance of spatial awareness in enhancing driving practices. The compare feature allowed drivers to assess their performance relative to peers, positively influencing behavior ($\beta = 1.215$, $p < 0.001$). Additionally, the competition & challenges feature proved highly effective ($\beta = 2.053$, $p < 0.001$) by motivating drivers to adopt safer driving behaviors through gamified elements and rewards for safe driving.

In terms of driving behavior metrics, driver telematics feedback significantly reduced the percentage of speeding time ($\beta = -0.214$, $p < 0.001$) and harsh braking events per 100km ($\beta = -0.027$, $p < 0.001$). However, an increase in harsh accelerations per 100km ($\beta = 0.026$, $p < 0.001$) suggests the need for further refinement of feedback systems to address unintended consequences. Exposure factors also played a key role in shaping driver behavior, with morning peak exposure correlating with increased risk-taking ($\beta = 2.473$, $p < 0.001$), likely driven by time pressure during commuting hours. Conversely, afternoon peak exposure was associated with less aggressive behavior ($\beta = -1.360$, $p < 0.001$), providing insights into temporal variations in driving patterns.

Regression analysis confirmed these findings, highlighting the interplay between exposure and feedback features. While exposure positively influenced speeding ($\beta = 0.326$, $p < 0.001$), feedback features effectively mitigated this behavior. The competition & challenges feature, in particular, showed promise in moderating harsh accelerations ($\beta = -0.001$, $p < 0.001$). Harsh braking incidents were also significantly reduced by feedback, reinforcing the role of feedback in promoting safer driving practices. Covariance analysis further revealed strong interrelationships between risky behaviors, such as speeding and harsh braking, underscoring the complexity of driver behavior patterns. These findings suggest that speeding often necessitates sudden corrections, like harsh braking, and both behaviors may stem from underlying traits such as risk-taking tendencies or aggressive driving habits.

The practical implications of these findings are substantial. Feedback features, particularly those leveraging personalized scorecards, spatial tools, and gamification elements, hold great promise for improving driver safety. Tailored interventions targeting specific behaviors and times of day could further enhance the efficacy of these systems. However, limitations such as the exclusion of mobile phone use from the final model and potential selection biases due to the voluntary nature of participation should be addressed in future research.

9.2.3 Post-feedback effect on long-term driver behavior

Survival analysis techniques were applied to a dataset of 24,904 trips from 31 car drivers, each contributing at least 20 trips in the post-feedback phase, to investigate the long-term effects of driver telematics feedback on driving behavior. The analysis focused on relapse patterns in mobile phone use, speeding, harsh braking, and harsh accelerations. The methods utilized included Kaplan-Meier curves, Cox-PH models with frailty, Weibull Accelerated Failure Time (AFT) models incorporating clustered heterogeneity, and Random Survival Forests. The findings demonstrate the effectiveness of feedback interventions in achieving significant short-term behavioral improvements during the feedback phase. However, the post-feedback phase reveals varied relapse tendencies, emphasizing the need for sustained interventions to maintain these improvements over time.

The Kaplan-Meier survival analysis emphasized relapse trends, showing a steady decline in improved behavior over successive trips in the post-feedback phase. For harsh accelerations, survival probabilities dropped from 84.8% at 50 trips to 49.2% by 150 trips. Similar trends were observed for harsh braking and speeding, with survival probabilities declining to approximately 40.3% and 46.8%, respectively, by the 150-trip mark. These patterns underscore the transient nature of feedback effects and the need for continuous reinforcement mechanisms. Mobile phone use showed slightly greater resilience, with survival probabilities remaining above 80% at 100 trips, but the gradual relapse was evident over time.

Among the survival analysis models applied, the Weibull Accelerated Failure Time (AFT) model consistently emerged as a robust performer across the examined indicators, balancing predictive accuracy and interpretability. The concordance index (C-index) values ranged between 0.677 and 0.773, with the model achieving the highest predictive ability for mobile phone use relapse (C-index = 0.773), indicating strong discriminative capacity in identifying drivers most at risk of relapse. Key predictors such as age group [35-54] ($\beta = 0.165$, $p = 0.041$), trip duration ($\beta = -0.022$, $p < 0.001$), and self-reported aggressiveness (approaching significance at $p = 0.089$) were highlighted, providing actionable insights into relapse behavior. The model also captured heterogeneity across drivers by incorporating random effects, with frailty effects showing significant variability in survival times.

For speeding relapse, the Weibull AFT model achieved a C-index = 0.700 with significant predictors including trip duration ($\beta = -0.022$, $p < 0.001$) and morning peak hours ($\beta = -0.096$, $p = 0.004$). Trip duration, in particular, emerged as the dominant predictor, consistently reducing survival time across all relapse indicators, underscoring the role of prolonged driving in behavioral regression. Similarly, in the analysis of harsh braking relapse, the model achieved a moderate predictive accuracy (C-index = 0.724) with significant contributions from variables such as age group [35-54] ($\beta = 0.360$, $p = 0.010$) and vehicle engine capacity (>1400cc) ($\beta = -0.508$, $p = 0.012$). These findings highlight that younger age groups and drivers of larger-engine vehicles are more prone to relapse.

The Random Survival Forest (RSF) model demonstrated superior predictive performance in some examined indicators, excelling in capturing non-linear interactions and complex relationships between predictors. With Root Mean Squared Error (RMSE) values as low as 85.87 and out-of-bag (OOB) prediction errors of 24.3% for mobile phone use relapse, RSF identified critical predictors such as trip duration, aggressive driving tendencies, and vehicle engine size. Its flexibility in handling diverse predictors and uncovering nuanced dynamics makes RSF an invaluable tool for predictive analyses. However, the model's "black box" nature and reliance on larger datasets limit its interpretability and applicability for explanatory purposes.

Overall, comparing the models, the Weibull AFT model stands out for balancing interpretability and predictive accuracy, making it particularly suited for contexts requiring actionable insights into survival dynamics. The Cox model offers a useful compromise with its interpretability and ability to handle frailty, however repeatedly failed to meet model assumptions. The RSF model is most appropriate for predictive tasks where capturing non-linear relationships and complex interactions is critical, though its lack of transparency limits its utility in understanding the underlying behavioral mechanisms. These findings emphasize the importance of aligning model configuration with research objectives. For studies focused on understanding behavioral dynamics and guiding intervention design, the Weibull AFT model provides robust insights. Conversely, when predictive accuracy is paramount, RSF offers a superior alternative.

While this study offers valuable insights into the dynamics of driver feedback and relapse, limitations such as the relatively small sample size, exclusion of traffic conditions, and macroscopic focus should be noted. Future research could incorporate more granular data, such as traffic dynamics and moment-to-moment driver decisions, to provide a deeper understanding of behavioral patterns. Employing advanced modeling techniques, such as random parameters with heterogeneity-in-means, could further enhance the analysis by accounting for driver-specific variability.

9.3 Innovative Scientific Contributions

The innovative contributions of this doctoral dissertation consist of five original scientific contributions, as described below, and illustrated in Figure 9.1.

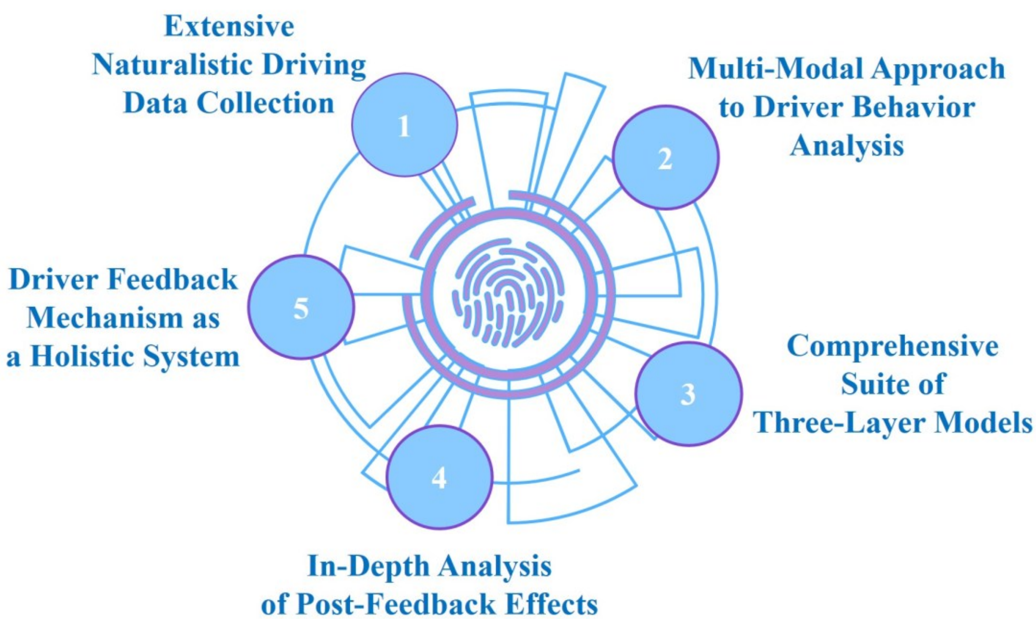


Figure 9.1: Innovative contributions of the doctoral dissertation

9.3.1 Extensive naturalistic driving data collection

The present dissertation represents a significant step forward in naturalistic driving (ND) research by leveraging non-intrusive data collection methods that rely on smartphone sensors. Unlike traditional approaches, this methodology minimizes disruption to participants, enabling the unobtrusive capture of real-world driving behaviors. The data spans a large sample size of drivers (230) across diverse road environments and vehicle types, including car drivers, motorcyclists, and professional van drivers. This inclusivity ensures that findings are not only representative but also account for variations across driver demographics and vehicle categories. The dataset's rich temporal resolution provides detailed insights into driving behaviors at the trip level, offering a granular perspective on driver behavior dynamics.

Moreover, the long-term data collection 21-month period, spanning multiple feedback phases and covering various road environments, adds unique value. By capturing behavior changes over time, the study bridges a critical gap in existing ND research, which often relies on short-term observations. This long-term perspective enables the assessment of sustained behavior modifications and relapse tendencies, providing a robust foundation for developing adaptive and sustainable interventions to improve road safety. The methodology sets a new benchmark for ND experiments, paving the way for more scalable, cost-effective, and technologically advanced driving behavior studies.

9.3.2 Multi-modal approach to driver behavior analysis

This dissertation takes a multi-modal approach, emphasizing the importance of understanding driving behaviors across diverse road user groups and environments. By including car drivers, motorcyclists, and professional van drivers, the research recognizes the critical need to study vulnerable road users, such as motorcyclists, who face heightened risks, and professional drivers, who spend extended hours on the road. This inclusive focus ensures a comprehensive evaluation of driver telematics feedback, highlighting their relevance across varying risk profiles and exposure levels.

