INCIDENT OCCURRENCE AND RESPONSE ON URBAN FREEWAYS

Ecole Doctorale Ville, Transports et Territoires (Université Paris-Est)
Laboratory of Railways and Transport (NTUA)
Génie des Réseaux de Transport et Informatique Avancée (INRETS)

Dissertation

for the degree of Doctor of Philosophy of

- the École des Ponts (ENPC-Paris Tech) and the
- National Technical University of Athens (NTUA)

Discipline: Transport

Written and defended by under the supervision of

Zoi CHRISTOFOROU Simon COHEN / Matthew KARLAFTIS

Paris, December 2010

A CASE STUDY ON THE A4-A86 JUNCTION IN ILE-DE-FRANCE, FRANCE

Jury Members

Jean-Pierre ORFEUIL	Professor, Université Paris-Est, France	*Président*
Marc GAUDRY	Emeritus Professor, University of Montreal, Canada	*Rapporteur*
Georgios MINTSIS	Professor, Aristotle University of Thessaloniki, Greece	*Rapporteur*
Antonios STATHOPOULOS	Professor, National Technical University of Athens, Greece	*Membre*
Georgios YANNIS	Associate Professor, National Technical University of Athens, Greece	*Membre*
Simon COHEN	Professor, ParisTech, France Research Director at INRETS	*Directeur*
Matthaios KARLAFTIS	Associate Professor, National Technical University of Athens, Greece	*Directeur*

Then, tomorrow was another day
The morning found me miles away
With still a million things to say

Now, when twilight dims the sky above
Recalling thrills of our love
There's one thing I'm certain of
Return I will to old Brazil

Frank Sinatra

Les statistiques sont une forme d'accomplissement de désir, tout comme les rêves.

Jean Baudrillard

I am thankful to all Institutions and Laboratories involved in the present thesis:

- the Ecole des Ponts-ParisTech and the National Technical University of Athens for having honored me by accepting my PhD Candidate application
- the INRETS for having accepted to support my research
- the Ecole Doctorale Ville, Transports et Territoires (ENPC) and the Department of Transportation Engineering and Planning of the School of Civil Engineering (NTUA)

I am grateful to the Hellenic State Scholarships Foundation (IKY) for the funding of my research throughout these years.

Preface

The completion of the present thesis has been an intensive exercise requiring great effort, discipline and time. It is said that the effort invested in a PhD thesis elaboration is never paid back. Personally, I feel already compensated for my endeavors during the past four years as I believe that I have evolved in both an academic and a personal level. Most importantly, within the thesis context and throughout its elaboration, I have been very fortunate to have met and spent time with important people; distinguished scientists and beloved friends. All these individuals –in France and in Greece- have considerably helped in completing the work undertaken and in overcoming difficulties of all kinds. Without their contribution, I wouldn't be able to accomplish this highly demanding project. By the present note, I would like to acknowledge their contribution and express my deepest gratitude along with my wish to continue our collaboration or/and friendship.

I wish to thank, first and foremost, my supervisors, Simon COHEN and Matthew KARLAFTIS. Professor COHEN is the person who introduced to me the field of scientific research and who motivated me to undertake a PhD thesis. He has honored me with his trust and respect since the beginning of our collaboration and guided me throughout both my DEA and my PhD theses. He has showed extreme patience with my mistakes, level of French, inclination to verbosity, inability to attend some of our meetings, along with all other particularities of my character; while always being an excellent pedagogue and mentor. Professor KARLAFTIS closely supervised my work and extensively supported my research efforts. His contribution in all statistical analyses performed has been of critical importance. He also entrusted me with several assignments within numerous research projects and academic activities. He has shown patience with my weaknesses, while having repeatedly advised me on several issues and occasions as a real friend. I will be always thankful to both Professors for their continuous help, support and understanding during these years.

I am also grateful to all Professors and Researchers that have contributed to the analysis and writing of the dissertation by providing useful comments and contributions. In this context, I am truly indebted and thankful to all six members of the jury for their time and effort. Both reviewers, Professors Gaudry and Mintsis, have

provided very fruitful comments that considerably helped in improving the thesis. Apart from participating in the jury, Professors Orfeuil and Leurent have been my teachers during my post-graduate studies in France and significantly influenced my academic interests. I have also been fortunate with Professors Stathopoulos and Yannis who have continuously been my teachers since my under-graduate studies and up to present and have encouraged me in many ways and on many different occasions. Professor Stathopoulos has been presiding the jury for my diploma thesis, supervising my PhD thesis (for IKY), directing the Laboratory of Railways and Transport to which I am attached, and -even more recently- teaching at my Urban Planning post-graduate studies. Professor Yannis has inspired my interest towards road safety since my under-graduate studies, he has motivated and supported my studies in France, and he continues to guide my academic steps.

Furthermore, I would like to express my gratitude to the director of GRETIA laboratory in INRETS, Mr. Schemama, along with all GRETIA staff and, especially Mrs. Seidowski, Mr. Aron, and Mrs. Bergel that have contributed at different stages of the dissertation. Without their help and input, this dissertation would not have been possible. Maurice has been patiently answering my questions since my DEA thesis, while Regine has provided me with valuable moral support and chocolate bars when most needed.

I am also obliged to many Professors and colleagues from the NTUA as well. Professors Kanellaidis and Golias have decisively facilitated my efforts, while Professor Limberis has inspired my interest towards transportation engineering. My colleague Kostas Kepaptsoglou has shared my disappointments and successes and morally supported me throughout all these years. More recently, Christina Milioti has decisively helped me in overcoming difficulties and in remaining optimist under all circumstances. It would have been extremely difficult to find the strength to continue without her presence.

I am also thankful to Mrs. Theofanidou (NTUA), Mrs. Metaxa (Hellenic State Scholarships Foundation - IKY), and Mrs. Alcouffe (ENPC) for their continuous support and assistance regarding all administrative procedures that proved to be extremely cumbersome.

Turning to my friends and beloved ones in France, I owe sincere thankfulness to Nikos for his love, respect, and extreme patience. I also thank Vasso for being a true 'copinada' in all possible meanings. Anastasia and Dominique have morally supported me when being in France. Giorgos has additionally supported me by providing an administrative address, continuous interpreter service, and numerous valuable cynical comments. Furthermore, I thank all my Athenian friends that were obliged to tolerate my 'turbulences' during the thesis elaboration on an everyday basis. I would like to mention the special contribution of Dafni, Natalia, Isidora, and of my sister Maria; they all embraced me with love and called me to order when they had to. Most importantly, I should mention my family, my mother Effie and Thodoris, without whose care, support and love I wouldn't have found the strength and resources to complete the dissertation. Finally, I would like to thank my father. His omnipresent absence throughout all these years makes it impossible to consider any other dedication of the thesis but to his memory.

Abstract

Research on road safety has been of great interest to engineers and planners for decades. Regardless of modeling techniques, a serious factor of inaccuracy - in most past studies - has been data aggregation. Nowadays, most freeways are equipped with continuous surveillance systems making disaggregate traffic data readily available; these have been used in few studies. In this context, the main objective of this dissertation is to capitalize highway traffic data collected on a real-time basis at the moment of accident occurrence in order to expand previous road safety work and to highlight potential further applications. To this end, we first examine the effects of various traffic parameters on type of road crash as well as on the injury level sustained by vehicle occupants involved in accidents, while controlling for environmental and geometric factors. Probit models are specified on 4-years of data from the A4-A86 highway section in the Ile-de-France region, France. Empirical findings indicate that crash type can almost exclusively be defined by the prevailing traffic conditions shortly before its occurrence. Increased traffic volume is found to have a consistently positive effect on severity, while speed has a differential effect on severity depending on flow conditions. We then establish a conceptual framework for incident management applications using real-time traffic data on urban freeways. We use dissertation previous findings to explore potential implications towards incident propensity detection and enhanced management.

Key Words: road safety; crash type; severity; incident management; real-time traffic data; probit.

Executive Summary

An incident can be defined as any occurring event that causes some disruption or deviation to a system's normal operational conditions. Traffic incidents are unplanned events that occur randomly in time and space. They cause a reduction to roadway capacity or an abnormal increase in demand and are associated with high economic and social impact. An estimated 1.2 million of people are killed worldwide each year, while another 50 million are injured (WHO, 2004). Apart from the loss of human lives, accidents have multiple collateral effects: delays, congestion, material damage, environmental damage, pain to society, loss of productivity, impact on freight transport, health costs, and so on. Increases in incident related costs along with sustainable development concern have turned countries and international organizations towards accident mitigation programs and policies.

'*Safety is the number of accidents (crashes), or accident consequences, by kind and severity, expected to occur on the entity during a specified period of time'* (Hauer, 1997). Research on road safety has also attracted considerable research interest in the past three decades. Major factors known to affect safety – in terms of both incident occurrence and severity - are driver characteristics, vehicle features, exposure to risk (e.g. traffic volumes), traffic control, weather conditions, and roadway design characteristics. These measurable factors do not completely explain accident occurrence and, so, stochastic models (including a disturbance error term) are typically used.

Regardless of modeling techniques, a serious factor of inaccuracy - in most past studies - has been data aggregation. The Average Annual Daily Traffic (AADT) has been the most commonly used measure to reflect traffic conditions. However as most freeways are equipped with continuous surveillance systems, disaggregate traffic data collection is possible as well as readily available. While detailed vehicle movement data in a section would be the best data source, traffic data from several consecutive detectors in a section can be a good surrogate to identifying traffic dynamics that may lead to accidents. Disaggregate traffic data have been used in only a limited number of studies.

In this context, the main thesis research question is to explore the effect of actual traffic conditions on accident patterning and consequences. The thesis objective is to use highway traffic data collected on a real-time basis in order to: a) explore the effects of traffic parameters on type of road crash, b) investigate the influence of traffic parameters on the injury level sustained by vehicle occupants, and to c) explore possible implications in incident management strategies. To this end, four main research activities are undertaken: a) a literature review, b) an empirical investigation on incident type propensity using real-time traffic data on freeways, c) an empirical investigation of vehicle occupant injury severity on freeways using real-time traffic data, and d) the development of a conceptual framework towards introducing real-time traffic data in incident management and response.

In the first research activity, a literature overview on related studies is performed. The overview indicates that, due to the complexity of the road system and its management, road safety analysis necessarily involves numerous scientific disciplines. The state of the art in related research is summarized; the large body of literature is organized on the basis of both methodological and thematic criteria. These criteria include: a) the method employed, b) the level of analysis assumed, c) the scope of the performed analysis, and d) the accident phase considered. The dissertation field of interest is, then, defined with respect to the taxonomy established.

Road Safety Literature Organization

Classification Criteria		Controlled Experiment		Field Observational		In-depth investigation		Data Observational	
		A	D	A	D	A	D	A	D
Generating	Descriptive	-	✓	✓	✓	-	✓	✓	✓
	Predictive	-	-	-	-	-	-	✓	-
Patterning	Descriptive	-	✓	✓	✓	-	✓	✓	✓
	Predictive	-	-	-	-	-	-	✓	-
Response	Descriptive	-	✓	-	✓	-	✓	✓	✓
	Predictive	-	-	-	-	-	-	✓	-
Consequences	Descriptive	-	✓	✓	-	-	✓	✓	✓
	Predictive	-	-	-	-	-	-	✓	-

*A: aggregate
D: disaggregate

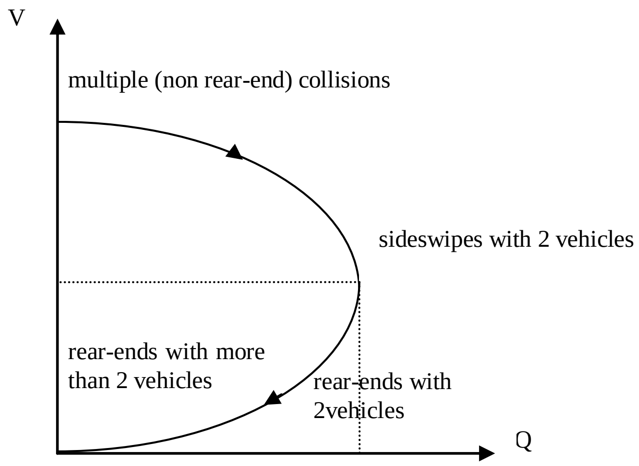
dissertation field of interest

In the present dissertation, we conduct a data observational study within a descriptive scope of analysis. Stochastic modeling is used in a rather disaggregate context of analysis. Accident outcomes - in terms of either crash type or severity – serve as dependent variables. Crash type refers to accident patterning, while severity is linked to accident consequences. Independent variables include road user attributes, weather and lighting conditions, vehicle type and age, traffic data, and so on. To this end, real-time traffic data are extracted from continuous loop measurements at the time of the accident occurrence (aggregate field observations). Results provide probability estimations for accident outcomes, given that these accidents occur under specific circumstances; if combined with frequency models, they could additionally provide prediction estimations. Finally, we examine potential implications of the developed models in optimizing incident management techniques; the latter being related to accident response phase.

In the second research activity, we examine the effects of various parameters on type of road crash as there is strong empirical evidence that accident characteristics are crash type-specific. Several authors underlined the importance of by-crash-type analysis, particularly when it comes to real-time risk assessment. They suggested that the conditions preceding crashes are expected to differ by type of crash and, therefore, any approach towards proactive traffic management should be type-specific in nature.

Multivariate Probit models are specified on 4-years of data (2000-2002, 2006) from the A4-A86 highway section in the Ile-de-France region, France. Traffic parameters are collected real-time both at – and prior to - the time of the accident and include measurements of volume, speed, and density over 6-minute intervals. Empirical results indicate a diverse effect of accident contributing factors to each crash type, along with interdependencies that would be neglected under a univariate analysis context. It has to be noted that previous studies adopt univariate approaches.

Rear-end crashes involving two vehicles are found to be more probable for relatively low values of both speed and density, while rear-ends involving more than two vehicles appear to be more probable under congestion. Two-vehicle sideswipe accident probability increases with increasing volume, while multi-vehicle sideswipe crashes are more probable at high speeds, during daytime, and on flat freeway segments. Overall, multi-vehicle crashes tend to occur under low or very high speeds, while single-vehicle crashes appeared to be largely geometry-dependent. Qualitative results from the Multivariate Probit model application are illustrated in the following diagram. The fundamental diagram depicts the relationship between traffic volume (Q) and speed (V) on a given freeway segment; each crash type (whose probability is traffic-dependent) is related to a particular traffic regime which corresponds to a specific part of the diagram.



In the third research activity, we extend research on the factors influencing the level of accident severity by including traffic data from the moment of the accident. Results from previous research indicate that low speeds and high traffic volumes decrease accident severity, while high speeds and low traffic volume produce the opposite effect; a result largely based on mean annual traffic values. However, few studies have investigated the association between traffic accident severity and actual traffic characteristics (traffic volume, speed) collected real-time during the time of the accident occurrence. A random parameters ordered probit model is applied to explore the influence of speed and traffic volume on the injury level sustained by vehicle occupants involved in accidents on the A4-A86 junction in the Paris region. The random parameters specification allows for heterogeneity amongst road users; otherwise neglected under a fixed parameters approach. It has to be noted that all previous studies use a fixed parameters approach.

Empirical results indicate that travelling on 2 wheels and at nighttime significantly increases the probability of getting involved in more severe accidents. In contrast, travelling in heavy vehicles, on weekends or on dry pavement surfaces reduces the probability of severe accidents. Less experienced drivers seem to encounter problems in dealing with adverse weather conditions and related potential dangers. Most importantly, results indicate that there is a significant relationship between the severity outcome and the traffic characteristics at the time of the accident. Traffic volume was found to have a consistently positive effect, while speed appears to have a differential effect on severity depending on flow conditions. While in higher traffic volumes higher speeds aggravate severity outcome, in lower traffic volumes speed does not significantly influence severity in a consistent pattern.

In the fourth research activity, we investigate the introduction of incident analysis outcomes in an integrated incident management scheme. To this end, a synthesis of related incident management analyses is performed. Further, crash data studies using traffic data collected on a real-time basis at the time of the incident occurrence are analyzed. The synthesis indicates that real-time traffic data has not been fully utilized as they are only used for emergency vehicles' travel-time estimation. However, they could be used as a criterion for location and allocation if appropriately combined with road safety analysis outcomes. Finally, we use dissertation previous findings to explore such potential implications towards incident propensity detection and enhanced management.

The present dissertation demonstrated the importance and magnitude of the effect of prevailing traffic conditions on accident occurrences, while it provided additional insight in accident mechanism of occurrence. From a methodological standpoint, the use of disaggregate real-time traffic data provided better probability estimates. Integrating such data in incident management strategies shows great potential towards accident mitigation and enhanced management. Overall, the thesis significantly contributes to research state of the art regarding an issue that has scarcely been examined in the past. Furthermore, it provides the appropriate theoretical framework, along with necessary supporting models for better utilizing disaggregate traffic data.

ESTIMATION DES PROBABILITES D'INCIDENTS ET TRAITEMENT DE LEUR REPONSE SUR DES AUTOROUTES URBAINES

Résumé

Les recherches en matière de sécurité routière suscitent largement l'intérêt des chercheurs. Indépendamment des techniques de modélisation, un facteur important d'imprécision -qui caractérise les études dans ce domaine- concerne le niveau d'agrégation des données. Aujourd'hui, la plupart des autoroutes sont équipées de systèmes permanents de surveillance qui fournissent des données désagrégées. Dans ce contexte, l'objectif de la thèse est d'exploiter les données trafic recueillies en temps réel au moment des accidents, afin d'élargir le champ des travaux précédents et de mettre en évidence un potentiel d'applications innovantes. À cette fin, nous examinons les effets du trafic sur le type d'accident ainsi que sur la gravité subie par les occupants des véhicules, tout en tenant compte des facteurs environnementaux et géométriques. Des modèles Probit sont appliqués aux données de trafic et d'accidents enregistrés pendant quatre années sur le tronc commun aux autoroutes A4 et A86 en Ile-de-France. Les résultats empiriques indiquent que le type d'accident peut être presque exclusivement défini par les conditions de trafic prévalant peu avant son occurrence. En outre, l'augmentation du débit s'avère exercer un effet constamment positif sur la gravité, alors que la vitesse exerce un effet différentiel sur la gravité en fonction des conditions d'écoulement. Nous établissons ensuite un cadre conceptuel pour des applications de gestion des incidents qui s'appuie sur les données trafic recueillies en temps réel. Nous utilisons les résultats de la thèse afin d'explorer des implications qui ont trait à la propension et à la détection des incidents, ainsi qu'à l'amélioration de leur gestion.

Mots clés : sécurité routière; type de collision; gravité; gestion d'incidents; données trafic en temps réel; probit.

Résumé Substantiel

L'incident peut être défini comme tout événement qui cause une certaine perturbation ou déviation des conditions normales de fonctionnement d'un système. Les incidents de trafic sont des événements non planifiés qui se produisent de manière aléatoire dans le temps et dans l'espace. Ils sont la cause soit d'une réduction de la capacité routière soit d'une augmentation irrégulière de la demande et sont associés à un impact économique et social élevé. Au niveau mondial, on estime qu'environ 1,2 million de personnes sont tuées chaque année, tandis qu'encore 50 millions sont blessées (WHO, 2004). Outre la perte de vies humaines, les accidents ont des multiples effets collatéraux: embouteillages, retards subis par les usagers, dégâts matériels et environnementaux, impact social, perte de productivité, difficultés pour le transport de marchandises, frais médicaux, etc. Les coûts qui sont associés, ainsi que les préoccupations concernant le développement durable ont incité des pays et des organisations internationales à adopter des politiques spécifiques et à appliquer des programmes de réduction de l'insécurité routière.

« *La sécurité est le nombre d'accidents, ou les conséquences des accidents, par type et gravité, qui sont susceptibles de se produire sur l'entité au cours d'une période spécifique* » (Hauer, 1997). La recherche sur la sécurité routière a suscité l'intérêt des ingénieurs et des planificateurs pendant des décennies. Les facteurs importants connus pour leur effet sur la sécurité - à la fois en termes d'occurrence et de gravité – sont, entre autres, les caractéristiques du conducteur et du véhicule, l'exposition au risque (par exemple les volumes de trafic), le contrôle de la circulation, les conditions météorologiques, et les caractéristiques géométriques du tronçon routier. Ces facteurs mesurables ne peuvent pas expliquer l'occurrence des accidents d'une manière exhaustive. Par conséquent, des modèles stochastiques (comprenant un terme de perturbation) sont généralement utilisés.

Indépendamment des techniques de modélisation, un facteur important d'incertitude - dans la plupart des études précédentes - concerne le niveau d''agrégation des données. Le trafic journalier moyen annuel (TMJA) a été un des indicateurs le plus fréquemment utilisé afin de représenter les conditions d'écoulement du trafic. De nos jours, la plupart des autoroutes sont équipées de systèmes permanents de surveillance

qui rendent disponibles des données trafic désagrégées; celles-ci ont été utilisées dans certaines études. Tandis que les données détaillées du mouvement du véhicule sur une section routière seraient la meilleure source de données, les données trafic provenant de plusieurs détecteurs consécutifs sur une section peuvent servir comme alternative pour l'identification des régimes de trafic menant aux accidents. Néanmoins, les données trafic désagrégées ont été très peu utilisées dans des études précédentes.

Dans ce contexte, l'objectif principal de la dissertation est d'exploiter les données de trafic des autoroutes collectées en temps réel afin de : a) explorer les effets des paramètres de trafic sur le type d'accident routier, b) étudier l'influence des paramètres de trafic sur le niveau de gravité subi par les occupants des véhicules, c) explorer les implications potentielles dans des stratégies de gestion des incidents. Nous abordons les questions de recherche par quatre étapes de travaux : a) revue bibliographique, b) étude empirique des effets du trafic sur le type d'accident, c) étude empirique des effets du trafic sur la gravité, d) élaboration d'un cadre conceptuel pour l'intégration des données en temps réel à la gestion des incidents.

Tout d'abord, nous effectuons une revue bibliographique des études sur le sujet. La littérature indique qu'en raison de la complexité inhérente au système routier et à sa gestion, l'analyse de sécurité routière implique nécessairement le recours à de nombreuses disciplines scientifiques. Nous proposons le système suivant de classification de la littérature de la sécurité routière. Ce système est construit sur la base de quatre critères principaux : a) la méthode employée, b) le niveau de l'analyse escomptée, c) l'objectif de l'analyse effectuée, et d) la phase des accidents considérée. Puis, nous définissons le champ d'intérêt de la thèse par rapport à la classification proposée.

Organisation de la littérature de sécurité routière

Critères de classification		Essai Contrôlé		Observation au champ		Investigation approfondie		Observation des données	
		A	D	A	D	A	D	A	D
Genèse	Descriptive	-	✓	✓	✓	-	✓	✓	✓
	Prédictive	-	-	-	-	-	-	✓	-
Structuration (Evolution/ achèvement)	Descriptive	-	✓	✓	✓	-	✓	✓	✓
	Prédictive	-	-	-	-	-	-	✓	-
Réponse	Descriptive	-	✓	-	✓	-	✓	✓	✓
	Prédictive	-	-	-	-	-	-	✓	-
Conséquences	Descriptive	-	✓	✓	-	-	✓	✓	✓
	Prédictive	-	-	-	-	-	-	✓	-

*A: agrégé
D: désagrégé

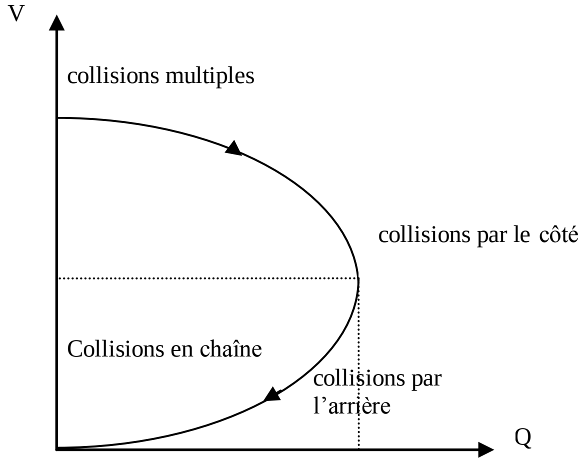
Champs d'intérêt de la thèse

Dans la thèse, nous réalisons une étude d'observation des données de nature descriptive. Une technique de modélisation stochastique est utilisée afin d'aborder les questions principales de recherche dans un contexte d'analyse plutôt désagrégée. Les résultats des accidents -en termes à la fois de type et de gravité- servent comme variables dépendantes. Le type d'accident fait référence à la structuration des accidents, tandis que la gravité est liée à leurs conséquences. Les variables indépendantes incluent les attributs d'usager de la route, les conditions météorologiques et d'éclairage, le type et l'ancienneté des véhicules, des données trafic etc. Dans cet objectif, des données du trafic au moment de l'occurrence de l'accident ont été extraites de la base des enregistrements continus issus des boucles (des observations agrégées et au champ). Les résultats fournissent des estimations de probabilité, étant donné que les accidents se réalisent sous des conditions spécifiques. Si on combine ces estimations avec des modèles de fréquence, on peut obtenir des estimations prédictives. Nous examinons également les conséquences favorables des modèles élaborés sur l'optimisation des techniques de gestion des incidents; cette dernière est liée à la phase de réponse aux accidents.

En second lieu, nous choisissons d'étudier les effets de différents paramètres sur le type d'accident routier puisqu'il y a des preuves empiriques fortes indiquant que les caractéristiques des accidents sont largement dépendantes du type d'accident. Plusieurs auteurs ont souligné l'importance de l'analyse par type d'accident, particulièrement quand il s’agit des évaluations du risque en temps réel. Ils suggèrent que les conditions qui précédent l’occurrence des accidents diffèrent par type d'accident. En conséquence, toute tentative de gestion « proactive » de la circulation doit être faite par type d’accident.

Des modèles Probit multivariés ont été appliqués sur des données de trafic et d’accidents relevées pendant quatre années (2000-2002, 2006) sur le tronc commun des autoroutes A4 et A86 en Ile-de-France. Les paramètres de trafic ont été collectés en temps réel avant et au moment exact de l’accident et incluent des mesures de débit, de vitesse, et de densité pendant des séquences de 6 minutes. Les résultats empiriques ont indiqué que les facteurs contribuant à l’occurrence des accidents exercent un effet différencié par type d'accident considéré. De plus, les résultats ont révélé des interdépendances parmi les variables dépendantes qui seraient négligées dans un contexte d'analyse univariée. Il faut noter que les études précédentes sur le sujet se réalisent dans un contexte d’analyse univariée.

Les collisions par l’arrière se sont avérées plus probables pour des valeurs relativement basses de vitesse et de densité, tandis que les collisions en chaîne paraissent plus probables en congestion. La probabilité de collisions latérales (de deux véhicules) augmente avec les hausses du débit. Les collisions multiples sont plus probables en régime des vitesses élevées, pendant la journée et sur des tronçons autoroutiers plats. En résumé les collisions multiples se produisent plutôt en régime de vitesses soit faibles soit très élevées, tandis que la géométrie routière s’avère être l’indicateur unique des accidents sans collision. Les résultats qualitatifs acquis par l'application du modèle Probit multivarié sont illustrés dans le diagramme qui suit. Le diagramme fondamental représente la relation entre le débit (Q) et la vitesse (V) sur un segment donné. Chaque type d'accident dont la probabilité est dépendante du trafic est lié à un régime de trafic précis qui correspond à une partie spécifique du diagramme.



Troisièmement, nous examinons les facteurs ayant une influence sur le niveau de gravité des blessures. Les résultats des recherches antérieures convergent sur le fait que les vitesses faibles et les débits élevés diminuent la gravité des accidents, tandis que les vitesses élevées et les débits faibles produisent l'effet inverse. Ces résultats ont été –en grande partie- obtenus sur la base des valeurs moyennes annuelles du débit. Néanmoins, rares sont les études qui ont tenté d'estimer la relation liant la gravité des accidents aux caractéristiques de la circulation (débit, vitesse, densité) enregistrées au moment de l'occurrence des accidents. Dans la thèse, nous explorons les facteurs ayant un effet sur la gravité des accidents à partir des données trafic enregistrées à l'instant d'occurrence des accidents. En particulier, nous cherchons à estimer l'influence de la vitesse et du débit sur le niveau de gravité des blessures que subissent les occupants des véhicules impliqués dans les accidents survenant sur le tronc commun des autoroutes A4 et A86, en région parisienne. Dans cet objectif, nous appliquons un modèle Probit ordonné avec paramètres aléatoires. Ce modèle permet l'hétérogénéité des usagers de la route, négligée dans le cas d'une approche avec des paramètres fixes. Il faut noter que la totalité des études précédentes adoptent l'approche des paramètres fixes.

Les résultats empiriques indiquent que voyager en 2 roues et pendant la nuit augmente la probabilité d'implication dans des accidents plus graves. En revanche, voyager en poids lourd, en week-end ou sur des chaussées sèches réduit la probabilité des accidents graves. De plus, il apparaît que les conducteurs les moins expérimentés rencontrent des difficultés face aux mauvaises conditions météorologiques et aux dangers que celles-ci impliquent. Ainsi, les résultats indiquent une relation significative entre la gravité des accidents et les caractéristiques d'écoulement du trafic au moment de leur occurrence. L'augmentation du débit s'avère avoir un effet constamment positif sur la gravité (en la diminuant), tandis que la vitesse semble exercer un effet différentiel par rapport au volume du trafic. Plus précisément, sous des régimes de débit élevé, les vitesses élevées aggravent les accidents. En revanche, sous des régimes de débit faible, nous n'avons pas détecté d'influence de la vitesse sur la gravité.

Dans la quatrième étape d'analyse, nous étudions l'introduction, dans un schéma intégral de gestion des incidents, des résultats issus des étapes précédentes. Nous effectuons une synthèse des études sur ce sujet afin d'établir un cadre conceptuel pour des applications de gestion des incidents sur la base des données recueillies en temps réel sur des autoroutes urbaines. La synthèse indique que les données de trafic en temps réel n'ont pas été entièrement exploitées et qu'elles s'utilisent exclusivement pour l'estimation du temps de parcours des véhicules d'urgence. Néanmoins, elles pourraient bien être utilisées comme critère pour la localisation et l'affectation des unités d'urgence si elles étaient combinées proprement avec des résultats des analyses de sécurité routière. Enfin, nous utilisons des résultats de la thèse afin d'explorer des applications potentielles portant sur la propension à la détection des incidents ainsi que sur l'amélioration de leur gestion.

La thèse contribue à une compréhension approfondie du mécanisme d'occurrence des accidents routiers. En ce qui concerne la méthodologie appliquée, l'utilisation des données de trafic désagrégées et recueillies en temps réel fournit une meilleure estimation de probabilité. L'intégration de ces données dans des stratégies de gestion d'incidents montre une capacité considérable de réduction des accidents et d'amélioration de leur gestion. En somme, la thèse contribue à la recherche sur un sujet qui a été peu étudié dans le passé. En même temps, elle fournit le cadre théorique nécessaire et les outils mathématiques indispensables pour mieux exploiter les données trafic désagrégés.

Thesis Outline

Table of Contents

List of Figures

List of Tables

Abbreviations

AADT Average Annual Daily Traffic

AIS Abbreviated Injury Scale

ANN Artificial Neural Network

BAAC Bulletins d'Analyse des Accidents Corporels

CCTV Closed-circuit television

EMS Emergency Medical Services

ENPC Ecole Nationale des Ponts et Chaussées

ETSC European Transport Safety Council

EU Emergency Unit

GLM General Linear Model

GNP Gross National Product

GPS Global Positioning System

ICD International Classification of Diseases

IHL Involuntary High-consequence Low-probability

IMS Integrated Management System

INRETS Institut National de REcherche sur les Transports et leur Sécurité

ISS Injury Severity Score

ITS Intelligent Transportation Systems

LSCP Location Set Covering Problem

MCLP Maximal Coverage Location Problem

NLCCA Nonlinear Canonical Correlation Analysis

NN Neural Network

NTUA National Technical University of Athens

PCA Principal Components Analysis

PDO Property Damage Only

SISER Service Interdépartemental de Sécurité et d'Exploitation Routière

VMS Variable Message Signs

VLH Voluntary Low-consequence High-probability

Chapter 1

Introduction

This chapter is an introduction to incident occurrence; it aims at discussing several key-terms and at providing a concise state-of-the-art analysis concerning road accidents. The road safety problem is pointed out, while incident management theory and techniques are overviewed. The analysis main objectives are presented and the potential interest of results is pointed out. Finally, the structure of the dissertation remainder is provided.

1.1 Generic incidents

1.1.1 Operational incident

An incident can be defined as any occurring event that causes some disruption or deviation to a system's normal operational conditions. The disruption scale may vary from minor discontinuities to complete failure. Incidents are involuntary, random events whose effects are commonly unpleasant. They may be generally expected to happen and processes to best address them may have been designed; however, their definite occurrence remains uncertain and the moment of that occurrence is always unknown. The entity that provokes an incident is not necessarily the main recipient of its consequences; incident objects and incident subjects do not coincide. Nevertheless, consequence intensity is generally greater for the objects that caused the incident occurrence. Incidents may have collateral positive effects, but the predominant ones remain negative. Incident theories have been developed and applied in several fields such as industrial management and design, response to natural disasters, and so on. By the term 'accident', we often refer to incidents with severe outcomes. However, in injury prevention and epidemiology the term 'accident' is commonly avoided in an attempt to highlight the predictable and preventable nature of most injuries.

1.1.2 Causal factors

The causal factors leading to incident occurrences are partially or completely unknown. A full understanding of these factors would make the incident foreseeable and, thus, avoidable. Knowing all incident contributing factors is not adequate for preventing incidents. The most crucial issue is the form of the relationship between contributing factors and incident occurrence as well as factors' quantified impact. This relationship form may vary from a simple linear causal chain to very complex interactions and can be approximated by controlled experiments or observational studies. Assumptions are made for the triggering event that initiates the incident mechanism; however triggering events remain random and are often associated to human mistakes or sudden mechanical failures. Analysts try to explore all possible outcomes that may follow several hypothetical triggering events. The objective of such investigations is either to minimize incident occurrences or to mitigate incident consequences and interrupt their mechanism as soon as possible.

1.1.3 Characteristics

Incident characteristics include incident probability of occurrence, type of triggering event, type and magnitude of unpleasant outcomes, type and availability of appropriate response, implicated costs, and so on. Apparently, strategies to deal with incidents are incident characteristics-dependent. Perrow (1999) illustrates the risk of technical systems by the use of a four-quadrant taxonomy. Those systems having the characteristics of tight coupling and high complexity are most at risk of system accidents while those that are less tightly coupled and less complex are not as likely to experience such incidents. Rare events (having low probability of occurrence) are generally not examined as thoroughly as common ones. Nevertheless, if a rare event causes an environmental disaster (type and magnitude of outcome), it is probably ‘worth’ investigating. On the other hand, if incident prevention is more expensive than its consequences (cost implications), maybe no effort will be made to prevent it from happening.

1.1.4 Safety and risk

Safety can be regarded as a compromise between requirements and economic necessity (Petroski, 1994). Therefore, the level of risk tolerated depends on an utilitarian calculus that safety is desirable but costly and that organizations choose a level of safety by balancing the benefits of safety reduction against the costs of safety improvement (Marcus and Nichols, 1996). The level of safety achieved is not the highest technically and humanly possible, but rather, depends on resource availability. Under this scope, incidents that should be investigated may either be involuntary, high-consequence, low-probability (IHL) events (like nuclear plant meltdowns) or voluntary, low-consequence, high probability (VLH) events (like road accidents as defined by Naveh and Marcus (2002)).

1.1.5 Incident management

Over the last decades, incident management has been of great interest to both researchers and practitioners. Incident management includes a variety of applications under the objective of best addressing an incident occurrence (as well as its consequences) in various fields such as industrial failures, natural disasters, and so on. While reactive approaches to incident management include all research performed in the area of being prepared to deal with the occurrence of a specific incident and of its consequences (given that it occurs), proactive incident management includes all investigations made in the aim of finding ways to prevent an incident from occurring.

In many fields (e.g. industry, medicine), the concept of prevention is commonly described by a division into sub-concepts, each of which is intended to represent one main preventive strategy (Andersson and Menckel, 1995). The most widely employed classification in medicine was launched by Gjestland (1955). According to this classification, preventive activities are divided into primary, secondary, and tertiary activities that are related to different periods in time in the course of a disease. Primary prevention is taken in advance, while secondary and tertiary actions are taken later on. Primary prevention can be further divided into proactive and reactive (Catalano and Dooley, 1980). Proactive activities are designed to deter or limit exposure, while reactive are aimed at the promotion of coping or increasing adaptation in response to an exposure that has already taken place (Catalano and Dooley, 1980). Thus, proactive actions are taken before exposure, while reactive actions can be taken either before or after exposure but are always designed to have an effect after exposure (Andersson and Menckel, 1995).

1.1.6 Transportation

In the field of transportation, incidents have been occurring since the construction of the first transportation system. Incidents in a transportation system include all infrastructure, operational, or vehicle dysfunctions due to human, natural, mechanical, or other causes. We distinguish among a large spectrum of incident types in respect to the means of transportation and the infrastructure concerned (maritime, air, railway, and so on). Most generic incident theory principles apply also in transportation safety and risk analyses. Thus, air crashes - that happen rarely but count up to a hundred or two of fatalities – can be considered as IHL events and are thoroughly investigated. On the contrary, road accidents cause few fatalities, but they are very frequent events (VLH events); as such they deserve research focus.

