

A MACROLEVEL TABU-SEARCH BASED APPROACH TO DETERMINE
THE MAJOR FACTORS IN BUILT ENVIRONMENT-TRAFFIC SAFETY
RELATIONSHIP

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY



BY
ONUR ALIŞAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
CITY AND REGIONAL PLANNING

MARCH 2021

Approval of the thesis:

**A MACROLEVEL TABU-SEARCH BASED APPROACH TO DETERMINE
THE MAJOR FACTORS IN BUILT ENVIRONMENT-TRAFFIC SAFETY
RELATIONSHIP**

submitted by **ONUR ALIŞAN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in City and Regional Planning, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences**

Prof. Dr. Serap Kayasü
Head of the Department, **City and Regional Planning**

Prof. Dr. Ela Babalık
Supervisor, **City and Regional Planning, METU**

Examining Committee Members:

Prof. Dr. Emine Yetişkul Şenbil
City and Regional Planning, METU

Prof. Dr. Ela Babalık
City and Regional Planning, METU

Prof. Dr. Burcu Özüduru
City and Regional Planning, Gazi University

Assoc. Prof. Dr. Hediye Tüydeş Yaman
Civil Engineering, METU

Assoc. Prof. Dr. Fikret Zorlu
City and Regional Planning, Mersin University

Date: 05.03.2021



I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: Onur Alişan

Signature:

ABSTRACT

A MACROLEVEL TABU-SEARCH BASED APPROACH TO DETERMINE THE MAJOR FACTORS IN BUILT ENVIRONMENT-TRAFFIC SAFETY RELATIONSHIP

Alişan, Onur
Doctor of Philosophy, City and Regional Planning
Supervisor: Prof. Dr. Ela Babalık

March 2021, 318 pages

Traffic crashes are a significant contributing factor to fatalities globally, with an increasing pace of crashes occurring in urban areas. Thus, this study aims to analyze the relationship between urban built environment and traffic safety (BE-TS) to reveal the factors that threaten TS so that they could be avoided in early planning phases. BE was operationalized through the *D*-classification (e.g., density and diversity) whereas, TS was operationalized through varying crash categories (e.g., severity) and representation (i.e., frequency and density). A GIS-based process was applied to assign crashes and BE variables at the census block group level, followed by a dimension reduction process to eliminate correlated variables in an extensive set. Then the relationship between BE and TS was examined through a series of generalized linear regression models with several statistical distributions. A tabu-search-based model generation was applied to each BE-TS model to identify the set of BE variables that minimizes the modeling error. Moreover, variables were traced during the algorithm to identify consistent and then critical variables. Those models were analyzed for varying case studies, two cases within the developed country context, from the US, single county (Leon), and tri-county (Broward, Miami-Dade,

and Palm-Beach) cases; and a case within developing country context, from Turkey (Konya). According to the results, exposure variables (e.g., bus stops and traffic volume) have positive associations with all crash types. However, there are variables with mixed relations (e.g., primary road percentage has positive relations with crash types except for vulnerable road user involved crashes, and employment density has positive or negative impacts according to the local economic structure). Finally, non-motorized modes are found to have positive, and motorized modes are found to have negative associations with all crash types within the developed country context, which are not common in safety research.

Keywords: Built Environment, Traffic Safety, Crash Frequency, Subset Selection, Tabu-Search

ÖZ

YAPILI ÇEVRE-TRAFİK GÜVENLİĞİ İLİŞKİSİNDE BAŞLICA FAKTÖRLERİN BELİRLENMESİ İÇİN TABU-ARAMA TABANLI BİR YAKLAŞIM

Alişan, Onur
Doktora, Şehir ve Bölge Planlama
Tez Yöneticisi: Prof. Dr. Ela Babalık

Mart 2021, 318 sayfa

Trafik kazaları, küresel olarak ölüm sebepleri içerisinde önemli bir paya sahiptir. Bununla birlikte kentsel alanlarda meydana gelen trafik kazalarının sayısında artış yaşanmaktadır. Buradan hareketle bu çalışmanın amacı, trafik güvenliğini tehdit eden yapıli çevre unsurlarını planlamanın erken aşamalarında önleyebilmek için, kentsel yapıli çevre ile trafik güvenliği (YÇ-TG) arasındaki ilişkiyi çözümlemektir. YÇ'nin operasyonel tanımları *D*-sınıflandırması (ör. yoğunluk) aracılığıyla yapılırken, TG'nin operasyonel tanımları, farklı kaza kategorileri (ör. kaza şiddeti) ve kazaların temsili (görölme sıklığı ve yoğunluğu) aracılığıyla yapılmıştır. Trafik kazalarını ve YÇ değişkenlerini nüfus blok grubu düzeyinde atayabilmek için CBS tabanlı bir süreç uygulanmıştır. Bu süreci çok sayıda elemanı olan değişken kümesinde birbiriyle ilişkili değişkenlerin elenmesi için uygulanan boyut azaltımı işlemi takip etmiştir. Sonrasında YÇ-TG arasındaki ilişkiler farklı istatistiksel dağılımları kullanan bir dizi genelleştirilmiş doğrusal regresyon modeli aracılığıyla incelenmiştir. Her YÇ-TG modelinde, modelleme hatasını en aza indirecek YÇ değişkenleri alt kümesini belirlemek için tabu-arama-tabanlı bir model üretim yöntemi uygulanmıştır. Ayrıca, tutarlı ve sonrasında kritik değişkenleri