The investigation also considers the influence of urban, rural, and highway environments, acknowledging the distinct challenges posed by each road type. This contextual approach reveals that feedback effectiveness is not uniform; behaviors like mobile phone use or speeding respond differently to interventions depending on the driving environment. For instance, motorcyclists may benefit more from feedback targeting situational awareness, while professional drivers might require tailored interventions addressing fatigue and repetitive exposure to high-risk scenarios.

By integrating this diversity of user groups and contexts, the dissertation provides actionable insights for policymakers, road safety advocates, and technology developers. It emphasizes the importance of developing tailored feedback systems that cater to the unique needs of vulnerable road users, such as motorcyclists, and professional drivers, who contribute significantly to road traffic activity. This comprehensive approach supports the creation of adaptive, context-sensitive interventions, ultimately improving road safety for all users.

9.3.3 Comprehensive suite of three-layer models

This dissertation employs a comprehensive suite of advanced statistical and machine learning models, tailored to address the multifaceted nature of driving behavior analysis. By incorporating Generalized Linear Mixed-Effects Models, Structural Equation Models, and Survival Analysis techniques (e.g., Weibull AFT, Cox-PH with frailty, and Random Survival Forest), the study provides a rigorous analytical framework capable of uncovering both linear and non-linear relationships between variables. Each model is carefully selected to align with the research objectives, balancing predictive accuracy with interpretability to ensure actionable insights.

This model suite also enables the exploration of complex phenomena, such as the interplay between feedback features, driving behaviors, and contextual factors like time of day or road type. For example, survival models uniquely capture relapse dynamics, offering novel insights into post-feedback behavioral tendencies. Machine learning techniques further enhance the study by capturing nuanced, non-linear interactions, ensuring that the models are equipped to handle the complexity of real-world driving data. This innovative analytical framework not only elevates the scientific rigor of the research but also demonstrates the potential of combining traditional statistical methods with state-of-the-art machine learning approaches for driver behavior studies.

9.3.4 In-depth analysis of post-feedback effects

This dissertation is among the first to analyze thoroughly post-feedback effects on driver behavior using advanced statistical and machine learning techniques, addressing a critical gap in existing research. Through survival analysis methods, such as Weibull AFT and Random Survival Forest, the study evaluates long-term behavior changes and relapse patterns after feedback withdrawal. These techniques enable a detailed exploration of the factors influencing relapse in risky behaviors like speeding, harsh events, and mobile phone use, providing actionable insights for the design of sustained intervention strategies.

The findings reveal the importance of adaptive feedback systems that can maintain behavior improvements over time. For example, survival analysis showed that trip duration and time of day significantly influence relapse dynamics, emphasizing the need for context-aware feedback mechanisms. This innovative focus on the post-feedback phase provides a novel framework for understanding the longevity of feedback-induced improvements, allowing for more durable and impactful road safety interventions. It also sets a precedent for future research to integrate long-term perspectives into the evaluation of driving behavior modification strategies.

9.3.5 Driver feedback mechanism as a holistic system

This dissertation uniquely approaches the feedback mechanism as a holistic system, examining its full lifecycle through a multiparametric analytical framework. By systematically analyzing the pre-feedback, feedback, and post-feedback phases, the study offers a comprehensive understanding of how feedback influences driver behavior across time. The integration of diverse feedback features, such as scorecards, maps, comparison tools, and competition elements, enables the evaluation of their individual and combined impacts on behavior modification. This multi-phase perspective not only captures immediate behavior changes but also sheds light on long-term patterns and relapse tendencies.

Furthermore, the holistic framework provides valuable insights into the synergies and trade-offs between different feedback features. For instance, while scorecards and competition elements are highly effective in reducing speeding, their impact on other behaviors like harsh accelerations requires further refinement. This systemic approach advances the field by moving beyond isolated feedback evaluations, offering a scalable, data-driven framework for designing and implementing telematics-based interventions. The findings emphasize the potential of adaptive feedback systems to improve driving behavior sustainably, ultimately contributing to safer road environments.

9.4 Challenges Ahead

While this thesis provides significant contributions to understanding driver behavior and the role of telematics-based feedback mechanisms, several limitations should be acknowledged. One notable limitation lies in the lack of traffic data, which could have provided essential context for interpreting driving behavior. Traffic density, flow patterns, and congestion levels are known to significantly influence driver decisions, such as speeding, harsh braking, and aggressive maneuvers. The absence of such data limits the ability to fully account for external conditions that may contribute to risky driving behaviors. Future studies should aim to integrate traffic data from external sources, such as real-time traffic monitoring systems or connected infrastructure, to provide a more holistic understanding of driver behavior in various traffic conditions.

A related limitation is the geographic and sample constraints of the study. The data collection was conducted within specific regions, focusing on a sample of drivers that, while diverse in terms of vehicle types and road environments, may not represent the global driving population. Cultural, legal, and infrastructural differences across regions may affect driving behaviors and the effectiveness of feedback mechanisms. To enhance the generalizability of the findings, future research should replicate the study across multiple geographic locations, incorporating diverse road environments, legal frameworks, and driving cultures.

The study also faced challenges related to self-selection bias, inherent in naturalistic driving experiments. Participants voluntarily agreed to take part, which may introduce bias as these individuals might already be more safety-conscious or willing to improve their driving behavior. This could lead to an overestimation of the effectiveness of feedback mechanisms. Future research should aim for broader recruitment strategies, such as random sampling or targeted incentives, to ensure a more representative sample of the driving population.

The technological constraints of using smartphone sensors for data collection pose another limitation. While the use of smartphones provides a low-cost, non-intrusive method, they have inherent limitations in data accuracy and coverage. For example, harsh event detection, such as accelerations and brakings, is sensitive to the smartphone's placement and calibration. Moreover, smartphones lack the ability to capture critical parameters such as traffic conditions, weather, and external environmental factors, which could significantly influence driving behavior. Future studies should explore integrating data from multiple sources, such as connected vehicle systems, external traffic monitoring systems, and weather data, to create a richer and more robust dataset.

The study's temporal scope, although extensive compared to many prior studies, is still limited in capturing the full lifecycle of behavior changes induced by feedback. While the research examines pre-feedback, feedback, and post-feedback phases, the long-term sustainability of these changes remains unclear. Behavioral relapses may evolve over years, requiring more longitudinal studies to understand how feedback interventions impact behavior over extended periods. This challenge is particularly pertinent as driving technologies and environments continue to evolve, potentially altering the dynamics of driver behavior.

A further limitation arises from the complexity of modeling and interpretation. Advanced statistical and machine learning models were used to analyze the data, each with its strengths and limitations. While machine learning models, such as Random Survival Forests, excel in predictive accuracy, they lack transparency, making it challenging to derive actionable insights. Conversely, traditional models like the Cox Proportional Hazards model offer interpretability but may fail to capture complex, non-linear relationships. Future research should explore hybrid modeling approaches that combine the strengths of these methodologies, enhancing both predictive power and practical applicability.

Lastly, the integration of feedback with other road safety pillars, such as vehicle technology and road infrastructure, remains underexplored. While the thesis emphasizes driver behavior, the interaction between drivers, vehicles, and the built environment is a critical area for future exploration. Emerging technologies, including connected and autonomous vehicles, will further complicate these dynamics. Adaptive feedback systems that incorporate real-time traffic and infrastructure data will be necessary to address these complexities. Additionally, studying feedback interventions for vulnerable road users, such as pedestrians and cyclists, can extend the applicability of these findings to broader road safety contexts.

Addressing these limitations will require innovative methodologies, expanded datasets, and interdisciplinary collaborations. The increasing availability of advanced sensors, telematics platforms, and artificial intelligence provides an opportunity to overcome many of these challenges. By incorporating traffic data, expanding geographic and temporal scope, and leveraging advanced analytical techniques, future research can build upon this thesis to further advance road safety and driver behavior analysis.

References

1. Aarts, L., & Van Schagen, I. (2006). Driving speed and the risk of road crashes: A review. Accident Analysis and Prevention, 38(2), 215–224. https://doi.org/10.1016/j.aap.2005.07.004
2. Abdelkader, G., Elgazzar, K., & Khamis, A. (2021). Connected Vehicles: Technology Review, State of the Art, Challenges and Opportunities. Sensors 2021, Vol. 21, Page 7712, 21(22), 7712. https://doi.org/10.3390/S21227712
3. Adavikottu, A., Velaga, N. R., & Mishra, S. (2023). Modelling the effect of aggressive driver behavior on longitudinal performance measures during car-following. Transportation research part F: traffic psychology and behavior, 92, 176-200.
4. Afghari, A. P., Haque, M. M., Washington, S., & King, M. J. (2020). A survival analysis approach to investigating duration of distractions among drivers: Implications for road safety. Accident Analysis & Prevention, 137, 105463.
5. Ahmed, M. M., Khan, M. N., Das, A., & Dadvar, S. E. (2022). Global lessons learned from naturalistic driving studies to advance traffic safety and operation research: A systematic review. Accident Analysis & Prevention, 167, 106568. https://doi.org/10.1016/J.AAP.2022.106568
6. Aidman, E., Chadunow, C., Johnson, K., & Reece, J. (2015). Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance. Accident Analysis & Prevention, 81, 8–13. https://doi.org/10.1016/J.AAP.2015.03.041
7. Akritidou, S., Kontaxi, A., Ziakopoulos, A., & Yannis, G. (2023). Critical factors affecting mobile phone use while driving through the exploitation of smartphone sensor data. Transportation Research Procedia, 72, 919–925 https://doi.org/10.1016/J.TRPRO.2023.11.518
8. Albalate D, Fernández-Villadangos L. Motorcycle injury severity in Barcelona: the role of vehicle type and congestion. Traffic Inj Prev. 2010; (6):623-31. doi:10.1080/15389588.2010.506932.
9. Ali, Y., Haque, M. M., Zheng, Z., & Afghari, A. P. (2022). A Bayesian correlated grouped random parameters duration model with heterogeneity in the means for understanding braking behavior in a connected environment. Analytic methods in accident research, 35, 100221.
10. Amado, S., Arikan, E., Kaça, G., Koyuncu, M., & Turkan, B. N. (2014). How accurately do drivers evaluate their own driving behavior? An on-road observational study. Accident Analysis & Prevention, 63, 65–73. https://doi.org/10.1016/J.AAP.2013.10.022
11. Amarasinghe, M., Kottegoda, S., Arachchi, A. L., Muramudalige, S., Bandara, H. M. N. D., & Azeez, A. (2016). Cloud-based driver monitoring and vehicle diagnostic with OBD2 telematics. 15th International Conference on Advances in ICT for Emerging Regions, ICTer 2015 - Conference Proceedings, 243–249. https://doi.org/10.1109/ICTER.2015.7377695
12. Anderson, K. M. (1991). A nonproportional hazards Weibull accelerated failure time regression model. Biometrics, 281-288.
13. Archer, J. C. (2010). State of the science in health professional education: effective feedback. Medical Education, 44(1), 101–108. https://doi.org/10.1111/J.1365-2923.2009.03546.X
14. Backer-Grøndahl, A., & Sagberg, F. (2011). Driving and telephoning: Relative accident risk when using hand-held and hands-free mobile phones. Safety Science, 49(2), 324–330. https://doi.org/10.1016/j.ssci.2010.09.009
15. Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. R package version, 1(7), 1-23.