1.2 Road incidents

1.2.1 Traffic incidents and accidents

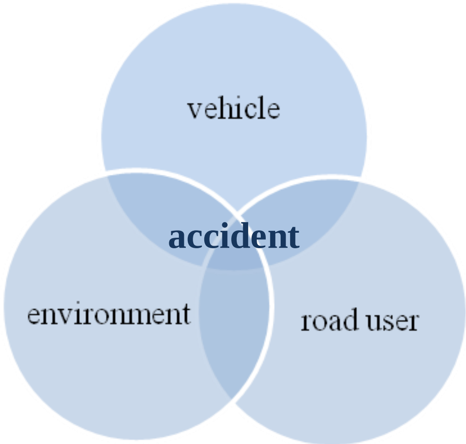
Traffic incidents are unplanned events that occur randomly in time and space. They cause a reduction to roadway capacity or an abnormal increase in demand and are associated with high economic and social impact. Such events include traffic crashes, disabled vehicles, spilled cargo, highway maintenance and reconstruction projects, and special non-emergency events (FHWA, 2000). Incidents' effects include congestion and delays that result in increased cost of goods and vehicle maintenance, productivity reduction, increased fuel consumption, environmental impacts, and so on. Most importantly, incidents trigger secondary crashes whose severity is often greater than that of the original incident (VNTSC, 1995).

Among all road incidents, traffic accidents are the most commonly occurring events. Traffic accidents are incidents in which a road user (pedestrian, bicyclist, car driver etc.) or its vehicle collides with anything that causes damage to other road users (pedestrians, animals, drivers, passengers), vehicles, and roadway features, or in which the driver loses control of the vehicle and gets off the roadway or rolls over. Traffic accidents are sometimes equally referred to as 'crashes', 'traffic collisions', 'motor vehicle accidents'. There is an on-going discussion about the semantic difference between 'accident' and 'crash', in that a 'crash' indicates in a simple factual way what is observed, while 'accident' in addition seems to suggest a general explanation of why it occurred. Some authors state that 'accident' has connotations of it being an unavoidable event. In the present dissertation, all terms are used interchangeably.

1.2.2 Causal factors

A great body of literature deals with identifying factors contributing to accident occurrence as well as with quantifying the impact of such factors. Accidents are commonly viewed as the result of a complex interaction among driver, vehicle, and environmental factors (Figure 1). Driver attributes influencing accident occurrence and severity include: driving experience, level of alert, restraint system use, years of age, alcohol consumption, and so on. Vehicle characteristics playing an important role in accident mechanism of occurrence are: years of age, size and weight, technical characteristics, safety equipment availability and condition, and so on. Finally, environmental and infrastructure-related factors include a variety of factors from weather and lighting conditions to pavement quality, road geometry, speed limits, traffic characteristics, and so on. An exhaustive list of such factors can be found in Gaudry and Lassare (2000).

Figure 1 Accident causal factors



Road accidents occur as a result of a potentially very large number of causal factors exercising their influence at the same location and time. Apart from vehicle-, driver-, and environment-related characteristics that are endogenous to the system, other factors may also influence accident occurrence such as i) factors external to the system (oil price, population), ii) socioeconomic factors (taxation) that are subject to political intervention, iii) factors related to the transportation policy applied (accident countermeasures), iv) the size and structure of the transportation sector, and v) sheer randomness.

1.2.3 Characteristics

Although accidents are the results of human choices and behavior, they are not chosen to occur. On the contrary, when an accident happens, it is because certain road users failed in avoiding it, although they wanted to. Accidents are the unintentional side effects of certain actions taken for reasons other than that of causing injury or damage. They are random and unpredictable in the sense that had they been anticipated, they would probably not have happened. In this context, each single accident is unpredictable by definition.

More detailed explorations indicate various characteristics that differentiate road accidents from each other; such characteristics are used for classification purposes. Targeted studies are then undertaken to investigate and mitigate specific accident categories in respect with their properties. Classification criteria found in the literature include number and type of vehicles involved (trucks, motorcycles), pedestrian involvement, type of collision (rear-end, fixed object), severity outcomes (injury, property damage only), professional driving, accident site geometry (intersection, grade), road network (rural, highway), and so on.

1.2.3 Consequences and cost

Road accidents are an issue of major concern for all countries independently to their level of development; highly developed countries have 60% of the total motor vehicle fleet but they contribute only to 14% of the global road accident deaths (Jacobs et al, 2000). An estimated 1.2 million of people are killed worldwide each year, while another 50 million are injured (WHO, 2004). Road accidents account for 95% of total transportation fatalities and figure as the leading cause of death among young people. Road traffic injuries ranked as the ninth leading cause of the global burden of disease and injury. WHO estimates that road accidents will become the third leading cause of death by the year 2020 (after heart disease and deaths linked to mental illness) if no effective actions and efficient measures are taken.

Apart from the loss of lives, accidents have multiple collateral effects: material damage, environmental damage, pain to society, loss of productivity, impact on freight transport, health costs, delays, congestion, and so on. All these effects correspond to a certain cost, which cumulatively results to be extremely high. This cost is paid by insurance companies and health care systems and is, finally, covered by citizens. Road traffic injuries’ cost accounts for 1% to 2% of the gross national product of low- and middle-income countries (WHO, 2004). In 1997, the ETSC estimated the total cost of transport accidents in Europe at 166 billion euros (ETSC, 1997). 97% of these costs were directly related to road transport. However, policies that yield the largest reductions in road accident counts are not necessarily the most effective. The most cost-effective policy would be one yielding the highest net social benefits. Theoretically, the optimal road safety program has a marginal social cost that equals net marginal social benefit. The methods for evaluating the socio-economic cost of road accidents vary among countries; cost elements taken into account include medical costs, non-medical rehabilitation, lost productive capacity, human costs, damage to property, administrative costs, and other costs such as congestion.

1.2.4 Road safety and risk

Increased implicated costs along with sustainable development concerns have turned countries and international organizations towards adopting accident mitigation programs and policies. Road safety refers to the level of safety achieved on a roadway segment, network or a country. Inversely, risk (i.e. 'road unsafety') is the danger to which the user is exposed when travelling. The level of safety implicates both accident frequency and severity and is measured in many ways (such as absolute number of crashes or deaths, crash rates over vehicle kilometers travelled, crash probability of occurrence, response efficiency); however no integral index meeting general acceptance has been established.

General policy may indirectly affect road accident rate in a number of ways: legislation (speed limits etc.), courts and police enforcement (e.g. vehicle inspections), police deterrence (point system), traffic police management, publicity and education programs, and tax levies (e.g. gasoline taxes). Engineering is also crucial in terms of urban design and land use, road design, vehicle safety design, road maintenance, safety improvement measures, traffic planning, and so on. Medical service response and health care system efficiency greatly affect accident outcomes. Fire departments and all emergency vehicles interfere in accident clearance. Most importantly, drivers and other road users play a decisive role in accident occurrence as they define the so-called *behavioral causal factors*. Consequently, road safety can be regarded as a common objective of psychology, statistics, engineering, policy planning, and so on.

1.2.5 Incident management

In transportation research, incident management is defined as the systematic, planned, and coordinated use of human, institutional, mechanical, and technical resources to reduce the duration and impact of incidents, and improve the safety of motorists, crash victims, and incident responders (FHWA, 2000). On uninterrupted flow facilities equipped with continuous surveillance systems, research and operators mainly focus on minimizing incident overall duration; the latter including detection, response and clearance times. Important factors affecting these times – and consequently the overall incident duration – are (i) the operator ability to promptly detect an incident occurrence, and (ii) the location of emergency stations (police, ambulance). The benefits from minimizing incident duration are numerous and concern highway operators (e.g. cost, road safety performance), crash victims (e.g. time to hospital), other road users (e.g. delays, secondary incidents), and society (e.g. incident externalities).

(i) Incident detection involves the analysis of patterns in the traffic surveillance data observed just after the incident in order to develop models that can separate real-time traffic conditions resulting from incidents from free-flow and/or recurring congestion (Abdel-Aty and Pande, 2007). Incident detection analysis is reactive in nature and attempts to detect incidents so that their impact can be minimized, while it does not search to prevent incidents from happening.

(ii) Emergency station location (e.g. police, fire stations) analysis falls into location analysis; term that refers to the modeling, formulation, and solution of a class of problems that can best be described as sitting facilities in some given space (ReVelle and Eiselt, 2005). Obviously, emergency unit location is important to overall incident duration. In particular, the time needed to reach an incident scene is of great concern to emergency medical services (EMS) in order to mitigate incident consequences on people. In a real-time context, emergency authorities are faced with two main problems: an allocation problem and a redeployment problem (Gendreau et al., 2001). The allocation problem consists of determining which unit must be sent to answer a call. The redeployment problem consists of relocating available units to the potential location sites when calls are received; emergency units are assigned to potential sites to provide coverage.

1.3 State of the Art

Research on road safety has attracted considerable research interest in the past three decades. Major factors known to affect safety are driver characteristics, vehicle features, exposure to risk (e.g. traffic volumes), traffic control, weather conditions, and roadway design characteristics. To predict the safety of transportation systems traffic, engineers model crash rate or frequency as a function of the above mentioned factors. These measurable factors do not completely explain accident occurrence and, so, models typically used are stochastic models including a disturbance error term. So, despite their frequent application, the ability of such models to reliably identify important accident predictors is open to question (Davis, 2004).

Accident occurrence remains unexplained to a certain extent. Some of the problems frequently held responsible are: a) accident underreporting (mainly for property damage only), b) miscounts for accidents or exposure measurements, c) inaccuracies due to misclassification or misjudgment, d) conflicts among different databases (aggregate exposure-discrete accident data, weather, differences among countries in the way they register and count variables), c) time lag between the reporting and registration of accidents or a site lag, d) unaccounted factors that strongly affect the outcome, e) factors whose influence is badly estimated, f) unaccounted interrelations between factors that are taken into account, g) other modeling assumptions and restraints, and h) aggregation of the data used.

Regardless of modeling techniques, a serious factor of inaccuracy – in most past studies – has been data aggregation (Lord and Mannering, 2010) and sample size insufficiency (Pande and Abdel-Aty, 2006). Traditionally, models were macroscopic in nature, where researchers mainly used summary statistics rather than microscopic measures to develop the models. The Average Annual Daily Traffic (AADT) has been most commonly used to reflect prevailing traffic conditions (Kim et al., 2006; Mouskos et al., 1999; Qin et al., 2004). AADT is an aggregate measure of exposure; the use of AADT to approximate vehicle kilometers traveled at a site might reduce the natural variance that exists in exposure data and may result in heavy underdispersion (Pasupathy et al., 2000). Later, many authors used aggregated data over shorter periods of time (month or day) for developing the same models; others used deduced hourly traffic characteristics by combining AADT and a 1-day hourly traffic profile for the site analyzed (Ivan et al., 2000). Nevertheless, even hourly measures cannot consider the short-term variation of traffic flow and are rather not well suited for application to real-time operations.

As most freeways are equipped with continuous surveillance systems, disaggregate traffic data collection is possible as well as readily available. Disaggregate traffic data have been used in only a limited number of studies (Abdel-Aty et al., 2007; Kockelman and Ma, 2007; Lee et al., 2003, 2002; Madanat and Liu, 1995). While detailed vehicle movement data in a section would be the best data source, traffic data from several consecutive detectors in a section can be a good surrogate to identifying traffic dynamics that may lead to accidents (Oh et al., 2001).

1.4 Research question and objectives

The main research question of the thesis is whether and how traffic parameters affect accident patterning, consequences, and response. Thus, the thesis objective is to use highway traffic data collected on a real-time basis in order to:

a. explore the effects of traffic parameters on type of road crash,
b. investigate the influence of traffic parameters on the injury level sustained by vehicle occupants, and to
c. explore possible implications in incident management strategies.

1.5 Interest of thesis

Road safety analyses are of particular interest to societies because of accident intense consequences and increased cost. A markedly extensive body of literature deals with accident frequency and severity, while numerous studies have been addressing incident management techniques optimization. Nevertheless, research efforts on the integration of road safety tools along with incident management techniques remain few. Such efforts would introduce results obtained from safety analyses in an integrated incident management scheme including both proactive and reactive considerations.

Furthermore, real-time traffic data have little been utilized in road safety and incident management analyses. The exploration of the influence of real time traffic variables on accident patterning (in terms of crash type) could provide significant insight in the accident mechanism of occurrence, while proving highly beneficial to traffic and incident managers. Pande and Abdel-Aty (2006) underlined the importance of by-crash-type analysis, particularly when it comes to real-time risk assessment. They suggested that the conditions preceding crashes are expected to differ by type of crash and, therefore, any approach towards proactive traffic management should be type-specific in nature.

Real-time data integration to accident severity analyses offers the possibility to associate accident attributes to the actual traffic flow characteristics at the time of the accident. Based on the analysis of historical data, typical traffic patterns recorded prior to accidents may then act as real-time identifiers (Abdel-Aty and Pande, 2007). Such explorations are useful for both researchers and practitioners in estimating accident and congestion external costs and in transportation planning. Further, such analyses may enable practitioners and authorities to locate hazardous – on severity grounds – spots on the road networks. Finally, they may provide additional insight regarding the factors that contribute to higher probabilities context for severe injuries (given that an accident occurs).

In conclusion, the attempt to further study and develop accident models, and in particular the integration of real-time data, can significantly contribute to the elaboration of a better-structured incident response system with predictive power. Thus, accident counts would decrease and accident consequences would be mitigated. Apart from human lives saved, an economic burden would be taken off from societies; non-recurrent congestion would be decreased, while environmental gains would accumulate.

1.6 Dissertation structure

The dissertation main text is organized as follows:

- Chapter 2 includes a theoretical background related to road safety research. Definitions of key terms are provided, while a classification of road safety literature is attempted. Emphasis is given on issues of interest to the analysis further performed such as data observational studies, disaggregate investigations, and so on.
- In Chapter 3, we examine the effects of various traffic parameters on type of road crash. Multivariate Probit models are specified on 4-years of data (200-2002, 2006) from the A4-A86 highway section in the Ile-de-France region, France. Empirical findings indicate that crash type can almost exclusively be defined by the prevailing traffic conditions shortly before its occurrence.
- In Chapter 4, we apply a random parameters ordered probit model to explore the influence of speed and traffic volume on the injury level sustained by vehicle occupants involved in accidents on the A4-A86 junction in the Paris region. Results indicate that increased traffic volume has a consistently positive effect on severity, while speed has a differential effect on severity depending on flow conditions.
- In Chapter 5, we investigate the introduction of road safety analysis outcomes in an integrated incident management scheme. A synthesis of related studies is performed so as to establish a conceptual framework for incident management applications using real-time traffic data on urban freeways. We use dissertation previous findings to explore potential implications towards incident propensity detection and enhanced management.
- Chapter 6 summarizes dissertation major findings and provides overall conclusions regarding the analysis performed. The overall contribution of the thesis is discussed, while indications for future research are given.

The dissertation main text is followed by a complete list of references, an annex summarizing in Greek language the analysis performed, and an annex providing indicative model outputs.

Chapter 2

Road Safety Literature

This chapter aims at establishing a theoretical background for the upcoming analysis and at summarizing the state of the art in road safety research. Several key terms are defined and basic assumptions are made. The large body of road safety literature is organized on the basis of both methodological and thematic criteria. Finally, the dissertation field of interest is defined with respect to the taxonomy established.

2.1 Introduction

The large body of literature on road safety includes many disciplines, various methods, units and levels of analysis. A general classification of road safety studies is attempted in the effort of establishing a rigorous taxonomy of previous research efforts.

2.1.1 Road safety of an entity

In performing an experiment, the number of successes achieved largely depends on the number of trials performed (*exposure*). In that sense, the number of accidents occurring on a roadway segment largely depends on the amount of travel performed. The exposure is then expressed in terms of vehicle-kilometers, vehicles, and so on. Chapman (1973) defined exposure as *'the amount of or opportunity for accidents that driver or traffic system experiences'.* Carroll (1971) proposed that *exposure* is '*the frequency of traffic events which create a risk of accident*'. Hauer (1982) defined a unit of exposure as '*a trial in which the outcomes are an accident (possibly of several types) or a non-accident*'.

If we know exposure, we can differentiate high accident occurrence due to high risk from high accident occurrence due to high exposure. Unfortunately exposure measurements are expensive and carried out much too seldom while not specifically for road safety purposes. A fundamental problem in studies of accident occurrence is how to combine exposure and accident data in a meaningful and consistent way so that the contribution of individual factors to accident risk can be identified (Jovanis and Chang, 1989).

It is an open question whether safety should be measured in units of dangerous situations (exposure) or accidents (outcome). The traffic conflicts technique assumes proportionality between exposure and accidents. On the contrary, Wolfe (1982) notes that exposure-based definitions suffer from two major limitations: a) they ignore exposure to accidents implicating pedestrians and fixed objects (such as bridge abutments, utility poles, and parked cars), and b) one can be exposed to risk while not participating in dangerous situations (e.g. while being in a parked vehicle). Overcoming such constraints, Hauer (1997) proposed the following definition: '*Safety* *is the number of accidents (crashes), or accident consequences, by kind and severity, expected to occur on the entity during a specified period of time'* (Hauer, 1997).

Described this way, the safety of an entity is a series of expected numbers or frequencies. These expected numbers change in time and are distinct from accident counts, which are a reflection of the underlying expected number that enable us to estimate what the expected number or frequency at some point in time is or was. *Un-Safety (as measured by accident occurrence) is the product of the probability of having an accident (risk) and the number of exposure units (*Hauer, 1997*).*Factors contributing to accident risk are thus conceptualized as affecting the probability of an accident.

2.1.2 Disciplines

Due to the complexity of the road system and its management, road safety analysis necessarily involves numerous scientific disciplines, in order to effectively treat problems of:

- driver behavior in risk situations:
  - economics (insurance, decisions under uncertainty)
  - psychology (perception of danger, choice under uncertainty, driver training)
  - ergonomics (information gathering, man/machine interaction, road/driver task adaptation)
  - physiology (capacities, handicaps for driving)
  - psycho-sociology (attitudes and judgment when confronted with risk, controls and social norms)
  - sociology (cultural and organizational aspects, enforcement system)
- accident mechanism of occurrence:
  - mechanics (traction, vehicle structure)
  - traffic engineering (infrastructure and operations)
  - ergonomics (understanding the road traffic system)
  - physiology (fatigue, alcohol, physical capacities)
- traumatisms during collisions:
  - bio-mechanics (shock resistance)
  - medicine (traumatism severity)

In what follows, four main criteria are used to classify road safety literature in groups of studies; each group is then briefly described, along with some basic assumptions. More elaborate descriptions are provided for methods further used in the present thesis. We note that references made are indicative and not exhaustive. Thorough literature reviews on the thesis field of interest are provided in Chapters 3 to 5. Road safety literature organization enables for interesting remarks that are presented in the end of Chapter 2.

2.1.3 Classification criteria

We propose road safety literature classification by the use of four main criteria: a) the method employed, b) the level of analysis assumed, c) the scope of the performed analysis, and d) the accident phase considered.

a) Method employed

The method employed refers to the scientific approach that is selected with regard to the analysis objectives and to data availability. This may be:

- controlled experiment,
- field observational,
- multidisciplinary in-depth investigation, or
- data observational

For each of these approaches, specific methodologies are developed to appropriately treat available data.

b) Level of analysis

The level of performed analyses is also dependent upon the analysis objectives and data availability. We distinguish between:

- disaggregate and
- aggregate investigations.

Generally, disaggregate analyses over perform aggregate analyses as they provide more accurate results; however they have higher data requirements.

c) Scope of analysis

Some researchers are mostly interested in explaining a present or past situation/event, while others attempt to predict future situations/events. In that sense, an analysis may be either:

- *descriptive* of past and present situations or
- *predictive* of future changes.

We note that the term 'descriptive' is commonly used to refer to explanatory models that may equally refer to present or future conditions. However, in the context of the present thesis, this term is used differently as mentioned above.

d) Accident phase considered

Some researchers focus on the study of accident contributing factors and triggering events, while others focus on accident evolution mechanism. Many researchers have studied different accident response techniques, whereas others mainly worked on accident consequences. Thus, four discrete accident phases may be considered:

- generating,
- patterning,
- response, and
- consequences.

2.2 Method employed

2.2.1 Controlled experiment (laboratory)

Laboratory experiments or simulations are used to study in detail driver and vehicle actions and reactions that may be linked to accidents but are difficult – or impossible- to be observed in the field. Laboratory experiments commonly study actions such as steering wheel movement, lateral and longitudinal position, speed estimation, breathing rate, vigilance, and so on. Driver reactions to external environment stimulus have been simulated and subjected to in-depth analysis. Driving simulators are used for this purpose ; they allow for useful conclusions that would otherwise not have been possible such as changes in the field of eyesight vision. Furthermore, statistical models have been developed to simulate collision kinematical characteristics (Šušteršič et al., 2007). However, the most common safety experiment is the one undertaken by vehicle industries ('crash test'); it measures crashworthiness rates. In all these experiments or simulations, the relationship between independent variables and intermediate measures is directly applied. Then, inferences are made about the effect of independent variables on highway accident risk.

The advantages of laboratory experiments include safety of the subjects, control of some confounding results, and possibly reduced costs compared to field observation. Among all shortcomings of laboratory experiments, the most restrictive is the difficulty in generalizing laboratory findings to actual highway environments. Although laboratory experiments allow us to obtain individual disaggregate and detailed performance data, they are limited in their ability of providing insight in the process of accident occurrence and, obviously, do not contain data on actual involvements. Trials are conducted during the experiments, but they reflect 'pseudo' exposure as no actual risk system exposure is undertaken. In order to relax this constraint, some researchers have tried to combine laboratory results with observational data in order to test their real credibility (Jones and Whitfield, 1988). Hauer (1997) points out the non-existence of road safety field experiments and argues that real on-the-road experiments should be considered.

2.2.2 Field observational

On uninterrupted flow facilities equipped with continuous surveillance systems, researchers applied field observational techniques in order to promptly detect incident occurrences. Incident detection involves the analysis of patterns in the traffic data observed just after the incident in order to develop models that can separate real-time traffic conditions resulting from incidents from free-flow and/or recurring congestion (Abdel-Aty and Pande, 2007). These patterns are identified through continuous loop measurements.

Research on actual driving conditions includes unobtrusive observation of individual drivers and vehicles, and on-road measurements of drivers in instrumented vehicles. Emerging technologies, such as CCTV, are used in this purpose. In Liu (2007), laser speed guns were used to measure the speed of oncoming vehicles. Cameras have also been used to record driver behavior and characteristics in an effort to associate approaching speed with driver and vehicle characteristics. However, accidents are rare

events, while continuous field observations are expensive. Thus, data on accident occurrences are not usually used; instead inferences about risk are made.
The traffic conflicts technique assumes proportionality between the frequency of dangerous situations and accident occurrences. It is based on defining near accidents ('conflicts'), which are typically expressed as time to collision of the involved road users. The advantage of this method is that data can be collected quite quickly. The disadvantage is that the validity is lower compared to real accident analyses. In order to estimate the number of accidents based on conflicts registration, fixed ratios between the number of conflicts and accident counts are used. The latter implies that conflicts can also be regarded as a measurement of exposure as well as an indirect estimate of the number of accidents.

The major advantage of on-the-road research is that results obtained are disaggregate and readily applicable to highway environment. A major disadvantage is that many variables are not under strict experimental control, and some results may be due to uncontrolled variables and/or be limited to the specific location where the study was conducted. In addition, it is not always possible to directly relate such observations to accident occurrences. Exposure to risk remains limited as accidents are rarely observed, while budget constraints make it difficult to conduct more comprehensive research (i.e. continuous surveillance over long highway segments and periods of time). On the other hand, incident detection techniques have become rather obsolete as the road users can promptly contact (by mobile phone etc.) highway authorities in case of emergency.

2.2.3 Multidisciplinary in-depth investigation

In-depth investigation mainly refers to on-site investigation by specially trained technicians, who rush to the accident site immediately after its occurrence. In-depth investigation is held by multidisciplinary teams and a compilation of results is made based upon their findings (OECD, 1988). Emphasis is always given on individual accident analysis. However, we should note that the analysis may take place long after the accident occurrence. In this context, clinical studies of individual accidents (Shinar, 1998) are also part of this category. Reconstruction of the accident and of the actual circumstances under which it took place is attempted and thoroughly studied. Within this procedure, individual road accidents are treated as essentially deterministic events; although incomplete information can leave one uncertain about how exactly an accident happened (Davis, 2004).

The main difference between multidisciplinary in-depth investigations and field observational studies is that the first do not deal with exposure and a priori assume incident occurrences. Also, they are unable to follow the actual incident patterning since generating; instead, they attempt to – a posteriori – reconstruct and approximate accident actual circumstances of occurrence. Such investigations require significant human resources and a wealth of data that is rarely available – or traceable – in road incident occurring. They seem more suitable for less frequent and more elaborately recorded events such as industrial incidents or airplane crashes.

2.2.4 Data observational

Data observational studies are hypothetical experiments conducted through econometric modeling. Econometric modeling techniques were initially developed for assessing complex economic relationships and can be viewed as a scientific substitute for perfectly controlled experiments. They are applicable in cases where it is impossible to vary one independent variable at a time, while keeping all others constant. They may be defined as the use of statistical reasoning and methods as means to establish data-based descriptions of economic phenomena and empirically based counterparts for, and tests of, economic theories (Smelser et al., 2001).

The essence of econometrics lies upon the mix of subject-matter theory, mathematical statistics, and empirical data. It is not a technique designed to explore the kind of empirical relationships that might possibly exist between the arbitrary set of variables, but to estimate the parameters of a given theory. This theory must come from somewhere else than the data at hand themselves, or the whole analysis will be more or less invalidated due to circularity of argument (Heckman and Leamer, 2007). Errors due to confounding effects can be avoided only to the extent that the relevant explanatory factors are included in the model.

Usually, analysts combine accident data with controlled exposure and test the hypothesis of interest. Regression and multivariate models have been employed to describe accident occurrence. Studies about the safety effects of interventions are usually retrospective quasi-experiments. To assess the effect of some treatment on the safety of an entity, one must predict what would have been the safety of an entity, had it been left untreated. In observational studies, this can be done in a variety of ways (e.g. comparison group).

2.2.4.1 Models

Models are formulated in order to increase our understanding of observed phenomena. Most analytical models contain some basic assumptions from which conclusions are logically deduced. As such, models are a restricted form of general theories, often containing hypotheses, postulates, and assumptions used to test these theories empirically. A theory is formulated to explain and predict regularities. The research cycle could start with observations from which regularities, through induction, are formulated into theories. Alternatively, theories are formulated through deductive processes into testable hypotheses, subsequently verified with observations. A model deriving from an acceptable theoretical construct is expected to possess the following basic characteristics (Hakim et al., 1991):

- description of the phenomenon
- explanation of the phenomenon
- prediction of the phenomenon
- incorporate policy variables.

In reality, it is possible to build partial models that are able to produce accurate predictions without necessarily explaining the phenomenon. Alternatively, some less sophisticated models could describe the phenomenon without actually being able to explain or predict it.

2.2.4.2 Statistical analysis

A statistical analysis is essentially a logical argument, where assumptions about the process generating data are combined with observation statements in order to derive statements about quantities not directly observed. Differences between what has been assumed and how the data have actually been produced could then invalidate any conclusions drawn from the analysis (underlying assumptions).

A random experiment is an observation activity that can be repeated under identical conditions, where the set of possible outcomes is known, but where the outcome of any particular repetition is unknown in advance. Hacking (1964) notes that there are situations which tend to reproduce, under repeated operation, stable relative frequencies of outcomes. He calls this tendency chance, and the corresponding situation a chance set-up. Probability theory then provides a logic for reasoning about chance. As Hauer (1982) noted, one can readily apply Hacking's treatment to the study of road accidents by assuming, in the simplest case, that a section of road or an intersection can be modeled as a chance set-up. Individual vehicle traversals provide the trials to which the set-up is subjected; each trial is assumed to have a chance of resulting in an accident. This chance may vary with roadway, traffic or environmental conditions. Thus, we can empirically study how an intervention affects road safety and the chances of accidents without knowing the details of any particular accident.

Statistical approaches assume that road accidents are individually unpredictable, chancy phenomena although aggregates of accidents can show predictable statistical regularities. Therefore, accidents are treated as individually random, although the parameters governing their probability distributions may be modeled deterministically (Davis, 2004). The majority of road safety statistical studies are based upon this assumption, even though some objections have been raised.

One problem is that without prior knowledge concerning the underlying mechanism generating the aggregated data, it may be difficult – or even impossible – to correctly interpret aggregated results. The Simpson's paradox (Simpson, 1951) arises in the interpretation of contingency tables when an association between variables observed in sub-populations is attenuated or even reversed when the sub-populations are aggregated (Freedman, 1999). Davis (2004) suggests that the statistical regularities observed in accident data have no independent status, but are simply the result of aggregating particular types and frequencies of mechanisms. Based on statistical applications in other fields, Davis (2004) proposes an alternative to the prevailing assumption in accident statistical analysis. This alternative treatment begins with the idea of a population of individual deterministic mechanisms, rather than the idea of repeated trials of a chance set-up.

A comprehensive literature review on accident statistical modeling can be found in Garber and Wu (2001). The authors state that statistical models have been used in road safety research for three main purposes:

a) to identify the major contributing factors,
b) to establish relationships between crashes and explanatory variables, and
c) to screen covariates.

Statistical models can be divided into deterministic and stochastic. In deterministic modeling (Foldvary, 1979; Gwynn, 1967; Hall and Pendleton, 1989; Lundy, 1965), the influence of a variable is often examined by keeping all other parameters fixed (single-variate models). In other cases, multiple linear regression, robust regression and multivariate ratio of polynomials models were applied in multivariate analyses of the causal factors (e.g. Garber and Ehrhart, 2000; Mohamedshah et al., 1993). All deterministic models have a strong underlying assumption; they assume that the error of the independent variable is normally distributed with a constant variance. Also, they assume that the dependent variable is continuous.

Stochastic models amplify the number of independent variables treated. Despite the stochastic treatment of the dependent variables, the link functions (that connect the mean of number of accidents or the severity outcome to contributing factors) remain deterministic. Several modeling specifications have been applied.

<u>In frequency analyses:</u>

- Poisson (Ivan et al, 1999; Ivan and O'Mara, 1997; Lord et al., 2005; Oh et al. 2004, 2006; Saccomanno and Buyco, 1988; Vogt and Bared, 1998),
- Negative Binomial (Abdel-Aty and Radwan, 2000; Carson and Mannering, 2001; Donnell and Mason, 2006; Garber and Wu, 2001; Hadi et al., 1995; Hiselius, 2004; Karlaftis and Tarko, 1998; Knuiman et al., 1993; Lord, 2006; Lord et al., 2005; Maher 1991; Maher and Summersgill, 1996; Martin, 2002; Miaou, 1994; Milton and Mannering, 1998; Oh et al. 2004, 2006; Persaud et al., 2002; Poch and Mannering, 1996; Shankar et al., 1995; Tunaru, 1999),

- Zero Inflated Poisson and Zero Inflated Negative Binomial (Garber and Wu, 2001; Kumara and Chin, 2003; Lee and Mannering, 2002; Miaou, 1994; Shankar et al., 1997).

An overview of such studies can be found in Xie et al. (2007).

In severity analyses:

- Linear regression (Boufous et al., 2008),
- Ordered Probit model (Chimba and Sando, 2009; Xie et al., 2009; Gray et al., 2008; Pai and Saleh, 2008; Lee and Abdel-Aty, 2005; Yamamoto and Shankar, 2004; Kockelman and Kweon, 2002; Quddus et al., 2002; Zajac and Ivan, 2002),
- partial proportional odds model (Quddus et al., 2009; Wang and Abdel-Aty, 2008),
- Logit model (Abdel-Aty, 2003 ; Eluru et al., 2008 ; Kim et al., 2008 ; Milton et al., 2008 ; Savolainen and Mannering, 2007 ; Shankar et al., 1996).

An overview of such studies can be found in Chapter 4 of the present dissertation.

2.2.4.3 Non-parametric methods

Data mining can be defined as the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in large amounts of data (Fayyad et al., 1996). From a statistical perspective, it can be viewed as a computer automated exploratory data analysis of (usually) large complex data sets (Friedman, 1997). However, in contrast to statistical techniques, the problems and methods of data mining have some distinct features of their own. First, data sets may be much larger than in typical statistical analyses. Second, data mining pays much less attention to the large-scale asymptotic properties of its inferences; instead, emphasis is given on the general philosophy of ‘learning’, including consideration of the complexity of models and the computations they require (Hosking et al., 1997). Furthermore, data mining has tackled with problems such as what to do in situations where the number of variables is so large that looking all pairs of variables is computationally infeasible (Mannila, 2000). Additionally, in contrast with statistics, data mining is typically a form of secondary analysis: the data has been collected for some other purpose than for answering a specific data analytical question (Geurts et al. 2005).

Data mining techniques can be divided in two: the computational (supervised learning) and the non-computational (unsupervised learning) techniques. Each technique has its strengths and weaknesses in terms of representation language, classification power, descriptive abilities and expert knowledge required.

Supervised learning

Rule induction is used to identify rule sets representing interesting subgroups in accident data (Kavsek et al., 2002). The learning is based on past experience, and the learned knowledge is used to classify new data. Applications of computational techniques in accident modeling include:

- Decision trees (Clarke et al., 1998a ; Strnad et al., 1998),
- Neural networks
  - Artificial NN (Abdel-Aty and Pande, 2005; Abdelwahab and Abdel-Aty, 2002; Awad and Jason, 1998; Chang, 2005; Delen et al., 2006; Mussone et al., 1999; Mussone et al., 1996; Riviere et al., 2006),
  - Probabilistic NN (Abdel-Aty and Pande, 2005),
  - Back Propagation NN (Chang , 2005),
  - Time delay NN (Zhong et al., 2004),
  - Bayesian NN (Riviere et al., 2006 ; Xie et al., 2007),
  - Genetic Algorithms (Gas) (Clarke et al., 2005, 1998b),
  - Spatial data mining (Zeitouni and Chelghoum, 2001),
  - Association algorithm (Geurts et al., 2005),
- Fuzzy methods (Jilani and Burney, 2007; Song and Chissom, 1993; Vaija, 1987),
- Hybrid methods (Neuro-fuzzy) (Awad and Jason, 1998).

Unsupervised learning

Non-computational learning is based on the statistical regularity of the patterns recognized in data. Clustering techniques are used to discover frequent patterns in accident data (Ljubic et al., 2002). Tree-based methodologies have been shown as useful tools to obtain homogeneous data sets in accident analysis and to establish the empirical relationship between traffic accidents and independent variables. They result from the statistical regularity of the patterns recognized in data and are used for explaining and/or predicting either a categorical or a continuous response. They have commonly been applied to reduce the heterogeneity in accident data (Abdel-Aty et al., 2005; Chang and Chen, 2005; Hakkert et al., 1996; Karlaftis and Golias, 2002; Maggazù et al., 2006; Park and Saccomanno, 2005; Stewart, 1996). The tree structure is very helpful in clarifying the relationships between independent variables and accidents, along with the interactions among independent variables. The ability of graphically displaying results is very advantageous.

2.2.4.4 Parametric vs. non-parametric models

Parametric models have been widely preferred from analysts because they have explicit theoretical foundations, they can produce interpretable coefficients for each explanatory variable of the model, and they can be easily estimated. Nevertheless, parametric procedures require the functional form of the model to be specified in advance, they are not invariant with respect to monotone transformation of the variables, they are easily and significantly influenced by outliers, they do not handle well discrete independent variables with more than two levels, and they are adversely affected by multicollinearity among independent variables (Hadi et al., 1995; Karlaftis and Tarko, 1998).

The application of non-parametric models for accident data modeling has received much less attention. The primary reason is the complexity in estimating these models. Other criticism that has impeded on ANN broad use concerns the over-fitting observed in small samples (Vogt and Bared, 1998). This constraint limits their transferability and applicability for crash predictions, even though these models possess better linear and nonlinear approximation abilities than statistical regression methods. We should note, though, that approaches able to alleviate this problem do exist. The most important problem that analysts confront with ANN is that they essentially work as black-boxes and do not generate interpretable parameters for each explanatory variable. In some studies (Delen et al., 2006; Fish and Blodgett, 2003; Xie et al., 2007), sensitivity analysis was performed to deduct relationships between dependent and explanatory variables (interdependent or not to each other). Sensitivity analysis is empirical and cannot be considered equivalent to the GLM statistical inference. Moreover, it is usually time consuming to develop an ANN model, because it is made by experimentation; the computation time is also superior to regression models and greatly depends on the size of the training data set. The most significant advantage of ANN methods is that they perform without having to establish the functional form linking the dependent and explanatory variables. Furthermore, they can approximate any continuous function defined on a compact set with arbitrary accuracy, though this strong ability may lead to over-fitting. Another advantage is that they can handle interrelation problems between independent variables.

Tree-based methodologies also present theoretical and practical advantages compared to parametric models as they neither require the functional form of the model to be specified in advance. They can handle collinearity problems, while the assumption of additive relationship between risk factors is not required. Outliers are isolated into a node and do not contribute to splitting, so they do not affect the coefficient estimates, as in GLM models. They are adaptable in dealing with high dimensional and non-homogeneous data sets. However, as discussed by Harrel (2001), they do not utilize continuous and ordinal variables effectively. Also, they have the disadvantage of over-fitting in three directions: searching for predictors, the best splits, and multiple searches. They do not provide a probability level or confidence interval for the risk factors and predictions. When they are employed to analyze a new data set, the lack of formal statistical inference procedures is a critical issue (Chang and Chen, 2005). In addition, the simple binary tree appears to have difficulty in handling the interactions between risk factors. A further drawback is the difficulty in doing elasticity analysis in order to acquire information on the marginal effects of the variables.