belirleyebilmek için algoritma sırasında YÇ değişkenleri takip edilmiştir. Bu modeller, ikisi gelişmiş ülke bağlamında ABD'den ve biri gelişmekte olan ülke bağlamında Türkiye'den olmak üzere olmak üzere üç farklı vaka çalışma alanı için çözümlenmiştir. ABD örnekleri tekil (Leon) ve üçlü (Broward, Miami-Dade ve Palm-Beach) bölgesel birimlerden oluşurken, Türkiye örneği Konya ilinden oluşmuştur. Sonuçlara göre, maruz kalma değişkenlerinin (ör. otobüs durakları ve trafik hacmi) tüm kaza türleriyle pozitif bir ilişkisi vardır. Bununla birlikte karma ilişkileri olan değişkenler vardır (ör. birinci kademe yol oranının, korumasız kullanıcıların dahil oldukları dışındaki kazalarla pozitif bir ilişkisinin olması ve istihdam yoğunluğunun yerel ekonomik yapıya göre pozitif veya negatif etkisinin olması). Son olarak, gelişmiş ülke bağlamında, tüm kaza türlerinin motorsuz ulaşım türleriyle pozitif ve motorlu ulaşım türleriyle negatif ilişkili olduğu bulunmuştur ki bu bulgular güvenlik araştırmaları açısından yaygın değildir.

Anahtar Kelimeler: Anahtar Kelimeler: Yapılı Çevre, Trafik Güvenliği, Kaza Sıklığı, Alt Küme Seçimi, Tabu-Arama



To Godot

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my supervisor, Dr. Ela Babalık, for her concern, support, positive attitude, and invaluable guidance throughout this study. I would also like to thank Dr. Hediye Tüydeş-Yaman for her guidance and contributions to this thesis study. Also, I would like to thank Dr. Emine Yetişkul Şenbil for her contributions and criticisms during this study. Moreover, I would like to thank the dissertation committee members Dr. Burcu Özuduru and Dr. Fikret Zorlu for their criticisms, detailed evaluations, and commitments to this final work.

I am also grateful to Dr. Eren Erman Özgüven and Dr. Michael Duncan for their support and contributions; Dr. Adnan Barlas and Dr. Melih Pınarcıoğlu for their guidance.

I don't know how to express my feelings for my family, but at least I would like to thank them here for their endless patience and support. Thank you Feyza, Müfit, Zeynep, Umut and Damla.

I would like to thank The Scientific and Technological Research Council of Turkey (TÜBİTAK) for the financial support during my doctoral study with the National Scholarship Program for Graduate Students (2211-A).

I would also like to thank the Turkish Fulbright Commission for supporting me during my visit to Florida State University and giving me the opportunity to broaden my perspective of life.

Finally, I would like to thank The General Directorate of Police-Traffic Department (Trafik Başkanlığı) for sharing crash data that contributed to this study.

TABLE OF CONTENTS

ABSTRACT.....	v
ÖZ	vii
ACKNOWLEDGMENTS	x
TABLE OF CONTENTS.....	xi
LIST OF TABLES	xv
LIST OF FIGURES	xviii
CHAPTERS	
1 INTRODUCTION	1
2 BACKGROUND: UNDERSTANDING THE RELATIONSHIPS BETWEEN THE BUILT ENVIRONMENT, TRAVEL, AND TRAFFIC SAFETY FROM SUSTAINABLE DEVELOPMENT AND URBAN GROWTH PARADIGMS.....	7
2.1 Sustainable Development.....	7
2.2 Sustainable Development for Urban Growth: Compactness vs. Sprawl Debate	10
2.2.1 Urban built environment	21
2.2.1.1 Dimensions of the built environment	22
2.2.2 Urban transportation	25
2.2.2.1 Travel demand	26
2.2.2.2 Traffic safety.....	27
2.2.3 Background on the relationships between travel, built environment, and traffic safety	28
2.2.3.1 Built environment and travel	29
2.2.3.2 Travel and traffic safety.....	36
2.2.3.3 Built environment and traffic safety	40