16. Bell, J. L., Taylor, M. A., Chen, G. X., Kirk, R. D., & Leatherman, E. R. (2017). Evaluation of an in-vehicle monitoring system (IVMS) to reduce risky driving behaviors in commercial drivers: Comparison of in-cab warning lights and supervisory coaching with videos of driving behavior. Journal of Safety Research, 60, 125–136. https://doi.org/10.1016/j.jsr.2016.12.008
17. Bellinger, D.B., Budde, B.M., Machida, M., Richardson, G.B., Berg, W.P. (2009). The effect of cellular telephone conversation and music listening on response time in braking. Transportation Research Part F 12 (6), pp. 441-451.
18. Birrell, S. A., & Fowkes, M. (2014). Glance behaviors when using an in-vehicle smart driving aid: A real-world, on-road driving study. Transportation Research Part F: Traffic Psychology and Behavior, 22, 113–125. https://doi.org/10.1016/j.trf.2013.11.003
19. Bo, C., Jian, X., Li, X. Y., Mao, X., Wang, Y., & Li, F. (2013, September). You're driving and texting: detecting drivers using personal smart phones by leveraging inertial sensors. In Proceedings of the 19th annual international conference on Mobile computing & networking (pp. 199-202).
20. Bolderdijk, J. W., Knockaert, J., Steg, E. M., & Verhoef, E. T. (2011). Effects of Pay-As-You-Drive vehicle insurance on young drivers ' speed choice : Results of a Dutch field experiment. Accident Analysis and Prevention, 43(3), 1181–1186. https://doi.org/10.1016/j.aap.2010.12.032
21. Bonsall, P., Liu, R., & Young, W. (2005). Modelling safety-related driving behavior—impact of parameter values. Transportation Research Part A: Policy and Practice, 39(5), 425-444.
22. Boquete, L., Rodríguez-Ascariz, J. M., Barea, R., Cantos, J., Miguel-Jiménez, J. M., & Ortega, S. (2010). Data acquisition, analysis and transmission platform for a pay-as-you-drive system. Sensors, 10(6), 5395-5408.
23. Boylan, J., Meyer, D., & Chen, W. S. (2024). A systematic review of the use of in-vehicle telematics in monitoring driving behaviors. Accident Analysis & Prevention, 199, 107519. https://doi.org/10.1016/J.AAP.2024.107519
24. Braun, D., Reiter, E., & Siddharthan, A. (2015). Creating textual driver feedback from telemetric data. Proceedings of the 15th European Workshop on Natural Language Generation (ENLG), 156–165.
25. Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. Journal of the American statistical Association, 88(421), 9-25.
26. Brouwer, R. F. T., Stuiver, A., Hof, T., Kroon, L., Pauwelussen, J., & Holleman, B. (2015). Personalised feedback and eco-driving: An explorative study. Transportation Research Part C: Emerging Technologies, 58(PD), 760–771. https://doi.org/10.1016/j.trc.2015.04.027
27. Caird, J. K., Simmons, S. M., Wiley, K., Johnston, K. A., & Horrey, W. J. (2018). Does Talking on a Cell Phone, With a Passenger, or Dialing Affect Driving Performance? An Updated Systematic Review and Meta-Analysis of Experimental Studies. https://doi.org/10.1177/0018720817748145
28. Caird, J. K., Willness, C. R., Steel, P., & Scialfa, C. (2008). A meta-analysis of the effects of cell phones on driver performance. Accident Analysis and Prevention, 40(4), 1282–1293. https://doi.org/10.1016/j.aap.2008.01.009
29. Camden, M. C., Soccolich, S. A., Hickman, J. S., & Hanowski, R. J. (2019). Reducing risky driving: Assessing the impacts of an automatically-assigned, targeted web-based instruction program. Journal of Safety Research, 70, 105–115. https://doi.org/10.1016/j.jsr.2019.06.006
30. Cash, P., Stanković, T., & Štorga, M. (2016). An Introduction to Experimental Design Research. Experimental Design Research: Approaches, Perspectives, Applications, 3–12. https://doi.org/10.1007/978-3-319-33781-4_1

31. Castignani, G., Derrmann, T., Frank, R., & Engel, T. (2015). Driver behavior profiling using smartphones: A low-cost platform for driver monitoring. IEEE Intelligent Transportation Systems Magazine, 7(1), 91–102. https://doi.org/10.1109/MITS.2014.2328673
32. Chan, T. K., Chin, C. S., Chen, H., & Zhong, X. (2020). A Comprehensive Review of Driver Behavior Analysis Utilizing Smartphones. IEEE Transactions on Intelligent Transportation Systems, 21(10), 4444–4475. https://doi.org/10.1109/TITS.2019.2940481
33. Charlton, S. G. (2009). Driving while conversing: Cell phones that distract and passengers who react. Accident Analysis and Prevention, 41(1), 160–173. https://doi.org/10.1016/j.aap.2008.10.006
34. Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. Journal of Economic Behavior & Organization, 81(1), 1–8. https://doi.org/10.1016/J.JEBO.2011.08.009
35. Chen, W., & Donmez, B. (2021). A Naturalistic Driving Study of Feedback Timing and Financial Incentives in Promoting Speed Limit Compliance. IEEE Transactions on Human-Machine Systems, 52(1), 64–73. https://doi.org/10.1109/THMS.2021.3117234
36. Chen, W., & Donmez, B. (2022). A Naturalistic Driving Study of Feedback Timing and Financial Incentives in Promoting Speed Limit Compliance. IEEE Transactions on Human-Machine Systems, 52(1), 64–73. https://doi.org/10.1109/THMS.2021.3117234
37. Choudhary, P., & Velaga, N. R. (2017). Modelling driver distraction effects due to mobile phone use on reaction time. Transportation Research Part C: Emerging Technologies, 77, 351-365.
38. Choudhary, P., Velaga, N.R., 2019. Effects of phone use on driving performance: A comparative analysis of young and professional drivers. Saf. Sci. 111, 179–187.
39. Choudhary, V., Shunko, M., & Netessine, S. (2021). Does immediate feedback make you not try as hard? A study on automotive telematics. Manufacturing and Service Operations Management, 23(4), 835–853. https://doi.org/10.1287/MSOM.2020.0924
40. Clark, T. G., Bradburn, M. J., Love, S. B., & Altman, D. G. (2003). Survival analysis part I: basic concepts and first analyses. British journal of cancer, 89(2), 232-238.
41. Cox, D. R. (1972). Regression models and life-tables. Journal of the Royal Statistical Society: Series B (Methodological), 34(2), 187-202.
42. Dahlinger, A., Tiefenbeck, V., Ryder, B., Gahr, B., Fleisch, E., & Wortmann, F. (2018). The impact of numerical vs. symbolic eco-driving feedback on fuel consumption – A randomized control field trial. Transportation Research Part D: Transport and Environment, 65, 375–386. https://doi.org/10.1016/j.trd.2018.09.013
43. de Rome L, Fitzharris M, Baldock M, Fernandes R, Ma A, Brown J. The prevalence of crash risk factors in a population-based study of motorcycle riders. Injury. 2016; 47(9):2025-33. doi: 10.1016/j.injury.2016.03.033.
44. DiClemente, C. C., Marinilli, A. S., Singh, M., & Bellino, L. E. (2001). The Role of Feedback in the Process of Health Behavior Change. American Journal of Health Behavior, 25(3), 217–227. https://doi.org/10.5993/AJHB.25.3.8
45. Dijksterhuis, C., Lewis-Evans, B., Jelijs, B., De Waard, D., Brookhuis, K., & Tucha, O. (2015). The impact of immediate or delayed feedback on driving behavior in a simulated Pay-As-You-Drive system. Accident Analysis and Prevention, 75, 93–104. https://doi.org/10.1016/j.aap.2014.11.017
46. Dingus, T. A., Neale, V. L., Klauer, S. G., Petersen, A. D., & Carroll, R. J. (2006). The development of a naturalistic data collection system to perform critical incident analysis: An investigation of safety and fatigue issues in long-haul trucking. Accident Analysis & Prevention, 38(6), 1127–1136. https://doi.org/10.1016/J.AAP.2006.05.001

47. Donmez, B., Boyle, L. N., & Lee, J. D. (2007). Safety implications of providing real-time feedback to distracted drivers. Accident Analysis and Prevention, 39(3), 581–590. https://doi.org/10.1016/j.aap.2006.10.003
48. Du, B., Zhang, C., Qu, T., Wang, Q., Spring Zhou, Q., Cui, T., Perez, P., & Astell-Burt, T. (2024). Impact of the COVID-19 pandemic on daily travel: Findings from New South Wales, Australia. Travel Behavior and Society, 36, 100798. https://doi.org/10.1016/J.TBS.2024.100798
49. Elamrani Abou Elassad, Z., Mousannif, H., Al Moatassime, H., & Karkouch, A. (2020). The application of machine learning techniques for driving behavior analysis: A conceptual framework and a systematic literature review. Engineering Applications of Artificial Intelligence, 87(March 2019), 103312. https://doi.org/10.1016/j.engappai.2019.103312
50. Ellison, A. B., Bliemer, M. C. J., & Greaves, S. P. (2015). Evaluating changes in driver behavior: A risk profiling approach. Accident Analysis and Prevention, 75, 298–309. https://doi.org/10.1016/j.aap.2014.12.018
51. ELSTAT (The Hellenic Statistical Authority) database 2018. Available at https://www.statistics.gr/en/statistics/-/publication/SDT04/2018. Accessed 1 April 2020
52. Elvik, R., Christensen, P., & Amundsen, A. (2004). Speed and road accidents. An evaluation of the Power Model. TØI report, 740, 2004
53. Elvik, R., Vadeby, A., Hels, T., & Van Schagen, I. (2019). Updated estimates of the relationship between speed and road safety at the aggregate and individual levels. Accident Analysis & Prevention, 123, 114-122.
54. Elvik. (2014). Rewarding safe and environmentally sustainable driving: Systematic review of trials. Transportation Research Record, 2465(0349), 1–7. https://doi.org/10.3141/2465-01
55. Espié S, Boubezoul A, Aupetit S, Bouaziz S. Data collection and processing tools for naturalistic study of powered two-wheelers users' behaviors. Accid Anal Prev. 2013; 58:330-9. doi: 10.1016/j.aap.2013.03.012.
56. European Commission. Traffic Safety Basic Facts on Motorcycles & Mopeds 2018, European Commission, Directorate General for Transport. Available at https://ec.europa.eu/transport/road_safety/sites/roadsafety/files/pdf/statistics/dacota/bfs201motomoped.pdf. Accessed 10 April 2021
57. Farah, H., Musicant, O., Shimshoni, Y., Toledo, T., Grimberg, E., Omer, H., & Lotan, T. (2014). Can providing feedback on driving behavior and training on parental vigilant care affect male teen drivers and their parents? Accident Analysis and Prevention, 69, 62–70. https://doi.org/10.1016/j.aap.2013.11.005
58. Farmer, C. M., Kirley, B. B., & McCartt, A. T. (2010). Effects of in-vehicle monitoring on the driving behavior of teenagers. Journal of Safety Research, 41(1), 39–45. https://doi.org/10.1016/j.jsr.2009.12.002
59. Feng, J., & Donmez, B. (2013). Designing feedback to induce safer driving behaviors: a literature review and a model of driver-feedback interaction. Human Factors.
60. Fitch, G. M., Bartholomew, P. R., Hanowski, R. J., & Perez, M. A. (2015). Drivers' visual behavior when using handheld and hands-free cell phones. Journal of safety research, 54, 105-e29.
61. Fleming, J. M., Allison, C. K., Yan, X., Lot, R., & Stanton, N. A. (2019). Adaptive driver modelling in ADAS to improve user acceptance: A study using naturalistic data. Safety Science, 119, 76–83. https://doi.org/10.1016/J.SSCI.2018.08.023
62. Frantzola, E.-K., Kontaxi, A., & Yannis, G. (2022). Impact of Road and Traffic Characteristics on Driver Behavior and Safety Using Data from Smartphones. Proceedings of the 6th International Symposium on Highway Geometric Design, Amsterdam, 26-29 June 2022.