2.3 Level of analysis

The level of analysis is of great importance in road safety explorations regardless the unit of the analysis that may itself be highly differentiated. For example, Jovanis and Delleur (1981) and Mountain et al. (1998) have analyzed specific accident locations such as links or intersections. Other researchers such as Saccomanno and Buyco (1988), Chirachavala and Cleveland (1985), and Woods and Simms (2002) have studied specific vehicle types (e.g. trucks). The analysis level refers to data aggregation that affects both dependent and independent variables. A common problem in road safety analyses is to successfully combine disaggregate dependent variables – such as accident counts – with aggregate regressors such as weather data.

2.3.1 Disaggregate

Disaggregate explorations consider an individual observation as the unit of the analysis. The individual observation may be a moving automobile, a single driver, an accident count, and so on. Disaggregate methodologies may be applied in case studies analyses, experiments or simulations. Markedly, disaggregate level of analysis refers not only to the dependent variables, but to regressors as well.

2.3.1.1 Unit of the analysis

The unit of analysis could be an intersection or a roadway segment, a specific road user category (e.g. pedestrians) or all road users, any vehicle involved in accidents or all vehicles on the roadway, and so on.

Fixed-length sections versus homogeneous road sections

An important issue of concern is road network segmentation for the analysis purposes; two are the most common alternatives: a) fixed-length sections or b) homogeneous sections (in terms of geometric or other characteristics).

In order to account directly for the effects of highway geometric characteristics on accidents, homogeneous sections should be preferred. Homogeneous sections are often considered to limit the observed heterogeneity. Miaou and Lum (1993) pointed out several problems in using homogeneous sections; this segmentation technique can result to short road sections (especially curved and graded ones) that may have undesirable impacts on the estimation of the linear regression. The homogeneity requirement may exacerbate potential heteroskedasticity problems and lead to losses in estimation efficiency. The resulting increase in the standards error of model coefficients could lead the analyst to draw erroneous inferences with regard to the effects of model covariates.

On the other hand, fixed-length sections limit potential heteroskedasticity due to unequal sample sizes, while being much easier to be defined. In addition, the migration of accidents is better accounted for. Nevertheless, heterogeneity problems among segments may arise and the impact of road geometry cannot be easily observed. In any case, the disadvantages of using fixed-length sections, relative to homogeneous sections, are far less severe (Shankar et al., 1995).

2.3.1.2 Dependent and independent variables

The dependent variable is chosen with regards to the study specific objectives. In the greatest part of the literature, dependent variables cluster in two large categories: a) crash counts-related, and b) severity-related. In both cases, research focuses on modeling relationships between these dependent variables and independent variables; independent (or explanatory) variables are factors thought to be related to accident occurrence and whose influence is being investigated (i.e. the nature and magnitude of their effect).

The choice of explanatory variables depends upon many criteria such as theoretical assumptions made, overall study objectives, data availability, and so on. Another issue of interest is the number of independent variables considered; small numbers offer a gain in transferability of results, whereas large numbers offer a considerable statistic gain on the detriment of generalization. Any variable whose value is not supposed to be held constant during the hypothetical experiment should not be included among regressors. Independent variables should not be affected by road user decisions; otherwise, endogeneity bias occurs. On the contrary, omitted variable bias occurs whenever a regressor is correlated to some relevant explanatory variable not included in the model. A third issue of concern is the correlation among independent variables. The problem of multicollinearity arises when one independent variable can be expressed as a linear function of others. In that case, the influence of each variable cannot be easily estimated as regressors tend to vary together within the sample. A common practice is to omit (if possible) some of them. Another way to address the problem is by applying multiplicative decompositions; alternative expressions of variables (e.g. log) in order to eliminate the linearity property.

Crash counts-related models

Studies that explore the influence of several factors on accident counts are the most common road safety analyses. The dependent variable is either an expected frequency (counts/unit of time) or an expected rate (counts/unit of exposure). The expected frequency is often expressed in accident counts per year, month, week, or day with respect to the aggregation level considered. The expected rate can be expressed in accident counts per capita, per vehicle, per road segment, and – even more accurately – per vehicle kilometers traveled. In that context, previous research has dealt with modeling relationships between accident occurrences and:

- geometric elements (Garber and Wu, 2001; Karlaftis and Golias, 2002; Knuiman et al., 1993; Lundy, 1965; Miaou et al., 1992; Okamoto and Koshi, 1989; Shankar et al., 1995; Vogt and Bared, 1998; Wong and Nicholson, 1992),
- prevailing traffic conditions (Carson and Mannering, 2001; Chang and Chen, 2005; Frantzeskakis and Iordanis, 1987; Garber and Wu, 2001; Hall and Pendleton, 1989; Lave, 1985; Oh and Chang, 1999; Oh et al., 2000),
- weather (Chang and Chen, 2005; Fridstrom and Ingebrigtsen, 1989; Ivey et al., 1981; Jovanis and Delleur, 1981),
- roadway environment in terms of lighting conditions, warning signs, pavement characteristics, and so on (Carson and Mannering, 2001; Karlaftis and Tarko, 1998; Lee and Mannering, 2002; Martin, 2002, Taylor et al., 2000), and
- public policy in terms of traffic regulations, speed limit, enforcement, hours-of-service for professional drivers, and so on (Hall and Mukherjee, 2008; McCarthy, 1999; Naveh and Marcus, 2007; Navon, 2003; Richter et al., 2004; Yannis et al., 2007).

Severity-related models

Numerous studies explore the factors having an impact on accident severity. The dependent variable is usually depicted on a 3- or 4-point ordinal scale (e.g. no injury, severe injury, fatal). The unit of analysis varies across studies and depends on the study objective; units include non-motorized road users (Ballesteros et al., 2004; Kim et al., 2008; Sze and Wong, 2007), crashes (Chang and Wang, 2006; Eluru et al., 2008; Gray et al., 2008; Wang and Abdel-Aty, 2008; Yamamoto and Shankar, 2004), and so on.

The independent variables considered usually include:

- driver (or rider) characteristics (Boufous et al., 2008; Dupont et al., 2010; Helai et al., 2008; Kim et al., 2008; Lapparent, 2008; Pai, 2009; Sze and Wong, 2007; Wang and Abdel-Aty, 2008; Yamamoto and Shankar, 2004),
- vehicle characteristics (Ballesteros et al, 2004 ; Dupont et al., 2010 ; Helai et al., 2008 ; Kim et al., 2008 ; Pai, 2009),
- road geometry (Al-Ghamdi, 2002; Milton et al., 2008; Savolainen and Mannering, 2007; Shankar et al., 1996),
- crash characteristics related to the exact circumstances under which the accident occurred (Chang and Wang, 2006; Gray et al., 2008; Helai et al., 2008; Wang and Abdel-Aty, 2008; Yamamoto and Shankar, 2004; Yannis et al., 2010), and
- other variables such as speed limit, day of the week, time of day, AADT, and traffic conditions (Abdel-Aty, 2003; Conroy et al., 2008; Gray et al., 2008; Helai et al., 2008; Kim et al., 2008; Milton et al., 2008; Pai and Saleh, 2008; Pai, 2009; Savolainen and Mannering, 2007; Sze and Wong, 2007; Yamamoto and Shankar, 2004; Zajac and Ivan, 2002).

2.3.2 Aggregate

Aggregate studies cluster individual observations in order to infer statistical properties for the whole group. Clustering may refer to accident counts (under similar conditions, on the same day, and so on), road segments (neighboring or sharing similar characteristics), individuals (drivers or other road users involved in the same accident and so on), prevailing conditions (traffic, weather, lighting, and so on). Aggregation may equally refer to the period of reference considered. Efforts are made to shorten the period of reference in order to gain in accuracy; daily, hourly, or minute-intervals averaged measures are preferred over annual ones.

In aggregate accident data, the random effects (noise, disturbance) which have a decisive impact at the micro level, are 'evened out' by the virtue of the law of large numbers. The causal process determines the expected number of accidents, as a function of all factors making up the causal set. The variation due to causal factors is systematic and can be influenced by policy measures. Natural data variation may result in heavy underdispersion (Pasupathy et al., 2000). Furthermore, ecological fallacy may arise whenever an observed statistical relationship between aggregated variables is falsely attributed to the units over which they are aggregated (Davis, 2002). Generally, aggregation makes results less detailed; nevertheless, aggregate studies are commonly performed because of data limitations such as sample data insufficiency.

Aggregate observational accident studies are based on cross-sectional (spatial) or time-series (temporal) variation, or a combination of both (panel data).

2.3.2.1 Cross-sectional models

Cross-sectional models link (frequency- or severity-related) outcomes to entities-specific characteristics by making use of the variation among entities; the variation is observed at the same point of time. The 'entity' can be any kind of geographically defined unit, or any sort of identifiable physical or institutional object such as a person, a family, a company, a vehicle, or a group of such micro-units exhibiting certain common characteristics. Cross-sectional analysis is based upon a very restrictive assumption; i.e. the entities are not different in any other way than what is captured by the variables of the model. However, not all variables – that vary across sites and affect road accidents – are identifiable and measurable. If unmeasured, such variation may cause biased estimations. Cross-sectional data sets have beneficial properties that help in parameter estimation such as a) variation that exists in the explanatory variables across observations (often without strong covariation) and b) large population. Therefore, analysts perform cross-sectional studies for description and comparison purposes. We note, however, that cross-sectional studies have suffered severe criticism as to their use for prediction purposes.

2.3.2.2 Time-series models

A time-series is a chronological sequence of observations of the same phenomenon; each observation referring to the same period of time (hour, day, month, year, and so on). Time-series modeling is based on the assumption that the historical values of a variable provide an indication of its value in the future (Box and Jenkins, 1970). Road safety analysts explore time-series data in an effort to detect patterns in accident occurrences that could serve for prediction purposes. In time-series modeling, variation among observations may be too small if the data show collinearity among regressors. Additionally, time-series models may exhibit autocorrelation (correlation between successive disturbance terms) due to omitted variables. Furthermore, in some cases, an observation may be dependent upon previous observations (autoregression). Several techniques have been developed to address autocorrelation and autoregression problems. In macroscopic road safety evaluations, time-series models are often considered superior to cross-sectional models, because the latter do not take into account the geographical and cultural differences between countries, states or provinces (Page, 2001).

2.3.2.3 Panel data models

Panel data models exploit combined cross-sectional and time-series data sets; i.e. repeated chronological observations on a given cross-section of entities. Special techniques have been developed for panel data treatment (Hsiao, 1993). Panel databases are the richest source of information. Panel data analyses show several advantages over cross-sectional or time-series analyses (Hsiao, 1993). From a statistical perspective, by increasing the number of observations, panel data have higher degrees of freedom and less collinearity (particularly in comparison with time-series data); the efficiency of parameter estimates is improved. Moreover, panel data allow the researcher to distinguish between time and sectional trends within data (Karlaftis and Tarko, 1998). However, panel data may suffer from heterogeneity and heteroscedasticity resulting to decreased estimation efficiency.

2.4 Scope of analysis

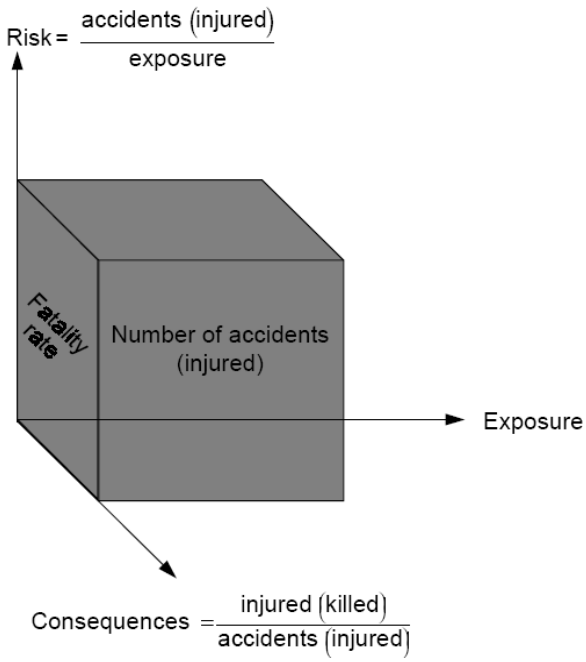
The scope of road safety analyses may be to describe a present situation (in terms of safety performance) of a road network on a local, regional, national or international level. Alternatively, the analysis scope may emphasize on predicting future safety levels with regards to vehicle fleet evolution, road safety measures' implementation, infrastructure upgrade, and so on. In light of the above, the scope of road safety analyses is either descriptive or predictive.

2.4.1 Descriptive

Accurate descriptions and safety evaluations (over a road segment, a network, and so on) can prove useful in many ways. They help in assessing any progress made and, also, in establishing quantified targets. An integral index could be of use in describing actual conditions and in performing comparisons between networks, regions or countries; however, no such index meeting general acceptance has been defined. Microscopic risk analysis and description on the individual level are also important in exploring accident occurrences. Connections between macroscopic theories and individual models remain limited.

Rumar (1999) proposed a three-dimensional cube to describe the road safety problem (Figure 2); the three dimensions being a) *exposure* (magnitude of the activity that results in accidents), b) *accident risk*, and c) *consequence*. The magnitude of the problem is then the product of these factors. Exposure is measured in various ways such as population size, vehicles, network size, drivers, vehicle kilometers traveled, and so on. Risk is defined as the probability of accident occurrence per units of exposure, while consequences refer to injuries. Factors influencing exposure include: economic situation, GDP per capita, urban population density and other demographic factors, modal split, travel route, trip length, traffic decomposition, and so on. Factors influencing risk are related to driver (speed, alcohol, age, gender), to groups of road users (protection), to vehicles (vehicle type), to roads (type of road, maintenance), to environment (darkness, weather), and so on. Finally, factors effecting severity may be related to human factors (speed, alcohol), vehicle (active and passive safety), crash protective roadsides, guardrails, barriers, emergency care, health system, and so on.

Figure 2 The Road Safety Problem (Rumar, 1999)



Descriptive analyses that are being exclusively based on absolute numbers of accidents, injuries or fatalities can be reality distorting. Fatalities may increase while risk is decreasing just because the decrease in risk does not fully compensate the effects of exogenous factors (Page, 2001). In order to be able to compare and rank road safety problems, it is necessary to estimate the magnitude and character of the activities that generate the problems – i.e. the exposure. Exposure can be directly extracted from traffic measurements aggregated in time and space. Alternatively, it may be estimated by induced exposure techniques. In these techniques, the relative weights of both exposure and risk level are interpreted and estimated from accident data or by case control studies, where the risk is estimated directly by comparing samples with equal exposure. Quasi-induced exposure techniques – exclusively based on accident data – have also been developed. The quasi-induced exposure method was first introduced by Haight (1971) and its applications have received considerable attention ever since. The attractiveness of the quasi-induced exposure method lies upon the simplistic nature of the theory and its independence from data requirements associated with traditional exposure measures. Nevertheless, this method has suffered severe criticism and analysts are advised to use it very carefully (Jiang and Lyles, 2007).

2.4.1.1 Individual risk evaluations

In a microscopic level of analysis, risk factor models have been developed to analytically describe risk on the individual level. These bottom-up models try to describe processes of individual behavior or to demonstrate interactions among different elements of the transportation system. The main objective is to identify all technical and human failures in the traffic- vehicle- road interaction that lead to collisions, as well as to quantify their influence. Approaches to risk quantification can be found in various disciplines such as psychology, sociology, ergonomics, medicine, biomechanics, and physics. Overall, we can distinguish between technical models and human-factor models; the vehicle (e.g. size, brakes, stability), the road (e.g. geometry, surface, intersections), and the traffic (e.g. volume, speed, gaps) being considered as situational stimuli to driver behavior.

Behavioral models

Behavioral models are human-factored models that focus on road user risk perception and response; they search to identify if the failure occurred at perceiving, accepting, or at controlling the risk, while associating accident involvement to specific behaviors. Behavioral models can be classified into: input-output (action) models; task analysis models (taxonomic); functional models (e.g. cognitive, motivational, adaptive, mechanical). Action models are behavioral models based on variables related to user disposition, user assimilation, or user situation. Most action models are structured following Rasmussen's hierarchical model that differentiates among knowledge-, rule-, and skill-based errors.

In road safety literature, the most frequently discussed behavioral risk model is Wilde's (1982) risk homeostasis model; a model relating risk perception to risk acceptance. According to this theory, risk taking behavior involves an attempt to balance perceived risk and desired risk. In particular, people seem to adjust their behavior in response to changes in perceived risk (Adams, 1985). Risk-homeostatic theory leads to the conclusion that the only measures having a permanent effect on accidents are those that alter attitudes to risk taking. We should note though that this theory has been debated extensively in the literature.

2.4.1.2 National and international evaluations

Numerous factors influence the safety level as measured on a country scale; some being endogenous to road safety performance, while others being exogenous. Such factors are concerned with road safety policy, distribution and crashworthiness of vehicle fleet, road network characteristics, human behavior and attitudes, and so on (Brenac, 1989; Brühning, 1995; Sivak, 1996; Vanlaar and Yannis, 2006). The evaluation of safety actions by international comparisons consists in evaluating the performance of a country; i.e. the effectiveness of endogenous factors in each country, the exogenous factors being neutralized (Pages, 2001). The performance is the ability of a road safety policy to be effective and the ability of a population to accept and respect this policy.

National crash counts and absolute number of fatalities do not provide the necessary information in order to perform comparative evaluations. Even exposure measurements (exogenous factor) are not considered to be sufficient for data standardization (Andreassen, 1991). A common practice in international comparisons is to use time-series of accident or fatality rates in an effort to reveal safety improvements, while assuming constant all exogenous factors. Nevertheless, time-series comparisons are dependent upon the initial level of mileage and on the origin and end periods of comparison (Andreassen, 1991).

Several researchers (Bester, 2001; Oppe, 1989) have worked on the issue of assessing safety performance on a large scale; a comprehensive literature review can be found in Al-Haji (2005). Most researchers used national motorization level as the main independent variable in their analysis. Smeed (1949) compared data from twenty countries and found an inverse relationship between traffic risk (expressed as fatality per motor vehicle) and motorization level (number of vehicles per inhabitant). Since 1949, many studies have been based on Smeed's formula (Jacobs and Hutchinson, 1973; Mekky, 1985); an overview of these studies can be found in Elvik and Vaa (2004). However, several authors have criticized Smeed's simplistic assumption about motorization being the only independent variable in the model (Broughton, 1988; Koornstra and Oppe, 1992). Thus, more elaborate models have been developed in order to include the influence of additional variables. Al-Haji (2005) identified a set of eleven macro-indicators that affect road safety level and, thus, defined a composite (multidimensional) index. Page (2005) established a country-performance indicator by comparing the mean performance of 21 OECD countries and over a 15-years period.

2.4.1.3 Before-after studies

Effectiveness evaluations (before-after studies) investigate and compare the safety level of an entity before and after the implementation of a safety measure. The entity may be a road segment, a highway or a whole network. Safety measures vary in scale accordingly; from simple road lighting of a crossroad to enhanced enforcement of the national network. Before-after studies provide an estimate of the effectiveness achieved under specific conditions and possess very limited predictive power.

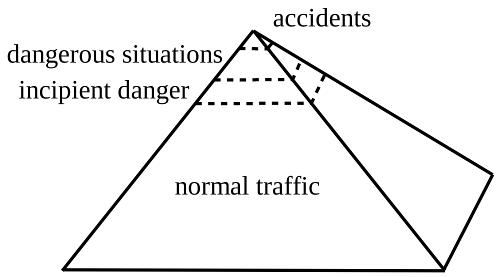
2.4.2 Predictive

Predictive models are used in cases where a large number of factors are involved and/or when these factors cannot be controlled through experimental design. The objective is to estimate an equation that relates independent variables of interest to accidents. Thus, any future change in independent variables has a measurable effect on accident occurrence; thus, predictions can be made. However, accidents are random events and, consequently, all analyses should be based on explicitly probabilistic models. Single events may occur at random intervals, but their long-term frequency may remain constant. Therefore, single events are impossible to be predicted, but their overall frequency may be estimated.

2.5 Accident phase

Any accident may be viewed as a continuum of interrelated consecutive events; road safety explorations focus on different phases of an accident mechanism of occurrence. The sequence of such events begins with the occurrence of some type of danger or conflict. This danger may be due to weather, traffic, or other causal factors. Accidents are a subset of these dangerous situations and, thus, possess lower frequencies of occurrence. Hauer (1997) proposed a four-level pyramid (Figure 3) to represent the continuum of events leading to accidents; each level volume indicating frequency of occurrence. A triggering event – such as driver misestimating potential dangers – provokes the transition from dangerous situations to accident occurrences and initiates accident mechanism of occurrence. *Accident generating* is the phase beginning from any dangerous situation and ending with an accident occurrence.

Figure 3 Accident generating (Hauer, 1997)



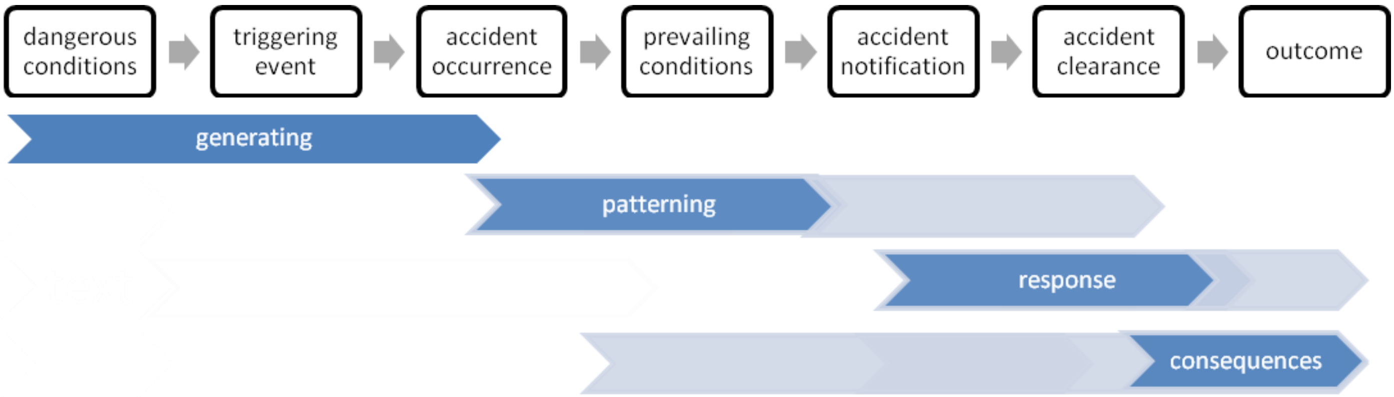
Given that an accident occurs, prevailing conditions at the proximity of the accident site play an important role to its patterning. Prevailing conditions refer to a wide range of characteristics such as weather, road geometry, traffic decomposition, and so on. On a high-speed traffic stream, for example, a collision may be more severe compared to low-speed environments. Prevailing conditions may also act as triggering events for new accident occurrences, i.e. secondary accidents. Accident patterning is the phase during which the accident is evolving with respect to prevailing conditions and lasts approximately until accident notification.

Accident patterning is followed by a response phase starting from accident notification and lasting until the complete accident clearance and the full restoration of roadway capacity to its normal level. This phase includes dispatching of a response unit (RU) to be assigned to service the incident, its arrival on the site, and all actions needed to restore roadway capacity. These actions may be related to site management, traffic management, motorist information, clearance, and so on (Bunn and Savage, 2003). Response phase is critical in determining accident consequences.

Accident consequences occur during all accident phases and include road infrastructure deterioration, delays and congestion, property damage, human losses, and so on. However, accident consequences are measured and examined mainly after the accident clearance, when accident duration can be accurately estimated and all consequences are revealed. Police reports register the main consequences suffered by vehicles and road users. Medical services examine and record data on the health condition of implicated persons.

In accordance to the above, accident mechanism of *events* can be graphically represented as follows (Figure 4).

Figure 4 Accident mechanism of events



<u>2.5.1 Generating</u>

In the accident generating process, the interest lies upon determining *if* and *how* the risk system affects whether or not an accident will occur. Analysts explore the successive transition from normal traffic conditions to dangerous conditions and from dangerous conditions to accident occurrences. In this context, emphasis is given on conflicting traffic conditions, adverse weather, driver fatigue, and all parameters that may create potential hazards. Moreover, triggering events are investigated in an effort to understand accident mechanism of occurrence and to implement appropriate measures for their mitigation. A large number of studies attempt to identify accident causal factors and to quantify their impact. In great part of the literature, every accident is considered to result from a combination of: driver, vehicle and roadway/operational environment mistakes or failures. However, the assumption of time-increasing danger (up to the accident occurrence) holds in all of these studies; while the triggering event is considered to be random.

Instead of this causal chain-based approach, some authors addressed the problem in a more integrated way by adopting systemic approaches. Jovanis and Chang (1989) formulated a model of accident occurrence using principles from survival theory.

They viewed the accident as a 'system failure' and introduced hazard models in road safety research. The latter had been widely used in biomedical studies to detect the effect of specific medical treatments on the lifetimes of observed patients. The probability of failure at any time is determined by the total hazard contributed by the level of each risk component at that moment. This model is called a latent system model because the cause of failure is not specifically known and each component has some latent effect on the risk system. Further, the cumulative hazard of some risk components (e.g. fatigue of driver) can be considered by specifying the system hazard as a function of time. The model can estimate the probability of having an accident at any time given survival until that time.

2.5.2 Patterning

Accident patterning is the process by which the risk system determines the type of accident outcome; it is primarily determined by instantaneous risk factors (those at or near the accident scene). During the patterning process, the risk system interacts with the accident occurrence and impacts on its evolution. For example, a two-vehicle crash may be fatal or not depending on traffic stream speed. The same crash may lead to multiple collisions due to low visibility. Accident patterning approximately lasts until the accident notification. However, in some cases, the patterning procedure may continue until the complete clearance of the accident; secondary accidents may occur while e.g. waiting for the emergency units. Jovanis and Chang (1986) affirm that accident patterning may be examined with or without detailed exposure data.

2.5.2.1 Crash type

Crash type can be viewed as the outcome of a circumstance of type-specific dangerous conditions. Road safety studies that do not distinguish among different collision types are aggregate in nature; different crash types may occur under substantially different circumstances and may be associated with predictor variables in different ways. Crash taxonomy includes a wide variety of parameters such as crash location (e.g. on/off-road), collision type (head-on, sideswipe, at angle, rear-end collision), number of vehicles involved (single-, two-, multi-vehicle crashes), and maneuvers of vehicles prior to the accident (lane-changing etc.).

All these parameters are expected to interact with traffic conditions and to influence crash outcome. Kim et al. (2006) argued that crash type models are useful for at least three reasons: (a) the need to identify sites that may be of high risk with respect to specific crash type, but that may not be revealed through total crash modeling, (b) countermeasures are likely to affect only a subset of all crashes, and (c) different crash types are usually associated with road geometry, the environment, and traffic variables in different ways. Pande and Abdel-Aty (2006) underlined the importance of by-crash-type analysis, particularly when it comes to real-time risk assessment; Golob and Recker (2004) argued that there may be a direct correspondence between level of service (a traffic performance measure) and crash typology (a traffic safety measure).

Several authors reported that different types of highway crashes occur under markedly different circumstances (e.g. roadway or light conditions), particularly with respect to traffic volume. In most of these studies, the analysis is performed after segregating accidents in two large categories: single- and multi-vehicle crashes (Ceder and Livneh, 1982; Garber and Subramanyan, 2001; Ivan et al., 2000; Pasupathy et al., 2000; Persaud and Mucsi, 1995; Qin et al., 2004; Zhou and Sisiopiku, 1997). In few other studies, the segregation is made by distinguishing between rear-ends and sideswipes (Golob et al., 2004; Lee et al., 2006a).

Summarizing the above, there is strong empirical evidence that accident characteristics are crash type-specific (Golob et al., 2008; Khattak et al., 1998; Lee et al., 2006a; Shankar et al., 1995). However, different types of collisions are generally not distinguished in most research, except in severity analyses. A possible reason for this is the difficulty in collecting the necessary data; in addition, most accidents on freeways are thought to be rear-ends, but substantial number of crashes are not (Lee et al., 2006a). The differentiation becomes clearer when it comes to considering traffic parameters such as speed; ignoring the differential effect of traffic parameters on crash occurrence may introduce serious bias in the results (Mensah and Hauer, 1998). Smeed (1955), for example, argued that the annual amount of exposure does not influence – in practice – the annual accident rate for total accidents. Nonetheless, he pointed out that the accident rate for single-vehicle crashes tends to decrease when exposure increases, while the opposite is observed in the case of multivehicle accidents.

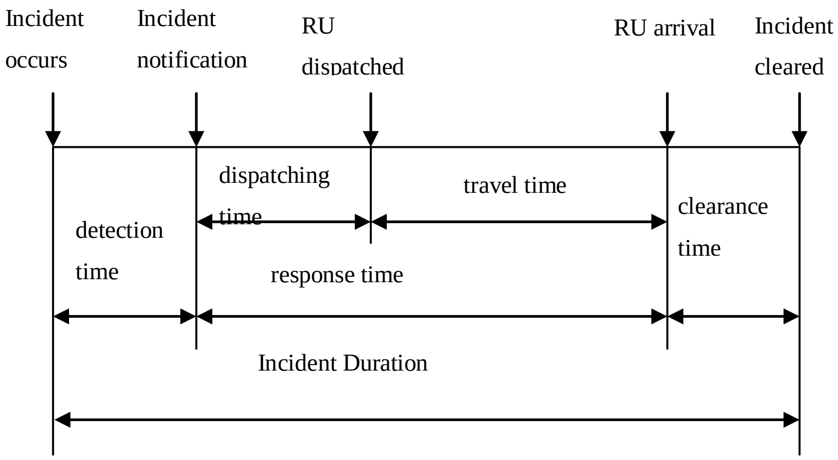
2.5.3 Response

Response to road incidents refers to all actions taken in order to best address incident occurrence (as well as its consequences) and restore roadway capacity to its normal level. The response phase includes incident detection and verification, motorist information, emergency units' allocation and redeployment, site management, traffic management, and clearance (Bunn and Savage, 2003). The proper identification and prioritization of factors that contribute to emergency management services response and clearance times result in better usage of taxpayer resources (Lee and Fazio, 2005). Empirical evidence suggests that environmental factors such as weather or roadway conditions has minimal effect on response times, while day of the week, urban or rural area, off or opposing-lane crash location, number of vehicles involved, heavy vehicle involvement, and response time significantly affect clearance time during peak periods (Lee and Fazio, 2005).

Incident duration is the time elapsed between occurrence of an incident and the clearance of the incidence and restoration of the roadway capacity to its normal level (Zografos et al., 2002). The total incident duration can be segregated in four elements (Figure 5):

- Detection/reporting time (*T1*): from the incident occurrence to the incident detection and verification
- Dispatching/preparation time (*T2*): from the incident detection to the dispatching of a response unit (RU) to be assigned to service the incident
- Travel time (*T3*): from the assignment of a response unit to its arrival on the site
- Clearance time (*T4*): the time required to clear the incident and restore roadway capacity.

Figure 5 The components of incident duration (Zografos et al., 2002)



Incident management is defined as the systematic, planned, and coordinated use of human, institutional, mechanical, and technical resources to reduce the duration and impact of incidents, and to improve the safety of motorists, crash victims, and incident responders (FHWA, 2000). On highways, research and operators focus on minimizing incident overall duration; important factors that affect duration are: (i) the operator ability to promptly *detect* an incident occurrence, and (ii) the *location* of emergency stations (police, ambulance). The benefits of minimizing incident duration are numerous and concern highway operators (e.g. cost, road safety performance), crash victims (e.g. time to hospital), other road users (e.g. delays, secondary incidents), and society (e.g. incident externalities).

Emergency station location (e.g. police, fire stations) analysis falls into location analysis; term that refers to the modeling, formulation, and solution of a class of problems that can best be described as sitting facilities in some given space (ReVelle and Eiselt, 2005). Obviously, emergency unit location is important to overall incident duration. In particular, the time needed to reach an incident scene is of great concern to emergency medical services (EMS) in order to mitigate incident consequences on people. In a real-time context, EMS managers are faced with two main problems: an allocation problem and a redeployment problem (Gendreau et al., 2001). The allocation problem consists of determining which ambulance must be sent to answer a call. The redeployment problem consists of relocating available ambulances to the potential location sites when calls are received; ambulances are assigned to potential sites to provide coverage. Covering constraints may be either absolute or relative.

Absolute constraints require that all demands are satisfied within r_2 minutes, while relative constraints require that a proportion of demand a is also satisfied within r_1 minutes ($r_2 > r_1$).

2.5.4 Consequences

A large and multi-disciplinary body of literature focuses on the measurement and mitigation of accident consequences. A lot of studies emphasized on accident consequence mitigation rather than accident risk limitation. Consequences can be reduced in many ways: by changes in the environment or in vehicles, by the use of protective devices, by driver training, rescue procedures, treatment or rehabilitation routines.

Consequences are measured long or short after the accident clearance (e.g. 30 days for fatalities) and may be described in many different ways depending on the point of view of interest. They primarily concern the efficiency of medical services provided after the accident occurrence (the treatment of injured persons). Secondarily, they concern explorations on accident severity contributing factors such as weather conditions or speeding. In that sense, consequences phase may begin from the actual time of the accident occurrence. A third category of consequence analyses attempt to estimate the overall accident monetary cost. In accordance with the above, three types of indicators are used to describe consequences: a) health-related, b) contributing factors investigations, and c) monetary indicators.

2.5.4.1 Medical services

Biomechanical models have been developed through experiments and simulated collisions. The human body can be simulated for most of the common collisions and the injuries caused can be studied using computers. There are various ways to describe the injury level. A commonly used classification scheme is the International Classification of Diseases (ICD); it describes the type of injury and its location but it does not capture injury severity, which is a very important variable. Thus, the Abbreviated Injury Scale (AIS) is often preferred to ICD. The AIS measures severity and varies from 1: *minor injury* to 6: *maximum injury*; however, it cannot represent multiple injuries. In such cases, another scale is used – the Injury Severity Score (ISS). The latter indicates the severity in terms of long-term disability and goes from

0: no long-term impairment to 6: lifetime serious impairment. Finally, an injury cost scale has been elaborated by Zeidler et al. (1993).

2.5.4.2 Severity contributing factors

The outcome of an accident is often measured as the level of injury sustained by the most severely injured vehicle occupant (Chang and Mannering, 1999). This typically includes severity levels of: no injury (property damage only), evident injury, disabling injury, and fatality.

Many factors influence accident consequences; they have been modeled and quantified to a significant extent. A review of such models can be found in Hakim et al. (1991). The main factors found to affect severity are: type and age of traffic element (user/vehicle) involved in the accident, accident maneuver type, speed, vehicle mass, road and roadside design, use of protective equipment, alcohol and drug consumption, traffic characteristics, intervention policies, weather, day of the week, speed limit, and so on.

Various methodological approaches have been applied: Logistic regression (Jones and Whitfield, 1988; Lui et al., 1988; Shankar and Mannering, 1996; Yau, 2004), multivariate time-series approaches to predict severity (Lassarre, 1986), bivariate models of injury outcomes (Saccomanno et al., 1996; Yamamoto and Shankar, 2004), standard multinomial logit models (Carson and Mannering, 2001; Savolainen and Mannering, 2007), nested logit for shared unobservables among severity categories (Khorashadi et al., 2005; Lee and Mannering, 2002; Martin, J.-L., 2002; Shankar et al. 1995; Ulfarsson and Mannering, 2004) and mixed logit structures (Milton et al., 2008).

2.5.4.3 Monetary indicators

Apart from the loss of lives, accidents have multiple collateral effects: material damage, environmental damage, pain to society, loss of productivity, impact on freight transport, health costs, delays, congestion, and so on. Also, they may trigger secondary crashes whose severity is often greater than that of the original incident (VNTSC, 1995). All these effects correspond to a certain cost, which cumulatively results to be extremely high. This cost is paid by insurance companies and health care systems and is, finally, covered by citizens. In most countries, accident economic losses reach 1 or 2% of GNP (Page, 2001). In 1997, the ETSC estimated the total cost of transport accidents in Europe at 166 billion euros (ETSC, 1997). 97% of these costs were directly related to road transport.

The methods for evaluating the socio-economic cost of road accidents vary significantly across countries; cost elements taken into account include medical costs, non-medical rehabilitation, lost productive capacity, human costs, damage to property, administrative costs, and other costs such as congestion. Accident prevention strategies and safety measures' choice are often based on costing accident consequences; the aim being to relate accident cost reductions to road safety investments.

2.6 Summary of findings

2.6.1 Literature organization

Road safety literature was organized by making use of four main criteria: a) the method employed, b) the level of analysis assumed, c) the scope of the performed analysis, and d) the accident phase considered. A four-dimension matrix (Table 1) illustrates road safety literature organization and summarizes analysis major findings.

Table 1 Road Safety Literature Organization

Classification Criteria		Controlled Experiment		Field Observational		In-depth investigation		Data Observational	
		A	D	A	D	A	D	A	D
Generating	Descriptive	-	✓	✓	✓	-	✓	✓	✓
	Predictive	-	-	-	-	-	-	✓	-
Patterning	Descriptive	-	✓	✓	✓	-	✓	✓	✓
	Predictive	-	-	-	-	-	-	✓	-
Response	Descriptive	-	✓	-	✓	-	✓	✓	✓
	Predictive	-	-	-	-	-	-	✓	-
Consequences	Descriptive	-	✓	✓	-	-	✓	✓	✓
	Predictive	-	-	-	-	-	-	✓	-

*A: aggregate
D: disaggregate

	dissertation field of interest

Controlled experiments allow for individual disaggregate observations such as single driver behavior or vehicle performance. They may refer to accident generating, patterning, response, or consequences. By simulating actual conditions, analysts explore individual reactions to external stimuli (e.g. in driving simulators); these reactions refer to accident generating and patterning processes. Additionally, there are also applications in incident management; for example, traffic simulators can help in estimating emergency units travel times. Accident outcomes can also be modeled in controlled experiments as in ‘crash tests’. In general, laboratory experiments allow for in-depth investigations to accident causal factors; however, they do not contain data on actual accident involvements. Most importantly, controlled experiments do not consider exposure to risk and, consequently, lack predictive power.

Field observational studies mainly focus on disaggregate individual observations (e.g. driver behavior as recorded on CCTV); however, they sometimes treat aggregate groups of individuals (e.g. loop measurements). They include analyses of accident generating, patterning, and response phases. To the best of our knowledge no field observational research exists on accident consequences. Disaggregate response studies include, for example, CCTV incident duration registrations. Road accidents are rare events and, so, sample size inefficiency is a common problem. On the other hand, budget constraints do not allow for long continuous surveillance and, thus, inferences about exposure are made. As a result, field observational studies lack predictive power.

In-depth investigations are rare in road safety analysis; they take place after accident occurrences and – by definition – refer to individual observations. They cannot reproduce accident actual circumstances of occurrence, but they approximate accident causal factors and patterning through examining its consequences; examples include vehicle examinations by engineers. In-depth investigations may also include accident response explorations as in IMS evaluation studies. In-depth investigations are observation-specific and do not account for any exposure measure – actual or inferred. As such, they remain only descriptive and cannot be used in accident predictions.

In data observational studies, the unit of the analysis may be a single accident occurrence, but independent variables –such as weather – may be aggregate in nature. Most importantly, exposure measurements (either used as independent variables or in the dependent variable expression) are generally aggregate. As a result, data observational studies remain aggregate to an extent; although efforts towards disaggregation are constantly made. As to accident phasing, observational studies may equally refer to all accident phases; i.e. generating, patterning, response, and consequences. In addition, statistical modeling enables for both descriptive and predictive analyses given that exposure is taken into account. Accidents are random and unpredictable; however, robust estimations about future exposure are possible. Consequently, if the analyst approximates the relationship between accident occurrence and exposure, he may then infer accident propensities. Data observational studies offer more possibilities compared to all other study types and can be easily applied, while data requirements are not extremely high. Nevertheless, their results do not possess the level of detail, certainty and precision that other study types provide.

2.6.2 Field of interest

In the present dissertation, we conduct a data observational study within a descriptive scope of analysis. Stochastic modeling is used in a rather disaggregate context of analysis. Accident outcomes – in terms of either crash type or severity – serve as dependent variables. Crash type refers to accident patterning, while severity is linked to accident consequences. Independent variables include road user attributes, weather and lighting conditions, vehicle type and age, traffic data, and so on. To this end, real-time traffic data are extracted from continuous loop measurements at the time of the accident occurrence (aggregate field observations). Results provide probability estimations for accident outcomes, given that these accidents occur under specific circumstances; if combined with frequency models, they could additionally provide prediction estimations. Finally, we examine potential implications of the developed models in optimizing incident management techniques; the latter being related to accident response phase.

Chapter 3

Identifying crash type propensity using real-time traffic data on freeways

In this Chapter, we examine the effects of various traffic parameters on type of road crash. Multivariate Probit models are specified on 4-years of data from the A4-A86 highway section in the Ile-de-France region, France. Empirical findings indicate that crash type can almost exclusively be defined by the prevailing traffic conditions shortly before its occurrence.

*This chapter has been accepted for publication to the Journal of Safety Research.

3.1 Introduction

Crash data analysis is the most frequently used tool for assessing the safety performance of a transportation facility (Abdel-Aty and Pande, 2007). Accidents may be viewed as the result of the interaction of multiple variables including road geometry (e.g. curvature), driver characteristics and behavior (e.g. gender, age), traffic conditions (e.g. speed limit, volume), environmental factors (e.g. weather), and so on. The conventional approach to crash data analysis has been to establish relationships between these variables and crash frequency (Ceder, 1982; Garber and Ehrhart, 2000; Yan et al., 2009) or severity (Abdelwahab and Abdel-Aty, 2002; Al-Ghamdi, 2002; Srinivasan, 2002). Crash frequency data have been analyzed using a number of modeling techniques that have ranged from conventional regression (Garber and Ehrhart, 2000; Mountain et al., 1996) to artificial neural network models (Abdel-Aty and Pande, 2005; Abdelwahab and Abdel-Aty, 2002). For a more in-depth discussion on methodological advances in crash data analyses, see Abdel-Aty and Pande (2007); for shortcomings and a discussion see Lord and Mannering (2010) and Songchitruksa and Tarko (2006).

Regardless of the modeling technique used, a serious factor of inaccuracy – in most past studies – has been data aggregation (Lord and Mannering, 2010) and sample size sufficiency (Pande and Abdel-Aty, 2006). Nowadays, most freeways are equipped with continuous surveillance systems making disaggregate traffic data readily available; these have been used in some studies (Abdel-Aty et al., 2007; Kockelman and Ma, 2007; Lee et al., 2002, 2003; Madanat and Liu, 1995). While detailed vehicle movement data in a section would be the best data source, traffic data from several consecutive detectors in a section can be a good surrogate to identifying traffic dynamics that may lead to accidents (Oh et al., 2001).

Aggregation also refers to crash type as different crash types may occur under substantially different circumstances and may be associated with predictor variables in different ways. Crash taxonomy includes a wide variety of parameters such as crash location (e.g. on/off-road), type of collision (frontal etc.), number of vehicles involved, and maneuvers of vehicles prior to the accident (lane-changing etc.). All these parameters are expected to interact with traffic conditions and to influence crash outcome; however, crash prediction models investigating different crash types have not been developed possibly due to the difficulty in collecting the necessary data. Kim et al. (2006) argued that crash type models are useful for at least three reasons: (a) the need to identify sites that are high risk with respect to specific crash types but that may not be revealed through total crash modeling, (b) countermeasures are likely to affect only a subset of all crashes, and (c) different crash types are usually associated with road geometry, the environment, and traffic variables in different ways. Pande and Abdel-Aty (2006) underlined the importance of by-crash-type analysis, particularly when it comes to real-time risk assessment. They suggested that the conditions preceding crashes are expected to differ by type of crash and, therefore, any approach towards proactive traffic management should be type-specific in nature.

In this chapter we focus on examining the effects of various traffic parameters collected real-time both at – and prior to – the time of the accident on type of crash. Multivariate Probit models are specified on 4-years of data from the A4-A86 highway section in the Ile-de-France region, France. We use a disaggregate approach in which the units of analysis are the crashes themselves (rather than aggregations of crashes over time), and traffic data are measurements of volume, speed, and density over 6-minute intervals. Such an analysis offers a wide variety of potential benefits; from a methodological standpoint, disaggregation minimizes possible bias (Davis, 2002), while additional light can be shed on the causal relationship between accidents and several contributing factors such as geometry, traffic conditions, and so on.

3.2 Background

Traditionally, crash prediction models were macroscopic in nature, where researchers mainly used summary statistics rather than microscopic measures to develop the models. The Average Annual Daily Traffic (AADT) has been the most commonly used measure in the literature to reflect traffic conditions (Kim et al., 2006; Mouskos et al., 1999; Qin et al., 2004). AADT is an aggregate measure of exposure; however, the use of AADT to approximate vehicle kilometers traveled at a site might reduce the natural variance that exists in exposure data and may result in heavy underdispersion (Pasupathy et al., 2000). Many researchers reported a U-shaped relationship between traffic volume and accident rate (Gwynn, 1967; Leutzbach, 1966), with multi-vehicles accidents increasing with flow and single-vehicle accidents decreasing with increasing flow (Ceder and Livneh, 1982; Zhou and Sisiopiku, 1997).

Later, many authors used aggregated data over a month, a week or a day for developing the same models; others used deduced hourly traffic characteristics by combining AADT and a 1-day hourly traffic profile for the site analyzed (Ivan et al., 2000). Noticeably, a number of studies used aggregated congestion measures (v/c ratio in Frantzeskakis and Iordanis (1987), Level of Service in Pasupathy et al. (2000), Persaud and Nguyen (1998), Zhou and Sisiopiku (1997)) instead of the AADT, with most authors reporting increased crash probability under congested traffic. Hourly traffic measures, when first utilized, were considered to be disaggregate in nature compared to annual measurements; however, even hourly measures cannot consider the short-term variation of traffic flow and are rather not well suited for application to real-time operations.

Regardless of the aggregation level, there is strong empirical evidence that accident characteristics are crash type-specific (Golob et al., 2008; Khattak et al., 1998; Lee et al., 2006a; McCartt et al., 2004; Shankar et al., 1995). However, various types of collisions are generally not distinguished in most research, except in some severity analyses. A possible reason for this is the difficulty in collecting the necessary data; in addition, most accidents on freeways are thought to be rear-ends, but substantial number of crashes are not (Lee et al., 2006a). The differentiation becomes clearer when it comes to considering traffic parameters such as speed; ignoring the differential effect of traffic parameters on crash occurrence may introduce serious bias in the results (Mensah and Hauer, 1998). Smeed (1955), for example, argued that the annual amount of exposure does not influence – in practice – the annual accident rate for total accidents. Nonetheless, he pointed out that the accident rate for single-vehicle crashes tends to decrease when exposure increases, while in the case of multivehicle accidents the opposite is observed. Golob and Recker (2004) argued that there may be a direct correspondence between level of service (a traffic performance measure) and crash typology (a traffic safety measure).

Several authors reported that different types of highway crashes occur under markedly different circumstances (e.g. roadway or light conditions), particularly with respect to traffic volume. In most of these studies, the analysis is performed after segregating accidents in two large categories: single- and multi-vehicle crashes (Ceder and Livneh, 1982; Garber and Subramanyan, 2001; Ivan et al., 2000; Pasupathy et al., 2000; Persaud and Mucsi, 1995; Qin et al., 2004; Zhou and Sisiopiku, 1997). In some studies, crash occurrence contributing factors have been explored while considering the type of primary collision and distinguishing between rear-ends and sideswipes (Golob et al., 2004; Lee et al., 2006a). Most studies use aggregate traffic measures (Ceder and Livneh, 1982; Pasupathy et al., 2000; Zhou and Sisiopiku, 1997), while few of the studies perform a by-crash-type analysis utilizing traffic data collected real time shortly before the accident's occurrence (Abdel-Aty et al., 2007; Golob and Recker, 2001; Golob et al., 2004, 2008).

Golob and Recker (2001) performed nonlinear canonical correlation analysis (NLCCA) with three sets of variables concerning accidents that occurred on South California highways during 1998. Results indicated that the type of collision is strongly related to median traffic speed and to temporal variations in speed on the left and interior lanes. Further, using the same datasets, Golob et al. (2004) and Golob and Recker (2001) developed a classification scheme by which traffic flow conditions on an urban freeway can be classified into mutually exclusive clusters that differ as much as possible in terms of likelihood of type of crash. In Golob et al. (2004) vehicle exposure to each regime was estimated by drawing a random sample of traffic flow measurements for the period of the analysis. Results suggested that lane-change crashes tend to occur under conditions in which there is the highest variability in speeds, while rear-end crashes tend to cluster where there is both lower speed variation and lower speeds. In Golob and Recker (2001), 21 traffic regimes for three different ambient conditions were defined and each of these regimes was shown to have a unique profile in terms of the type of crashes that are most likely to occur. Daylight conditions were shown to be related to collision type rather than the number of vehicles involved, while the scaling for nighttime conditions was based on the number of vehicles involved. Golob et al. (2008) captured the relationships between traffic flow and type of accidents that occur under different types of traffic flow conditions; results indicate that accidents involving a single vehicle are predominantly associated with late-night hit-object and run-off-the-road accidents. Also, two-vehicle accidents are more likely when volumes are similar in all lanes; large-scale accidents

(4 or more vehicles) are more likely to occur when volumes are similar in all lanes and there are high levels of variation in these volumes.

Abdel-Aty and Pande (2005) used real-time traffic and accident data from the I-4 corridor in Orlando to identify and classify crash propensity factors; variations in speed at least 10 to 15 minutes prior to an accident's occurrence was found to be the best classifier and results showed that at least 70% of the crashes on the evaluation dataset could be identified. In Pande and Abdel-Aty (2006), the need to further distinguish crash propensity by crash type was recognized and the authors developed classification models using historical crash data and information on real-time traffic parameters obtained from the same site. High average speed downstream along with low average speed upstream, low average differences between adjacent lane occupancies upstream at high speeds (up and downstream), and high standard deviation of volume and speed downstream were found to increase the likelihood of lane-change related collisions.

Abdel-Aty et al. (2007) developed real time crash risk assessment models for rear-end and lane-change crashes in an effort to reliably assess the crash risk on a real-time basis using historical crash and loop detector data obtained from the I-4 corridor in Orlando from 1994 through 2003. Rear-end crashes were separated into two groups – congested and high speed – based on prevailing speeds at surrounding stations 5 to10 minutes before the crash. Under congestion, rear-end crashes were found to be more probable, particularly if speed variation and average occupancy are elevated. Under free flow, rear-end crashes seem to have a greater likelihood of occurrence when the average difference between occupancy of adjacent lanes and the average upstream speeds are high.

Lee et al. (2002, 2003, 2006b, 2006c) extensively worked on establishing real-time risk indicators. In Lee et al. (2006a), 4-years of accident data from the I-880 freeway in Hayward (California) were used to identify real-time indicators of sideswipes versus rear-ends. Using logistic regression models, they showed that other traffic related factors such as variation in flow and peak/off-peak periods are important factors that are correlated with sideswipe crashes.

Simultaneous analysis of accident frequency by crash type and vehicle involvement using real-time traffic data has remained sparse (Golob and Recker, 2001), particularly in Europe. The papers previously discussed have provided very useful insights toward understanding accident occurrence; by using prevailing traffic flow data just prior to the time of the accident, we attempt to overcome the problems of *argument* and *function averaging* (Mensah and Hauer, 1998). The first is a result of using aggregate flow data rather than data measuring traffic conditions at the time of the accident, while the second is caused by using the same functional relationship for all types of collisions under all conditions.

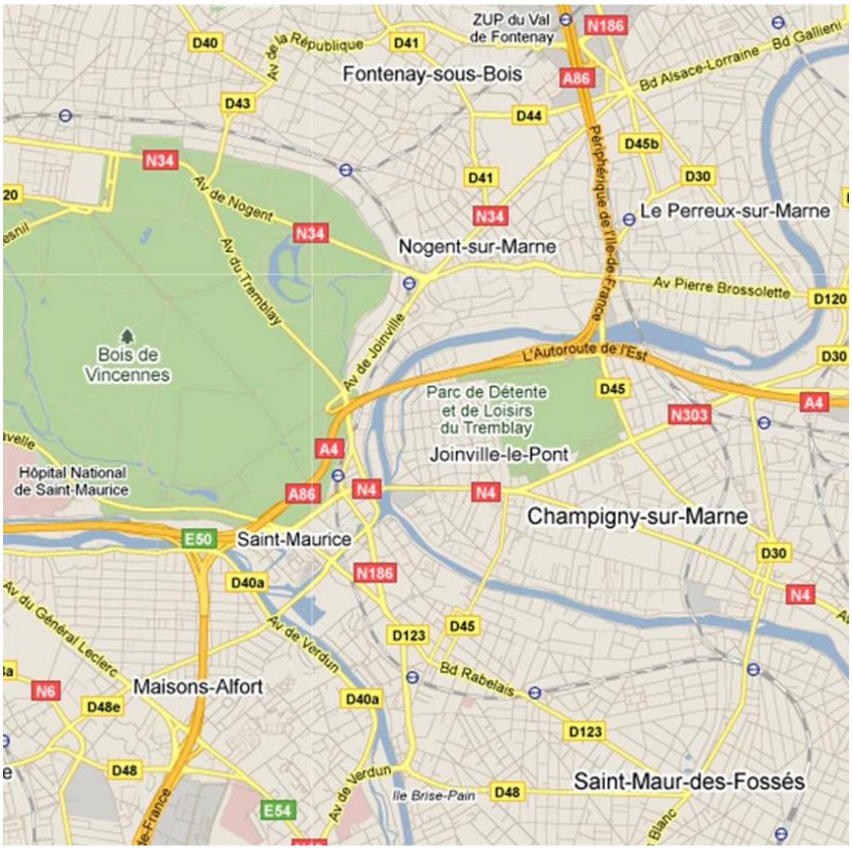
3.3 Data and Methodology

3.3.1 The Data

To explore the factors that determine accident occurrence by crash type, the A4-A86 highway section from a dense urban area a few miles to the east of Paris was selected (Figure 6). The A4-A86 junction has a length of 2.3 kilometres and includes four lanes per direction (to and from Paris). In particular, we used measurements from 3 stations per direction situated at kilometres 5.50, 6.00, 7.05 (direction to Paris) and 5.50, 6.14, 7.03 (direction from Paris). The A4-A86 junction is a particular site as it is the point where the Ile-de-France Ring Road (Périphérique-A86) coincides with the Autoroute de l’Est (A4) and merging is prevalent; five lanes are reduced to four on each direction. All stations are situated on the common part of the two highways; i.e. after the merging and before their separation.

Accident data were extracted from B.A.A.C. (Bulletins d’Analyse des Accidents Corporels) along with the Verbal Proceedings from an INRETS study *(*Aron and Seidowsky, 2004). The BAAC files provide a wealth of useful information such as crash type for all accidents, location and time, lighting conditions, and infrastructure characteristics such as road curvature and alignment. Detailed weather data are available on a 30-minute basis. We extracted such data directly from the closest meteorological station and for the 30-minute interval into which the reported time of the accident occurred. In total, 381 accidents were recorded during the period 2000-2002 and 2006. We statistically checked and found no significant difference between the 2000-2002 and 2006 registrations.

Figure 6 The A4-A86 junction



Traffic data (flow, speed, occupancy) were provided as part of the same INRETS study (Aron and Seidowsky, 2004), and cover the period 2000-2002 and 2006; data are recorded on 6-minute intervals. Such intervals may be too large to capture short-term variations; however, data averaged on shorter intervals are not available. Besides, several authors (Abdel-Aty and Pande, 2005; Oh et al., 2000; Pande and Abdel-Aty, 2006) have used 5-minute intervals to perform similar analyses. For each 6-minute period, the traffic database provides a series of speed, volume and occupancy measurements for each lane. The recorded traffic volume and speed – used in the thesis – were for the six-minute period ending 6 minutes before the accident (from the closest downstream detector). This time lag was used to avoid the impact of the crash itself on the traffic variables and as a buffer to compensate for any 'inaccuracies' in the exact time of the accident. For example, if an accident occurred at 9:00h, the traffic data considered were obtained from the 8:48-8:54 period. Similar techniques have been applied in other real-time data analyses (Abdel-Aty et al., 2007; Lee et al., 2006a).

Loop detectors often suffer from problems that may result in unreasonable values for speed, volume, and occupancy. We reviewed all data sequences based on time series deviations, deviations across lanes, and logical rules derived from reasonable volume, occupancy and speed relationships. Aberrant values (e.g. speed>200 km/h or speed>0 along with flow=0) were discarded from the database. Accidents with traffic data unavailability were also discarded. Each observation in the dataset is a record of the crash type of each accident, the corresponding traffic conditions, and various external factors.

Table 2 presents the descriptive statistics of all crash types considered. Frontal crashes ('Type 1') rarely occur on freeways. A number of accidents were recorded as belonging to none of the described categories ('Type 7'). Rear-ends account for almost 36% of total accidents which contradicts the common belief that almost all freeway accidents are rear-ends. Finally, accident distribution appears to be rather balanced; this indication suggests the need for detailed analysis by crash type.

Table 2 Dependent variables in crash-type analysis

Crash type	Type	Summary Statistics[1]	Description
Type 2	binary	F(1)=21.3%	=1 if rear-end with 2 vehicles; =0 otherwise
Type 3	binary	F(1)=16.6%	=1 if sideswipe with 2 vehicles; =0 otherwise
Type 4	binary	F(1)=14.5%	=1 if rear-end with more than 2 vehicles; =0 otherwise
Type 5	binary	F(1)=14.1%	=1 if multiple collisions; =0 otherwise
Type 6	binary	F(1)=11.8 %	=1 if single-vehicle crash; =0 otherwise

[1]F: frequency

Table 3 presents a definition for each independent variable considered together with its type, some summary statistics, and a short description. Most independent variables are defined as extracted from the data bases used, with the exception of 'pointe', 'SDQ2', 'SDV2'. 'pointe' is a binary variable introduced to represent if the accident occurred under congested or free-flow traffic regime. Peak and off-peak hours per direction were based on previous analyses and are site-specific. 'SDQ2' equals the standard deviation of traffic volume across lanes over 6 minutes, while 'SDV2' equals the standard deviation of speed across lanes over 6 minutes. These variables were intended to capture driver lane change behavior as there is empirical evidence that such variables are associated with sideswipe crashes (Oh et al., 2001; Pande and Abdel-Aty, 2006).

Table 3 Explanatory variables in crash-type analysis

Variable	Type	Summary Statistics[1]	Description
General accident information			
Road direction			
(sens)	binary	F(1)=47.1%	=1from Paris; =2 to Paris
Type of day			
(tjour)	binary	F(1)=83.0%	=1 if weekday, Saturday; =2 if Sunday, holiday
Peak/off-peak period			

(pointe)	dummy	F(0)=53.2%	=0 if peak hours; =1otherwise
			If (sens=1 and tjour=weekday) peak hours=7.00-19.00
			If (sens=1 and tjour=Saturday) peak hours=8.00-17.00
			If(sens=1 and tjour=2) peak hours=17.00-20.00
			If(sens=2 and tjour=weekday) peak hours=8.00-20.00
			If (sens=2 and tjour=Saturday) peak hours=11.00-18.00
			If (sens=2 and tjour=2) peak hours=0

Weather conditions			
(meteo)	dummy	F(0)=83.0%	=0 if weather is fine or cloudy; =1 otherwise
Lighting conditions			
(jour)	dummy	F(0)=65.5 %	=0 if daylight, dawn or dusk; =1 otherwise
Road geometry			
Road curvature			
(tplan)	dummy	F(0)=39.1%	=0 if straight line; =1 otherwise
Gradient			
(profil)	dummy	F(0)=75.9%	=0 if flat; =1 otherwise
Traffic characteristics			
Traffic volume			
(Q2)	continuous	M=112.4, SD=55.3	Average traffic volume per lane and over 6 minutes (in vehicles)
(SDQ2)	continuous	M=22.35, SD=13.5	Standard deviation of traffic volume across lanes over 6 minutes
Speed			
(V2)	continuous	M=73.3,SD=31.7	Average speed for all lanes and over 6 minutes (in km/h)
(SDV2)	continuous	M=12.2,SD=7.9	Standard deviation of speed across lanes over 6 minutes
Density			
(D2)	continuous	M=2.4, SD=2.3	Average traffic density per lane over 6 minutes (in veh/km)

[1] F: frequency
M: average value
SD: Standard Deviation

3.3.2 Methodology

Probit models have been widely used to analyze dependent variables of discrete nature (0,1). The relationship among the dependent and the independent variables does not lead to the estimation of a value for the dependent variable, but to the estimation of a probability that one of the two alternatives will occur (Washington et al., 2003). Here, we use Probit models to estimate the factors that affect the occurrence of a given crash type versus all other crash types considered. In this sense, increased probability of a crash type occurrence under increased values of a specific independent variable would indicate that the latter contributes to the mechanism of occurrence of the specific crash type.

The general specification for a univariate Probit model (for an event *n* resulting in an outcome *i*) can be expressed as (Washington et al., 2003):

$$Y_{in} = \beta_i X_{in} + \varepsilon_{in} \qquad \text{(Equation 3.1)}$$

Where:

Y_{in} defines an unobserved variable representing the latent utility (or propensity) for alternative *i*,

X_{in} is a vector of observed characteristics determining the outcome of the event *n*,

β_i represents a vector of unknown coefficients to be estimated for the alternative *i*, and

ε_{in} represents a vector of error terms.

If further assumed that ε_{in} follows the normal distribution and that *i=1,2*, we obtain the specification of the binomial Probit model (Washington et al., 2003):

$$P_n(1) = P(\beta_1 X_{1n} - \beta_2 X_{2n} \geq -\varepsilon_{1n} + \varepsilon_{2n}) \qquad \text{(Equation 3.2)}$$

Equation 3.2 estimates the probability of occurrence of alternative *1* for event *n*. The terms $\varepsilon_{1n}, \varepsilon_{2n}$ are normally distributed with zero mean and variances σ_{12} , σ_{22}.

Any road accident can be regarded as an event whose outcome is the type of crash that finally occurred (rear-end, side-swipe etc.). The binomial Probit model of Equation 3.2 can be used for estimating factors contributing or preventing a specific crash type versus all other types. Under this assumption, Equation 3.2 provides the probability of occurrence of a crash type (alternative 1) for each of the *n* accidents.

The multivariate probit model is one form of a multivariate discrete choice model that simultaneously estimates the influence of independent variables on (more than one) dependent variables and allows for the error terms to be freely correlated. The multivariate Probit model is based on the multivariate normal distribution and is recommended in cases where the dependent variables may be reasonably assumed as being correlated (Greene, 2003). Road accidents are very complex events; the contributing factors of each crash type may as well have a – positive or negative – influence to the occurrence of other crash types. Independence among different crash types may not be a valid assumption. The multivariate Probit model was used in the analysis to jointly identify traffic patterns that contribute to accident occurrence of all crash types while controlling for geometry and environmental conditions.

Multivariate probit models have been used in transportation research in a number of cases mainly focusing on travel demand and mode choice (Choo and Mokhtarian, 2008; Goulias et al., 1998). The general specification for a multivariate probit model of *n* dependent variables (alternatives) can be expressed as (Greene, 2003):

$$Y_i^* = \beta_i X_i + \varepsilon_i \text{ }, i = 1, \ldots, n$$ (Equation 3.3)

Where:

Y_i^* defines an unobserved variable representing the latent utility (or propensity) for alternative *i=1,2,*

X_i is a vector of observed characteristics determining alternative *I*,

β_i represents a vector of unknown coefficients to be estimated, and

ε_i represents a vector of error terms that are normally distributed with zero mean and constant variance.

The variance-covariance matrix of the error terms is given as follows:

$$\Sigma = \begin{bmatrix} \mathbf{1} & \cdots & \rho\mathbf{1}n \\ \vdots & \ddots & \vdots \\ \rho n\mathbf{1} & \cdots & \rho nn \end{bmatrix}$$ (Equation 3.4)

where ρ is a measure of the correlation among the latent utilities.

The main barrier for extending the univariate probit model to a multivariate setting lies in the evaluation of higher-order multivariate normal integrals (Viswanathan et al., 2000). It has been suggested that it is possible to approximate multivariate normal probabilities by random sampling (Lerman and Manski, 1981). Further, in cases where errors of estimates vary randomly, an estimate of the log-likelihood and its derivatives can be obtained; these estimates are close to the one that results from the actual computation of the integral (McFadden, 1989). In light of the above, multivariate probit models are estimated using simulation methods – most frequently Monte Carlo integration – rather than conventional numerical approaches.

3.4 Empirical results

Separate binomial Probit models were applied for each dependent variable considered; the statistical software Limpdep (v.8) was used for all applications. Parameters β of Equation 3.1were estimated using maximum likelihood. Univariate estimation results are presented in Table 4. A multivariate Probit model was also applied to jointly estimate the probabilities of occurrence of each crash type, considering that the contributing factors are interrelated. The multivariate Probit estimation results are presented in Table 6, while Table 7 shows the correlation coefficients among the five equations (crash types) that are significant at the 95% level. All omitted variables were discarded on the grounds of low statistical significance. In annex 2 of the dissertation, model outputs are attached.

Table 4 Model estimation for univariate Probit models

	Dependent Variables									
	Type2		Type3		Type4		Type5		Type6	
Independent Variables	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat	Coef.	t-stat
constant	1.069	1.46	-1.482	-5.70	-1.789	-5.00	-1.275	-4.67	-0.784	-4.03
V2	-0.171	-2.21					0.006	1.68		
D2	-0.222	-2.47			0.076	1.79				
Q2			0.004	1.92						
Jour	-0.321	-1.51					-0.410	-1.75		
Tjour					0.443	1.76				
profil			0.034	1.54			-0.473	-1.75	-0.616	-2.01
tplan									-0.509	-2.13
observations	235		235		235		235		235	
p-value for overall model significance (x^2 test)	0.040		0.068		0.007		0.039		0.039	

Table 4 provides model estimation results for rear-end crashes involving two vehicles ('type2'); significant factors were the average vehicle speed ('V2'), traffic density ('D2'), and lighting conditions ('jour'). In particular, 2-vehicle rear-end crashes seem to be more probable (compared to all other crash types) during daytime compared to nighttime; this finding is similar to the work of Golob and Recker (2001) who reported that rear-end collisions are more likely to occur on dry roads during daylight. Persaud and Mucsi (1995) and Ivan et al. (2000) concluded that all types of multi-vehicle crashes occur mainly during the daytime when light conditions are good. This finding was supported by suggesting that during daytime, congestion is formed and sudden traffic decelerations are frequent. However, in our analysis, the variable reflecting congestion ('pointe' in Table 3) was not found to be statistically significant revealing that this assumption remains an open question. A possible explanation for the latter may be the presence of the variable 'D2' (traffic density) that absorbs all relevant variance. The use of lights during nighttime makes deceleration visible from a longer distance and reaction times may increase for on-coming drivers. Therefore, as drivers perceive potential dangers earlier, they have more time to reduce their speed or perform other last-minute maneuvers to avoid the crash; the latter could explain the increased probabilities of rear-end crashes under daylight conditions in our data.

Regarding traffic flow parameters, rear-end crashes were found to be more probable for lower values of density and average speed. This result may seem contradictory to previous finding – and appear surprising – as several studies including Abdel-Aty et al. (2007), Golob and Recker (2004), Golob et al. (2008) indicated congestion as being among the most robust precursors of rear-end crash occurrence. However, when considering that the specific site has heavy traffic during daytime, it can be reasonably assumed that these traffic conditions (low density and speed) most probably reflect the critical transition from free-flow to congestion (at least in the context of this thesis). In this context, rear-end crashes seem to cluster at a traffic flow regime when average speed starts to decrease, but traffic density is still not high; under this regime, queues are not yet formed, but sudden decelerations may occur at any time.

Table 4 includes model estimation results for sideswipe crashes involving two vehicles ('type3') that were found to be positively associated with both average traffic volume ('Q2') and road gradient ('profil'). This suggests that the probability of occurrence of such a sideswipe crash (versus all other crash types) increases on 'non-flat' road segments and for high volumes of traffic. As traffic flow increases, variation in speeds between adjacent lanes becomes more probable, lane change maneuvers become more frequent, and sideswipe accidents may occur. In addition, road gradient makes lane change maneuvers even more difficult due to visibility restrictions and difficulties in maintaining the vehicle's control. Moreover, on grades, speed variation among drivers may be higher compared to flat road segments (due to different vehicle mechanical properties), the latter indicating a higher propensity for lane-change maneuvers.

Table 4 also provides model estimation results for the occurrence of rear-end crashes involving more than two vehicles ('type4') that were found to be positively associated with average traffic density ('D2') and type of day ('tjour'). In particular, rear-ends involving more than two vehicles are more probable to occur (compared to all other crash types) on Sundays and on holidays and for high levels of traffic density. This finding is similar to other findings that rear-ends happen under congestion as various authors have reported (Abdel-Aty et al., 2007; Lee et al., 2002; McCartt et al., 2004). Under congestion, queues are formed and on-coming drivers have to adjust their speeds (or even immobilize their vehicle) on short time and distance; several reasons such as driving at inappropriate speeds do not always allow for the appropriate actions to be taken in order to avoid an accident from occurring. Further, while in queues, other drivers do not have the possibility to react and, thus, multi-vehicle chain accidents are inevitable. This finding is further highlighted by the fact that, in contrast to the above, rear-ends involving two vehicles are not probable under congestion but rather before its formation.

Table 4 further provides model estimation results for the occurrence of multi-vehicle collisions other than rear-ends ('type5'); significant factors were average traffic speed ('V2'), lighting conditions ('jour'), and road gradient ('profil'). In particular, they seem more probable to occur at high speeds, during daylight conditions, and on flat road segments. It can be assumed that this crash category mainly includes multi-vehicle crashes related to lane-change maneuvers; results indicate that under 'normal' driving conditions (in terms of lighting and road gradient), when drivers chose their speed freely (high speeds at free-flow regime), multi-vehicle collisions have a higher propensity of occurring. One possible explanation could be that, under such conditions, driver attention decreases and possible fatigue may result in longer reaction times. Besides, the risk homeostasis hypothesis suggests that under 'dangerous' conditions drivers adapt their behavior and compensate their exposure to increased risk (Wilde, 1982); as a result, drivers do not react promptly in order to avoid their implication in an accident occurring at their vicinity, particularly at high speeds. Besides, at lower speeds, multi-vehicle rear-end crashes are more probable to occur (as already mentioned). We do note however that this suggestion needs to be further investigated.

Table 4 also includes model estimation results for the occurrence of single-vehicle collisions ('type6') that were found to be exclusively associated with road geometry ('profil', 'tplan'). In particular, single-vehicle crashes seem to be more probable (compared to all other crash types) on straight and flat road segments. In the literature, single-vehicle accidents are reported under light-traffic conditions (Golob et al., 2008) and after sunset (Ivan et al., 2000; Persaud and Mucsi, 1995), while no association to road geometry has been reported. However, this was not found here, as single-vehicle accidents seem to cluster on straight and flat road segments. Several possible explanations can justify the above; first, as in the case of multi-vehicle collisions, this finding can be attributed to lower driver attention under such geometric conditions. Second, the risk homeostasis hypothesis suggests that under conditions perceived as 'dangerous' (such as ascending, descending or curved road segments) drivers adapt their behavior and compensate their exposure to increased risk (Wilde, 1982). As a result, drivers may lose the vehicle's control due to fatigue or other reasons. However, all other drivers, because of favorable geometric features, react on time to avoid their implication and, thus, occurring accidents remain single-vehicle ones. On the other hand, higher probability for single-vehicle accident occurrence on flat and straight roadway segments could be attributed to drivers avoiding other vehicles. In any case, this finding needs further investigation.

Table 5 Qualitative results for univariate Probit models

Dependent variables	Independent Variables							
	constant	V2	D2	Q2	jour	tjour	profil	tplan
Type 2	+	-	-		-			
Type 3	-			+			+	
Type 4	-		+			+		
Type 5	-	+			-		-	
Type 6	-						-	-

A comparative overview of the univariate Probit models qualitative results allows for interesting remarks (Table 5). Surprisingly, weather was not found to be significant in any of the models estimated; probably real-time traffic observations ‘include’ weather influence and partially ‘absorb’ its effect on accident occurrence. In addition, the standard deviations of average traffic measures among lanes (‘SDV2’, ‘SDQ2’) were discarded from the final models due to low statistical significance (below 90%). However, we note that in the literature, there is strong empirical evidence that the difference in traffic characteristics among adjacent lanes significantly affects crash occurrence (Abdel-Aty et al., 2007; Golob and Recker, 2001; Lee et al., 2002).

Noticeably, traffic-related independent variables were found significant in almost all models. This finding suggests the importance of considering such variables in crash frequency analyses; in particular, results indicate that two-vehicle crashes (sideswipes and rear-ends) seem to cluster at traffic regimes close to transition from free-flow to congestion. Multi-vehicle rear-ends are most probable under congestion, while all other multi-vehicle crashes most frequently occur under free-flow. Single-vehicle accidents (hit-object, roll-overs etc.) cannot be attributed to ‘specific’ traffic patterns as it is possible that many other exogenous factors (such as alcohol use and fatigue) may intervene.

Multivariate model results (Table 6) reveal no difference in the manner in which independent variables affect crash outcome (crash type). However, not all variables were found to be statistically significant; most non-traffic-related factors resulted in low t-statistics for all crash types with the exception of single-vehicle accidents. The outcome of a crash (involving at least two vehicles), in terms of resulting crash type, seems to be almost exclusively defined by the prevailing traffic conditions shortly before its occurrence. This finding would be neglected under a univariate analysis per crash type. Further, the correlation coefficients (Table 7) reveal significant shared effects among different crash types' mechanism of occurrence and suggest the validity of the proposed multivariate investigation, particularly in this setting where limited data are available. The negative sign of all correlation coefficients implies that each crash type occurs under different traffic regimes; the latter coming to support the necessity for by-crash-type analyses. As previous findings of this chapter (concerning traffic parameters) also suggest, single-vehicle crashes appear to be the less related to other crash type patterns of occurrence. However, they appear to happen under markedly different conditions compared to rear-ends involving 2 vehicles. Finally, the shared effects among all other crash types result to be of similar magnitude.