63. Furlan, A. D., Kajaks, T., Tiong, M., Lavallì Ere G, M., Campos, J. L., Babineau, J., Haghzare, S., Ma, T., Vrkljan, B., & Furlan, A. D. (2020). Advanced vehicle technologies and road safety: A scoping review of the evidence. Accident Analysis and Prevention, 147, 105741. https://doi.org/10.1016/j.aap.2020.105741
64. Ghamari, A., Rezaei, N., Malekpour, M. R., Azadnajafabad, S., Jafari, A., Ahmadi, N., Bhalla, K., Shahraz, S., Bose, D., & Farzadfar, F. (2022). The effect of non-punitive peer comparison and performance feedback on drivers' behavior using the telematics: The first randomized trial in Iran. Journal of Safety Research, 82, 430–437. https://doi.org/10.1016/j.jsr.2022.07.010
65. Global status report on road safety 2023. Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO. (n.d.).
66. Grimberg, E., Botzer, A., & Musicant, O. (2020). Smartphones vs. in-vehicle data acquisition systems as tools for naturalistic driving studies: A comparative review. Safety Science, 131(August), 104917. https://doi.org/10.1016/j.ssci.2020.104917
67. Gu W, Zhou Z, Liu Y, Spanos CJ, Zhou Y, Zhang L. BikeSafe: Bicycle behavior monitoring via smartphones. UbiComp/ISWC 2017 - Adjun Proc 2017 ACM Int Jt Conf Pervasive Ubiquitous Comput Proc 2017 ACM Int Symp Wearable Comput. 2017:45-48. doi:10.1145/3123024.3123158
68. Gu, Y., Wang, Q., & Kamijo, S. (2019). Intelligent Driving Data Recorder in Smartphone Using Deep Neural Network-Based Speedometer and Scene Understanding. IEEE Sensors Journal, 19(1), 287–296. https://doi.org/10.1109/JSEN.2018.2874665
69. Gündüz, G., Yaman, Ç., Peker, A. U., & Acarman, T. (2018). Prediction of Risk Generated by Different Driving Patterns and Their Conflict Redistribution. IEEE Transactions on Intelligent Vehicles, 3(1), 71-80.
70. Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? - A literature review of empirical studies on gamification. Proceedings of the Annual Hawaii International Conference on System Sciences. https://doi.org/10.1109/HICSS.2014.377
71. Han, W., & Zhao, J. (2020). Driver behavior and traffic accident involvement among professional urban bus drivers in China. Transportation Research Part F: Traffic Psychology and Behavior, 74, 184–197. https://doi.org/10.1016/j.trf.2020.08.007
72. Hanowski, R. J., Perez, M. A., & Dingus, T. A. (2005). Driver distraction in long-haul truck drivers. Transportation Research Part F: Traffic Psychology and Behavior, 8(6), 441-458.
73. Haque, M. M., & Washington, S. (2014). A parametric duration model of the reaction times of drivers distracted by mobile phone conversations. Accident Analysis & Prevention, 62, 42-53.
74. Hari, D., Brace, C. J., Vagg, C., Poxon, J., & Ash, L. (2012). Analysis of a driver behavior improvement tool to reduce fuel consumption. Proceedings - 2012 International Conference on Connected Vehicles and Expo, ICCVE 2012, 208–213. https://doi.org/10.1109/ICCVE.2012.46
75. Harrell, F. E. (2015). Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis. Springer.
76. Hattie, J., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77(1), 81--112.
77. Hickman, J. S., & Hanowski, R. J. (2011). Use of a video monitoring approach to reduce at-risk driving behaviors in commercial vehicle operations. Transportation Research Part F: Traffic Psychology and Behavior, 14(3), 189–198. https://doi.org/10.1016/J.TRF.2010.11.010
78. Hosmer, D. W., Lemeshow, S., & May, S. (2008). Applied Survival Analysis: Regression Modeling of Time-to-Event Data. Wiley.
79. Hougaard, P. (2000). Analysis of Multivariate Survival Data. Springer.

80. Hsieh CH, Tsai HM, Yang SW, Lin S De. Predict scooter's stopping event using smartphone as the sensing device. Proc - 2014 IEEE Int Conf Internet Things, iThings 2014, 2014 IEEE Int Conf Green Comput Commun GreenCom 2014 2014 IEEE Int Conf Cyber-Physical-Social Comput CPS 2014. 2014; (iThings):17-23. doi:10.1109/iThings.2014.12
81. Husnjak, S., Peraković, D., Forenbacher, I., & Mumdziev, M. (2015). Telematics System in Usage Based Motor Insurance. Procedia Engineering, 100(January), 816–825. https://doi.org/10.1016/J.PROENG.2015.01.436
82. Ishwaran, H., Kogalur, U. B., Blackstone, E. H., & Lauer, M. S. (2008). Random survival forests. The Annals of Applied Statistics, 2(3), 841–860.
83. ITF, 2018. Speed and Crash Risk. OECD/ITF, Paris.
84. Jansen, R. J., & Wesseling, (2018). SWOV. Harsh Braking by Truck Drivers: A Comparison of Thresholds and Driving Contexts Using Naturalistic Driving Data. In Proceedings of the 6th Humanist Conference.
85. Johnsson, C., Laureshyn, A., & De Ceunynck, T. (2018). In search of surrogate safety indicators for vulnerable road users: a review of surrogate safety indicators. Transport Reviews, 38(6), 765-785.
86. Jöreskog, K. G. ., & Sörbom, Dag. (2001). LISREL 8 : user's reference guide. 378. https://books.google.com/books/about/LISREL_8.html?id=9AC-s50RjacC
87. Kaplan, E. L., & Meier, P. (1958). Nonparametric estimation from incomplete observations. Journal of the American statistical association, 53(282), 457-481.
88. Katrakazas, C., Michelaraki, E., Sekadakis, M., Ziakopoulos, A., Kontaxi, A., & Yannis, G. (2021). Identifying the impact of the COVID-19 pandemic on driving behavior using naturalistic driving data and time series forecasting. Journal of Safety Research, 78, 189–202. https://doi.org/10.1016/J.JSR.2021.04.007
89. Khan, I., Rizvi, S. S., Khusro, S., Ali, S., & Chung, T. S. (2021). Analyzing drivers' distractions due to smartphone usage: evidence from AutoLog dataset. Mobile Information Systems, 2021.
90. Klein, J. P., & Moeschberger, M. L. (2003). Survival Analysis: Techniques for Censored and Truncated Data. Springer.
91. Koesdwiady, A., Soua, R., Karray, F., & Kamel, M. S. (2017). Recent Trends in Driver Safety Monitoring Systems: State of the Art and Challenges. IEEE Transactions on Vehicular Technology, 66(6), 4550–4563. https://doi.org/10.1109/TVT.2016.2631604
92. Kontaxi A., Ziakopoulos A., Yannis G. (2025) "Examining the Impact of Feedback on Traffic and Safety Behavior of Car Drivers in a Naturalistic Driving Study" Proceedings of the Transportation Research Board (TRB) 104th Annual Meeting, Washington, 5-9 January 2025
93. Kontaxi, A., Tzoutzoulis, D. M., Ziakopoulos, A., & Yannis, G. (2023). Exploring speeding behavior using naturalistic car driving data from smartphones. Journal of Traffic and Transportation Engineering (English Edition), 10(6), 1162–1173. https://doi.org/10.1016/J.JTTE.2023.07.007
94. Kontaxi A., Ziakopoulos A., Yannis G. (2022) "Discovering the influence of feedback on driver behavior through a multiphase experiment based on a smartphone application" Proceedings of 8th Road Safety & Simulation International Conference, Athens, Greece, 08-10 June, 2022
95. Kontaxi, A., Ziakopoulos, A., & Yannis, G. (2021a). Investigation of the speeding behavior of motorcyclists through an innovative smartphone application. Traffic Injury Prevention, 22(6), 460–466. https://doi.org/10.1080/15389588.2021.1927002