Table 6 Model estimation for multivariate Probit model

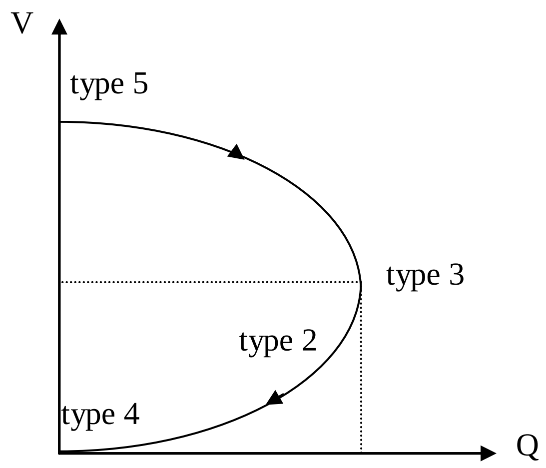
Independent Variables	Dependent Variables									
	Type2		Type3		Type4		Type5		Type6	
	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat	Coef	t-stat
constant	0.927	0.10	-1.476	-5.01	-1.703	-4.39	-1.226	-4.62	-0.526	-3.38
V2	-0.016	-1.85					0.006	1.73		
D2	-0.201	-1.54			0.075	1.57				
Q2			0.004	1.62						
Jour	-0.24	-1.06					-0.363	-1.70		
Tjour					0.387	1.45				
profil			0.240	0.98			-0.633	-2.42	-0.529	-2.05
tplan									-0.645	-3.89
Number of observations : 235										
p-value for overall model significance (x^2 test) : 0.0328										

Table 7 Correlation coefficients for multivariate Probit model

Correlation Coefficients		
	coefficient	t-stat
R(type2,type3)	-0.278	-2.24
R(type2,type4)	-0.327	-2.02
R(type3,type4)	-0.335	-1.80
R(type2,type5)	-0.288	-5.59
R(type3,type5)	-0.224	-1.52
R(type4,type5)	-0.278	-2.86
R(type2,type6)	-0.426	-2.60
R(type3,type6)	-0.876	-0.88
R(type4,type6)	-0.003	-0.03
R(type5,type6)	-0.186	-1.08

Qualitative results from the Multivariate Probit model application are illustrated in Figure 7. The fundamental diagram depicts the relationship between traffic volume (Q) and speed (V) on a given freeway segment; each crash type (whose probability is traffic-dependent) is related to a particular traffic regime which corresponds to a specific part of the diagram. The simultaneous analysis by vehicle involvement and maneuver possibly indicates that aggregation in by-crash-type analyses may lead to erroneous estimations. Rear-end crashes are more probable under congestion, while side-swipes are more probable under 'intermediate' density traffic regimes. Similar findings were reported by Golob and Recker (2004).

Figure 7 Crash type distribution



3.5 Concluding remarks

We assessed the effects of traffic variables (as obtained on a real-time basis) on crash type while controlling for road geometry and environmental factors. Empirical results indicated a diverse effect of accident contributing factors to each crash type, along with interdependencies that would be neglected under a univariate analysis context. Rear-end crashes involving two vehicles were found to be more probable for relatively low values of both speed and density, while rear-ends involving more than two vehicles appear to be more probable under congestion. Two-vehicle sideswipe accident probability increases with increasing volume, while multi-vehicle sideswipe crashes are more probable at high speeds, during daytime, and on flat freeway segments. Overall, multi-vehicle crashes tend to occur under low or very high speeds, while road geometry was found to be the single crash type indicator for single-vehicle accidents. In particular, single-vehicle accidents were found to be more probable on flat and straight road segments.

Weather and traffic's standard deviation across lanes were not found to be statistically significant indicators of crash type. Empirical results seem promising in establishing real-time crash type predictors. However, we recognise that this analysis suffers from limitations needing further investigation; first, loop detectors aggregate counts and occupancies over 6-min intervals. Possible uncertainty in the exact time of the accident occurrence should also be examined. Further, we did not distinguish in our analysis among freeway lanes, and only separated traffic regimes in two (peak and off-peak). Golob and Recker (2004), after performing a similar analysis, provided important evidence that it is significant (i) to capture variations in speed and flows separately across lanes and (ii) to more strictly define traffic regimes.

The potential benefits of integrating empirical results in a real-time traffic management application are numerous. Once a 'location' (i.e. a specific set of design, operational and travel characteristics) is identified as being susceptible to a given crash type occurrence, it may be flagged with warnings through variable message signs (VMS). Further, the concept of variable speed limits could be used to intervene on driver behavior and to reduce speed variation. In addition to real-time monitoring of safety levels, a safety performance tool can be used in project evaluation and planning. Safety aspects of costs and benefits can be assessed by comparing the levels of safety estimated before and after implementation of a treatment (Golob and Recker, 2004). Finally, a procedure that uses real-time data on traffic flow, speed, and occupancy and the relationship between these variables and crash-type occurrence could be used to develop congestion mitigation strategies that incorporate safety (Garber and Subramanyan, 2001).

Chapter 4

Vehicle Occupant Injury Severity on Highways: An empirical Investigation

In this Chapter, we apply a random parameters ordered probit model to explore the influence of speed and traffic volume on the injury level sustained by vehicle occupants involved in accidents on the A4-A86 junction in the Paris region. Results indicate that increased traffic volume has a consistently positive effect on severity, while speed has a differential effect on severity depending on flow conditions.

*This chapter has been published to Accident Analysis and Prevention 42 (2010) 1606-1620.

4.1 Introduction

Accident severity investigations are of particular concern to both decision makers and researchers and the literature has indicated several factors as significantly influencing crash-injury severity level sustained by road users (Abdel-Aty, 2003; Chimba and Sando, 2009; Gray et al., 2008; Kockelman and Kweon, 2002; Lapparent, 2008; Lee and Abdel-Aty, 2005; O'Donnell and Connor, 1996; Pai and Saleh, 2008; Quddus et al., 2002; Xie et al., 2009; Zajac and Ivan, 2002). Among the most important factors are driver age, collision type, weather and lighting conditions that have been extensively explored as to their effect on severity.

Results from previous research indicate that low speeds and high traffic volumes decrease accident severity, while high speeds and low traffic volume produce the opposite effect (see for example Martin, 2002); a result largely based on mean annual traffic values. However, few studies have investigated the association between traffic accident severity and actual traffic characteristics (traffic volume, speed) collected real-time during the time of the accident's occurrence (Golob and Recker, 2003; Quddus et al., 2009).

We extend research on the factors influencing the level of accident injury severity by including traffic data from the moment of the accident. Thus, the purpose of this research effort is to elaborate a model that associates traffic characteristics to the severity outcome of freeway accidents. Real-time traffic data are nowadays available for most freeway networks. Their integration to road safety analyses offers the possibility to associate accident attributes to the actual traffic flow characteristics at the time of the accident. Based on the analysis of historical data, typical traffic patterns recorded prior to accidents may then act as real-time identifiers (Abdel-Aty and Pande, 2007). Such research is useful for researchers and practitioners in estimating accident and congestion external costs and in transportation planning. Further, it may enable practitioners and authorities to locate hazardous – on severity grounds – spots on the road networks by utilizing real-time data widely available. Finally, it may provide additional insight regarding the factors that may contribute to higher probabilities context for severe injuries (given that an accident occurs). The model controls for various driver, vehicle, and crash characteristics along with real-time traffic and weather, by exploring possible associations between these factors and the severity outcome sustained by individuals involved in accidents.

4.2 Background

There has been extensive literature documenting links between accident characteristics and the severity levels sustained by road users; studies have investigated the effect of road and vehicle characteristics, driver attributes, weather etc. (Table 8 provides a summary of previous research efforts in the injury severity area).

In all modeling efforts, the *dependent* variable is severity, usually depicted on a 3- or 4-point ordinal scale (e.g. no injury, severe injury, fatal). The *unit of analysis* varies across studies and depends on the objective; units include, for example:

- non-motorized road users (pedestrians and/or bicyclists) (Ballesteros et al., 2004; Kim et al., 2008; Lee and Abdel-Aty, 2005; Sze and Wong, 2007),
- crashes (Chang and Wang, 2006; Eluru et al., 2008; Gray et al., 2008; Wang and Abdel-Aty, 2008; Yamamoto and Shankar, 2004),
- motorcycle occupants (Pai and Saleh, 2008; Pai, 2009),
- drivers (Abdel-Aty et al., 1998; Abdel-Aty, 2003; Conroy et al., 2008; Kockelman and Kweon, 2002) or
- any car occupant (Lapparent, 2008).

The *independent variables* considered usually include driver, vehicle, crash and road characteristics. Driver (or rider) characteristics most commonly refer to driver age, gender, alcohol consumption, and safety equipment usage (Boufous et al., 2008; Helai et al., 2008; Kim et al., 2008; Lapparent, 2008; Pai, 2009; Sze and Wong, 2007; Wang and Abdel-Aty, 2008; Yamamoto and Shankar, 2004). Vehicle characteristics refer to the type of vehicles (car, heavy vehicle, etc.) and the number of vehicles involved in the crash (Ballesteros et al., 2004; Helai et al., 2008; Kim et al., 2008; Pai, 2009). Road factors include curvature, number of lanes, type of road (e.g. rural and urban), surface conditions, junction control and so on (Al-Ghamdi, 2002; Milton et al., 2008; Savolainen and Mannering, 2007; Shankar et al., 1996). Crash characteristics are related to the exact circumstances under which the accident occurred (vehicle maneuvering before collision, crash's main cause and so on) (Chang and Wang, 2006; Gray et al., 2008; Helai et al., 2008; Wang and Abdel-Aty, 2008; Yamamoto and Shankar, 2004). Other variables such as speed limit, day of the week, time of day, AADT, weather and traffic conditions have also been examined regarding their influence on accident severity (Abdel-Aty, 2003; Conroy et al., 2008; Gray et al., 2008; Helai et al., 2008; Kim et al., 2008; Milton et al., 2008; Pai and Saleh, 2008; Pai, 2009; Savolainen and Mannering, 2007; Sze and Wong, 2007; Yamamoto and Shankar, 2004; Zajac and Ivan, 2002).

Findings from previous studies are, to a large extent, consistent. Factors most commonly found to increase severity are:

- increased driver or rider age (Sze and Wong, 2007; Xie et al., 2009)
- driving while intoxicated (Kim et al., 2008; Savolainen and Mannering, 2007; Zajac and Ivan, 2002).
- head-on-collisions (Eluru et al., 2008; Savolainen and Mannering, 2007),
- crashes with heavy vehicles and motorcycles (Lee and Abdel-Aty, 2005; Yamamoto and Shankar, 2004),
- poor lighting conditions (Chimba and Sando, 2009; Gray et al., 2008; Helai et al., 2008),
- vertical and horizontal curvature (Savolainen and Mannering, 2007; Xie et al., 2009),
- rural versus urban areas (Lapparent, 2008), and
- speeding (Boufous et al., 2008; Lee and Abdel-Aty, 2005; Pai and Saleh 2008)

In contrast to the above, the use of restraint systems (helmet or seat belt) appears to significantly decrease the level of injuries sustained by road users (Wang and Abdel-Aty, 2008; Yamamoto and Shankar, 2004).

However, research reports conflicting findings on some occasions, particularly when it comes to factors such as gender, intersections type, road surface conditions and seating position. For example, O'Donnell and Connor (1996) and Sze and Wong (2007) concluded that female drivers are associated with increased severity, while Shankar et al. (1996) and Yamamoto and Shankar (2004) report the opposite. Milton et al. (2008) and Xie et al. (2009) found that interchanges and junctions decrease injury levels, Lapparent (2008) and Al-Ghamdi (2002) both found that intersection presence decreases severity, while Helai et al. (2008) and Boufous et al. (2008) report that severity increases in more complex intersection arrangements (such as Y and T). Yamamoto and Shankar (2004) observed that icy pavements are associated with lower severities in urban areas and Shankar et al. (1996) argued that in single vehicle crashes wet pavements increase severity, while icy or snow-covered pavements tend to decrease it. Xie et al. (2009) found that both icy and wet pavements decrease accident severity, while in Quddus et al. (2002), wet surface was found to decrease the severity outcome of motorcycle accidents. Seating on the left-rear position is more dangerous according to O'Donnell and Connor (1996). Lapparent (2008) found that all car seating positions are safer than the driver's seat.

Interestingly, in most studies, fine weather was found to increase severity (Eluru et al., 2008; Gray et al., 2008; Kim et al., 2008; Pai and Saleh, 2008; Pai, 2009; Quddus et al., 2009; Yamamoto and Shankar, 2004; Xie et al., 2009), while other weather conditions (rain and snow) tend to decrease it; still, various authors claim the opposite (Abdel-Aty, 2003; Lee and Abdel-Aty, 2005), while some conclude that weather has no significant effect on severity (Abdel-Aty, 2003). Drivers seem to rather adjust their behavior to inclement weather by decreasing the speed which has a positive (decreased) effect on injury-severity (Quddus et al., 2009; Shankar et al., 1996). Further, adverse weather may alarm drivers who tend to be more vigilant and become more conservative in their driving behavior as argued in the offset hypothesis, which predicts that users adapt to innovations that improve safety by becoming less vigilant about safety. For example, Winston et al. (2006) tested the offset hypothesis using disaggregate data to analyze the effects of airbags and antilock brakes on automobile safety and found that safety-conscious drivers are more likely than other drivers to acquire airbags and antilock brakes; however, these safety devices were not found to have a significant effect on collisions or injuries, suggesting drivers trade off enhanced safety for speedier trips.

Average traffic characteristics such as truck percentage and traffic volume have been used as explanatory variables in a number of studies (Abdel-Aty et al., 1998; Milton et al., 2008; Wang and Abdel-Aty, 2008; Zajac and Ivan, 2002); these studies were limited to average values of the parameters such as the annual truck percentage or the Average Annual Daily Traffic (AADT). Traffic data at the moment of the accident have seldom been utilized to explore crash severity (Quddus et al., 2009).

For example, in Abdel-Aty et al. (1998), the odds for severe crashes for Average Daily Traffic>20,000 were found much higher for young and middle aged drivers and much smaller for very old drivers (over 75), while the middle age group (25-64 years old) was found to be the ‘safest’ in terms of injury severity. In Zajac and Ivan (2002), AADT was not found to significantly affect injury severity in pedestrian crashes. Wang and Abdel-Aty (2008) examined left-turn crash-injury severity at signalized intersections and concluded that neither the total approach traffic volume nor the entire intersection volume, but rather specific vehicle trajectories, affected crash injury significantly. In Milton et al. (2008), findings were not consistent across segments, suggesting that the effect of traffic (AADT) on injury-severity outcomes cannot be assumed uniform across geographic locations. Quddus et al. (2009) used real-time traffic data from the time of the accident; a 30-minutes time-lag was applied to avoid the traffic impact of the crash itself and traffic flow was found to be important in explaining the severity of each road accident, with traffic increases leading to crash severity decreases. In general, despite the large number of research efforts on the topic of accident severity, papers investigating the impact of traffic characteristics on accident severity remain few.

Table 8 Summary of injury severity modeling studies

Study	Injury severity scale	Data source	Method of analysis employed	Unit of analysis	Independent variables considered	Summary of findings
Chimba and Sando (2009)	(1) Property damage only (2) Incapacitating and fatal	2003 Florida (USA) high crash location database	(1) ANN backpropagation (2) ordered probit model	Driver	- Driver characteristics[a] -Speed limit -General info[b] -Weather -Road geometry[c]	Factors increasing severity • No daylight • Cloudy sky • Curved sections • Higher speed limit • Alcohol use • Turning movement Other finding • ANN performs better than OP
Pai (2009)	(1) Slight injury (2) Serious injury (3) Fatal injury	1991-2004 UK two- or more-vehicles crashes, occurring at T-junctions, involving at least one motorcycle and resulting to at least one slight injury	Binary logistic model	Motorcyclist	-Driver characteristics[a] -Motorcyclist characteristics[a] - Motorcycle characteristics[d] - Crash characteristics[e] -General info[b] -Weather	Factors increasing severity • Rider being over 60 • Engine size over 125cc • Heavy goods vehicle involvement • ≥2 vehicles involved, • Fine weather, • Non built-up roads • Right-of-way-violation Other finding • Injuries were greatest when a travelling straight motorcycle on the main road crashed into a right-turn car from a minor road, particularly at stop-/yield-controlled junctions
Quddus et al. (2009)	(1) Slight injury (2) Serious injury (3) Fatal injury	2003-2006 UK M25 motorway crash and traffic dataset	(1) Ordered logit (2) Heterogeneous choice model (HCM) (3) Generalized ordered logit (4) Partial	Crash leading to at least one slight injury	- General info[b] -Traffic characteristics[f] -Congestion -Weather -Road geometry[c]	Factors increasing severity • Fine weather versus rain • Weekdays and darkness Factors not influencing severity uniformly or at all • Congestion • Traffic flow • Snow

			proportional odds			
Xie et al. (2009)	(1) No injury (2) Possible injury (3) Non-incapacitated injury (4) Capacitated injury (5) Fatal injury	2003 US national crash database (automobiles, SUVs, vans)	(1) Ordered probit model (2) Bayesian ordered probit model	Driver	-Driver characteristics [a] - Vehicle characteristics[d] -General info[b] -Road geometry[c] - Crash characteristics[e] -Weather	Factors increasing severity • The age of the driver and of the vehicle, drunk driving, curvy road alignments, inadequate light conditions, initial impact points on the left side area, rollover and fire Factors decreasing severity • Crashes related to junctions or interchanges, icy or wet surfaces and adverse weather Other findings • For large sample data, the two models produce similar results • For small sample data and proper prior setting, the BOP overperforms the OP
Boufous et al. (2008)	Survival risk ratio (the number of patients with a certain injury code who have not died in the hospital and the total number of patients diagnosed with that code)	2000-2001 New South Wales (Australia) crash accidents that resulted in the hospitalization of drivers aged over 50	Linear regression	Driver	-driver characteristics[a] - vehicle characteristics[d] - crash characteristics[e] -general info[b] -road geometry[c] -area type	Factors increasing severity • Complex intersections (Y or T junctions and roundabouts) • High speed limit • Not wearing seat belt • Driver's errors (e.g. disobeying traffic control) • Speeding • Rural areas
Conroy et al. (2008)	Injury severity score (0-75)	1997-2006 US head-on frontal crashes resulting in at least one serious injury and in which car occupants used safety	logistic regression	Driver	-driver characteristics[a] - crash characteristics[e] -vehicle	Factors increasing severity • Drivers with intrusion into their position • Driving a passenger vehicle Factor decreasing severity • Drivers in wide impacts versus narrow impacts

		belts and vehicles' age was under 8 years			damage	Other finding • Drivers in wide frontal impacts were almost four times more likely to have a head injury
Eluru et al. (2008)	(1) No injury or possible injury (2) Non-incpacitating injury (3) Incapacitating injury (4) Fatal injury	2004 NHTSA (US) crashes involving non motorists (pedestrians and cyclists)	(1) Ordered-response logit model (2) Mixed generalized ordered-response logit	Crash involving non motorists	- Driver characteristics [a] - Vehicle characteristics[d] - Crash characteristics[e] - Road geometry[c] - Non-motorist characteristics[g] - Weather	Factors increasing severity • Being pedestrian versus being cyclist • Higher speed limits • Being male and older • Crashes with vehicles other than passenger cars • Frontal impact crash • Evening and late night periods Factors decreasing severity • Snow • Signalized intersections
Gray et al. (2008)	(1) Slight injury (2) Serious injury (3) Fatal injury	Subset of 1991-2003 accidents in Great Britain involving male drivers aged 17-25	Ordered probit model	Crash	- Young male driver characteristics[a] - Crash characteristics[e] - General info[b] - Road geometry[c] - Weather	Factors increasing severity • Driving in darkness • Trips in early morning and towards the end of the week • Driving on main roads • Overtaking maneuvers • Weather other than 'fine no high winds' • Speed limit of 60mph • Passing the site of a previous accident versus other carriageway hazards
Helai et al. (2008)	(1) Low individual severity (2) High individual severity	2003-2005 Singapore crashes at signalized intersections	Hierarchical binomial logistic model	Driver-vehicle unit	- Driver characteristics[a] - Vehicle characteristics[d] - Crash characteristics[e] - General info[b] - Road geometry[c] - Weather	Factors increasing severity • Night versus daytime • Y or T intersections • Right-most lane • Bad street lighting • Red light camera presence • 2-wheel vehicle • Being aged under 25 or over 65

Kim et al. (2008)	(1) Possibe or no-injury (2) Non-incapacitating injury (3) Incapacitating injury (4) Fatality	1997-2000 North Carolina (US) pedestrian-vehicle crashes involving one pedestrian and one vehicle	Heteroskedastic logit model	Pedestrian	- Driver characteristics[a] - Vehicle characteristics[d] - Crash characteristics[e] - General info[b] - Road geometry[c] - Pedestrian characteristics[g] - Area type - Weather	Factors increasing severity • Intoxicated driver, darkness with or without streetlights, greater pedestrian age, sport utility vehicle, truck, freeway, state-route, and speeding Factors decreasing severity • Intoxicated driver, PM peak (15:00-17:59), traffic signal control, and inclement weather Other finding • Pedestrian's age was a significant contributor to heteroskedasticity by increasing the variance of the error terms across pedestrians with age, but the gender did not affect the variation
Lapparent (2008)	(1) No injury (2) Light injury (3) Severe injury (4) Fatal injury	2003 crashes at France	Bivariate ordered probit model	Car user	- Car user characteristics[a] - Crash characteristics[e] - General info[b] - Road geometry[c]	Factors increasing severity • Increasing age • Crash at intersection for drivers • Crash on secondary roads and being front-seat passenger • Crash out of the city Factors decreasing severity • Crash at intersections and being front- or rear-seat passenger • Crash on highways and being rear-seat passenger Other finding • Safety belt use reduces injury level whatever the position of the car occupant
Milton et al. (2008)	(1) Property damage only (2) Possible injury (3) Injury (evident, disabling or fatality)	- Washington State highway segments accident database - Western Regional Climate Centre weather database - WSDOT traffic data	Mixed (random parameters) logit model	The most severely injured person	- ADT[h] - Road geometry[c] - Traffic characteristics[f] - Weather	Factors decreasing severity • Grade brakes, interchanges and horizontal curves Factors not influencing severity uniformly • Increasing truck ADT • ADT • Average annual snowfall • Truck %

Pai and Saleh (2008)	(1) No injury (2) Slight injury (3) Serious injury or fatality	1991-2001 GB accidents at T-junctions involving motorcycles and resulting to at least one injury	Ordered probit model	Motorcycle occupant	- Motorcyclist characteristics[a] - Motorcycle characteristics[d] - Crash characteristics[e] - General info[b] - Road geometry[c] - Weather	Factors increasing severity • Rider being male or over 60 • Increasing engine size • Crash partner other than motorcycle • Darkness, fine weather, spring/summer, midnight/early morning, and weekend riding • Speeding Other finding • The effects of some variables on injury levels vary across different crash types
Wang and Abdel-Aty (2008)	(1) No injury (2) Possible injury (3) Non-incapacitating injury (4) Incapacitating injury (5) Fatality	2000-2005 vehicular left-turn crashes at 4-legged signalized intersections in Central Florida (US)	Partial proportional odds model with logit and probit functions	Accident	- Driver characteristics[a] - Vehicle characteristics[d] - Crash characteristics[e] - General info[b] - Road geometry[c] - AADT	Factors increasing severity • Crashes involving motorcycle, with drivers ejected from vehicle, front impact, conflicting with near-side approaching vehicle, and driver being intoxicated Factors decreasing severity • Driver being young or using safety equipment, crashes occurring at night at intersections with street lights Factor not influencing severity • AADT
Savolainen and Mannering (2007)	(1) No injury or possible injury (2) Non-incapacitating injury (3) Incapacitating injury (4) Fatality	2003-2005 Indiana (US) motorcycle accidents	Nested logit model Standard Multinomial logit model	Motorcycle operator	- Driver and rider characteristics[a] - Vehicle characteristics[d] - Crash characteristics[e] - General info[b] - Road geometry[c] - Speed	Factors increasing severity in single-vehicle crashes • Increasing age, speeding, April and July, darkness, collisions with roadside object, being female, and alcohol involvement Factors decreasing severity in single-vehicle crashes • Helmet use, motorcycle less than 5 years, and wet pavement and intersection crashes
Sze and Wong (2007)	(1) Slightly injured (2) Killed or seriously injured	1991-2004 Hong Kong police crash database (crashes involving pedestrians)	Logistic regression	Pedestrian	- General info[b] - Road geometry[c] - Non-motorist	Factors increasing severity • Age above 65, head injury, a crash at a crossing or on a road section with a speed limit above 50km/h or at a signalized intersection

					characteristics[g] - Speed limit - Congestion	Factors decreasing severity • Being male, aged below 15, being on an overcrowded or obstructed path and being involved in a daytime crash on a road section with severe or moderate congestion
Chang and Wang (2006)	(1) No injury (2) Injury (3) Fatality	2001 accidents in the Taipei area (Taiwan)	Classification and regression tree model	Crash	- Driver characteristics[a] - Vehicle characteristics[d] - Crash characteristics[e] - General info[b] - Road geometry[c] - Speed limit - Weather	Factors increasing severity • Collisions with pedestrians, motorcycle and bicycle riders Other finding • Collision type, contributing circumstance and driver/vehicle action are critical in determining severity
Lee and Abdel-Aty (2005)	(1) No injury (2) Possible injury (3) Non-incpacitating injury (4) Incapacitating injury (5) Fatal injury	1999-2002 Florida (US) pedestrian accidents at intersections	Ordered probit model	pedestrian	- Driver characteristics[a] - Vehicle characteristics[d] - General info[b] - Road geometry[c] - Pedestrian characteristics[g] - Area type - Weather	Factors increasing severity • Pedestrian being old, female or intoxicated • High vehicle speed • Adverse weather and dark lighting • Vans, buses and trucks versus passenger cars • Rural versus urban areas • Intersections without traffic control

Ballesteros et al. (2004)	(1) Non fatal (2) Fatal (1) ISS<16 (2) ISS>=16	(1) 1995-1999 Maryland (US) crashes involving at least one pedestrian having been treated or dead and either a conventional car, a sports utility vehicle, a pickup truck or a van (2) Trauma registries by Emergency Medical Services (3) 1995-1999 fatalities from medical examiner	Logistic regression	Pedestrian	- Vehicle characteristics[d] - Speed limit	Factor increasing severity • SUVs and Pus compared to conventional cars Factor not influencing severity • Vans compared to conventional cars
Yamamoto and Shankar (2004)	(1) Property damage only (2) Possible injury (3) Evident injury (4) Fatality	1993-1996 Washington State (US) single vehicle collisions with fixed objects	Bivariate ordered probit model	Accident	- Driver characteristics[a] - Vehicle characteristics[d] - Crash characteristics[e] - General info[F4] - Road geometry[c] - Number of passengers - Speed - Weather	Factors increasing driver's injury severity in urban areas • Off-roadway, collisions with trees, driver age, speeding, falling asleep, and being sober Factors decreasing driver's injury severity in urban areas • Intersection, rain, icy pavement, restraint systems use, vehicle age, driving a motorcycle or truck, being male, collisions with sign posts or concrete barrier Factors increasing passenger's injury severity in urban areas • Driver's age, driver being intoxicated, or falling asleep, increasing number of passengers, the driver being male Factors decreasing passenger's injury severity in urban areas • Intersection, rain, restraint systems use, being a truck passenger, collisions with sign posts or bridge face
Abdel-Aty (2003)	(1) Property damage only (2) Possible/ evident injury (3) Severe/fatal injury	1999-2000 Central Florida (US) crashes near toll plaza	(1) Ordered probit model (2) Multinomial logit model (3) Nested logit model	Driver	- Driver characteristics[a] - Vehicle characteristics[d] - Crash characteristics[e] - General info[b] - Road	Factors increasing severity • Being over 65 • Being female • Impact on the side • No seatbelt use • e-pass use • Number of impacts • Adverse weather

					geometry[c] - Area type - Toll characteristics[i] - Weather	• Not driving a truck Other findings • OPM performs than MLM • The NLM performs slightly better than the OPM but is more difficult to estimate
Al-Ghamdi (2002)	(1) Non-fatal (2) Fatal	A subset of 1997-1998 accidents on urban roads in Riyadh (S. Arabia) leading to at least one slight injury	Logistic regression	Accident	- Driver characteristics[a] - Vehicle characteristics[d] - Crash characteristics[e] - General info[b] - Road geometry[c]	Factors increasing severity • Non-intersection versus intersection • Wrong-way versus all other causes (violations) Other finding • Among all independent variables, only the location and cause of the accident were found to significantly affect severity
Kockelman and Kweon (2002)	(1) No injury (2) Not severe injury (3) Severe injury (4) Death	1998 US dataset including property damages, injury crashes and fatal crashes	Ordered probit model	Driver in 1. single-vehicle crash 2. two-vehicle crash 3: all crashes	- Driver characteristics[a] - Vehicle characteristics[d] - Crash characteristics[e] - General info[b]	Factor increasing severity • Under single-vehicle crash conditions, the use of pickups and sport utility versus passenger cars Factor decreasing severity • In two-vehicle crashes, driving a pickups or sport utility vehicle versus being occupant of other collision partners
Quddus et al. (2002)	(1) Slight injury (2) Serious injury (3) Fatal injury	1992-2000 Singapore accidents involving motorcycles	Ordered probit model	The most severely injured person	- Motorcyclist characteristics[a] - Motorcycle characteristics[d] - Crash characteristics[e] - General info[b] - Road geometry[c]	Factors increasing severity • The motorcyclist having non-Singaporean nationality • Motorcycle increased engine capacity • Motorcycle headlight not turned on during daytime • Collisions with pedestrians and fixed objects • Having a pillion passenger • Riding during early morning hours • The motorcyclist being at fault Factor decreasing severity • Wet surface Factor not influencing severity • Motorcyclist age

Zajac and Ivan (2002)	(1) No injury (2) Probable injury, but not visible (3) Not disabling injury, but visible (4) Disabling injury (5) Fatality	1989-1998 rural Connecticut (US) highway accidents involving pedestrians crossing the road at locations with no traffic control	Ordered probit model	Crash	- Driver characteristics[a] - Vehicle characteristics[d] - General info[b] - ADT[h] - Non-motorist characteristics[g] - Speed limit - Area type - Weather	Factors increasing severity • Driver alcohol consumption • Pedestrian age 65years and older • Pedestrian alcohol consumption • Village, downtown fringe and low-density residential areas versus compact residential, low-density commercial and medium density commercial areas Factor not influencing severity • On-street parking
Abdel-Aty et al. (1998)	(1) No injury (2) Injury (3) Fatal	1994-1995 Florida crash and ADT files	Log-linear model with 3 variables and two-way interactions	Drivers by age group	- Driver characteristics[a] - ADT[h]	Factor increasing fatality • Being very old (80+) versus being old (65-79) Factor decreasing severity • Belonging to middle age group (25-64)
O'Donnell and Connor (1996)	(1) Non-treated injury (2) Treated injury (3) Admitted injury (4) Death	1991 NSW (Australia) motor vehicle accident victims' file	(1) Ordered logit model (2) Ordered probit model	Motor-vehicle occupant	- Vehicle characteristics[d] - Crash characteristics[e] - Occupant's characteristics[j]	Factors increasing severity • Seating on the left-rear position • Being female • Being involved in head-on collisions Other finding • The effects of increasing the age of casualty from 33 to 50 are greater than effects of a speed increase from 42 to 100 km/h
Shankar et al. (1996)	(1) Property damage only (2) Possible injury (3) Evident injury (4) Disabling injury or fatality	1988-1993 accidents on rural freeways in Washington State (US)	Nested logit model	Accident	- Driver characteristics[a] - Vehicle characteristics[d] - Crash characteristics[e] - General info[b] - Road geometry[c] - Weather	Factors increasing severity in single-vehicle crashes • Wet-pavement rear-end collisions, curves' length number, all drivers being male, vehicle-mass difference indicator Factors decreasing severity in single-vehicle crashes • Icy or snow-covered pavement, restraint system use

[a]Under ‘driver (rider/car user) characteristics’ category, several different variables related to the drivers (riders/car users) –implicated in the accidents treated- are considered in each study. E.g. alcohol consumption and sex.

[b]Under ‘general info’ category, several different variables related to the crash are considered. E.g. daylight and time.

[c]Under ‘road geometry’ category, several different variables related to the geometry of the road segment –on which the accident took place are considered in each study. E.g. curvature and intersection.

[d]Under ‘vehicle characteristics’ category, several different variables related to the vehicles –implicated in the accidents treated- are considered in each study. E.g. age and type.

[e]Under ‘crash characteristics’ category, several different variables related to the crash incident are considered in each study. E.g. collision type or speed.

[f]Under ‘traffic characteristics’ category, several different variables related to traffic are considered in each study. E.g. traffic flow and truck %.

[g]Under ‘non motorist characteristics’ category, several different variables related to pedestrians and cyclists –implicated in the accidents treated- are considered in each study. E.g. age and sex.

[h]Average daily traffic volume.

[i]Under ‘toll characteristics’ category, several different variables related to toll plazas and road users are considered. E.g. e-pass user or not and plaza structure.

[j]Under ‘occupant’s characteristics’ category, several different variables related to each occupant –of the vehicles involved in crashes- are considered in each study. E.g. position in vehicle and age.

4.3 Data and Methodology

4.3.1 The data

To explore the factors that determine occupant injury severity, the highway A4-A86 section from a dense urban area a few miles to the east of Paris was selected. Accident, weather, and traffic data were extracted from the same databases as in the crash type analysis (3.3.1 The Data).

The unit of the analysis adopted was *any vehicle occupant* involved in an accident (rider, driver or passenger) resulting in at least one person being slightly injured. The severity levels considered were:

- ‘no injury’,
- ‘slight injury’,
- ‘severe or fatal injury’.

Each observation in the dataset is a record of the level of injury sustained by each vehicle occupant involved in the accident and by various external factors. By the term ‘vehicle occupant’, we refer to either drivers or passengers. Eventually, a single accident corresponds to various observations; whose number equals the number of all persons involved in the accident. Table 9 presents a definition for each variable together with its type, some summary statistics, and a short description.

Table 9 Explanatory variables in severity analysis

Variable	Type	Summary Statistics[1]	Description
General accident information			
Road direction			
(sens)	Binary	F(1)=47.1%	=1from Paris; =2 to Paris
Type of day			
(tday)	Binary	F(1)=63%	=1 if weekday; =2 if weekend
Time of the day			
(heure)	Continuous	M=11.7,SD=7.9	time of accident (24 h clock)
Lighting conditions			
(daylight)	Dummy	F(0)=40.4%	=0 if daylight; =1 otherwise
(crepus)	Dummy	F(0)=12.7%	=0 if dawn or dusk; =1 otherwise
(nsansep)	Dummy	F(0)=1.5%	=0 if nighttime with no public lighting; =1 otherwise
(nepallum)	Dummy	F(0)=45.4%	=0 if nighttime with public lighting; =1 otherwise
Road surface condition			
(normal)	Dummy	F(0)=63.0%	=0 if dry; =1 otherwise
(mouille)	Dummy	F(0)=35.4%	=0 if wet; =1 otherwise
Pavement condition			
(bon)	Dummy	F(0)=99.3%	=0 if good or comfortable; =1 otherwise

Road curvature			
(ligne)	Dummy	F(0)=60.0%	=0 if straight line; =1 otherwise
(flat)	Dummy	F(0)=65.2%	=0 if flat; =1 otherwise
Weather conditions			
(normale)	Dummy	F(0)=77.2%	=0 if weather is fine or cloudy; =1 otherwise
(pluie)	Dummy	F(0)=22.5%	=0 if raining; =1 otherwise
(neige)	Dummy	F(0)=0%	=0 if snowing; =1 otherwise
(brouilla)	Dummy	F(0)=0.3%	=0 if weather is foggy; =1 otherwise
Road user attributes			
Socio-professional status			
(c_chom)	Dummy	F(0)=6.4%	=0 if unemployed; =1 otherwise
(c_prof)	Dummy	F(0)=2.8%	=0 if professional driver; =1 otherwise
(retrait)	Dummy	F(0)=1.9%	=0 if retired; =1 otherwise
Sex			
(masculine)	Dummy	F(0)=93.9%	=0 if male; =1 otherwise
Age			
(age)	Continuous	M=34.3,SD=11.6	Road user's age in years
Travel purpose			
(travail)	Dummy	F(0)=32.5%	=0 if work or university; =1 otherwise
(prof)	Dummy	F(0)=17.8%	=0 if professional use; =1 otherwise
Restraint system use			
(utilse2)	Binary	F(0)=83.9%	=1 if yes; =2 no
Alcohol consumption			
(alcool)	Binary	F(1)=95.4%	=1 if legal; =2 if illegal
Road user			
(conduct)	Dummy	F(0)=78.9%	=0 if driver; =1 otherwise
Driver's experience			
(anciperm)	Continuous	M=9.4,SD=9.7	Years of driving license holding
(ancienw1)	Continuous	M=8.8,SD=9.2	=anciperm*normale
Vehicle's characteristics			
Vehicle type			
(moto)	Dummy	F(0)=20.2%	=0 if 2wheels; =1 otherwise
(VL)	Dummy	F(0)=71.9%	=0 if car; =1 otherwise
(HV)	Dummy	F(0)=4.4%	=0 if heavy vehicle; =1 otherwise
Vehicle age			
(ancient)	Continuous	M=6.0,SD=4.3	Vehicle's age
Traffic characteristics			
Traffic volume			
(Q6minv)	Continuous	M=104.2,SD=41.3	Average traffic volume per lane and over 6 minutes (in vehicles)
(Q1)	Dummy	F(0)=67.9%	=0 if Q6minv<112; =1 otherwise
(Q2)	Dummy	F(0)=32.1%	=0 if Q6minv>112; =1 otherwise
Speed			
(Vmoy)	Continuous	M=82.6,SD=31.4	Average speed for all lanes and over 6 minutes (in km/h)
(VQ1)	Continuous	M=75.7,SD=31.1	=Vmoy*Q1 (in km/h)
(VQ2)	Continuous	M=85.7,SD=31.1	=Vmoy*Q2 (in km/h)

[1] F: frequency
M: average value
SD: Standard Deviation

Almost all independent variables are defined as extracted from the data bases used, with the exception of 'ancienw1', 'Q1', 'Q2', 'VQ1', and 'VQ2'. 'Ancienw1' equals the years of driving license holding under weather conditions other than normal. This variable was defined to explore the combined effect of adverse weather and driving experience. The mean value and the standard deviation of 'ancienw1' were estimated after omitting all zero values (77.2% of the observations). ‘Q1’ and ‘Q2’ were defined to separate traffic flow in two regimes; that is respectively over or under 112 vehicles per lane in 6-minutes time. The 112 value resulted after performing various trials to significantly separate the two regimes. ‘VQ1’ and ‘VQ2’ are intended to reflect the differential effect of speed on severity under different traffic flow regimes. It has to be noted that –as in the case of ‘ancienw1’- the descriptive statistics computations (for both VQ1 and VQ2) were made after discarding of all zero values.

4.3.2 Methodology

Ordered-response models recognize the indexed nature of dependent variables and assume the existence of an underlying continuous latent variable – related to a single index of explanatory variables- and an error term (Greene, 2003). In an ordered probit model the random error associated with this continuous variable is assumed to follow a normal distribution. Because the injury severity outcome of traffic accidents is ordered (0 for no injury, 1 for slight injuries, and 2 for severe injury/fatality), ordered probit models have been extensively used to model the marginal probability effects of several contributory factors on severity (Abdel-Aty, 2003; Gray et al., 2008; Kockelman and Kweon, 2002; Lee and Abdel-Aty, 2005; O’Donnell and Connor, 1996; Pai and Saleh, 2008; Quddus et al., 2002; Xie et al., 2009; Zajac and Ivan, 2002). Markedly, all such models are multinomial choice models (polytomous) as they provide more than two available possibilities.

The normality assumption made by the ordered probit models is not restrictive, since ‘shifting’ the thresholds would alter the probabilities of observing each severity outcome (Washington et al., 2003); further, it allows conditional heteroskedasticity to be captured more easily than with other specifications. An additional attractive property of the ordered probit model –versus other models of discreteness- that makes it appropriate for exploring severity is that the differences between the ordinal categories of the dependent variable (no injury, slight injury, and fatality) are not assumed to be equal (McKelvey and Zavoina, 1975).

The severity function determining the severity level for each individual (*in*) can be defined as follows (Greene, 2003):

$$y^*_{in}=\beta X_{in}+\varepsilon_{in} \quad \text{(Equation 4.1)}$$

where:

y^*_{in} denotes the latent injury risk propensity for each individual involved in a road accident, X_{in} is a vector of the independent variables considered, β is the vector of estimable coefficients, and ε_{in} is a random error term assumed to follow the standard normal distribution across individuals.

It is reasonable to assume that unobserved values of injury y_{in}^* correspond to observed values of injury y_{in} as follows (Greene, 2003):

$$y_{in}=\begin{cases} 0 \ \text{if} -\infty<y_{in}^*<\mu_1 & \text{(no injury)} \\ 1 \ \text{if} \ \mu_1<y_{in}^*<\mu_2 & \text{(slight injury)} \\ 2 \ \text{if} \ \mu_2<y_{in}^*<+\infty & \text{(severe or fatal injury)} \end{cases}$$

Thresholds μ_1, μ_2 ($\mu_1 < \mu_2$) are constant and to be estimated along with the β.

Then, the predicted probability of the injury level l ($l = 0,1,2$) for given X_{in} is given by (Greene, 2003):

$$\begin{aligned} \hat{P}(y_{in}=0|X_{in}) &= F(\widehat{\mu_1}\_X_{in}\hat{\beta}) \\ \hat{P}(y_{in}=1|X_{in}) &= F(\widehat{\mu_2}\_X_{in}\hat{\beta}) - F(\widehat{\mu_1}X_{in}\hat{\beta}) \\ \hat{P}(y_{in}=2|X_{in}) &= 1 - F(\widehat{\mu_2}\_X_{in}\hat{\beta}) \end{aligned} \quad \text{(Equation 4.2)}$$

However, some assumptions of the standard probit model pose limitations to its application, including marginal probability effects that (marginal effects) change their sign exactly once when moving from the smallest to the largest outcome and that possible *heterogeneity* among observations is not properly addressed; these shortcomings are a result to the following properties:

(i) independent variables coefficients $\hat{\beta}$ are fixed across injury levels,

(ii) the thresholds $\hat{\mu}_l$ are fixed across observations,

(iii) the probability function (Equation 4.2) is single –indexed,

(iv) the error term is normally distributed.

Much research in the statistics and econometrics literature has gone into addressing these shortcomings. To address the fixed coefficients assumption (i), Everitt (1988) proposed a *finite mixture model* that accounts for heterogeneity between groups of individual observations by clustering. In the case of ordered data, it provides a very flexible way of modeling heterogeneity among groups of individuals (Boes and Winkelmann, 2006), while also improving on the marginal effect estimation. In the same case (i), Boes and Winkelmann (2006) proposed a *random coefficients model* as a more flexible specification. These models randomize the parameters of interest by introducing an error term correlated with the unobserved factors in ε_{in} (from Equation 4.1). In the case of fixed thresholds (ii), Terza (1985) and Maddala (1983) proposed the *generalized threshold model* which overcomes the limitation of fixed thresholds by allowing them to be dependent on covariates; this generalization also makes the analysis of marginal effects more flexible, but includes additional parameters to be estimated. In the case of the probability function (iii), Boes and Winkelmann (2006) proposed a *sequential ordered-response model* based on methods used in the literature on discrete time duration data. The probability function of the dependent variable is expressed as a sequence of binary choice models where each decision is made for a specific category *l* conditional on refusing all smaller categories (this is achieved by starting from the lowest category and moving stepwise to the highest). Under this generalization, even though marginal effects estimation is more flexible, it becomes computationally cumbersome.

In transportation research, some of these variances to the standard ordered response model have been employed; examples are the works of Anastasopoulos and Mannering (2009), Eluru et al. (2008) and Milton et al. (2008). In the present thesis, we use the random coefficient specification where road user is the unit of analysis. Since each individual has specific characteristics that may influence the severity outcome differentially, there is a possibility of (additional) heterogeneity in the model. However, in the standard ordered probit model the distributional assumption does not allow for additional heterogeneity between individual realizations; the random parameter models allow the influences of variables affecting accident injury-severity proportions to vary across observations. This is achieved by adding an error term that is correlated with the unobserved factors in ε_{in} and translating individual heterogeneity into parameter heterogeneity as follows (Greene, 2003):

$$\beta_{in}=\beta+\varphi_{in} \qquad \text{(Equation 4.3)}$$

where φ_i is a randomly distributed term.

The severity function now becomes (Greene, 2003):

$$y^*_{in}=\beta X_{in}+\varepsilon_{in}' \qquad \text{(Equation 4.4)}$$

where ε' is the new error term (Greene, 2003):

$$\varepsilon_{in}'=X_{in}\,\varphi_{in}+\varepsilon_{in} \qquad \text{(Equation 4.5)}$$

4.4 Empirical Results

Choice probabilities in random parameters (mixed) models of discrete choice take the form of a multidimensional integral over a mixing distribution (Brownstone and Train, 1999). The integral does not have a closed form and, so, it must be evaluated numerically. If the integral is approximated with random draws, a large number of draws is usually needed to assure low simulation error in the estimated parameters (Train, 2000). Bhat (2001) tested Halton sequences (‘intelligent’ draws) for mixed logit estimation and found them be vastly superior to random draws. In particular, he found that the simulation error in the estimated parameters was lower using 100 Halton numbers than 1000 random numbers. The reasons for the improvement are twofold. First, the Halton numbers are designed to give fairly even coverage over the domain of the mixing distribution. With more evenly spread draws for each observation, the simulated probabilities vary less over observations, relative to those calculated with random draws (Train, 2000).

The probit specification of Equation 4.5 was estimated using simulation-based maximum likelihood, as maximum likelihood estimation of the random parameters ordered probit model is computationally cumbersome (Anastasopoulos and Mannering, 2009). Halton draws were used to estimate the parameters that maximized the simulated log-likelihood function and normal, triangular and uniform distributions were considered for the functional form of the parameter density function. The statistical software Limpdep (v.8) has been used for all applications.

Model estimation results are shown in Table 10; omitted variables were removed from the final model on the basis of low statistical significance. All estimated parameters included in the final model are statistically significant (at a 95% confidence level) and the signs are plausible as discussed below. The standard deviation of the parameter distribution was significantly different from 0 for all but two of the variables included in the final model; average speed in low traffic conditions and traffic volume was found to have fixed parameters across the population of road users. The normal distribution was found to provide the best statistical fit for the density function of the random parameters.

Table 10 Model estimation results for random and fixed parameters ordered probit models

Variable	Fixed parameters model		Random parameters model		
	coefficient	t-statistics	coefficient	t-statistics	SD[a]
Constant	0.426	0.33	2.429	6.48	0.737
‘Tday’	-0.737	-7.71	-1.680	-9.62	1.132
‘daylight’	0.930	10.17	1.544	10.18	2.277
‘Normal’	-0.508	-5.32	-1.064	-6.39	0.574
‘Flat’	0.629	7.09	0.764	5.59	0.092
‘Ancienw1’	-0.009	-8.69	-0.067	-3.46	0.027
‘Moto’	-0.627	-6.74	-1.805	-10.79	0.110
‘HV’	0.626	6.74	0.481	1.69	0.030
‘VQ1’	0.017	1.26	0.015	8.59	0.000
‘Q6minv’	-0.001	-0.83	-0.012	-6.96	0.000
Thresholds					
$\mu 1$	0.988	22.36	1.934	13.52	
$\mu 2$	2.011	33.13	4.330	16.65	
Number of observations	893				
Log-likelihood with constant only LLI	-1207.487		-1207.487		
Log-likelihood at convergence LL(β)	-1045.860		-590.441		
ρ_c^2=(LLI-LL(b))/LL(c)	0.134		0.511		

[a]Standard Deviation of parameter distribution

Table 10 indicates that the random parameters model significantly overperforms the fixed parameters model based on both the log-likelihood at convergence ($LL(\beta)$) and the overall fitting (ρ_c^2 statistic) which improve noticeably when moving from the fixed to the random parameters specification (Washington et al., 2003). Also, the likelihood ratio test[1] yields a value of 910.84, indicating a confidence that the random parameter model performs ‘better’. This modeling approach allows for the possibility of heterogeneity being present as it can capture the different effects of the independent variables on the population of vehicle occupants. We note that, besides statistical fit, the two modeling specifications yield different qualitative and quantitative results for the parameter estimates. For example, the traffic related variables (‘Vq1’ and ‘Q6min’) were found to be statistically significant only in the random parameters analysis, which would yield on widely different policy recommendations depending on the model selected.

The *type of the day* (working day or weekend/holiday) on which the accident occurred was found to have an effect on the severity outcome and resulted in a normally distributed random parameter with a mean of -1.680 and a standard deviation of 1.132. In particular, accidents occurrence on weekdays seems to increase the severity outcome sustained by most vehicle occupants, but this effect varies across the population of vehicle occupants. This finding may result from the higher level of alert of the emergency response systems on weekends. Further, regular drivers that commute daily and follow the same itinerary may be over-confident, may over-estimate their abilities and under-estimate potential dangers. However, further research is needed to validate the above. Similar were the findings in Quddus et al. (2009), but not in Gray et al. (2008), where more severe crashes were found to occur on Fridays, Saturdays and Sundays. If we attempt to further interpret the distributional results, we find that approximately 93.1% of the individuals are more seriously injured on working days (only 6.9% of the distribution is above 0). The latter indicates that not all users react uniformly to the type-of-day variable. This conclusion

[1] The likelihood ratio test statistic is $-2[LL(\beta_{fixed})-LL(\beta_{random})]$ and is X^2 distributed with v the degrees of freedom (Washington et al., 2003); v is equal to the difference in the numbers of parameters in the fixed and random parameters models (here v=7).

would have been neglected under a fixed-parameter analysis, introducing a bias in the results because of unaccounted heterogeneity.

The *lighting conditions* were also found to be significant in the injury outcome. Specifically, daylight conditions were found to exercise a positive and variable influence on the severity outcome of road users involved in accidents. A random parameter with a mean of 1.544 and a standard deviation of 2.777 was estimated for the related indicator variable (0 if daylight, 1 otherwise). This suggests that 71% of the distribution is over zero. It can be assumed that under daylight, most drivers' vision is better and, eventually, they have more time to perceive and comprehend the road environment and to react correspondingly. Therefore, as they perceive potential dangers earlier, they have more time to reduce their speed or perform other last-minute actions that reduce the severity of an on-coming crash. Similar conclusions were drawn by Chimba and Sando (2009), Helai et al. (2008), Lee and Abdel-Aty (2005), Pai and Saleh (2008), Savolainen and Mannering (2007).

Even though weather conditions were not found to determine severity, *road surface conditions* seem to significantly affect it. The presence of normal surface conditions (normal=0) versus all others (wet or icy pavement) resulted in a normally distributed random parameter with a mean of -1.064 and a standard deviation of 0.574 implying that normal road surface conditions significantly increase accident severity (negative coefficient), but their effect is not the same across different individuals. Interestingly, the distribution is over 0 at almost 97% of its surface, suggesting that normal road surface conditions almost always aggravate the severity outcome. We note however that this effect is not the same across road users (as it would be in a fixed-parameters model), but has a varying magnitude; normal surface conditions were found to provoke more severe accidents in several studies (Savolainen and Mannering, 2007 for single-vehicle crashes; Quddus et al., 2002; Savolainen and Yamamoto and Shankar, 2004 for urban areas; Shankar et al., 1996; Xie et al., 2009). This effect can be explained by driver risk-adjusting behavior to the environment (risk offset hypothesis), where on wet or icy pavement drivers are more careful.

Further, as anticipated, no (*road) curvature* was found to reduce severity of individuals involved in crashes. In other words, vertical curvature is expected to increase the severity level sustained. The relative indicator variable (‘flat’) equals zero for no vertical curvature and 1 otherwise. Its estimated coefficient was found to be normally distributed with a mean of 0.764 and a standard deviation of 0.092, implying that practically all road users react uniformly to road curvature with respect to the severity outcome. Ascending or descending highway segments may limit the driver’s field of vision causing a reduction in the time available for reacting to potential dangers. This finding corroborates earlier observations by Savolainen and Mannering (2007) for single-vehicle crashes. However, the magnitude of this effect is variable across drivers and rather limited as freeways design meet high standards.

The variable capturing the combined effect of *rainy weather and drivers’ experience* (‘ancienw1’) was defined to explore the possible association between severity and the behavior of inexperienced drivers (holding recent driving licenses) under adverse weather conditions. Indeed, it was found that under severe weather conditions such as snow, severity levels increase for ‘recent driver licenses’ (inexperienced drivers). The corresponding coefficient was found to be normally distributed with a mean of -0.067 and standard deviation of 0.027. This finding implies that under fine weather conditions, there is no significant effect of the driver’s experience on severity outcome. On the contrary, increased experience causes a significant reduction in the probability of severe accidents in rainy weather.

Two of the explanatory variables (‘moto’, ‘HV’) examined refer to the type of vehicle in which the road user was travelling at the time of the accident. They were both found to follow the normal distribution. Specifically, the parameter distribution for the indicator variable for 2-wheelers (moto=0, 1 otherwise) was found to have a mean of -1.805 and a standard deviation of 0.110. Correspondingly, the parameter distribution for the indicator variable of heavy vehicles occupants (HV=0, 1 otherwise) was found to have a mean of 0.481 and a standard deviation of 0.030. This indicates that practically all 2-wheels riders (all distribution below zero) have significantly higher probabilities of getting severely injured if involved in accidents. In contrast, practically all heavy vehicle drivers and passengers (all distribution above zero) have significantly lower probability of suffering a severe injury when involved in crashes. This finding can be explained by the difference in the mass of heavy vehicles compared to other vehicles; in case of a collision, the lighter vehicles absorb the greatest part of the kinetic energy. The limited protection that motorcycles offer along with their reduced stability and the difficulty in being observed by other road users influences severity in this direction. Abdel-Aty (2003) claimed that passenger car occupants suffer more severe injuries than van and pickup occupants. Shankar et al. (1996) concluded that the vehicle-mass difference indicator is a factor significantly affecting severity, while Yamamoto and Shankar (2004) found that driving a motorcycle or truck in urban areas significantly decreases driver injury severity. We note that, in the UK, heavy goods involvement in motorcycle accidents occurring at T-junctions was also found to increase severity (Pai, 2009).

Turning to traffic characteristics, model estimation results indicate that they have fixed parameters across observations as the standard deviation was not found to be significantly different from zero. The negative sign (-0.012) of the traffic volume coefficient (Q6minv) implies that for lower traffic volumes, probability of more severe accidents is significantly higher. The latter comes to verify the common assumption that under free flow, drivers tend to travel at higher speeds and, thus, the severity level of potential accidents increases (Golob et al., 2008; Quddus et al., 2009; The Scottish Office Central Research Unit, 1997). Indeed, the average speed developed under dense traffic conditions (>1,120 vehicles/lane/hour) was found to have a significant and positive association with severity as the corresponding variable (‘VQ1’) has a positive parameter coefficient (0.015). This implies that beyond a given traffic volume level (1,120/lane/hour), higher speeds imply higher probability for more severe accidents. However, there appears to be no significant difference beneath this traffic volume; this can be attributed to the dispersion of speeds under free flow at a rather random manner and does not affect severity in a consistent pattern. We note that in previous work by Aron et al. (2009) in accident frequency estimation, the authors concluded that under ‘fluid’ traffic conditions only 29% of the accidents occur, while fewer injury accidents occur in ‘dense’ traffic.

4.5 Concluding remarks

Accident severity is of particular concern to both decision makers and researchers. Past studies have indicated several factors as significantly influencing crash-injury severity (e.g. driver age, weather conditions and so on). In the present thesis, using highway data from Paris, France, we found that travelling on 2-wheels, at night or on highway segments with curvature significantly increases the probability of getting involved in more severe accidents. In contrast, travelling in heavy vehicles, on weekends or on dry pavement surfaces, reduces the probability of severe accidents. Less experienced drivers seem to encounter problems in dealing with adverse weather conditions and related potential dangers.

Most importantly, the analysis illustrated that there is a significant relationship between the severity outcome and the traffic characteristics at the time of the accident. Traffic volume was found to have a consistently positive effect, while speed appears to have a differential effect with respect to traffic volume. While in higher traffic volumes higher speeds aggravate severity outcome, in lower traffic volumes speed does not significantly influence severity in a consistent pattern. This finding indicates that speed-reducing measures should be considered even in rather dense traffic highway segments (that allow however for speed variation among drivers) and should address speeds lower than the posted speed limits. In this context, real time adjustment of speed limits may prove very beneficial, though further research is needed to verify the latter.

The modeling approach presented in this chapter is a random parameters ordered probit model that offers the possibility of heterogeneity as it captures the differential effect of the independent variables on the population of road users. This differentiation does not only concern the magnitude of the effect on the population, but also the effect itself, whether it is positive or negative across the population. The use of a fixed parameters ordered probit model would lead to neglecting heterogeneity, biasing the results and making incorrect policy recommendations.

Introducing real-time collected data from the time of the accident can provide additional insight into the context that severe accidents occur and could also prove helpful in reducing severity of accidents. If further combined with frequency models, they could help in identifying appropriate safety enhancements, in estimating monetary gain and possibly in preventing severe accident occurrence.

Suggestions for further research would include the influence of real-time traffic data aggregation level. On the freeway treated, traffic data are collected for every 6 minutes, while on other infrastructures, the respective period is much shorter. For example, on most U.S. freeways the measurements are averaged and collected for every 30 seconds.

Chapter 5

Incident management using real-time traffic data on urban freeways

This Chapter investigates the introduction of road safety analysis outcomes in an integrated incident management scheme. To this end, we provide a synopsis of related incident management analyses. A synthesis is then performed in the effort of establishing a conceptual framework for incident management applications using real-time traffic data on urban freeways. We use dissertation previous findings to explore potential implications towards incident propensity detection and enhanced management.

5.1 Introduction

5.1.1 Incident management and duration

Over the last decades, incident management has been of great interest to both researchers and practitioners. Incident management includes a variety of applications under the objective of best addressing an incident occurrence (as well as its consequences) in various fields such as industrial failures, natural disasters, and so on. In transportation research, incident management is defined as *the systematic, planned, and coordinated use of human, institutional, mechanical, and technical resources to reduce the duration and impact of incidents, and improve the safety of motorists, crash victims, and incident responders* (FHWA, 2000).

Incident duration is the time elapsed between the occurrence of an incident and its complete clearance (Zografos et al., 2002). During this time interval, consecutive actions taken by the operators are: detection, verification, motorist information, response, site management, traffic management, and clearance (Bunn and Savage, 2003). Among other goals, incident management strategies aim at minimizing incident duration. An Integrated Management System (IMS) consists of the following three subsystems: i) incident detection, ii) incident response logistics, and iii) motorist information and traffic management (Zografos et al., 2002). The benefits of minimizing incident duration are numerous and concern highway operators (e.g. cost, road safety performance), crash victims (e.g. time to hospital), other road users (e.g. delays, secondary incidents), and society (e.g. incident externalities). Besides, the proper identification and prioritization of factors that contribute to emergency management services response and clearance times result in better usage of taxpayer resources (Lee and Fazio, 2005).

5.1.2 Emergency response and location analysis

Emergency station location (e.g. police, fire stations) analysis falls into location analysis; term that refers to the modeling, formulation, and solution of a class of problems that can best be described as sitting facilities in some given space (ReVelle and Eiselt, 2005). Obviously, emergency unit location is important to overall incident duration. In particular, the time needed to reach an incident scene is of great concern to emergency medical services (EMS) in order to mitigate incident consequences on people. In a real-time context, EMS managers are faced with two main problems: an allocation problem and a redeployment problem (Gendreau et al., 2001). The allocation problem consists of determining which ambulance must be sent to answer a call. The redeployment problem consists of relocating available ambulances to the potential location sites when calls are received; ambulances are assigned to potential sites to provide coverage. Covering constraints may be either absolute or relative. Absolute constraints require that all demands are satisfied within r_2 minutes, while relative constraints require that a proportion of demand a is also satisfied within r_1 minutes ($r_2 > r_1$).

There are four components that characterize location problems; these are 1) customers, who are presumed to be already located at points on routes, 2) facilities that will be located, 3) a space in which customers and facilities are located, and 4) a metric that indicates distances or times between customers and facilities (ReVelle and Eiselt, 2005). Facility location models have been widely applied in real life problems with examples that include the sitting of EMS, police and fire stations, bus garages and airline hubs (Current et al. 2002). Comprehensive reviews of such models can be found in Drezner and Hamacher (2002), Goldberg (2004), Revelle and Eiselt (2005), and Jia et al. (2007), while Brotcorne et al. (2003) provide a focused review of their application in emergency response services. Location models are distinguished in coverage and median type models (Berman and Krass, 2002). Coverage-type models attempt to locate servers so that adequate coverage is provided to demand points, implying that there is at least one server that can undertake demand for service in a position within a preset maximum distance. Median-type models minimize average or total travel cost between servers and demand and locate them accordingly.

Early efforts on emergency response service planning focused on two basic coverage models: the Location Set Covering Problem (LSCP) by Toregas et al. (1971) and the Maximal Coverage Location Problem (MCLP) by Church and Revelle (1974). The former consists in minimizing the number of facilities required to cover all demand nodes within a specified time or distance standard, while the latter assumes that the number of vehicles is less than the number needed to cover all demand nodes. Later efforts considered the case of several server types (Marianov and ReVelle, 1992; Schilling et al., 1979) and multiple coverage of demand for service (Gendreau et al., 1997, 2001; Hogan and ReVelle, 1986). On the other hand, the p-median problem, originally proposed by Hakimi (1964) was used by Calvo and Marks (1973), Carson and Batta (1990) and Paluzzi (2004) for planning emergency response services.

Basic location models are deterministic and, in that sense, do not capture inherent uncertainties often encountered in emergency response services (Brotcorne et al. 2003; Jia et al., 2007). As a result, dynamic models that involve uncertainty (mainly uncertainty regarding demand evolution) were developed (Revelle and Eiselt, 2005). Probabilistic methods and scenario approaches were employed to address such problems. Nonetheless, Revelle and Eiselt (2005) note that very few references exist on the subject of cyclic demands for emergency responses. Cyclic or periodic demands refer to relatively predictable demand patterns that fluctuate through a day or a year. Another approach derives from queuing, where customers will patronize facilities not only based on their proximity to them, but also on the expected congestion at facilities (Larson, 1974). If parameters are uncertain, and furthermore, no information about probabilities is known, we refer to robust optimization problems (Snyder, 2006) that often optimize the worst case performance of the system.

5.1.3 Ambulance location and relocation problems

Brotcorne et al. (2003) provided a substantial review on the ambulance location and relocation models proposed over the last 30 years. The authors note that advanced information technologies are often used to assist the ambulance management process in terms of road network surveillance, vehicle positioning systems, geographical information systems, and so on. Ideally, they remark, these systems should be fully integrated and interconnected within an ambulance relocation module. Besides, with newest models and algorithms, large scale problems can be solved rapidly and dynamically in real-time, with a high level of accuracy. Thus, a new ambulance redeployment strategy can be recomputed at any time *t*, using real-time available information. The authors trace only one such research effort (Gendreau et al., 2001) and note potential research interest in the field.

Carson and Batta (1990) developed a model where a single ambulance was relocated on the Amherst Campus of the State of New York according to the population movements throughout a day. The authors divided the day in four unequal time periods and solved one-medium problems on respective network states to determine locations for the ambulance. Gendreau et al. (2001) developed a model that – in addition to the standard coverage and site capacity constraints – included a number of practical considerations inherent to the dynamic nature of the problem such as avoiding long trips, and avoiding round trips. The objective was to maximize the backup coverage demand minus a relocation cost. The model was truly dynamic since it incorporated new information on the state of the system received at each period *t* that a call was registered. The authors developed a fast tabu search heuristic implemented on parallel processors. Rajagopalan et al. (2008) formulated a dynamic model with the objective of determining the minimum number of ambulances and their locations for each time cluster in which significant changes in demand pattern occurred while coverage availability requirements were met. More recently, Schmid and Doerner (2010) worked on ambulance location and relocation problems in urban environment. They developed a multi-period version, taking into account time-varying coverage areas, where they allowed vehicles to be repositioned in order to maintain certain coverage standard through the planning horizon. The total planning horizon of 24hours was equally split into 6 time intervals and time-dependent travel times were aggregated accordingly.

An overview of time-dependent ambulance location and relocation models reveals that dynamic models may refer to:

- fluctuations in demand (calls) (Rajagopalan et al., 2008)
- population movements throughout a day (Carson and Batta, 1990)
- changing fleet sizes (Repede and Bernardo, 1994)
- ambulance specific busy probabilities (Galvão et al., 2005)
- travel time variability (Schmid and Doerner, 2010)

5.2 Real-time traffic data in incident management

Conventionally, planning and resources coordination were predefined. Lately, technological advancements allow for resources co-ordination to be made real-time taking into account time-varying parameters such as traffic flow. Nowadays, most freeways are equipped with continuous surveillance systems making disaggregate traffic data readily available; these have been used in incident management applications. In particular, real-time traffic data have been largely utilized under the scope of: a) travel time estimation, and b) incident detection.

a) Travel time on freeways is largely traffic-dependent; minimized under free-flow, it reaches a maximum when congestion is formed. Non-recurrent congestion is related to incident occurrences and accounts for over 50% of the total delay suffered by road users (Lindley, 1987). Incident response strategies include travel time implications as they search to minimize incident duration; the latter including emergency units' travel time and the time to restore roadway capacity. Thus, some studies address the issue of locating emergency units under the objective of minimizing travel times with regards to changing traffic conditions (Schmid and Doerner, 2010), while others focus on traffic incident management on-the-scene, in order to minimize the time required to restore road capacity (Bunn and Savage, 2003). All such efforts remain reactive in nature as they attempt to minimize incident impacts, while they do not search to prevent incidents from happening (as in proactive investigations). They all take place after an incidence occurrence.

b) Real-time decision support systems have been identified as promising means for improving the decision making capabilities in incidence response logistics (Zografos, 2002). Towards this direction, various authors (Madanat et al., 1995; Stephanedes and Liu, 1995) developed incident detection algorithms in the effort to promptly detect incidents and reduce the time required to initiate traffic management actions and emergency response measures (Corby and Saccomanno, 1997). Various criteria were used as accident-detection parameters; change in speeds, vehicle occupancy, traffic volume, and so on. Incident detection involves the analysis of patterns in traffic surveillance data observed just after the incident. The analysis is performed in order to develop models that separate real-time traffic conditions resulting from incidents from free-flow and/or recurring congestion (Abdel-Aty and Pande, 2007). In the 1990s, incident detection analyses attracted research interest since real-time traffic data became widely available. Later, new technologies (the use of mobile phones, GPS, CCTV systems etc.) made incident detection rather obsolete as users communicate directly with road operators in the case of an emergency; with the latter being able of promptly verifying an incident occurrence. Incident detection analysis is also reactive in nature and – by definition – takes place after an incident occurrence.

5.3 Real-time traffic data in road safety

Madanat and Liu (1995) were of the first researchers to conceive the idea of a real-time incident likelihood prediction simulator to *proactively* address incidents. They used freeway geometric characteristics, segment-wide characteristics (e.g. weather, time of the day), and section-specific conditions (e.g. traffic volume, speed, speed variance) as inputs to their model. The outputs were the time-varying likelihoods of traffic incidents.

Oh et al. (2000) classified traffic conditions in two patterns: (a) *disruptive*, being the traffic condition potentially leading to an accident occurrence and (b) *normal*, being the traffic pattern not involved in accidents. The standard deviation of speed averaged over a 5-minute interval was the best indicator of disruptive traffic flow. They used the I-880 freeway (California) accident and traffic dataset and estimated the likelihood of given traffic observations (as described by speed variation) belonging to either disruptive or normal conditions.

Golob and Recker (2004, 2001) and Golob et al. (2008) extensively worked on the matching of traffic flow parameters and crash characteristics including injury severity. Accident and traffic data from South California highways were used in that purpose. Traffic flow was measured in terms of 30-sec observations from inductive loop detectors in the vicinity of the accident prior to the time of its occurrence (-2.5 to -30min). In Golob et al. (2004), a prototype software tool was generated in an effort to develop a real-time safety monitoring tool. Kockelman and Ma (2007) used a subset of the previous study dataset to associate the time distance of a traffic observation (from accident occurrence time) to crash likelihood. However, empirical findings indicated no relationship between speed, speed variance, and crash likelihood.

Abdel-Aty and Pande (2005), Pande and Abdel-Aty (2006) and Abdel-Aty et al. (2007) used real-time traffic and accident data from the I-4 corridor in Orlando to identify crash propensity factors. Results showed that at least 70% of the crashes on the evaluation dataset could be identified. Furthermore, the authors evaluated ITS strategies in a simulation environment (PARAMICS) for their potential benefits in improving real-time safety on the I-4 corridor (Abdel-Aty et al., 2007). It was found that under congestion, ramp metering and/or route deviation can yield significant reduction in real-time crash risk. However, variable speed limits may be more beneficial.

Lee et al. (2002) opposed the notion of ‘preemptive warning system’ to conventional incident detection reactive systems. The goal was twofold: to a) in real-time identify traffic conditions associated with high crash frequency, and to b) intervene to modify traffic conditions (e.g. variable speed limit). Crash potential was defined as *the long-term likelihood that a crash will occur for given traffic, environment, and roadway conditions.* Using data from Canadian freeways (Lee et al., 2002, 2003, 2006a, 2006b), the authors found that variation in speed and traffic density were statistically significant predictors of rear-ends’ frequency, while variation in flow and peak/off-peak periods were correlated with sideswipe crashes. The optimal observation time was found to be precursor-specific. In Lee et al. (2006b), a microscopic traffic simulation was used to realistically simulate changes in traffic conditions as an effect of variable speed limits. In Lee et al. (2006c), the effects of a safety measure (ramp metering) were quantified; results suggested that ramp metering would reduce crash potential by 5-37%.

Concluding, safety-oriented analyses using real-time traffic data attempt to identify appropriate crash precursors that could act as identifiers of potentially dangerous situations. Results seem very promising, but have not been extensively applied. Applications include traffic management measures such as ramp metering (Lee et al., 2006c) and have not been integrated in IMS schemes for enhanced traffic incident management.

5.4 Conceptual framework

5.4.1 Proactive vs. reactive approaches

In many fields (e.g. industry, medicine), the concept of prevention is commonly described by a division into sub-concepts, each of which is intended to represent one main preventive strategy (Andersson and Menckel, 1995). The most widely employed classification in medicine was launched by Gjestland (1955). According to this classification, preventive activities are divided into primary, secondary, and tertiary activities that are related to different periods in time in the course of a disease. Primary prevention is taken in advance, while secondary and tertiary actions are taken later on.

Primary prevention can be further divided into proactive and reactive (Catalano and Dooley, 1980). Proactive activities are designed to deter or limit exposure, while reactive activities are aimed at the promotion of coping or increasing adaptation in response to an exposure that has already taken place (Catalano and Dooley, 1980). Thus, proactive actions are taken before exposure (primary activities), while reactive actions can be taken either before or after exposure but are always designed to have an effect after exposure (Andersson and Menckel, 1995).

In line with the above, proactive incident management includes all investigations made in the aim of finding ways to limit dangerous conditions and to prevent incidents from happening. Reactive approaches to generic incident management include all research performed in the area of being prepared to deal with the occurrence of a specific incident and of its consequences. The following scheme summarizes all possible preventive activities (Figure 8**Error! Reference source not found.**).

Figure 8 Preventive Activities

	Preventive Strategies	
	Primary	Proactive
		Reactive
accident		
	Secondary	Reactive
	Tertiary	Reactive

5.4.2 Incident management taxonomy

Employing a similar terminology in transportation research, proactive traffic incident management would refer to all actions taken in advance in order to limit or deter road users' exposure in danger. In essence, all such actions aim at preventing incidents from happening and may include passive or active measures; the latter requiring human participation and initiative. In line with the above, proactive measures may refer to enhanced road design (a passive measure), driver additional training (an active measure), or even real-time applications such as variable speed limits. In accordance to Gjestland (1955) classification, proactive incident management would – by definition – be a *primary* prevention as it takes place before incident occurrences.

Similarly, reactive traffic incident management would include all conventional incident management approaches, along with all actions taking place after the incident and aiming at mitigating its consequences. However, to the best of our knowledge, no primary reactive techniques have been developed with traffic incident management analysis. Such techniques would take place *before* the accident occurrence; while still targeting consequence mitigation. If, for example, a hazardous situation is occurring due to high speeds, the operator could –*proactively*- warn drivers for potential dangers or directly decrease speed limits (a passive measure if drivers' compliance is considered given). However, if drivers do not react accordingly and/or risk level stays high, the operator could –*reactively*- relocate the position of emergency vehicles in order to respond faster to a *potential* incident occurrence.

To the best of our knowledge, reactive incident management approaches taking place *before* the incident occurrence while utilizing real-time traffic data are not yet considered. Table 11 summarizes the proposed framework for traffic incident management techniques with respect to their target (proactive vs. reactive), to whether human initiatives are required (active, reactive), and to the time they apply (primary, secondary or tertiary). Indicative examples are given for each emerging category.

Table 11 Traffic incident management taxonomy

	Proactive		Reactive	
	active	passive	active	passive
Primary	*e.g. warning message for potential dangers*	*e.g. variable speed limits*	*e.g. warning message for restraint system use*	*e.g. ambulance relocation*
Secondary	-	-	*e.g. warning message for incident occurrence*	*e.g. ambulance allocation*
Tertiary	-	-	-	*e.g. enhanced health treatment*

Concluding, even though other disciplines (including road safety and traffic management) consider both proactive and reactive strategies, traffic incident management has remained mainly reactive in nature. Furthermore, it mainly considers passive measures and strategies where road user initiatives are not needed. Moreover, incident management techniques mainly take place after an incident occurs excluding any primary prevention considerations. Finally, crash type-specific strategies have little been discussed.

Real- time traffic data availability enables for additional applications; incorporating such data in incident management enables for new perspectives in road safety techniques and seems beneficial in many ways. Road safety studies have dealt with minimizing road user exposure by adopting either active or passive approaches; however without incident management considerations. To the best of our knowledge, incident management dynamic models do not take full advantage of real-time traffic data availability. Under the light of the above, integrating road safety analyses (using real-time traffic data) in incident management would help in minimizing incident duration as well as in reducing incident occurrences.

5.4.3 Integrating road safety analyses to incident management

In Chapter 3, we examined the effects of various traffic parameters on type of road crash. Multivariate Probit models were specified on 4-years of data from the A4-A86 highway section in the Ile-de-France region, France. Results provided the propensity of each of the five crash types considered with regard to lighting conditions, road gradient, traffic density, average speed, and traffic volume. Crash type propensity could be used in an IMS as an additional constraint in EU location. First, there is strong empirical evidence that severity outcomes are crash type-dependent (Yau, 2004; Pai and Saleh, 2008). Second, incident management (in terms of type of EU needed, overall duration, and so on) is related to crash type (Drakopoulos et al., 2001); the model outcomes could be used to estimate on a real-time basis the location of crash-specific emergency units such as fire-fighting vehicles. Third, crash type propensity (if combined with a corresponding crash frequency analysis) could be used in crash prediction modeling.

In Chapter 4, we applied a random parameters ordered probit model to explore the influence of speed and traffic volume on the injury level sustained by vehicle occupants involved in accidents on the A4-A86 junction in the Paris region. The application estimated a predicted probability of several injury levels given that an incident occurs with regard to type of day, lighting conditions, pavement condition, weather, road curvature, driver experience, type of vehicle, traffic flow, and average speed.

Severity probability estimation could be used in IMS in various ways. Traffic crash severity has the most effect on response times; by assessing resources currently dedicated to insignificant factors, emergency management services can further improve response times to those casualties that crucially need emergency services (Lee and Fazio, 2005). Moreover, severity outcome probability could be utilized in relocating ambulances along the highway according to the probability of severe incident occurrence. As a result, ambulances could be relocated nearer to the point where they are most needed and incident response time and duration could be significantly reduced. To the best of our knowledge, no such dynamic ambulance relocation model using probabilistic demand on highways has been developed.