96. Kontaxi, A., Ziakopoulos, A., & Yannis, G. (2021b). Trip characteristics impact on the frequency of harsh events recorded via smartphone sensors. IATSS Research, 45(4), 574–583. https://doi.org/10.1016/J.IATSSR.2021.07.004
97. Kontaxi A., Ziakopoulos A., Papantoniou P., Yannis G., Kostoulas G. (2021). "Monitoring and Improving Driving Behavior of Motorcyclists Through an Innovative Smartphone Application", Proceedings of 7th Humanist Conference, Rhodes Island, Greece, 25-26 October 2021
98. Kontaxi A., Frantzola E., Ziakopoulos A., Kostoulas G., Yannis G. (2021) "Investigation of speeding and aggressive behavior of professional drivers on highways through an innovative smartphone application", Proceedings of the 10th International Congress on Transportation Research, Rhodes Island, Greece, 1-3 September 2021
99. Koppel, S., Bugeja, L., Hua, P., Osborne, R., Stephens, A. N., Young, K. L., Chambers, R., & Hassed, C. (2019). Do mindfulness interventions improve road safety? A systematic review. Accident Analysis and Prevention, 123(August 2018), 88–98. https://doi.org/10.1016/j.aap.2018.11.013
100. Kujala, T., Grahn, H., Mäkelä, J., Silvennoinen, J., & Tokkonen, T. (2024). Effects of context-sensitive distraction warnings on drivers' smartphone use and acceptance: A long-term naturalistic field study. https://doi.org/10.1016/j.ijhcs.2024.103247
101. Lattanzi, E., & Freschi, V. (2021). Machine Learning Techniques to Identify Unsafe Driving Behavior by Means of In-Vehicle Sensor Data. Expert Systems with Applications, 176, 114818. https://doi.org/10.1016/J.ESWA.2021.114818
102. Leik, R. K. (1997). Experimental Design and Analysis INTRODUCTION AND OVERVIEW. Experimental Design and Analysis, 1–7. https://books.google.com/books/about/Experimental_Design_and_Analysis.html?id=INggp-t2BNQC
103. Li, X., Oviedo-Trespalacios, O., Rakotonirainy, A., & Yan, X. (2019). Collision risk management of cognitively distracted drivers in a car-following situation. Transportation research part F: traffic psychology and behavior, 60, 288-298.
104. Lin MR, Kraus JF. A review of risk factors and patterns of motorcycle injuries. Accid Anal Prev. 2009 Jul; 41(4):710-22. doi: 10.1016/j.aap.2009.03.010.
105. London, M., & Smither, J. W. (2002). Feedback orientation, feedback culture, and the longitudinal performance management process. Human Resource Management Review, 12(1), 81–100. https://doi.org/10.1016/S1053-4822(01)00043-2
106. Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. Transportation research part A: policy and practice, 44(5), 291-305.
107. Machin, D., Cheung, Y. B., & Parmar, M. (2006). Survival analysis: a practical approach. John Wiley & Sons.
108. Mantouka, E. G., Fafoutellis, P., & Vlahogianni, E. I. (2021). Deep survival analysis of searching for on-street parking in urban areas. Transportation Research Part C: Emerging Technologies, 128, 103173.
109. Mantouka, E. G., Barmpounakis, E. N., & Vlahogianni, E. I. (2019). Identifying driving safety profiles from smartphone data using unsupervised learning. Safety Science, 119(January), 84–90. https://doi.org/10.1016/j.ssci.2019.01.025
110. Mantouka, E. G., Barmpounakis, E. N., Milioti, C. P., & Vlahogianni, E. I. (2019). Gamification in mobile applications: The case of airports. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, 23(5), 417–426. https://doi.org/10.1080/15472450.2018.1473157

111. Mazureck, U., & Hattem, J. van. (2006). Rewards for Safe Driving Behavior. Https://Doi.Org/10.1177/0361198106198000106, 1980(1), 31–38. https://doi.org/10.1177/0361198106198000106
112. McCulloch, C. E. (2003). Generalized linear mixed models. In NSF-CBMS regional conference series in probability and statistics (pp. i-84). Institute of Mathematical Statistics and the American Statistical Association.
113. McDonnell, K., Murphy, F., Sheehan, B., Masello, L., Castignani, G., & Ryan, C. (2021). Regulatory and Technical Constraints: An Overview of the Technical Possibilities and Regulatory Limitations of Vehicle Telematic Data. Sensors 2021, Vol. 21, Page 3517, 21(10), 3517. https://doi.org/10.3390/S21103517
114. McGehee, D. V., Raby, M., Carney, C., Lee, J. D., & Reyes, M. L. (2007). Extending parental mentoring using an event-triggered video intervention in rural teen drivers. Journal of Safety Research, 38(2), 215–227. https://doi.org/10.1016/J.JSR.2007.02.009
115. Merrikhpour, M., Donmez, B., & Battista, V. (2014). A field operational trial evaluating a feedback-reward system on speeding and tailgating behaviors. Transportation Research Part F: Traffic Psychology and Behavior, 27(PA), 56–68. https://doi.org/10.1016/j.trf.2014.09.002
116. Meuleners, L., Fraser, M., Stevenson, M., & Roberts, P. (2023). Personalized driving safety: Using telematics to reduce risky driving behavior among young drivers. Journal of Safety Research, 86, 164–173. https://doi.org/10.1016/J.JSR.2023.05.007
117. Michelaraki E., Kontaxi A., Papantoniou P., Yannis G., "Correlation of driver behavior and fuel consumption using data from smartphones", Proceedings of the 8th Transport Research Arena TRA 2020 Conference (Helsinki, Finland, 27-30 April 2020)
118. Michelaraki, E., Katrakazas, C., Yannis, G., Filtness, A., Talbot, R., Hancox, G., Pilkington-Cheney, F., Brijs, K., Ross, V., Dirix, H., Neven, A., Paul, R., Brijs, T., Fortsakis, P., Konstantina Frantzola, E., & Taveira, R. (2021). Post-trip safety interventions: State-of-the-art, challenges, and practical implications. https://doi.org/10.1016/j.jsr.2021.02.005
119. Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberatî, A., Petticrew, M., Shekelle, P., Stewart, L. A., & Group, P.-P. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. https://doi.org/10.1186/2046-4053-4-1
120. Molloy, O., Molesworth, B. R. C., & Williamson, A. (2018). Improving young drivers' speed management behavior through feedback: A cognitive training intervention. Transportation Research Part F: Traffic Psychology and Behavior, 54, 324–337. https://doi.org/10.1016/j.trf.2018.02.010
121. Molloy, O., Molesworth, B., Williamson, A., & Senserrick, T. (2023). Improving young drivers' speed compliance through a single dose of feedback. Transportation Research Part F: Traffic Psychology and Behavior, 95, 228–238. https://doi.org/10.1016/J.TRF.2023.04.006
122. Moujahid, A., Dulva Hina, M., Soukane, A., Ortalda, A., Di Torino, P., Mounir, I., Tantaoui, E. A., Khadimi, A. El, & Ramdane-Cherif, A. (n.d.). Machine Learning Techniques in ADAS: A Review.
123. Mullen, N. W., Maxwell, H., & Bédard, M. (2015). Decreasing driver speeding with feedback and a token economy. Transportation Research Part F: Traffic Psychology and Behavior, 28, 77–85. https://doi.org/10.1016/j.trf.2014.11.008
124. Nambi, A. U., Bannur, S., Mehta, I., Kalra, H., Virmani, A., Padmanabhan, V. N. & Raman, B. (2018, October). Hams: Driver and driving monitoring using a smartphone. In Proceedings of the 24th Annual International Conference on Mobile Computing and Networking (pp. 840-842).

125. Newnam, S., Lewis, I., & Warmerdam, A. (2014). Modifying behavior to reduce over-speeding in work-related drivers: An objective approach. Accident Analysis and Prevention, 64, 23–29. https://doi.org/10.1016/j.aap.2013.10.032
126. Nichols, A. L., & Maner, J. K. (2008). The Good-Subject Effect: Investigating Participant Demand Characteristics. The Journal of General Psychology, 135(2), 151–166. https://doi.org/10.3200/GENP.135.2.151-166
127. Nikolaou, D., Ziakopoulos, A., & Yannis, G. (2023). A Review of Surrogate Safety Measures Uses in Historical Crash Investigations. Sustainability 2023, Vol. 15, Page 7580, 15(9), 7580. https://doi.org/10.3390/SU15097580
128. Nilsson, G. (2004). Traffic safety dimensions and the power model to describe the effect of speed on safety (Doctoral dissertation, Univ.).
129. Niu, J., Wang, X., Liu, X., Wang, D., Qin, H., & Zhang, Y. (2019). Effects of mobile phone use on driving performance in a multiresource workload scenario. Traffic injury prevention, 20(1), 37-44.
130. O'Connor, S. S., Shain, L. M., Whitehill, J. M., & Ebel, B. E. (2017). Measuring a conceptual model of the relationship between compulsive cell phone use, in-vehicle cell phone use, and motor vehicle crash. Accident Analysis & Prevention, 99, 372-378.
131. Orlovska, J., Novakazi, F., Lars-Ola, B., Karlsson, M. A., Wickman, C., & Söderberg, R. (2020). Effects of the driving context on the usage of Automated Driver Assistance Systems (ADAS) -Naturalistic Driving Study for ADAS evaluation. Transportation Research Interdisciplinary Perspectives, 4, 100093. https://doi.org/10.1016/J.TRIP.2020.100093
132. Otte D, Jänsch M, Haasper C. Injury protection and accident causation parameters for vulnerable road users based on German In-Depth Accident Study GIDAS. Accid Anal Prev. 2012 Jan; 44(1):149-53. doi: 10.1016/j.aap.2010.12.006.
133. Paefgen, J., Staake, T., & Thiesse, F. (2013). Evaluation and aggregation of pay-as-you-drive insurance rate factors: A classification analysis approach. Decision Support Systems, 56, 192-201.
134. Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Systematic Reviews, 10(1), 1–11. https://doi.org/10.1186/S13643-021-01626-4/FIGURES/1
135. Papadimitriou, E., Argyropoulou, A., Tselentis, D. I., & Yannis, G. (2019). Analysis of driver behavior through smartphone data: The case of mobile phone use while driving. Safety Science, 119, 91–97. https://doi.org/10.1016/J.SSCI.2019.05.059
136. Papantoniou, P., Antoniou, C., Yannis, G., & Pavlou, D. (2019). Which factors affect accident probability at unexpected incidents? A structural equation model approach. Journal of Transportation Safety & Security, 11(5), 544–561. https://doi.org/10.1080/19439962.2018.1447523
137. Papantoniou, P., Kontaxi, A., Yannis, G., & Fortsakis, P. (2021). Investigating the Correlation of Mobile Phone Use with Trip Characteristics Recorded Through Smartphone Sensors. Advances in Intelligent Systems and Computing, 1278, 935–944. https://doi.org/10.1007/978-3-030-61075-3_90
138. Papantoniou, P., Papadimitriou, E., & Yannis, G. (2015). Assessment of driving simulator studies on driver distraction. Advances in Transportation Studies, 35, 129–144.