Chapter 6

Conclusions

This Chapter summarizes thesis major findings and provides overall conclusions regarding the analysis performed. The thesis contribution is discussed, while indications for future research are given.

6.1 Research undertaken

Increases in accident related costs along with sustainable development concern have turned countries and international organizations towards accident mitigation programs and policies. Incident occurrence and response have also attracted considerable research interest in the past three decades. Regardless of modeling techniques, a serious factor of inaccuracy – in most past studies – has been data aggregation. The Average Annual Daily Traffic (AADT) has been the most commonly used measure to reflect traffic conditions. However as most freeways are equipped with continuous surveillance systems, disaggregate traffic data collection is possible as well as readily available; such data have been used in only a limited number of studies.

In this context, the main research question of the thesis was whether and how traffic parameters affect accident patterning, consequences, and response. The thesis objective was to use highway traffic data collected on a real-time basis in order to: a) explore the effects of various traffic parameters on type of road crash, b) investigate the influence of traffic parameters on the injury level sustained by vehicle occupants, and to c) explore possible implications in incident management strategies. To this end, four main research activities were undertaken:

1) a literature review of relevant road safety research,
2) an empirical investigation on accident type propensity,
3) an empirical investigation of vehicle occupant injury severity, and
4) the development of a conceptual framework towards introducing real-time traffic data in incident management and response.

6.2 Summary of findings

In the first research activity (Chapter 2), we summarized the state of the art in road safety research; the large body of literature was organized on the basis of both methodological and thematic criteria. These criteria included a) the method employed, b) the level of analysis assumed, c) the scope of the performed analysis, and d) the incident phase considered. The dissertation field of interest was defined with respect to the taxonomy established. In particular, a data observational study was conducted within a descriptive scope of analysis. Stochastic modeling was used to address the main research questions in a rather disaggregate context of analysis. Incident outcomes – in terms of either crash type or severity – were the dependent variables considered. Incident type refers to accident patterning, while severity is linked to incident consequences. To this end, real-time traffic data were extracted from continuous loop measurements at the time of the incident's occurrence (aggregate field observations).

In the second research activity (Chapter 3), we examined the effects of various traffic parameters on incident type. Multivariate Probit models were specified on 4-years of data from the A4-A86 highway section in the Ile-de-France region, France. Empirical findings indicated that incident type could almost exclusively be defined by the prevailing traffic conditions shortly before its occurrence. Rear-end crashes, for example, involving two vehicles were found to be more probable for relatively low values of both speed and density, rear-end crashes involving more than two vehicles appeared to be more probable under congested conditions, while single-vehicle crashes appeared to be largely geometry-dependent.

In the third research activity (Chapter 4), we extended research on the factors influencing the level of incident severity by including traffic data from the moment of the accident. A random parameters ordered probit model was applied to explore the influence of speed and traffic volume on the injury level sustained by vehicle occupants involved in accidents on the A4-A86 junction in the Paris region. Results indicated that increased traffic volume had a consistently positive effect on severity, while speed had a differential effect on severity depending on flow conditions.

In the fourth research activity (Chapter 5), we investigated the introduction of incident analysis outcomes in an integrated incident management scheme. To this end, a synthesis of related incident management analyses was performed. Further, crash data studies using traffic data collected on a real-time basis at the time of the incident occurrence were analyzed. The synthesis led to establishing a conceptual framework for incident management applications using real-time traffic data on urban freeways. We used findings from the previous research activities to explore potential implications towards incident propensity detection and enhanced management.

6.3 Conclusions and thesis contribution

The main research question of the thesis was to explore the effect of various traffic parameters on accident type frequency and accident severity. All research activities indicated a strong and critical impact of prevailing traffic conditions upon accident occurrences. Traffic speed and volume were found to almost exclusively define crash type and to significantly affect the injury severity level sustained by vehicle occupants involved in accidents. This overall conclusion suggests that similar accident investigations should consider the actual traffic conditions at the moment of the accident occurrences.

The thesis offered an important potential gain to society as additional light was shed on the accident mechanism of occurrence. Road users may become able to recognize hazardous situations (i.e. traffic conflicts), to know the best way of addressing them, and to react appropriately towards mitigating the hazard. Also, road authorities may implement effective real-time measures and, thus, reduce accident probabilities. Emergency response authorities could benefit from results in order to provide a quicker and more effective treatment to injured persons and, consequently, reduce fatalities and heavy injuries due to crashes. All these potential benefits may significantly reduce traffic accident severity and mitigate related fatalities that figure among the leading causes of death, especially in developing countries and among young people.

The thesis contribution to research state of the art and practice is manifold as real-time traffic data are readily available in most freeways and show significant potential for research and applications. However, related work has been limited in the past years. In freeway incident research, most studies use aggregation of exposure data neglecting their natural variance which may result in heavy underdispersion. We used traffic data averaged over a 6-minute interval and collected real-time at the moment of the incident’s occurrence. From a methodological standpoint, such disaggregation minimizes possible bias and provides better estimates. Moreover, the exploration of the influence of real-time traffic variables on incident outcomes (in terms of both type and severity) provided significant insight in the incident’s mechanism of occurrence.

Based on the analysis of historical data performed, typical traffic patterns recorded prior to accidents may then act as real-time identifiers (Abdel-Aty and Pande, 2007). Such research is useful for researchers and practitioners in estimating accident and congestion external costs and in transportation planning. Further, it may enable practitioners and authorities to locate hazardous spots on the road networks by utilizing real-time data widely available. Once a location is identified as being susceptible to a given crash type occurrence, it may be flagged with warnings through variable message signs (VMS). Furthermore, the concept of variable speed limits could be used to intervene on driver behavior and to reduce speed variation. The presence of traffic police on the designated locations could also serve as a crash prevention measure.

In addition to real-time monitoring of safety levels, a safety performance tool could be developed and used in project evaluation and planning. Safety aspects of costs and benefits can be assessed by comparing the levels of safety before and after implementation of a treatment (Golob and Recker, 2004). Finally, a procedure that uses real-time data on traffic flow, speed, and occupancy and the relationship between these variables and crash-type occurrence could be used to develop congestion mitigation strategies that incorporate safety (Garber and Subramanyan, 2001).

In conclusion, the attempt to further study and develop accident models, and in particular the integration of real-time data, can significantly contribute to the elaboration of a better-structured incident response system with predictive power. Thus, accident counts would be decreased and their consequences would be further limited. Apart from human lives saved, an economic burden would be taken off from societies; non-recurrent congestion would be decreased, while environmental gains would occur.

6.4 Future research

We recognise that the analysis performed suffers from limitations needing further investigation; first, loop detectors aggregate counts and occupancies over 6-min intervals. Second, the set of influencing factors used as regressors was not exhaustive. Possible uncertainty in the exact time of the accident occurrence should also be examined. Further, we did not distinguish in our analysis among freeway lanes, and only separated traffic regimes in two (peak and off-peak). Golob and Recker (2004), after performing a similar analysis, provided important evidence that it is significant (i) to capture variations in speed and flows separately across lanes and (ii) to more strictly define traffic regimes.

Considering the above, suggestions for further research would include the influence of real-time traffic data aggregation level. On the freeway treated, traffic data are collected every 6 minutes, while on other infrastructures, the respective period is much shorter. For example, on most U.S. freeways the measurements are averaged and collected every 30 seconds. In addition, more causal factors should be included in future analyses in order to acquire a better understanding of accident mechanism of occurrence. Also, considering more traffic regimes may provide additional insight in the impact of traffic conditions on safety analyses.

Furthermore, we have assumed linearity of the utility functions in all model specifications. We note however that the possibility of a non-linear utility function cannot be rejected; recent research in micro-economics (Orro et al., 2010) provides evidence that linearity may not be always the case. To the best of our knowledge, in road safety, such an assumption has not been tested yet.

Also, we used a dummy variable to test for the homogeneity of data over the analysis period (2000-2002 and 2006) and found them to be homogeneous even though operational changes had indeed taken place (e.g. speed camera enforcement). The robustness of results was not tested for the case of other changes (such as extension of the calibration period); such stability evaluations could be made to further validate results. In addition, the site studied (A4-A86 junction) has very specific characteristics. As a result, any extrapolation and transferability assumption regarding the results obtained should be first tested.

Finally, real-traffic time data availability enables for primary preventive incident management applications; i.e. applications that take place before the incident occurrences. Incorporating real-time traffic data in incident management enables for new perspectives in road safety techniques and seems beneficial in many ways; however it has not been adequately considered yet. We should note that all such techniques and measures should be first tested for their efficiency in terms of road safety enhancement.

References

Abdel-Aty, M., 2003. Analysis of driver injury severity levels at multiple locations using ordered probit models. *Journal of Safety Research 34*, 597-603.

Abdel-Aty, M., Keller, J., Brany, P., 2005. Analysis of types of crashes at signalized intersections by using complete crash data and tree-based regression. *Transportation Research Record 1908*, 37-45.

Abdel-Aty, M., Pande, A. 2007. Crash data analysis: collective vs. individual crash level approach. *Journal of Safety Research 38*, 581-587.

Abdel-Aty, M., Pande, A., 2005. Identifying crash propensity using specific speed conditions. *Journal of Safety Research 36*, 97-108.

Abdel-Aty, M., Pande, A., Lee, C., Gayah, V., Cunningham, R., Dhindsa, A., Dilmore, J., 2007. Linking Crash Patterns to ITS-Related Archived Data: Phase II, Volume I: Real-time Crash Risk Assessment Models. Florida Department of Transportation, BD-550-5.

Abdel-Aty, M., Radwan, A., 2000. Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention 32*, 633-642.

Abdel-Aty, M.A., Chen, C.L., Schott, J.R., 1998. An assessment of the effect of driver age on traffic accident involvement using log-linear models. *Accident Analysis & Prevention 30*, 851-861.

Abdelwahab, H. T., Abdel-Aty, M., 2002. Artificial Neural Networks and logit models for traffic safety analysis of toll plazas. *Transportation Research Record 1784*, 115-125.

Adams, J., 1985. Risk and freedom – the record of road safety regulation. Transport Publishing Projects. Cardiff, UK.

Al-Ghamdi, A.S., 2002. Using logistic regression to estimate the influence of accident factors on accident severity. *Accident Analysis & Prevention 34*, 729-741.

Al-Haji, G., 2005. Towards a Road Safety Development Index (RSDI)- Development of an International Index to Measure Road Safety Performance. Licentiate Thesis No. 1174, ISBN 91-85299-70-7, Linköping University, Sweden.

Anastasopoulos, P.Ch., Mannering, F.L., 2009. A note on modeling vehicle accident frequencies with random-parameters count models. *Accident Analysis &*

Prevention 41, 153-159.

Andersson, R., Menckel, E., 1995. On the prevention of accidents and injuries: a comparative analysis of conceptual frameworks. *Accident Analysis & Prevention 17*, 757-768.

Andreassen, D., 1991. Population and registered vehicle data vs. road deaths. *Accident Analysis & Prevention 23*, 343-351.

Aron M., Seidowsky R., 2004. Diagnostic d'exploitation et de sécurité du tronc commun autoroutier A4-A86 ; Analyse statistique des accidents survenus entre 1996 et 2002. Rapport de convention SISER-INRETS.

Aron, M., Seidowsky R., Cohen S., 2009. The safety impact of hard shoulder use during congested traffic. The case of a managed lane operation on a French urban motorway. In: *Proceedings of the 13th European Working Group on Transportation Conference*, Padova, Italy.

Awad, W., Jason, B., 1998. Prediction models for truck accidents at freeway ramps in Washington State using regression and artificial intelligence techniques. *Transportation Research Record 1635*, 30-36.

Ballesteros, M.,F., Dischinger, P.C., Langenberg, P., 2004. Pedestrian injuries and vehicle type in Maryland, 1995-1999. *Accident Analysis & Prevention 36*, 73-81.

Berman, O., Krass, D., 2002. Facility Location problems with Stochastic Demand and Congestion. In: *Facility Location: Applications and Theory*, Eds: Z. Drezner, H.W. Hamacher, Springer-Verlag, NY.

Bester, C.J., 2001. Explaining national road fatalities. *Accident Analysis & Prevention 33*, 663-672.

Bhat, C., 2001. Quasi-Random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model. *Transportation Research Part B: Methodological 35*, 677-693.

Boes, S., Winkelmann, R., 2006. Ordered response models. *Allgemeines Statistisches Archiv 90*, 167-181.

Boufous, S., Finch, C., Hayen, A., Williamson, A., 2008. The impact of environmental, vehicle and driver characteristics on injury severity in older drivers hospitalized as a result of a traffic crash. *Journal of Safety Research 39*, 65-72.

Box, G., Jenkins, J., 1970. Time series analysis: Forecasting and control. Holden-Day, San Francisco.

Brenac, T., 1989. Risque d'accident et caractéristiques des déplacements et des conducteurs. Thèse de doctorat. Ecole Nationale des Ponts et Chaussées. Spécialité Transport, Paris.

Brotcorne, L., Laporte, G., Semet, F., 2003. Ambulance location and relocation models. *European Journal of Operational Research 147*, 451-463.

Broughton, J., 1988. Predictive models of road accident fatalities. *Traffic Engineering Control 29*, 296–300.

Brownstone, D., Train, K., 1999. Forecasting New Product Penetration with Flexible Substitution Patterns. *Journal of Econometrics 89*, 109-129.

Brühning, E., 1995. Five years of IRTAD: Review and Perspectives. International Road Traffic and Accident Databases. Seminar proceedings, Helsinki, Finland.

Bunn, M., and G. Savage, 2003. Integrated traffic management and emergency response: success factors. University Transportation Center for Alabama, UTCA Report Number 01101.

Calvo, A., Marks, H. 1973. Location of health care facilities: An analytical approach. *Socio-economic Planning Sciences 7*, 407-422.

Carroll, P.S., 1971. The meaning of driving exposure. HIT LAB Rep., p. 1–3.

Carson, J., Mannering, F., 2001. The effect of ice warning signs on ice-accident frequency and severity. *Accident Analysis &Prevention 33*, 99-109.

Carson, Y., Batta, R., 1990. Locating an ambulance on the Amherst Campus of the State University of New York at Buffalo, *Interfaces 20*, 43-49.

Catalano, R., Dooley, D., 1980. Economic change in primary prevention. In: Price, R.; Ketterer, R.; Bader, B.; Monahan, J. (Eds). *Prevention in mental health*. Sage publications: London; 21-40.

Ceder, A., 1982. Relationship between road accidents and hourly traffic flow–II. *Accident Analysis & Prevention 14*, 34-44.

Ceder, A., Livneh, M., 1982. Relationship between road accidents and hourly traffic flow–I. *Accident Analysis & Prevention 14*, 19-34.

Chang, L.-Y., 2005. Analysis of freeway accident frequencies: Negative binomial regression versus artificial neural network. *Safety science 43*, 541-557.

Chang, L.-Y., Chen, W.-C., 2005. Data mining of tree-based models to analyze freeway accident frequency. *Journal of Safety Research 36*, 365-375.

Chang, L.-Y., Mannering, F., 1999. Analysis of injury severity and vehicle occupancy in truck- and non-truck-involved accident. *Accident Analysis & Prevention 31*, 579–592.

Chang, L-Y., Wang, H-W., 2006. Analysis of traffic injury severity: An application of non-parametric classification tree techniques. *Accident Analysis & Prevention 38*, 1019-1027.

Chapman, R., 1973. The concept of exposure. *Accident Analysis & Prevention 5*, 95–110.

Chimba, D., Sando, T., 2009. The prediction of highway traffic accident injury severity with neuromorphic techniques. *Advances in Transportation 19*, 17-26.

Chirachavala, T., Cleveland, D., 1985. Causal analysis of accidents involvements for the nation's large trucks and combination vehicles. *Transportation Research Record 847*, 37-42.

Choo, S., Mokhtarian, P., 2008. How do people respond to congestion mitigation policies? A multivariate probit model of the individual consideration of three travel-related strategy bundles. *Transportation 35*, 145-163.

Church, R. L., ReVelle, C., 1974. The maximal covering location problem. *Papers of the Regional Science Association 32*, 101-118.

Clarke, R., Forsyth, Wright, R., 1998a. Machine learning in road accident research: decision trees describing road-accidents during cross-flow turns. *Ergonomics 41*, 1060-1079.

Clarke, R., Forsyth, Wright, R., 1998b. Behavioral factors in accidents at road junctions: the use of a genetic algorithm to extract descriptive rules from police case files. *Accident Analysis & Prevention 30*, 223-234.

Clarke, R., Forsyth, Wright, R., 2005. A statistical profile of road accidents during cross-flow turns. *Accident Analysis & Prevention 37*, 721-730.

Conroy, C., Tominaga, G.T., Erwin, S., Pacyna, S., Velky, T., Kennedy,F., Sise, M., Coimbra, R., 2008. The influence of vehicle damage on injury severity of drivers in head-on motor vehicle crashes. *Accident Analysis & Prevention 40*, 1589-1594.

Corby, M.J., Saccomanno, F.F., 1997. Analysis of freeway accident detection. *Transportation Research Record 1603*, 80-89.

Current, J., M. Daskin, Schilling, D., 2002. Discrete network location models. In: *Facility location: applications and theory*, Eds: Z. Drezner, H.W. Hamacher, Springer-Verlag, NY.

Davis, G., 2004. Possible aggregation biases in road safety research and a mechanism approach to accident modeling. *Accident Analysis & Prevention 36*, 1119-1127.

Davis, G.A., 2002. Is the claim that 'variance kills' an ecological fallacy? *Accident Analysis & Prevention 34*, 343-346.

Delen, D., Sharda, R., Bessonov, M., 2006. Identifying significant predictors of injury severity in traffic accidents using a series of artificial neural networks. *Accident Analysis & Prevention 38*, 434-444.

Donnell, E.T., Mason, J.M., 2006. Predicting the frequency of median barrier crashes on Pennsylvania interstate highways. *Accident Analysis & Prevention 38*, 590-599.

Drakopoulos, A., Shrestha, M.R., Ertan Örnek, E., 2001. Freeway Crash Timeline Characteristics and Uses. *Transportation Research Record 1748*, 132-148.

Drezner, Z., Hamacher, H.W., 2002. Facility Location: Applications and Theory, Springer-Verlag, NY.

Dupont, E., Martensen, H., Papadimitriou, E., Yannis, G., 2010. Risk and protection factors in fatal accidents. *Accident Analysis & Prevention 42*, 645-653.

Eluru, N., Bhat, C.R., Hensher, D.A., 2008. A mixed generalized response model for examining pedestrian and bicyclist injury severity level in traffic crashes. *Accident Analysis and Prevention 40*, 1033-1054.

ETSC, 1997. *Transport Accident Costs and the Value of Safety*. European Transport Safety Council, Brussels, Belgium.

Everitt, B.S., 1988. A finite mixture model for the clustering of mixed-mode data. *Statistics and Probability Letters 6*, 305-309.

Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., 1996. From data mining to knowledge discovery: an overview. In: Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., Uthurusamy, R., Advances in Knowledge Discovery and Data Mining. AAAI Press/The MIT Press, 1-34.

FHWA, 2000. Traffic Incident Management Handbook. Federal Highway Administration, Office of Travel Management. Available at: http://www.iacptechnology.org/IncidentManagement/TrafficIncidentMgmtHand book.pdf

Fish, G.E., Blodgett, J.G., 2003. A visual method for determining variable importance in an artificial neural network model: an empirical benchmark study. Journal of Targeting, *Measurement and Analysis for Marketing 11*, 244-254.

Foldvary, L., 1979. Road accident involvement per miles traveled. *Accident Analysis & Prevention 11*, 79-99.

Frantzeskakis, J., Iordanis, D., 1987. Volume to capacity ratio and traffic accidents on interurban four-lane highways in Greece. *Transportation Research Record 112*, 29-38.

Freedman, D., 1999. From association to causation: some remarks on the history of statistics. *Statistical Science 14*, 243-258.

Fridstrom, L., Ingebrigtsen, S., 1989. An aggregate accident model based on combined close-sectorial and time-series data. Working paper E-1523, Institute of Transport Economics. Oslo, Norway.

Friedman, J.H., 1997. Data mining and statistics: what's the connection? In: Proceedings of the 29th Symposium on the Interface Between Computer Sciences and Statistics.

Galvão, R.D., Chiyoshi, F.Y., Morabito, R., 2005. Towards unified formulations and extensions of two classical probabilistic location models. *Computers and Operations Research 32*, 15-33.

Garber, N, Subramanyan, S., 2001. Incorporating crash risk in selecting congestion-mitigation strategies: Hampton Roads Area (Virginia) case study. *Transportation Research Record 1746*, 1-5.

Garber, N., Ehrhart, A., 2000. Effect of speed, flow, and geometric characteristics on crash frequency for two-lane highways. *Transportation Research Record 1717*, 76-83.

Garber, N., Wu, L., 2001. Stochastic models relating crash probabilities with geometric and corresponding traffic characteristics data. Research Report No: UVACTS-15-5-54, Virginia Transportation Research Council, University Station, Charlottesville, Virginia.

Gaudry, M., Lassarre, S., 2000. Structural road accident models – The international DRAG family, Pergamon.

Gendreau, M., Laporte, G., Semet, F., 2001. A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. *Parallel Computing 27*, 1641-1653.

Gendreau, M., Laporte, G., Semet, F., 1997. Solving an ambulance location model by Tabu search. *Location Science 5*, 75-88.

Geurts, K., Thomas, I., Wets, G., 2005. Understanding spatial concentrations of road accidents using frequent item sets. *Accident Analysis & Prevention 37*, 787-799.

Gjestland, T., 1955. The Oslo study of untreated syphilis. *Acta Dermato-Venereologica 34*.

Goldberg, J.B., 2004. Operations research models for the deployment of emergency services vehicles. *EMS Management Journal 1*, 20–39.

Golob, T., Recker, W., 2001. Relationship among urban freeway accidents, traffic flow, weather and lighting conditions. Institute of Transportation Studies, University of California, Berkeley, Report No. UCB-ITS-PWP-2001-19.

Golob, T., Recker, W., 2003. Relationships Among Urban Freeway Accidents, Traffic Flow, Weather, and Lighting Conditions. *Journal of Transportation Engineering 129 4*, 342-353.

Golob, T., Recker, W., 2004. A method for relating type of crash to traffic flow characteristics on urban freeways. *Transportation Research Part A 38*, 53-80.

Golob, T., Recker, W., Alvarez V., 2004. Freeway Safety as a function of traffic flow. *Accident Analysis & Prevention 36*, 933-946.

Golob, T., Recker, W., Pavlis, Y., 2008. Probabilistic models of freeway safety performance using traffic flow data as predictors. *Safety Science 46*, 1306-1333.

Goulias, K., Brog, W., Erl, E., 1998. Perceptions on Mode Choice Using Situational Approach. *Transportation Research Record 1645*, 82-93.

Gray, R.C., Quddus, M.A., Evans, A., 2008. Injury severity analysis of accidents involving young male drivers in Great Britain. *Journal of Safety Research 39*, 483-495.

Greene, W., 2003. Econometric Analysis, 5th Edition, Prentice-Hall Press, New York.

Gwynn, D., 1967. Relationship of accident rates and accident involvements with hourly volume. *Traffic Quarterly 21(3)*, 407-418.

Hacking, I., 1964. On the Foundations of Statistics. *British Journal for the Philosophy of Science 15 (57)*, 1-26.

Hadi, M., Aruldhas, J., Chow, L.F., Wattleworth, J., 1995. Estimating safety effects of cross-section design for various highway types using negative binomial regression. *Transportation Research Part C 3*, 247-260.

Haight, F., 1971. Indirect methods for measuring exposure factors as related to the incidence of motor vehicle traffic accidents prepared for National Highway Traffic Safety Administration. Pennsylvania Transportation and Traffic Safety Center, University Park.

Hakim, S., Shefer, D., Hakkert, A.S., Hockerman, I., 1991. A critical review of macro models for road accidents. *Accident Analysis & Prevention 23,* 379-400.

Hakimi S.L., 1964. Optimum locations of switching centers and the absolute centers and medians of a graph. *Operations Research 12,* 450-459.

Hakkert A., Hocherman, I., Mensah, A., 1996. Levels of safety on Interurban Roads. *Transportation Research Record 1553*, 95-102.

Hall, J., Pendleton, O., 1989. Relationship between v/c ratio and accident rates. Report FHWA-HPR-NM-88-02. FHWA reports, Washington, DC.

Hall, R., Mukherjee, A., 2008. Bounds on effectiveness of driver hours-of-service regulations for freight motor carriers. *Transportation Research Part E 44*, 298-312.

Harrell, F., 2001. Regression modeling strategies. New York: Springer.

Hauer, E., 1982. Traffic conflicts and exposure. *Accident Analysis & Prevention 14*, 359-364.

Hauer, E., 1997.Observational before-after studies in road safety. Pergamon.

Heckman, J., Leamer, E., 2007. *Handbook of Econometrics*, 6B, North-Holland.

Helai, H., Chor, C.H., Haque, Md.M., 2008. Severity of driver injury and vehicle damage in traffic crashes at intersections: a Bayesian hierarchical analysis. *Accident Analysis & Prevention 40*, 45-54.

Hiselius, L.W., 2004. Estimating the relationship between accident frequency and homogeneous and inhomogeneous traffic flows. *Accident Analysis & Prevention 36,* 985-992.

Hogan, K., ReVelle, C.S., 1986. Concepts and applications of backup coverage. *Management Science 34,* 1434-1444.

Hosking, J., Pednault, E., Sudan, M., 1997. A statistical perspective on data mining. *Future Gen. Comput. Syst. 13*, 117-134.

Hsiao, C., 1993. Analysis of Panel Data. Cambridge University Press, Cambridge, UK.

Ivan, J., O'Mara, P., 1997. Prediction of traffic accident rates using Poisson regression. In: *Proceedings of the 77th Transportation Research Board Meeting*, Washington, DC.

Ivan, J., Pasupathy, R., Ossenbruggen, P., 1999. Differences in causality factors for single- and multi-vehicle crashes on two-lane roads. *Accident Analysis & Prevention 31*, 695-704.

Ivan, J., Wang, C., Bernardo, N., 2000. Explaining two-lane highway crash rates using land use and hourly exposure. *Accident Analysis & Prevention 32*, 787-795.

Ivey, D., Griffin, L., Newton, T., Lytten, R., Hankins, K., 1981. Predicting wet weather accidents. *Accident Analysis &Prevention 13*, 83-89.

Jacobs G.D., Hutchinson, P., 1973. A study of accident rates in developing countries, Transport and Road Research Laboratory, Report LR546, Crowthorne, Berkshire, UK.

Jacobs, G. D., Aeron-Thomas, A. and A. Astrop (2000). Estimating global road fatalities. TRL Report 445. Transport Research Laboratory, Crowthorne.

Jia, H., Ordonez, F., Dessouky, M., 2007. A modeling framework for facility location of medical services for large-scale emergencies. *IIE Transactions 39*, 41-55.

Jiang, X., Lyles, R., 2007. Difficulties with quasi-induced exposure when speed varies systematically by vehicle type. *Accident Analysis & Prevention 39*, 649-656.

Jilani, T.A., Burney, S.M.A., 2007. Multivariate stochastic fuzzy forecasting models. *Expert Systems with Applications 37*, 691-700.

Jones, I., Whitfield, R., 1988. Predicting injury risk with 'New car assessment program' crashworthiness ratings. *Accident Analysis & Prevention 20*, 411-419.

Jovanis, P., Chang, H.-L., 1986. Modeling the relationship of accidents to miles traveled. *Transportation Research Record 1068*, 42-51.

Jovanis, P., Chang, H.-L., 1989. Disaggregate model of highway accident occurrence using survival theory. *Accident Analysis & Prevention 21*, 445-458.

Jovanis, P., Delleur, J., 1981. Exposure-based analysis for motor vehicle accidents. *Accident Analysis & Prevention 13*, 83-99.

Karlaftis, M., Golias, I., 2002. Effects of road geometry and traffic volumes on rural roadway accident rates. *Accident Analysis & Prevention 34*, 357-365.

Karlaftis, M., Tarko, A., 1998. Heterogeneity considerations in accident modeling. *Accident Analysis & Prevention 30*, 425-433.

Kavsek, B., Lavrac, N., Bullas, J.C., 2002. Rule induction for subgroup discovery: a case study in mining UK traffic accident data. In: *Proceedings of Conference on Data Mining and Warehouses*, Ljubljana, Slovenia, October 15.

Khattak, A., Kantor, P., Council, F., 1998. Role of adverse weather in key crash types on limited-access roadways. *Transportation Research Record 1621*, 10-19.

Khorashadi, A., Niemeier, D., Shankar, V., Mannering, F., 2005. Differences in rural and urban driver-injury severities in accidents involving large trucks: an exploratory analysis. *Accident Analysis & Prevention 37*, 910-921.

Kim, D-G, Washington, S., Oh, J., 2006. Modeling crash types: new insights into the effects of covariates on crashes at rural intersections. *Journal of Transportation Engineering 132*, 282-292.

Kim, J.-K., Ulfarsson, G.F., Shankar, V.N., Kim, S., 2008. Age and pedestrian injury severity in motor-vehicle crashes: A heteroskedastic logit analysis. *Accident Analysis & Prevention 40*, 1695-1702.

Knuiman, M., Council, F., Reinfurt, D., 1993. Association of median width and highway accident rates. Transportation Research Record 1401, 70-82.

Kockelman, K., Ma, J., 2007. Freeway speeds and speed variations proceeding crashes, within and across lanes. *Journal of the Transportation Research Forum 46*, 43-61.

Kockelman, K.M., Kweon, Y.-J., 2002. Driver Injury Severity: an application of ordered probit models. *Accident Analysis & Prevention 34*, 313-321.

Koornstra, M.J., Oppe, S., 1992. Predictions of road safety in industrialized countries and Eastern Europe, VTI Report 380A, PT1, Linkoeping, 47–70.

Kumara, S.P., Chin, H.S., 2003. Modeling accident occurrence at signalized tee intersections with special emphasis on excess zeros. *Traffic Injury Prevention 3*, 53-57.

Lapparent, M., 2008. Willingness to use safety belt and levels of injury in car accidents. *Accident Analysis & Prevention 40*, 1023-1032.

Larson, R.C., 1974. A hypercube queuing model for facility location and redistricting in urban emergency services. *Computers and Operations Research 1*, 67-95.

Lassarre, S., 1986. The introduction of the variables ‘traffic volume’, ‘speed’, and 'belt-wearing' into a predictive model of the severity of accidents. *Accident Analysis & Prevention 18*, 129-134.

Lave, C., 1985. Speed, coordination, and 55mph limit. The American Economy Review 1159-1164.

Lee, C., Abdel-Aty, M., 2005. Comprehensive analysis of vehicle-pedestrian crashes at intersections in Florida. *Accident Analysis & Prevention 37*, 775-786.

Lee, C., Abdel-Aty, M., Hsia, L., 2006a. Potential real-time indicators of sideswipe crashes on freeways. In: *Proceedings of the 85th Annual Meeting of the Transportation Research Board*, Washington DC.

Lee, C., Hellinga, B., Ozbay, K., 2006c. Quantifying the effects of ramp metering on roadway safety. *Accident Analysis & Prevention 38*, 279-288.

Lee, C., Hellinga, B., Saccomanno, F., 2003. Real-time crash prediction model for the application to crash prevention in freeway traffic. In: *Proceedings of the 82nd Annual Meeting of the Transportation Research Board*, Washington, DC.

Lee, C., Hellinga, B., Saccomanno, F., 2006b. Evaluation of variable speed limits to improve traffic safety. *Transportation Research Part C 14*, 213-228.

Lee, C., Saccomanno, F., Hellinga, B., 2002. Analysis of crash Precursors on Instrumented Freeways. In: *Proceedings of the 81st Annual Meeting of the Transportation Research Board*, Washington, DC.

Lee, J., Mannering, F., 2002. Impact on roadside features on the frequency and severity of run-off-roadway accidents: an empirical analysis. *Accident Analysis & Prevention 34*, 149-161.

Lee, J-T., Fazio, J., 2005. Influential factors in freeway response and clearance times by emergency management services in peak periods. *Traffic Injury Prevention 6*, 331-339.

Lerman, R., Manski C., 1981. On the use of Simulated Frequencies to Approximate Choice Probabilities. Structural Analysis of Discrete Data with Econometric Applications, C. Manski and D. Mc Fadden eds, MIT press, Cambridge, Mass.

Leutzbach, W., 1966. Traffic accidents and traffic flow on the Frankfurt – Mannheim motorway – investigation of the relationship between traffic accidents and traffic flow. Hessisches Ministerium fur Wirtschaft und Verkehr.

Lindley, J., 1987. Urban Freeway Congestion: Quantification of the Problem and Effectiveness of Potential Solutions, ITE JOURNAL.

Liu, B.-S., 2007. Association of intersection approach speed with driver characteristics, vehicle type and traffic conditions comparing urban and suburban areas. *Accident Analysis & Prevention 39*, 216-223.

Ljubic, P., Todorovski, L., Lavrac, N. and J.C. Bullas, 2002. Time series analysis of UK traffic accident data. In: *Proceedings of Conference on Data Mining and Warehouses*, Ljubljana, Slovenia.

Lord, D., 2006. Modeling motor vehicle crashes using Poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. *Accident Analysis & Prevention 38*, 751-766.

Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: a review and assessment of methodological alternatives. *Transportation Research Part A 44*, 291-305.

Lord, D., Washington, S.P., Ivan, J.N., 2005. Poisson, Poisson-Gamma and zero-inflated regression models for motor vehicle crashes: balancing statistical fit and theory. *Accident Analysis & Prevention 37*, 35-46.

Lui, K., McGee, D., Rhodes, P., Pollock, D., 1988. An application of a conditional logistic regression to study the effects of safety belts, principal impact points, and car weights on drivers' fatalities. *Journal of Safety Research 19*, 197-203.

Lundy, R., 1965. Effect of traffic volumes and number of lanes on freeway accident rates. Highway Research Board Number 99, Highway Research Board of the National Academy of Sciences, National Research Council, Washington, DC.

Madanat, S., Teng, H-L., Liu, P-C., 1995. Sequential hypothesis-testing based freeway incident response decision-making system. In: *Proceedings of the 4th International Conference on Applications of Advanced Technologies in Transportation Engineering*, Capri, Italy.

Madanat, S., Liu, P-C., 1995. A prototype system for real-time incident likelihood prediction. IDEA program final report, Transportation Research Board, US Department of Transportation.

Maddala, G., 1983. Limited-dependent and qualitative variables in Econometrics. *Cambridge University Press*, Cambridge.

Maggazù, D., Cornelli, M., Marinoni, A., 2006. Are car drivers holding a motorcycle license less responsible for motorcycle-car crash occurrence? A non-parametric approach. *Accident Analysis & Prevention 38*, 365-370.

Maher M.J., Summersgill I., 1996. A comprehensive methodology for the fitting of predictive accident models. *Accident Analysis & Prevention 28*, 600-610.

Maher, M., 1991. A new bivariate negative binomial model for accident frequencies. *Traffic Engineering and Control 32,* 422-425.

Mannila, H., 2000. Theoretical frameworks for data mining. *SIGKDD Explorations 1*, 30-32.

Marcus, A. and M. Nichols, 1996. Acquiring and Utilizing Knowledge in Response to Unusual Events in a Hazardous Industry. In: *Proceedings of the Annual Meeting of the Academy of Management*, Cincinnati, USA, p.5.

Marianov, V., ReVelle, C.S., 1992. The capacitated standard response fire protection sitting problem: deterministic and probabilistic models. *Annals of Operations Research 40,* 303-322.

Martin, J.-L., 2002. Relationship between crash rate and hourly traffic flow on interurban motorways. *Accident Analysis and Prevention 34*, 619–629.

McCarthy, P.S., 1999. Public policy and highway safety: a city-wide perspective. *Regional science and urban economics 29,* 231-244.

McCartt, A.T., Northrup, V.S., & Retting, R. (2004). Types and characteristics of ramp-related motor vehicle crashes on urban interstate roadways in Northern Virginia. *Journal of Safety Research 35*, 107-114.

McFadden, D., 1989. A method of simulated moments for estimation of discrete response models with numerical integration. *Econometrica 57*, 995-1026.

McKelvey, R.D., Zavoina, W., 1975. A statistical model for the analysis of ordinal level dependent variables. *J.Math. Sociol. 4*, 103-120.

Mekky, A., 1985. Effects of rapid increase in motorization levels on road fatality rates in rich developing countries, *Accident Analysis & Prevention 17,* 101–109.

Mensah, A., Hauer, E., 1998. Two problems of averaging arising from the estimation of the relationship between accidents and traffic flow. *Transportation Research Record 1635,* 37-43.

Miaou S.-P., 1994. The relationship between truck accidents and geometric design for road sections: Poisson versus Negative Binomial regressions. *Accident Analysis & Prevention 26*, 471-482.

Miaou S.-P., Lum H., 1993. Modeling vehicle accidents and highway geometric design relationships. *Accident Analysis & Prevention 25*, 689-709.

Miaou, S.-P., Hu, P., Wright, T., Rathi, A., Davis, S., 1992. Relationship between accidents and highway geometric design: a Poisson regression approach. *Transportation Research Record 1376*, 10-18.

Milton, J., Mannering, F., 1998. The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation 25*, 395-413.

Milton, J.C., Shankar, V.N., Mannering, F.L., 2008. Highway accident severities and the mixed logit model: an exploratory empirical analysis. *Accident Analysis & Prevention 40*, 260-268.