139. Parmet, Y., Meir, A., & Borowsky, A. (2014). Survival analysis: A fresh approach for analyzing response times in driving-related hazard perception paradigms. Transportation Research Part F: Traffic Psychology and Behavior, 25, 98-107.
140. Paruchuri, V., & Kumar, A. (2015). Detecting driver distraction using smartphones. In 2015 IEEE 29th International Conference on Advanced Information Networking and Applications (pp. 468-475). IEEE.
141. Peer, S., Muermann, A., & Sallinger, K. (2020). App-based feedback on safety to novice drivers: Learning and monetary incentives. Transportation Research Part F: Traffic Psychology and Behavior, 71, 198–219. https://doi.org/10.1016/j.trf.2020.04.005
142. Pérez, A., García, M. I., Nieto, M., Pedraza, J. L., Rodríguez, S., & Zamorano, J. (2010). Argos: An advanced in-vehicle data recorder on a massively sensorized vehicle for car driver behavior experimentation. IEEE Transactions on Intelligent Transportation Systems, 11(2), 463–473. https://doi.org/10.1109/TITS.2010.2046323
143. Petraki V., Ziakopoulos A., Yannis G., (2020) "Combined impact of road and traffic characteristic on driver behavior using smartphone sensor data". Accident Analysis & Prevention, Vol. 144.
144. Pozueco, L., Rionda, A., Pañeda, A. G., Sánchez, J. A., Pañeda, X. G., García, R., Melendi, D., & Tuero, A. G. (2017). Impact of on-board tutoring systems to improve driving efficiency of non-professional drivers. IET Intelligent Transport Systems, 11(4), 196–202. https://doi.org/10.1049/iet-its.2016.0079
145. Precht, L., Keinath, A., & Krems, J. F. (2017). Effects of driving anger on driver behavior – Results from naturalistic driving data. Transportation Research Part F: Traffic Psychology and Behavior, 45, 75–92. https://doi.org/10.1016/J.TRF.2016.10.019
146. Professional drivers | Mobility and transport. (n.d.). Retrieved May 12, 2021, from https://ec.europa.eu/transport/road_safety/users/professional-drivers_en
147. Ram, T., & Chand, K. (2016). Effect of drivers' risk perception and perception of driving tasks on road safety attitude. Transportation Research Part F: Traffic Psychology and Behavior, 42, 162–176. https://doi.org/10.1016/J.TRF.2016.07.012
148. Reagan, I. J., Bliss, J. P., Van Houten, R., & Hilton, B. W. (2013). The effects of external motivation and real-time automated feedback on speeding behavior in a naturalistic setting. Human Factors, 55(1), 218–230. https://doi.org/10.1177/0018720812447812
149. Reimers, I., & Shiller, B. R. (2019). The impacts of telematics on competition and consumer behavior in insurance. Journal of Law and Economics, 62(4), 613–632. https://doi.org/10.1086/705119
150. Richard, C. M., Lee, J., Atkins, R., & Brown, J. L. (2020). Using SHRP2 naturalistic driving data to examine driver speeding behavior. Journal of Safety Research, 73, 271–281. https://doi.org/10.1016/j.jsr.2020.03.008
151. Richard, C., Campbell, J., Brown, J., Lichty, M., Chrysler, S., & Atkins, R. (2013). Investigating speeding behavior with naturalistic approaches. Transportation Research Record, 1(2365), 58–65. https://doi.org/10.3141/2365-08
152. Rios-Torres, J., & Malikopoulos, A. (2016). An overview of driver feedback systems for efficiency and safety. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 667–674. https://doi.org/10.1109/ITSC.2016.7795625
153. Risteska, M., Kanaan, D., Donmez, B., & Chen, H. Y. W. (2021). The effect of driving demands on distraction engagement and glance behaviors: Results from naturalistic data. Safety science, 136, 105123.

154. Roidl, E., Frehse, B., & Höger, R. (2014). Emotional states of drivers and the impact on speed, acceleration and traffic violations A simulator study. Accident Analysis & Prevention, 70, 282-292.
155. Rolim, C., Baptista, P., Duarte, G., Farias, T., & Pereira, J. (2016). Impacts of delayed feedback on eco-driving behavior and resulting environmental performance changes. Transportation Research Part F: Traffic Psychology and Behavior, 43, 366–378. https://doi.org/10.1016/j.trf.2016.09.003
156. Rossetti, R. J. F., Almeida, J. E., Kokkinogenis, Z., & Goncalves, J. (2013). Playing transportation seriously: Applications of serious games to artificial transportation systems. IEEE Intelligent Systems, 28(4), 107–113. https://doi.org/10.1109/MIS.2013.113
157. Roussou, S., Garefalakis, T., Michelaraki, E., Katrakazas, C., Adnan, M., Khattak, M. W., Brijs, T., & Yannis, G. (2023). Examination of the Effect of Task Complexity and Coping Capacity on Driving Risk: A Cross-Country and Transportation Mode Comparative Study. Sensors 2023, Vol. 23, Page 9663, 23(24), 9663. https://doi.org/10.3390/S23249663
158. Saifuzzaman, M., Haque, M. M., Zheng, Z., & Washington, S. (2015). Impact of mobile phone use on car-following behavior of young drivers. Accident Analysis & Prevention, 82, 10-19.
159. Samani, A. R., & Mishra, S. (2024). How long the effect of take-over conditions Lasts? a survival analysis of Commercial Motor vehicle drivers' reaction time and driving behavior in Level 4 of automated vehicles. Transportation Research Part F: Traffic Psychology and Behavior, 107, 149-166.
160. Shabeer, H. A., & Wahidabanu, R. S. D. (2012). Averting mobile phone use while driving and technique to locate the mobile phone used vehicle. Procedia engineering, 30, 623-630.
161. Shahverdy, M., Fathy, M., Berangi, R., & Sabokrou, M. (2020). Driver behavior detection and classification using deep convolutional neural networks. Expert Systems with Applications, 149, 113240. https://doi.org/10.1016/J.ESWA.2020.113240
162. Shangguan, Q., Fu, T., & Liu, S. (2020). Investigating rear-end collision avoidance behavior under varied foggy weather conditions: A study using advanced driving simulator and survival analysis. Accident Analysis & Prevention, 139, 105499.
163. Sharma, A., Zheng, Z., Kim, J., Bhaskar, A., & Haque, M. M. (2020). Is an informed driver a better decision maker? A grouped random parameters with heterogeneity-in-means approach to investigate the impact of the connected environment on driving behavior in safety-critical situations. *Analytic methods in accident research*, *27*, 100127.
164. Shi, X., Wong, Y. D., Li, M. Z. F., Palanisamy, C., & Chai, C. (2019). A feature learning approach based on XGBoost for driving assessment and risk prediction. Accident Analysis & Prevention, 129, 170–179. https://doi.org/10.1016/J.AAP.2019.05.005
165. Singh, H., & Kathuria, A. (2021a). Analyzing driver behavior under naturalistic driving conditions: A review. Accident Analysis and Prevention, 150. https://doi.org/10.1016/J.AAP.2020.105908
166. Singh, S. (2015). Critical Reasons for Crashes Investigated in the National Motor Vehicle Crash Causation Survey. DOT HS 812 115.
167. Soleymanian, M., Weinberg, C. B., & Zhu, T. (2019). Sensor data and behavioral tracking: Does usage-based auto insurance benefit drivers? Marketing Science, 38(1), 21–43. https://doi.org/10.1287/mksc.2018.1126
168. Soriguera F, Miralles E. Driver feedback mobile app. Transp. Res. Procedia. 2016; 18:264-71. doi:10.1016/j.trpro.2016.12.036
169. Stanglmayr M, Bäumler M. 2020. Towards Safer Rides: Measuring Motorcycle Dynamics with Smartphones. Paper presented at the 13th International Motorcycle Conference; October 5-6, Cologne, Germany. DOI: 10.26128/2020.

170. Stephens, A. N., & Groeger, J. A. (2009). Situational specificity of trait influences on drivers' evaluations and driving behavior. Transportation Research Part F: Traffic Psychology and Behavior, 12(1), 29–39. https://doi.org/10.1016/J.TRF.2008.06.005
171. Stevenson, M., Harris, A., Wijnands, J. S., & Mortimer, D. (2021). The effect of telematic based feedback and financial incentives on driving behavior: A randomised trial. Accident Analysis and Prevention, 159, 106278. https://doi.org/10.1016/j.aap.2021.106278
172. Streiffer, C., Raghavendra, R., Benson, T., & Srivatsa, M. (2017). Darnet: A deep learning solution for distracted driving detection. In Proceedings of the 18th acm/ifip/usenix middleware conference: Industrial track (pp. 22-28).
173. Strömberg, H. K., & Karlsson, I. C. M. A. (2013). Comparative effects of eco-driving initiatives aimed at urban bus drivers - Results from a field trial. Transportation Research Part D: Transport and Environment, 22, 28–33. https://doi.org/10.1016/j.trd.2013.02.011
174. Su, Z., Woodman, R., Smyth, J., & Elliott, M. (2023). The relationship between aggressive driving and driver performance: A systematic review with meta-analysis. Accident Analysis & Prevention, 183, 106972. https://doi.org/10.1016/J.AAP.2023.106972
175. Sullman, M. (2020). Using in-Vehicle Data Monitors to Reduce Risky Driving among Sales Representatives. Policy and Practice in Health and Safety, 18(1), 55–64. https://doi.org/10.1080/14773996.2019.1667094
176. Takeda, K., Miyajima, C., Suzuki, T., Angkititrakul, P., Kurumida, K., Kuroyanagi, Y., Ishikawa, H., Terashima, R., Wakita, T., Oikawa, M., & Komada, Y. (2012). Self-coaching system based on recorded driving data: Learning from one's experiences. IEEE Transactions on Intelligent Transportation Systems, 13(4), 1821–1831. https://doi.org/10.1109/TITS.2012.2205917
177. Takeda, K., Miyajima, C., Suzuki, T., Kurumida, K., Kuroyanagi, Y., Ishikawa, H., Angkititrakul, P., Terashima, R., Wakita, T., Oikawa, M., & Komada, Y. (2011). Improving driving behavior by allowing drivers to browse their own recorded driving data. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, 44–49. https://doi.org/10.1109/ITSC.2011.6082893
178. Tarko, A. P. (2018). Surrogate measures of safety. In Safe mobility: challenges, methodology and solutions. Emerald Publishing Limited.
179. Theofilatos A, Yannis G. A review of powered-two-wheeler behavior and safety. Int J Inj Contr Saf Promot. 2015; 22(4):284-307. doi: 10.1080/17457300.2014.908224.
180. Theofilatos A, Ziakopoulos A. Examining Injury Severity of Moped and Motorcycle Occupants with Real-Time Traffic and Weather Data. J Transp Eng Part A Syst. 2018; 144(11):04018066. doi:10.1061/jtepbs.0000193
181. Thurlings, M., Vermeulen, M., Bastiaens, T., & Stijnen, S. (2013). Understanding feedback: A learning theory perspective. Educational Research Review, 9, 1–15. https://doi.org/10.1016/J.EDUREV.2012.11.004
182. Tivesten, E., & Dozza, M. (2015). Driving context influences drivers' decision to engage in visual–manual phone tasks: Evidence from a naturalistic driving study. Journal of safety research, 53, 87-96.
183. Toledo, G., & Shiftan, Y. (2016). Can feedback from in-vehicle data recorders improve driver behavior and reduce fuel consumption? Transportation Research Part A: Policy and Practice, 94, 194–204. https://doi.org/10.1016/j.tra.2016.09.001
184. Toledo, T., & Lotan, T. (2006). In-vehicle data recorder for evaluation of driving behavior and safety. Transportation Research Record, 1953, 112–119. https://doi.org/10.3141/1953-13

185. Toledo, T., Musicant, O., & Lotan, T. (2008). In-vehicle data recorders for monitoring and feedback on drivers' behavior. Transportation Research Part C: Emerging Technologies, 16(3), 320–331. https://doi.org/10.1016/j.trc.2008.01.001
186. Tselentis DI, Vlahogianni EI, Yannis G. Driving safety efficiency benchmarking using smartphone data. Transp Res Part C Emerg Technol. 2019. doi:10.1016/j.trc.2019.11.006
187. Tselentis, D. I., Yannis, G., & Vlahogianni, E. I. (2017). Innovative motor insurance schemes: A review of current practices and emerging challenges. Accident Analysis and Prevention, 98, 139–148. https://doi.org/10.1016/j.aap.2016.10.006
188. Tselentis, D., Kostoulas, G., Kontonasios, K.-N., Fortsakis, P., & Yannis, G. Driver feedback during naturalistic driving experiments: A review of types, methods and future challenges. Paper presented at Transport Research Arena Conference (TRA); April 27-30, 2020; Helsinki, Finland. (Conference cancelled)
189. Uddin, M., & Huynh, N. (2017). Truck-involved crashes injury severity analysis for different lighting conditions on rural and urban roadways. Accident Analysis and Prevention, 108, 44–55. https://doi.org/10.1016/j.aap.2017.08.009
190. Vaiana, R., T. Iuele, V. Astarita, M. V. Caruso, A. Tassitani, C. Zaffino, and V. P. Giofrè. Driving Behavior and Traffic Safety: An Acceleration-Based Safety Evaluation Procedure for Smartphones. Modern Applied Science. 2014 Feb 1;8(1):88.
191. van Schagen, I., & Sagberg, F. (2012). The Potential Benefits of Naturalistic Driving for Road Safety Research: Theoretical and Empirical Considerations and Challenges for the Future. Procedia - Social and Behavioral Sciences, 48, 692–701. https://doi.org/10.1016/J.SBSPRO.2012.06.1047
192. Vlahogianni EI, Yannis G, Golias JC. Overview of critical risk factors in Power-Two-Wheeler safety. Accid Anal Prev. 2012 Nov; 49:12-22. doi: 10.1016/j.aap.2012.04.009.
193. Wang, J., Chai, W., Venkatachalapathy, A., Tan, K. L., Haghighat, A., Velipasalar, S., Adu-Gyamfi, Y., & Sharma, A. (2022). A Survey on Driver Behavior Analysis From In-Vehicle Cameras. IEEE Transactions on Intelligent Transportation Systems, 23(8), 10186–10209. https://doi.org/10.1109/TITS.2021.3126231
194. Wang, X., Xu, R., Zhang, S., Zhuang, Y., & Wang, Y. (2022). Driver distraction detection based on vehicle dynamics using naturalistic driving data. https://doi.org/10.1016/j.trc.2022.103561
195. Wang, Y. (2006). Estimation of accelerated failure time models with random effects. Iowa State University.
196. Wang, Y., Zhang, W., & Abdel-Aty, M. (2021). Exploring driving volatility and crash propensity using parametric survival analysis. Safety Science, 134, 105070.
197. Washington, S., Karlaftis, M. G., Mannering, F., & Anastasopoulos, P. (2020). Statistical and econometric methods for transportation data analysis. Chapman and Hall/CRC.
198. Wesley, A., Shastri, D., & Pavlidis, I. (2010). A novel method to monitor driver's distractions. In CHI'10 Extended Abstracts on Human Factors in Computing Systems (pp. 4273-4278).
199. Wienke, A. (2011). Frailty Models in Survival Analysis. Chapman and Hall/CRC.
200. Wijayaratna, K. P., Cunningham, M. L., Regan, M. A., Jian, S., Chand, S., & Dixit, V. V. (2019). Mobile phone conversation distraction: Understanding differences in impact between simulator and naturalistic driving studies. Accident Analysis & Prevention, 129, 108–118. https://doi.org/10.1016/J.AAP.2019.04.017
201. Will, S., Metz, B., Hammer, T., Mörbe, M., Henzler, M., Harnischmacher, F., & Matschl, G. (2020). Methodological considerations regarding motorcycle naturalistic riding investigations based on the use of g-g diagrams for rider profile detection. Safety Science, 129(April), 104840. https://doi.org/10.1016/j.ssci.2020.104840

202. Williams, A. F. (2003). Teenage drivers: patterns of risk. Journal of Safety Research, 34(1), 5–15. https://doi.org/10.1016/S0022-4375(02)00075-0
203. Williams, V., Mclaughlin, S., & Atwood, J. (2016). Factors that Increase and Decrease Motorcyclist Crash Risk Organizational Mission. 1–20.
204. Wouters, P. I. J., & Bos, J. M. J. (2000). Traffic accident reduction by monitoring driver behavior with in-car data recorders. Accident Analysis & Prevention, 32(5), 643–650. https://doi.org/10.1016/S0001-4575(99)00095-0
205. Wu, K. F. (Ken), & Wang, L. (2021). Exploring the combined effects of driving situations on freeway rear-end crash risk using naturalistic driving study data. Accident Analysis & Prevention, 150, 105866. https://doi.org/10.1016/J.AAP.2020.105866
206. Yang, M., Bao, Q., Shen, Y., & Qu, Q. (2023). Thirty years of research on driving behavior active intervention: A bibliometric overview. Journal of Traffic and Transportation Engineering (English Edition), 10(5), 721–742. https://doi.org/10.1016/J.JTTE.2023.06.002
207. Yuan, Y., Yang, M., Guo, Y., Rasouli, S., Gan, Z., & Ren, Y. (2021). Risk factors associated with truck-involved fatal crash severity: Analyzing their impact for different groups of truck drivers. Journal of Safety Research, 76, 154–165. https://doi.org/10.1016/j.jsr.2020.12.012
208. Zhang, X., Wang, X., Bao, Y., & Zhu, X. (2022). Safety assessment of trucks based on GPS and in-vehicle monitoring data. Accident Analysis and Prevention, 168. https://doi.org/10.1016/j.aap.2022.106619
209. Zhao, G., & Wu, C. (2012). The effects of driver identity on driving safety in a retrospective feedback system. Accident Analysis and Prevention, 45, 354–365. https://doi.org/10.1016/j.aap.2011.08.002
210. Zhao, X., Xu, W., Ma, J., Li, H., & Chen, Y. (2019). An analysis of the relationship between driver characteristics and driving safety using structural equation models. Transportation Research Part F: Traffic Psychology and Behavior, 62, 529–545. https://doi.org/10.1016/J.TRF.2019.02.004
211. Ziakopoulos A, Theofilatos A, Yannis G, Margaritis D, Thomas P, Morris A, Brown L, Robibaro M, Usami DS, Phan V, Davidse R. A preliminary analysis of in-depth accident data for powered two-wheelers and bicycles in Europe. Paper presented at the International Research Council on the Biomechanics of Impacts (IRCOBI), September 12, 2018, Athens, Greece.
212. Ziakopoulos, A., Kontaxi, A., & Yannis, G. (2023). Analysis of mobile phone use engagement during naturalistic driving through explainable imbalanced machine learning. Accident Analysis and Prevention, 181(July 2022), 106936. https://doi.org/10.1016/j.aap.2022.106936
213. Ziakopoulos, A., Petraki, V., Kontaxi, A., & Yannis, G. (2022b). The transformation of the insurance industry and road safety by driver safety behavior telematics. https://doi.org/10.1016/j.cstp.2022.10.011
214. Ziakopoulos, A., Theofilatos, A., Papadimitriou, E., & Yannis, G. (2019). A meta-analysis of the impacts of operating in-vehicle information systems on road safety. IATSS Research, 43(3), 185–194. https://doi.org/10.1016/J.IATSSR.2019.01.003
215. Ziakopoulos, A., Tselentis, D., Kontaxi, A., & Yannis, G. (2020). A critical overview of driver recording tools. Journal of Safety Research, 72, 203–212. https://doi.org/10.1016/J.JSR.2019.12.021
216. Ziakopoulos, A., Vlahogianni, E., Antoniou, C., & Yannis, G. (2022a). Spatial predictions of harsh driving events using statistical and machine learning methods. Safety Science, 150, 105722. https://doi.org/10.1016/J.SSCI.2022.105722
217. Zöller, I., Abendroth, B., & Bruder, R. (2019). Driver behavior validity in driving simulators – Analysis of the moment of initiation of braking at urban intersections. Transportation

Research Part F: Traffic Psychology and Behavior, 61, 120–130. https://doi.org/10.1016/J.TRF.2017.09.008

Appendix I

List of publications produced within the framework of this dissertation

Publications in scientific journals with peer review

1. Kontaxi A., Yannis G. (2025). "Post-feedback effect on long-term driver behavior - A Survival Analysis Approach" (under review)
2. Kontaxi A., Ziakopoulos A., Yannis G. (2025). "Exploring the impact of driver feedback on safety: A systematic review of naturalistic driving studies" Journal of TRF: Traffic Psychology and Behavior (under review)
3. Kontaxi A., Ziakopoulos A., Yannis G. (2025). "Impacts of Smartphone-Based Feedback on Driver Behavior: Findings from a Multiphase Naturalistic Study" Travel Behavior and Society (under review)
4. Kontaxi, A., Ziakopoulos, A., & Yannis, G. (2021b). Trip characteristics impact on the frequency of harsh events recorded via smartphone sensors. IATSS Research, 45(4), 574–583. https://doi.org/10.1016/J.IATSSR.2021.07.004
5. Kontaxi, A., Ziakopoulos, A., & Yannis, G. (2021a). Investigation of the speeding behavior of motorcyclists through an innovative smartphone application. Traffic Injury Prevention, 22(6), 460–466. https://doi.org/10.1080/15389588.2021.1927002/

Publications in scientific conference proceedings (full papers with review)

1. Kontaxi A., Ziakopoulos A., Yannis G. (2025) "Examining the Impact of Feedback on Traffic and Safety Behavior of Car Drivers in a Naturalistic Driving Study" Proceedings of the Transportation Research Board (TRB) 104th Annual Meeting, Washington, 5-9 January 2025
2. Kontaxi A., Ziakopoulos A., Yannis G. (2022) "Discovering the influence of feedback on driver behavior through a multiphase experiment based on a smartphone application" Proceedings of 8th Road Safety & Simulation International Conference, Athens, Greece, 08-10 June, 2022
3. Kontaxi A., Ziakopoulos A., Papantoniou P., Yannis G., Kostoulas G. (2021). "Monitoring and Improving Driving Behavior of Motorcyclists Through an Innovative Smartphone Application", Proceedings of 7th Humanist Conference, Rhodes Island, Greece, 25-26 October 2021
4. Kontaxi A., Frantzola E., Ziakopoulos A., Kostoulas G., Yannis G. (2021) "Investigation of speeding and aggressive behavior of professional drivers on highways through an innovative smartphone application", Proceedings of the 10th International Congress on Transportation Research, Rhodes Island, Greece, 1-3 September 2021