Mohamedshah, Y., Paniati, J., Hobeika, A., 1993. Truck accident models for interstate and two-lane rural roads. *Transportation Research Record 1407*, 35-41.

Mountain, L., Fawaz, B., Jarrett, D., 1996. Accident prediction models for roads with minor junctions. *Accident Analysis & Prevention 28*, 695-707.

Mountain, L., Maher, M., Fawaz, B., 1998. Improved estimates of the safety effects of accident remedial schemes. *Traffic engineering and control 39*, 554-558.

Mouskos, K., Sun, W., Qu, T., 1999. Impact of Access Driveways on Accident Rates at Multilane Highways. New Jersey Department of Transportation NJ-99-008-NCTIP5, US Department of Transportation.

Mussone, L., Ferrari, A., Oneta, M., 1999. An analysis of urban collisions using an artificial intelligence model. *Accident Analysis & Prevention 31*, 705-718.

Mussone, L., Rinelli, S., Reitani, G., 1996. Estimating the accident probability of a vehicular flow by means of an artificial neural network. *Environment and Planning B23*, 667-675.

Naveh, E., Marcus, A., 2002. A Re-Assessment of Road Accident Data-Analysis Policy, Center for Transportation Studies, University of Minnesota.

Naveh, E., Marcus, A., 2007. Financial performance, ISO 9000 standard and safe driving practices effects on accident rate in the U.S. motor carrier industry. *Accident Analysis & Prevention 39*, 731-742.

Navon, D., 2003. The paradox of driving speed: two adverse effects on highway accident rate. *Accident Analysis & Prevention 35*, 361-367.

O'Donnell, C.J., Connor, D.H., 1996. Predicting the severity using models of ordered multiple choice. *Accident Analysis & Prevention 28*, 739-753.

OECD, 1988. *Road accidents: on-site inve*stigations. Bib ID: 377268. Organization for Economic Co-operation and Development, Paris.

Oh, C., Chang, M., 1999. Relationship between v/c and accident rate for freeway facility section. *Journal of Korean Society of Transportation 17*, 21-27.

Oh, C., Oh, J., Ritchie, S., Chang, M., 2001. Real time estimation of freeway accident likelihood. In: *Proceedings of the 80th Annual Meeting of Transportation Research Board*, Washington, DC.

Oh, C., Oh, J.-S., Ritchie, S., Chang, M., 2000. Real-time estimation of Freeway Accident Likelihood. Report Number: UCI-ITS-WP-00-21, Institute of Transportation Studies, University of California, Irvine, USA.

Oh, J., Washington, S.P., Choi, K., 2004. Development of accident prediction models for rural highway intersections. *Transportation Research Record 1897*, 18-27.

Oh, J., Washington, S.P., Nam, D., 2006. Accident prediction models for railway-highway interfaces. *Accident Analysis & Prevention 38*, 346-356.

Okamoto H., Koshi M., 1989. A method to cope with the random errors of observed accident rates in regression analysis. *Accident Analysis & Prevention 21*, 317-332.

Oppe, S., 1989. Macroscopic models for traffic and traffic safety. *Accident Analysis & Prevention 21*, 225-232.

Orro, A., Novales, M., Benitez, F., 2010. Box-Cox mixed Logit model for travel behaviour analysis. In: *Proceedings of the International Conference of Numerical Analysis and Applied Mathem*atics, *forthcoming.*

Page, Y., 2001. A statistical model to compare road mortality in OECD countries. *Accident Analysis & Prevention 33*, 371-385.

Pai, C.-W., 2009. Motorcyclist injury severity in angle crashes at T-junctions: Identifying significant factors and analyzing what made motorists fail to yield to motorcycles. *Safety Science 47*, 1097-1106.

Pai, C.-W., Saleh, W., 2008. Modelling motorcyclist injury severity by various crash types at T-junctions in the UK. *Safety Science 46*, 1234-1247.

Paluzzi, M., 2004. Testing a heuristic P-median location allocation model for sitting emergency service facilities". Presented at the Annual meeting of the Association of American Geographers, Philadelphia, PA.

Pande, A., Abdel-Aty, M., 2006. Assessment of freeway traffic parameters leading to lane-change related collisions. *Accident Analysis & Prevention 38*, 936-948.

Park, Y.J., Saccomanno, F., 2005. Collision frequency analysis using tree-based stratification. *Transportation Research Record 1908*, 121-129.

Pasupathy, R., Ivan, J., Ossenmbruggen, P., 2000. Single and multi-vehicle crash prediction models for two-lane roadways. USDoT, Region I University Transportation Center, Project UCNR9-8, Final Report.

Perrow, C., 1999. *Normal Accidents.* Princeton University Press, 2nd ed., USA.

Persaud, B., Lord, D., Palminaso, J., 2002. Issues of calibration and transferability in developing accident prediction models for urban intersections. *Transportation Research Record 1784*, 57-64.

Persaud, B., Mucsi, K., 1995. Microscopic accident potential models for two-lane rural roads. *Transportation Research Record 1635*, 37-43.

Persaud, B., Nguyen T., 1998. Disaggregate safety performance models for signalized intersections on Ontario provincial roads. *Transportation Research Record 1635*, 113-120.

Petroski, H., 1994. Design Paradigms: Case Histories of Error and Judgment in Engineering. Cambridge University Press, UK, p. 122.

Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering 122*, 105-113.

Qin, X., Ivan, J., Ravishanker, N., 2004. Selecting exposure measures in crash rate prediction for two-lane highway segments. *Accident Analysis & Prevention 36*, 183-191.

Quddus, M.A., Noland, R.B., Chor Chin, H., 2002. An analysis of motorcycle injury and vehicle damage using ordered probit models. *Journal of Safety Research 33*, 445-462.

Quddus, M.A., Wang, C., Ison, S.G., 2009. The impact of road traffic congestion on crash severity using ordered response models. In *TRB 2009 annual meeting* CD-ROM.

Rajagopalan, H., Saydam, C., Xiao, J., 2008. A multiperiod set covering location model for dynamic redeployment of ambulances. *Computers and Operations Research 35*, 814-826.

Rasmussen, J., 1983. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on*

Systems, Man, and Cybernetics 13, 257-266.

Repede, J.F., Bernardo, J., 1994. Developing and validating a decision support system for location emergency medical vehicles in Louisville, Kentucky. *European Journal of Operational Research 75*, 567-581.

ReVelle, C., Snyder, S., 1995. Integrated fire and ambulance sitting: a deterministic model. *Socio-Economic Planning Sciences 29*, 261-271.

ReVelle, C., Bigman, D., Schilling, D., Cohon, J., Church, R. 1977. Facility location: a review of context-free and EMS models. *Health Services Research 12,* 129–146.

ReVelle, C., Eiselt, H., 2005. Location analysis: A synthesis and survey. *European Journal of Operational Research 165,* 1-19.

Richter, E.D., Barach, P., Friedman, L., Krikler, S., Israeli, A., 2004. Raised speed limits, speed rollover, case-fatality, and road deaths in Israel: a 5-year follow-up. *American Journal of Public Health 94,* 568-574.

Riviere, C., Lauret, P., Ramsamy, J.F.M., Page, Y., 2006. A Bayesian neural network approach to estimating the energy equivalent speed. *Accident Analysis & Prevention 38*, 248-259.

Rumar, K. 1999. European Transport Safety Lecture: Transport Safety Visions, Targets and Strategies beyond 2000. ETSC, Brussels.

Rune, E., Vaa, T., 2004. The Handbook of Road Safety Measures. Amsterdam : Elsevier, ISBN 0-08-044091-6.

Saccomanno, F., Buyco, C., 1988. Generalized loglinear models of truck accident rates. *Transportation Research Record 1172*, 23-31.

Saccomanno, F., Nassar, S., Shortreed, J., 1996. Reliability of statistical road accident injury severity models. *Transportation Research Record 1542*, 14-23.

Savolainen, P., Mannering, F., 2007. Probabilistic models of motorcyclists' injury severities in single- and multi-vehicles crashes. *Accident Analysis & Prevention 39*, 955-963.

Schilling, D.A., Elzinga, DK, Cohon, J., Church, R.L., Revelle, C.S., 1979. The TEAM/FLEET models for simultaneous facility and equipment sitting. *Transportation Science 13*, 163-175.

Schmid, V., and K.F. Doerner, 2010. Ambulance location and relocation problems with time-dependent travel-times. European Journal of Operational Research. *In* *press*.

Shankar, V., Mannering, F., 1996. An exploratory multinomial logit analysis of single-vehicle motorcycle accident severity. *Journal of Safety Research 27*, 183-194.

Shankar, V., Mannering, F., Barfield, W., 1995. Effect of roadway geometrics and environmental factors on rural freeway accident frequencies. *Accident Analysis & Prevention 27*, 371-389.

Shankar, V., Mannering, F., Barfield, W., 1996. Statistical analysis of accident severity on rural freeways. *Accident Analysis & Prevention 28*, 391-401.

Shankar, V., Milton, J., Mannering, F., 1997. Modeling accident frequencies as zero-altered probabilities processes: an empirical inquiry. *Accident Analysis & Prevention 29*, 829-837.

Shinar, 1998. Speed and crashes: a controversial topic and an elusive relationship. Special Report 254, TRB, Washington D.C., pp. 221-276.

Simpson, E.H., 1951. The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society (Series B) 13*, 238–241.

Sivak, M., 1996. Motor-vehicle safety in Europe and in the USA: a public-health perspective. *Journal of Safety Research 27*, 225-231.

Smeed, R. J., 1955. Accident Rates. *International Road Safety & Traffic Revue 3(2)*, 30-34.

Smeed, R.J., 1949. Some statistical aspects of road safety research. *Journal of the Royal Statistical Society (Series A)* 112 (1), 1–34.

Smelser, N. J., Wright, J., Baltes, P. B., 2001. *International Encyclopedia of Social & Behavioral Sciences*. Pergamon.

Snyder, L., 2006. Facility location under uncertainty: a review. *Institute of Industrial Engineers Transactions 38*, 537-554.

Song Q., Chissom B.S., 1993. Forecasting enrollments with fuzzy time series-part I. *Fuzzy Sets and System 54*, 1-9.

Songchitruksa, P., Tarko, A., 2006. The extreme value theory approach to safety estimation. *Accident Analysis & Prevention 38*, 811-822.

Srinivasan, K., 2002. Injury severity analysis with variable and correlated thresholds: ordered mixed logit formulation. *Transportation Research Record 1784*, 132-142.

Stephanedes, Y., Liu, X., 1995. Artificial neural networks for freeway incident detection. *Transportation Research Record 1494*, 91-97.

Stewart, J., 1996. Applications of classification and regression tree methods in roadway safety studies. *Transportation Research Record 1542,* 1-5.

Strnad, M., Jovic, F., Vorko, A., Kovasic, L., Toth, D., 1998. Young children injury analysis by the classification entropy method. *Accident Analysis & Prevention 30*, 689-695.

Šušteršič, G., Grabec, I., Prebil, I., 2007. Statistical model of a vehicle-to-barrier collision. *International Journal of Impact Engineering 34*, 1585-1593.

Sze, N.N., Wong, S.C., 2007. Diagnostic analysis of the logistic model for pedestrian injury severity in crashes. *Accident Analysis & Prevention 39*, 1267-1278.

Taylor, M., Lynam, B., Baruya, A., 2000. The effects of drivers' speed on the frequency of road accidents. Project Report 421. Transportation Research Laboratory, Crowthorne, UK.

Terza, J., 1985. Ordered probit: a generalization. *Communications in Statistics – Theory and Methods 14*,1-11.

The Scottish Office Central Research Unit, 1997. Impact of the Road Network on Scotland's Accident Rates. ISBN 0-7480-6450-8 MVA Consultancy.

Toregas, C., Swain, R., Revelle, C., Bergman, L. 1971. The location of emergency service facilities. *Operations Research 19*, 1363–1373

Train, K., 2000. Halton Sequences for Mixed Logit. UC Berkeley: Department of Economics, UCB. Retrieved from:
http://www.escholarship.org/uc/item/6zs694tp.

Tunaru, R., 1999. Hierarchical Bayesian models for road accidents data. *Traffic Engineering and Control 40*, 318-324.

Ulfarsson, G., Mannering, F., 2004. Differences in male and female injury severity in sport-utility vehicle, minivan, pickup and passenger car accidents. *Accident Analysis & Prevention 36*, 135-147.

Vaija P., 1987. Application of fuzzy methods to process safety control and to accident analysis. Acta Polytechnical Scandinavica, Chemical Technology and Metallurgy Series. No 180. Heksinki, 26pp.

Vanlaar, W., Yannis, G., 2006. Perception of road accident causes. *Accident Analysis & Prevention 38*, 155-161.

Viswanathan, K., Goulias, K., Jovanis, P., 2000. Use of traveler information in the Puget Sound Region. *Transportation Research Record 1719*, 94-102.

VNTSC, 1995. Intelligent Transportation System Impact Assessment Framework: Final Report. Volpe National Transportation Systems Center.

Vogt, A., Bared, J., 1998. Accident models for two-lane rural segments and intersections. *Transportation Research Record 1635*, 18-29.

Wang, W., Abdel-Aty, M., 2008. Analysis of left-turn crash injury severity by conflicting pattern using partial proportional odds model. *Accident Analysis & Prevention 40*, 1674-1682.

Washington, S.P., Karlaftis, M.G., Mannering, F.L., 2003. Statistical end Econometric Methods for Transportation Data Analysis. Chapman & Hall/CRC.

WHO, 2004. *World report on road traffic injury prevention*. World Health Organization, Geneva.

Wilde, G., 1982. The theory of risk-homeostatic: implications for safety and health. *Risk analysis 2*, 209-255.

Winston, C., Maheshri, V., Mannering, F., 2006. An exploration of the offset hypothesis using disaggregate data: The case of airbags and antilock brakes. *Journal of Risk and Uncertainty 32*, 89–99.

Wolfe, A.C., 1982. The concept of exposure to the risk of a road traffic accident and an overview of exposure data collection methods. *Accident Analysis & Prevention 14*, 337-340.

Wong, Y., Nicholson, A., 1992. Driver behavior at horizontal curves: risk compensation at the margin of safety. *Accident Analysis & Prevention 24*, 425-436.

Woods, D., Simms, C., 2002. Car size and injury risk: a model for injury risk in frontal collisions. *Accident Analysis and Prevention 34*, 93-99.

Xie, Y., Lord, D., Zhang, Y., 2007. Predicting motor vehicle collisions using Bayesian neural network models: an empirical analysis. *Accident Analysis & Prevention 39*, 922-933.

Xie, Y., Zhang, Y., Liang, F., 2009. Crash injury severity analysis using Bayesian ordered probit model. *Journal of Transportation Engineering 135*, 18-25.

Yamamoto, T., Shankar, V.N., 2004. Bivariate ordered response probit model of driver's and passenger's injury severities in collisions with fixed objects.

Accident Analysis & Prevention 36, 869-876.

Yan, X., Radwan, E., Mannila, K. K., 2009. Analysis of Truck-Involved Rear-End Crashes Using Multinomial Logistic Regression. *Advances in Transportation Studies 17*, 39-52.

Yannis, G., Antoniou, K., Papadimitriou, E., 2007. Road casualties and enforcement: distributional assumptions of serially correlated count data. *Traffic Injury Prevention 8*, 300-308.

Yannis, G., Papadimitriou, E., Dupont, E., Martensen, H., 2010. Estimation of fatality and injury risk by means of in-depth fatal accident investigation data. *Traffic Injury Prevention 11*, 492-502.

Yau, K., 2004. Risk factors affecting the severity of single-vehicle traffic accidents in Hong-Kong. *Accident Analysis & Prevention 36*, 333-340.

Zajac, S., Ivan, J.N., 2002. Factors influencing injury severity of motor vehicle-crossing pedestrian crashes in rural Connecticut. *Accident Analysis & Prevention 35*, 369-379.

Zeidler, F., Pletchen, B., Scheunert, D., Mattern, B., Alt, B., Miksch, T., Eichendorf, W. and Reiss, S., 1993. Development of a new injury cost scale. *Accident Analysis & Prevention 25*, 675–687.

Zeitouni, K., Chelghoum, N., 2001. Spatial decision tree – application to traffic risk analysis. In: *ACS/IEEE International Conference on Computer Systems and Applications*, Beirut, Lebanon, June 26-29.

Zhong, M., Lingras, P., Sharma, S., 2004. Estimation of missing traffic counts using factor, genetic, neural, and regression techniques. *Transportation Research Part C 12*, 139-166.

Zhou, M, Sisiopiku, V., 1997. Relationship between volume-to-capacity ratios and accident rates. *Transportation Research Record 1581*, 47-52.

Zografos, K., K. Androutsopoulos, Vasilakis, G., 2002. *Transportation Research Part C 10*, 1-18.

Annex

Περίληψη

Ως *συμβάν* χαρακτηρίζεται οποιοδήποτε εκδηλούμενο γεγονός, το οποίο προκαλεί διαταραχή ή απόκλιση από τις κανονικές συνθήκες λειτουργίας ενός συστήματος. Τα οδικά συμβάντα είναι γεγονότα τα οποία σημειώνονται με τυχαίο τρόπο τόσο στο χρόνο, όσο και στο χώρο και τα οποία προκαλούν μείωση της κυκλοφοριακής ικανότητας των οδών, ενώ συνδέονται με υψηλό οικονομικό και κοινωνικό κόστος. Εκτός από την απώλεια ανθρώπινων ζωών, τα οδικά συμβάντα επιφέρουν πολλαπλές αρνητικές επιπτώσεις όπως κυκλοφοριακή συμφόρηση, καθυστερήσεις στους λοιπούς χρήστες του δικτύου, δυσχέρεια στη μεταφορά εμπορευμάτων, υλικές ζημιές, περιβαλλοντική επιβάρυνση, κοινωνικό αντίκτυπο, απώλεια παραγωγικότητας, καταβολή νοσηλίων κ.λπ. Το υψηλό συνεπαγόμενο κόστος σε συνδυασμό με το έντονο ενδιαφέρον για βιώσιμη ανάπτυξη έχουν στρέψει επιστήμονες, διεθνείς οργανισμούς και την Πολιτεία στην υιοθέτηση προσεγγίσεων για τη μείωση των οδικών συμβάντων και τον περιορισμό των συνεπειών τους.

Η μείωση και διαχείριση των οδικών συμβάντων έχει προσελκύσει έντονο επιστημονικό ενδιαφέρον. Διάφορες μεθοδολογίες προτυποποίησης έχουν εφαρμοστεί στην προσπάθεια κατανόησης του μηχανισμού πρόκλησης οδικών συμβάντων. Η ομαδοποίηση δεδομένων (data aggregation) αποτελεί σημαντικό παράγοντα ανακρίβειας στις περισσότερες πρότερες διερευνήσεις. Η Ετήσια Μέση Ημερήσια Κυκλοφορία (ΕΜΗΚ) είναι η πλέον χρησιμοποιούμενη ένδειξη για την αποτύπωση των κυκλοφοριακών συνθηκών στα οδικά δίκτυα. Ωστόσο, οι περισσότεροι αυτοκινητόδρομοι είναι πλέον εξοπλισμένοι με συστήματα συνεχούς παρακολούθησης, τα οποία καθιστούν διαθέσιμα μη ομαδοποιημένα κυκλοφοριακά δεδομένα. Η διαθεσιμότητα των δεδομένων αυτών δεν έχει επαρκώς αξιοποιηθεί σε προηγούμενες έρευνες.

Το κεντρικό ερώτημα της διατριβής είναι η επίδραση των κυκλοφοριακών συνθηκών στην πρόκληση ατυχημάτων. Το αντικείμενο της διατριβής είναι η αξιοποίηση των κυκλοφοριακών δεδομένων αυτοκινητοδρόμων που συλλέγονται σε πραγματικό χρόνο, ώστε να διερευνηθούν: α) η επίδραση κυκλοφοριακών παραμέτρων στον τύπο οδικού συμβάντος, β) η επίδραση κυκλοφοριακών παραμέτρων στη σοβαρότητα οδικού συμβάντος και γ) η ενσωμάτωση των προηγούμενων επιδράσεων στη βέλτιστη αντιμετώπιση των συνεπειών του συμβάντος. Η επίτευξη του στόχου της διατριβής πραγματοποιείται μέσω των εξής τεσσάρων επιμέρους εργασιών: 1) βιβλιογραφική ανασκόπηση, 2) εμπειρική διερεύνηση προδιάθεσης εκδήλωσης συγκεκριμένου τύπου συμβάντος, 3) εμπειρική διερεύνηση της σοβαρότητας συμβάντος και 4) ανάπτυξη πλαισίου για την ένταξη των κυκλοφοριακών δεδομένων στη διαχείριση έκτακτων συμβάντων.

Κατά την πρώτη ενότητα εργασιών, η βιβλιογραφία οργανώνεται με βάση μεθοδολογικά και θεματικά κριτήρια. Τα κριτήρια ταξινόμησης περιλαμβάνουν: α) τη μέθοδο (π.χ. πείραμα, παρατηρήσεις πεδίου), β) το επίπεδο ανάλυσης (π.χ. ομαδοποίηση ή μη ομαδοποίηση δεδομένων), γ) το σκοπό της ανάλυσης (π.χ. πρόβλεψη, περιγραφή) και δ) τη φάση του συμβάντος (π.χ. πρόκληση, έκβαση). Στη συνέχεια, η περιοχή ενδιαφέροντος της διατριβής ορίζεται ως προς την πραγματοποιηθείσα ταξινόμηση. Ειδικότερα, πραγματοποιείται πιθανοτική προτυποποίηση μέσω επεξεργασίας μη ομαδοποιημένων δεδομένων. Οι εκβάσεις των συμβάντων –νοούμενες είτε ως τύπος είτε ως σοβαρότητα συμβάντος - αποτελούν τις εξηρτημένες μεταβλητές της ανάλυσης. Για το σκοπό αυτό, χρησιμοποιούνται κυκλοφοριακά δεδομένα που αντιστοιχούν στη στιγμή εκδήλωσης συμβάντος από συνεχείς μετρήσεις φωρατών.

Στη δεύτερη ενότητα εργασιών, εξετάζεται η επίδραση αριθμού παραμέτρων στον τύπο οδικού συμβάντος. Εφαρμόζονται πολυμεταβλητά (multivariate) μοντέλα probit σε τετραετή δεδομένα συμβάντων από το κοινό τμήμα των αυτοκινητόδρομων A4-A86 στην περιοχή Ile-de-France της Γαλλίας. Τα κυκλοφοριακά δεδομένα συνελέγησαν σε πραγματικό χρόνο κατά τη διάρκεια των συμβάντων και περιλαμβάνουν μετρήσεις φόρτου, ταχύτητας και πυκνότητας κατά τη διάρκεια εξάλεπτων χρονικών διαστημάτων. Τα εμπειρικά αποτελέσματα καταδεικνύουν πως ο τύπος συμβάντος μπορεί –σχεδόν αποκλειστικά- να εκτιμηθεί από τις επικρατούσες κυκλοφοριακές συνθήκες. Για παράδειγμα, οι νωτομετωπικές εμπλοκές με δύο οχήματα εμφανίζονται πιο πιθανές για σχετικά χαμηλές τιμές ταχύτητας και πυκνότητας, ενώ οι νωτομετωπικές εμπλοκές με περισσότερα από δύο οχήματα φαίνονται πιθανότερες σε συνθήκες κυκλοφοριακής συμφόρησης. Η γεωμετρία της οδού φαίνεται να αποτελεί τη μοναδική ένδειξη υψηλής πιθανότητας πρόκλησης συμβάντος με εμπλοκή ενός μόνο οχήματος. Επισημαίνεται ότι η πολυμεταβλητή ανάλυση επιτρέπει τον εντοπισμό αλληλεπιδράσεων μεταξύ των εξηρτημένων μεταβλητών οι οποίες θα παραβλέπονταν με μονομεταβλητή (univariate) ανάλυση.

Στην τρίτη ενότητα, εξετάζονται οι παράγοντες που επηρεάζουν το επίπεδο σοβαρότητας συμβάντος με βάση κυκλοφοριακά δεδομένα από το κοινό τμήμα των αυτοκινητοδρόμων A4-A86 στο Παρίσι. Ειδικότερα, διερευνάται η επίδραση που ασκεί τόσο η αναπτυσσόμενη ταχύτητα, όσο και ο κυκλοφοριακός φόρτος. Εισάγεται, για πρώτη φορά στη σχετική βιβλιογραφία, διατεταγμένο μοντέλο probit τυχαίων παραμέτρων (random parameters ordered probit model). Η χρήση τυχαίων παραμέτρων προσφέρει τη δυνατότητα μοντελοποίησης και ποσοτικοποίησης της ετερογένειας μεταξύ των χρηστών του οδικού δικτύου, καθώς και τη δυνατότητα εξέτασης του κατά πόσον η ετερογένεια αυτή έχει σημαντικές επιπτώσεις στην αξιοπιστία των μοντέλων. Σημειώνεται ότι η ετερογένεια θα παραβλεπόταν στην περίπτωση εφαρμογής προτύπου σταθερών παραμέτρων, το οποίο χρησιμοποιείται σε όλη τη σχετική βιβλιογραφία. Από τα αποτελέσματα των μοντέλων προκύπτει ότι οι μετακινήσεις με δίκυκλο τη νύκτα προκαλούν σημαντική αύξηση της πιθανότητας εμπλοκής σε σοβαρότερο συμβάν. Αντιθέτως, οι μετακινήσεις με βαρέα οχήματα σχετίζονται με χαμηλότερη πιθανότητα εμπλοκής σε σοβαρότερα συμβάντα. Από τα αποτελέσματα τεκμαίρεται, επίσης, ότι η σοβαρότητα συνδέεται στενά με τις κυκλοφοριακές συνθήκες που επικρατούν τη στιγμή εκδήλωσης του συμβάντος. Η αύξηση του κυκλοφοριακού φόρτου ασκεί σταθερή επίδραση στη σοβαρότητα των συμβάντων, ενώ η επίδραση της ταχύτητας διαφοροποιείται ανάλογα με το επίπεδο του κυκλοφοριακού φόρτου.

Στο πλαίσιο της τέταρτης ενότητας εργασιών, διερευνάται η ένταξη των πορισμάτων των προηγούμενων εργασιών σε ολοκληρωμένο σχέδιο διαχείρισης έκτακτων συμβάντων. Πραγματοποιείται σύνθεση σχετικών ερευνών με στόχο τον καθορισμό πλαισίου εφαρμογών διαχείρισης συμβάντων με χρήση κυκλοφοριακών δεδομένων που συλλέγονται σε πραγματικό χρόνο από αστικούς αυτοκινητόδρομους. Η σύνθεση καταδεικνύει πως τα δεδομένα αυτά δεν έχουν πλήρως αξιοποιηθεί στο πεδίο της διαχείρισης συμβάντων, καθώς η χρήση τους περιορίζεται στον υπολογισμό των χρόνων ταξιδίου των οχημάτων έκτακτης ανάγκης. Ωστόσο, τα δεδομένα αυτά θα μπορούσαν επίσης να χρησιμοποιηθούν ως κριτήριο για τη χωροθέτηση και τον καταμερισμό των περιοχών ευθύνης μονάδων έκτακτης ανάγκης. Τέλος, τα πορίσματα της διατριβής αξιοποιούνται στη διερεύνηση εφαρμογών με απώτερο στόχο τον περιορισμό της προδιάθεσης πρόκλησης συμβάντων και τη βελτιωμένη διαχείρισή τους.

Η διατριβή συνεισφέρει στην καλύτερη κατανόηση του μηχανισμού πρόκλησης οδικών συμβάντων. Επίσης, η ένταξη των πορισμάτων σε στρατηγικές διαχείρισης έκτακτων συμβάντων δημιουργεί νέες δυνατότητες μείωσης των συμβάντων και βελτιωμένης διαχείρισής τους. Συνολικά, η διδακτορική διατριβή συμβάλλει στη διερεύνηση ενός ζητήματος το οποίο έχει περιορισμένα εξεταστεί στο παρελθόν, ενώ παρέχει το κατάλληλο θεωρητικό πλαίσιο και τα απαιτούμενα μαθηματικά εργαλεία για την περαιτέρω αξιοποίηση μη ομαδοποιημένων κυκλοφοριακών δεδομένων.

Examples of model outputs (Limpdep v.8)

```
--> probit;lhs=type4; rhs= ONE,tjour,k2 $
Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| Binomial Probit Model                       |
| Maximum Likelihood Estimates                |
| Model estimated: May 10, 2010 at 11:17:06AM.|
| Dependent variable               TYPE4      |
| Weighting variable                None      |
| Number of observations             235      |
| Iterations completed                 5      |
| Log likelihood function      -94.43988      |
| Number of parameters                 3      |
| Info. Criterion: AIC =          .82928      |
|   Finite Sample: AIC =          .82972      |
| Info. Criterion: BIC =          .87344      |
| Info. Criterion:HQIC =          .84708      |
| Restricted log likelihood    -97.14205      |
| McFadden Pseudo R-squared     .0278167      |
| Chi squared                   5.404346      |
| Degrees of freedom                   2      |
| Prob[ChiSqd > value] =        .6705965E-01 |
| Hosmer-Lemeshow chi-squared =  10.76788     |
| P-value=  .21520 with deg.fr. =      8      |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+---------
-+
|Variable| Coefficient  | Standard Error |b/St.Er.|P[|Z|>z]| Mean of
X|
+--------+--------------+----------------+--------+--------+---------
-+
---------+Index function for probability
 Constant -1.78863194      .35827108    -4.992   .0000
 TJOUR  |    .44331172      .25263488     1.755   .0793   1.17021277
 K2    |     .07607589      .04261042     1.785   .0742   2.41355111

+----------------------------------------+
| Fit Measures for Binomial Choice Model |
| Probit   model for variable TYPE4      |
+----------------------------------------+
| Proportions P0= .855319   P1= .144681  |
| N =     235 N0=     201   N1=      34  |
| LogL=      -94.440 LogL0=     -97.142  |
| Estrella = 1-(L/L0)^(-2L0/n) = .02305  |
+----------------------------------------+
|     Efron |  McFadden  |  Ben./Lerman  |
|    .02370 |    .02782  |        .75868 |
|    Cramer | Veall/Zim. |     Rsqrd_ML  |
|    .02482 |    .04967  |        .02273 |
+----------------------------------------+
| Information  Akaike I.C. Schwarz I.C.  |
| Criteria         .82928        .87344  |
+----------------------------------------+
```

```
--> MPROBIT;Lhs=TYPE2,TYPE3,TYPE4,TYPE5,TYPE6;
    Eq1=ONE,v2,k2,jour;
    Eq2=ONE,q2,profil;
    Eq3=ONE,k2,tjour;
    Eq4=ONE,V2,profil,jour;
    Eq5=ONE,profil,tplan$

Normal exit from iterations. Exit status=0.
+----------------------------------------------+
| Multivariate Probit Model:  5 equations.     |
| Maximum Likelihood Estimates                 |
| Model estimated: May 04, 2010 at 02:41:33PM.|
| Dependent variable             MVProbit      |
| Weighting variable                 None      |
| Number of observations              235      |
| Iterations completed                 42      |
| Log likelihood function       -446.7275      |
| Number of parameters                 27      |
| Info. Criterion: AIC =          4.03172      |
|   Finite Sample: AIC =          4.06281      |
| Info. Criterion: BIC =          4.42921      |
| Info. Criterion:HQIC =          4.19197      |
| Replications for simulated probs. =  20      |
+---------------------------------------------+
+--------+-------------+----------------+--------+--------+----------
+
|Variable| Coefficient |Standard Error |b/St.Er.|P[|Z|>z]|Mean of X|
+--------+--------------+----------------+--------+--------+---------
---------+Index function for TYPE2
 Constant|  .92657704       .93157796     .995   .3199
 V2    |    -.01614016       .00871664    -1.852   .0641   73.3042553
 K2    |    -.20092640       .13071583    -1.537   .1243   2.41355111
 JOUR  |    -.24011301       .22614560    -1.062   .2883    .34468085
---------+Index function for TYPE3
 Constant|   -1.47561169   .29450580    -5.010   .0000
 Q2    |     .00375210       .00231056     1.624   .1044   112.404752
 PROFIL|     .24023659       .24634945      .975   .3295    .25106383
---------+Index function for TYPE4
 Constant|   -1.70263483   .38816560    -4.386   .0000
 K2    |     .07535639       .04800296     1.570   .1165   2.41355111
 TJOUR |     .38705066       .26728168     1.448   .1476   1.17021277
---------+Index function for TYPE5
 Constant|   -1.22576893   .26534673    -4.619   .0000
 V2    |     .00590282       .00342004     1.726   .0844   73.3042553
 PROFIL|    -.63319856       .26183919    -2.418   .0156    .25106383
 JOUR  |    -.36314454       .21306959    -1.704   .0883    .34468085
---------+Index function for TYPE6
 Constant  -.52630315       .15578289   -3.378   .0007
 PROFIL|    -.52906800       .25835671    -2.048   .0406    .25106383
 TPLAN |    -.64455388       .16572314    -3.889   .0001    .60851064
---------+Correlation coefficients
 R(01,02)|    -.27817767       .12408676    -2.242   .0250
 R(01,03)|    -.32671084       .16161857    -2.021   .0432
 R(02,03)|    -.33545709       .18665474    -1.797   .0723
 R(01,04)|    -.28810840       .05151744    -5.592   .0000
 R(02,04)|    -.22396158       .14735068    -1.520   .1285
 R(03,04)|    -.27769381       .09720964    -2.857   .0043
 R(01,05)|    -.42603264       .16410021    -2.596   .0094
 R(02,05)|    -.08756017       .09923649     -.882   .3776
 R(03,05)|    -.00302254       .10815616     -.028   .9777
 R(04,05)|    -.18599181       .17283783    -1.076   .2819
```

```
--> create; if (Q6MINV<112) Q1=0 $
--> create; if (Q6MINV>112) Q1=1 $
--> create; if (Q6MINV<112) Q2=1 $
--> create; if (Q6MINV>112) Q2=0 $
--> create; vq1=vmoy*q1 $
--> create; vq2=vmoy*q2 $
--> create; ancienw=anciperm*pluie $
-->  ordered;lhs=grav2;rhs=one,tjour,  jour,  normal,  plat,  retrait,
ancienw, moto,pl, vq1, q6minv
    ;RPM;Pts=5;Halton
    ;Fcn=one(c),tjour(c),  jour(c),  normal(c),  plat(c),  retrait(c),
ancienw(n)...
    moto(c),pl(c), vq1(n), q6minv(n) $
Normal exit from iterations. Exit status=0.

+---------------------------------------------+
| Random Coefficients  OrdProbs Model         |
| Maximum Likelihood Estimates                |
| Model estimated: Jun 20, 2009 at 05:27:37PM.|
| Dependent variable                GRAV2     |
| Weighting variable                 None     |
| Number of observations              893     |
| Iterations completed                 20     |
| Log likelihood function       -563.2782     |
| Number of parameters                 16     |
| Info. Criterion: AIC =          1.29738     |
|   Finite Sample: AIC =          1.29807     |
| Info. Criterion: BIC =          1.38328     |
| Info. Criterion:HQIC =          1.33020     |
| Sample is  1 pds and     893 individuals.   |
| Ordered probability model                   |
| Ordered probit (normal) model               |
| LHS variable = values 0,1,..., 3            |
| Simulation based on   5 Halton draws        |
+---------------------------------------------+
+--------+--------------+----------------+--------+--------+---------
-+
|Variable| Coefficient   | Standard Error |b/St.Er.|P[|Z|>z]| Mean of
X|
+--------+--------------+----------------+--------+--------+---------
-+
---------+Means for random parameters
 Constant|    -.44301558       .59438251     -.745   .4561
 TJOUR   |  -.93295707       .13437103    -6.943   .0000    .21200750
 JOUR    |   .80057043       .11946992     6.701   .0000    .48217636
 NORMAL  |  -.71503572       .12310491    -5.808   .0000    .27204503
 PLAT    |   .93164857       .12137748     7.676   .0000    .32082552
 RETRAIT |  1.02478454       .52600775     1.948   .0514    .97936210
 ANCIENW |   .02977253       .00538850     5.525   .0000   9.14071295
 MOTO    |  -.69683858       .16128257    -4.321   .0000    .78986867
 PL      |   .44336970       .23218199     1.910   .0562    .94934334
 VQ1     |   .00520172       .00156140     3.331   .0009   29.7847717
 Q6MINV  |  -.00650266       .00159127    -4.086   .0000   107.034703
---------+Scale parameters for dists. of random parameters
 Constant|       .000000    ......(Fixed Parameter).......
 TJOUR   |       .000000    ......(Fixed Parameter).......
 JOUR    |       .000000    ......(Fixed Parameter).......
 NORMAL  |       .000000    ......(Fixed Parameter).......
 PLAT    |       .000000    ......(Fixed Parameter).......
 RETRAIT |       .000000    ......(Fixed Parameter).......
 ANCIENW |     .00366067       .00309254     1.184   .2365
```

```
 MOTO    |       .000000    ......(Fixed Parameter).......
 PL      |       .000000    ......(Fixed Parameter).......
 VQ1     |     .00176403       .00095787     1.842   .0655
 Q6MINV  |     .00099245       .00044707     2.220   .0264
---------+Threshold parameters for probabilities
 MU(1)   |    1.15355006       .09034273    12.769   .0000
 MU(2)   |    2.22566575       .14479618    15.371   .0000


Implied standard deviations of random parameters

Matrix S.D_Beta has 11 rows and  1 columns.
               1
        +--------------
       1|  .0000000D+00
       2|  .0000000D+00
       3|  .0000000D+00
       4|  .0000000D+00
       5|  .0000000D+00
       6|  .0000000D+00
       7|     .00366
       8|  .0000000D+00
       9|  .0000000D+00
      10|     .00176
      11|     .00099
```