List of additional publications in other research thematic areas

Publications in scientific journals with peer review

1. Nikolaou, D., Ziakopoulos, A., Kontaxi, A., Theofilatos, A., & Yannis, G. (2025). Spatial analysis of telematics-based surrogate safety measures. *Journal of Safety Research*, *92*, 98-108.
2. Roussou S., Petraki V., Deliali K., Kontaxi A., Yannis G. (2024) “Cost benefit analysis of reducing speed limits in Athens to 30 Km/h”, Case Studies on Transport Policy, 18, 101289
3. Kontaxi, A., Tzoutzoulis, D. M., Ziakopoulos, A., & Yannis, G. (2023). Exploring speeding behavior using naturalistic car driving data from smartphones. Journal of Traffic and Transportation Engineering (English Edition), 10(6), 1162–1173.
4. Ziakopoulos A., Kontaxi A., Yannis G. (2023) “Analysis of mobile phone use engagement during naturalistic driving through explainable imbalanced machine learning”, Accident Analysis and Prevention, Volume 181
5. Ziakopoulos A., Petraki V., Kontaxi A., Yannis G., (2022) “The transformation of the insurance industry and road safety by driver safety behavior telematics” Case Studies on Transport Policy
6. Katrakazas C., Michelaraki E., Sekadakis M., Ziakopoulos A., Kontaxi A., Yannis G. (2021). “Identifying the impact of the COVID-19 pandemic on Driving Behavior using naturalistic driving data and time series forecasting” Journal of Safety Research
7. Papantoniou P., Kontaxi A., Yannis G., Fortsakis P. (2020) “Investigating the Correlation of Mobile Phone Use with Trip Characteristics Recorded Through Smartphone Sensors”, Advances in Mobility-as-a-Service Systems.
8. Ziakopoulos A., Tselentis D.I., Kontaxi A., Yannis G. (2020). “A critical overview of driver recording tools.” Journal of Safety Research – Special Issue: Human Factors.

Publications in scientific conference proceedings (full papers with review)

1. Kontaxi A., Triantafyllou A., Yannis, G. “Investigation of the Impact of Eco-driving on Fuel Consumption Using Smartphone Data”, Proceedings of the 30th Intelligent Transportation Systems (ITS) World Congress, Dubai, 16-20 September 2024
2. Nikolaou D., Kontaxi A., Ziakopoulos, Yannis G., Fortsakis P., Frantzola E.K., Sigalos K., Kouridakis G. “Naturalistic Spatial Road Safety Analysis: The SmartMaps Project”, Proceedings of the Transport Research Arena TRA 2024 Conference, Dublin, 15-18 April 2024
3. Kontaxi A., Agourou C., Yannis G., “Analyzing Acceptance of Reduced Speed Limits on Greek Motorways: A Survey”, Proceedings of the Transport Research Arena TRA 2024 Conference, Dublin, 15-18 April 2024
4. Makrydakis K., Petraki V., Kontaxi A., Yannis G., “Cost Benefit Analysis of reducing speed limits at the Greek interurban road network”, Proceedings of the 11th International Congress on Transportation Research (Heraklion, Greece, 20-22 September 2023)
5. Nikolaou D., Kontaxi A., Ziakopoulos A., Yannis G., “Spatial analysis of telematics surrogate safety measures across road environments”, Proceedings of the 11th International Congress on Transportation Research (Heraklion, Greece, 20-22 September 2023)

6. Kontaxi A., Goulas E., Yannis G., "Free Public Transport in Athens: a stated preference approach", Proceedings of the Transport Research Arena TRA 2022 Conference, Lisbon, 14-17 November 2022
7. Kontaxi A., Akritidou S., Ziakopoulos A., Yannis G., "Critical Factors Affecting Mobile Phone Use While Driving Through the Exploitation of Data from Smartphone Sensors", Proceedings of the Transport Research Arena TRA 2022 Conference, Lisbon, 14-17 November 2022
8. Kyparissis I., Kontaxi A., Deliali A., Yannis G., "Electric or not? Factors affecting Greek Drivers' Preference when Purchasing a New Vehicle", Proceedings of the Transport Research Arena TRA 2022 Conference, Lisbon, 14-17 November 2022
9. Kontaxi A., Ziakopoulos A., Katrakazas C., Yannis G. (2022) "Measuring the impact of driver behavior telematics in road safety" Proceedings of FERSI Conference, The Hague, Netherlands, 06-07 October, 2022
10. Frantzola E., Kontaxi A., Yannis G., "Impact of Road and Traffic Characteristics on Driver Behavior and Safety Using Data from Smartphones", Proceedings of the 6th International Symposium on Highway Geometric Design, Amsterdam, 26-29 June 2022
11. Tzoutzoulis D., Kontaxi A., Ziakopoulos A., Yannis G. (2022) "Exploring critical driving parameters affecting speeding using data from smartphones" Proceedings of 8th Road Safety & Simulation International Conference, Athens, Greece, 08-10 June, 2022
12. Kokkali K., Ziakopoulos A., Kontaxi A., Yannis G., "Correlation of declared and revealed driver behavior using smartphone sensors", Proceedings of the 8th Road Safety and Simulation International Conference, Athens, 8-10 June 2022
13. Maragkoudakis V., Kontaxi A., Deliali K., Yannis G., "Public opinion on e-scooters in Athens: a stated preference approach", Proceedings of the 10th International Congress on Transportation Research (1-3 September 2021).
14. Priftis G., Kontaxi A., Yannis G., "Public opinion on Flying Autonomous Vehicles in Greece: a stated preference approach", Proceedings of the 10th International Congress on Transportation Research (1-3 September 2021).
15. Papantoniou P., Kontaxi A., Yannis G., Fortsakis P. (2020). "Investigating the correlation of mobile phone use with trip characteristics recorded through smartphone sensors", Proceedings of 5th Conference on Sustainable Urban Mobility – CSUM2020, June 17-19, 2020, Virtual Conference.
16. Ziakopoulos A., Kontaxi A., Yannis G., Tselentis D., Fortsakis P. "Advanced driver monitoring using smartphone applications: The BeSmart project", Proceedings of 8th Transport Research Arena TRA 2020 Conference (Helsinki, Finland, 27-30 April 2020)
17. Chaireti M., Kontaxi A., Pavlou D., Yannis G., "Investigation of the impact of weather conditions to young drivers' behavior and safety in cities", Proceedings of the 8th Transport Research Arena TRA 2020 Conference (Helsinki, Finland, 27-30 April 2020)
18. Michelaraki E., Kontaxi A., Papantoniou P., Yannis G., "Correlation of driver behavior and fuel consumption using data from smartphones", Proceedings of the 8th Transport Research Arena TRA 2020 Conference (Helsinki, Finland, 27-30 April 2020)
19. Kontaxi A., Ziakopoulos A., Tselentis D., Yannis G., "A Review of the Impact of Driver Distraction on Driving Behavior and Road Safety", Proceedings of the 9th International Congress on Transportation Research, organised by The Hellenic Institute of Transportation Engineers (HITE) and the Hellenic Institute of Transport (HIT/CERTH) (Athens, Greece, 24-25 October 2019)
20. Kokkinakis A., Kontaxi A., Tselentis D., Yannis G., "Identification of Critical Driving Parameters Affecting Speeding Using Data from Smartphones", Proceedings of the 9th

International Congress on Transportation Research, organised by The Hellenic Institute of Transportation Engineers (HITE) and the Hellenic Institute of Transport (HIT/CERTH) (Athens, Greece, 24-25 October 2019)

Appendix II

Consent Form for Participants

BESMART

Τομέας Μεταφορών και Συγκοινωνιακής Υποδομής, ΕΜΠ

oseven

Consent Form for Participants

1. I declare that I have read and understood the Participant Information Sheet provided for the BeSmart research project titled "Driver Behavior and Safety Support System for All Means of Transportation Using Mobile Phones." I also confirm that I have been given the opportunity to ask questions about the experiment procedure in which I will participate.

2. I understand that my participation is voluntary and that I am free to withdraw at any time without having to explain my reasons and without affecting my legal rights.

3. I consent to installing the BeSmart application, developed by OSeven as part of the BeSmart project, on my mobile phone, provided I have read and accepted the application's Terms of Use.

4. I authorize the designated members of the BeSmart research team to access all data recorded by the application installed on my mobile phone solely for the purpose of processing and analyzing it within the scope of the research project. The data recorded includes sensor values from the mobile phone (e.g., accelerometer, gyroscope, GPS).

5. I acknowledge full responsibility for my driving behavior and understand that the research team members bear no responsibility for the outcomes of my driving behavior.

Date: ....

Participant Name: ....

Email Address: ....

Vehicle Type:

☐ Private Car

☐ Commercial Vehicle

☐ Motorcycle

☐ Bicycle

Smartphone Type:

☐ Android

☐ iOS

Signature: ....

Appendix III

Driving behavior questionnaire

Driver Behavior Questionnaire

A. Driving Experience – Trips

1. Participant Email:

2. When did you obtain your car driver's license?

3. How many years of driving experience do you have, regardless of vehicle type?

4. How many days per week do you use your car? (1 to 7 options)

5. Approximate kilometers driven per week (<20 to 150+ options)

6. Average daily trips as a driver (1 to 5+ options)

7. Average daily trip length in kilometers (1-2 to 30+ options)

8. Approximate yearly kilometers driven (<5,000 to >20,000 options)

B. Vehicle

9. Ownership status (Personal, family-owned, rented, company vehicle, other)

10. Engine capacity (<1001cc to >2000cc options)

11. Vehicle age (<5 years to >15 years options)

12. Average fuel consumption (<5lt/100km to >15lt/100km options)

C. Driving Behavior

13. Accident history (last 3 years, with or without fault):

    a. Total number of accidents you have been involved in.

    b. Accidents with injuries.

    c. Accidents with only material damages.

14. Traffic violation fines in the last 3 years (0 to >3 options)

15. Statements on driving behavior (Never to Always scale):

    a. Exceeding speed limits

b. Harsh braking

c. Aggressive acceleration

d. Sudden turns

e. Mobile phone use while driving

16. Compliance with speed limits (1: Not at all, 5: Very much):

a. Highway

b. National roads

c. Urban roads

17. Driver self-assessment (1 to 5 scale):

a. How careful you perceive yourself to be?

b. How aggressive you perceive yourself to be?

D. Demographic Data

18. Gender (Male, Female, Other)

19. Age (ranges: 18-24 to ≥65)

20. Marital status (Single, Married, Divorced, Widowed)

21. Household size

22. Family annual income (<10,000 to >30,000 options or 'Prefer not to say')

23. Education level (Primary to Doctorate and Other)

24. Familiarity with smartphone applications (1: Very low, 5: Very high)