2.3	Summary.....	64
3	METHODOLOGY	67
3.1	Aim of the Study and Research Questions	67
3.2	Methodology.....	69
3.2.1	Extraction and classification of geospatial data	70
3.2.1.1	Crash frequency	73
3.2.1.2	Crash density	75
3.2.2	Dimension reduction	80
3.2.3	Subset selection	81
3.2.3.1	Tabu-search based subset selection.....	83
3.2.3.2	Fitting statistical distributions	87
3.2.4	Evaluation of the results	90
3.2.5	The frame of the methodology	93
3.3	Study Area	94
3.4	Data Extraction and Classification for the Study Area.....	97
3.4.1	Traffic safety dataset	97
3.4.1.1	Crash frequency	100
3.4.1.2	Crash density	105
3.4.2	Built environment dataset.....	111
3.4.3	Dimension reduction for the built environment dataset	118
3.4.4	Subset selection-tabu-search calibration	122
3.5	Summary.....	127
4	RESULTS FROM ACCIDENT FREQUENCY AND DENSITY MODELS, DEVELOPED COUNTRY CONTEXT.....	131
4.1	Background.....	131
4.2	Tabu-Search Based Solution Algorithm: Overview	132

4.3	Tabu-Search Based Solution Algorithm: Examples.....	134
4.4	Results of the Single County Case	139
4.4.1	Single county-crash frequency modeling results	139
4.4.2	Single county-crash density modeling results.....	142
4.4.3	Discussion of the single county crash modeling outcomes.....	145
4.4.4	Consistency of the single county crash modeling results	151
4.5	Results of the Tri-County Case	156
4.5.1	Tri-county-crash frequency modeling results	156
4.5.2	Tri-county-crash density modeling results.....	160
4.5.3	Discussion of the tri-county crash modeling outcomes	164
4.5.4	Consistency of the tri-county crash modeling results	171
4.6	The Key Built Environment Variables and Urban Development	174
5	RESULTS FROM ACCIDENTS FREQUENCY AND DENSITY MODELS, DEVELOPING COUNTRY CONTEXT	181
5.1	Introduction	181
5.2	Case Study Area, Konya	182
5.3	Data Extraction and Classification for Konya.....	183
5.3.1	Traffic safety dataset.....	183
5.3.2	Built environment dataset	188
5.4	BE-TS Modeling Results	191
5.5	Model Comparison	201
6	CONCLUSION.....	213
6.1	Overview of the Research	213
6.2	Summary of the Findings	215

6.3	Alternative Perspectives and the Flexibility of the Methodology	220
6.4	A Discussion for the Planning Practice	224
6.5	Limitations and Future Research Directions	231
REFERENCES		235
APPENDICES		
A.	Single County Case Results.....	271
B.	Tri-County Case Results.....	283
C.	Konya Case Results.....	295
D.	Results of Single and Tri-County Case Models with Key BE Variables ..	299
E.	Additional Goodness-of-Fit Statistics for Single and Tri-County Case Models	315
CURRICULUM VITAE		317

LIST OF TABLES

TABLES

Table 2.1 Characteristics of Urban Sprawl	15
Table 2.2 Viewpoints in Sprawl Debate	17
Table 2.3 Leading causes of death, all ages, 2016.....	27
Table 2.4 Review on Built Environment and Travel	31
Table 2.5 Review on Travel and Traffic Safety.....	38
Table 2.6 Review on Built Environment and Traffic Safety - Link Level	47
Table 2.7 Review on Built Environment and Traffic Safety - Zonal Level	52
Table 2.8 Review on Built Environment and Traffic Safety - Regional Level	61
Table 3.1 Applied Statistical Distributions and Their Parameters.....	89
Table 3.2 Fit Statistics for Regression Models	90
Table 3.3 Additional Goodness of Fit Statistics	92
Table 3.4 Research Frame.....	93
Table 3.5 Information on the Case Study Areas	94
Table 3.6 Crash Frequencies in the Study Areas	100
Table 3.7 Demographic and Socioeconomic Variables	114
Table 3.8 Density Variables.....	115
Table 3.9 Diversity Variables	115
Table 3.10 Design Variables.....	116
Table 3.11 Transit Variables.....	117
Table 3.12 Destination Accessibility Variables.....	117
Table 3.13 Travel Variables.....	117
Table 3.14 Descriptive Statistics of the Single County Case Dataset.....	119
Table 3.15 Descriptive Statistics of the Tri-County Case Dataset.....	120
Table 3.16 Results of Preliminary Experiments	123
Table 3.17 Test Results for Objective Function	126
Table 4.1 Organization of the Settings	132

Table 4.2 Optimal BE-TS Models for Single County Case (FQ severity).....	140
Table 4.3 Optimal BE-TS Models for Single County Case (FQ involvement).....	141
Table 4.4 Optimal BE-TS Models for Single County Case (KD severity)	143
Table 4.5 Optimal BE-TS Models for Single County Case (KD involvement)....	144
Table 4.6 Comparison of the IVs for Single County Case	146
Table 4.7 Consistency of the IVs for Single County Case	153
Table 4.8 Optimal BE-TS Models for Tri-County Case (FQ severity).....	157
Table 4.9 Optimal BE-TS Models for Tri-County Case (FQ involvement).....	158
Table 4.10 Optimal BE-TS Models for Tri-County Case (KD severity)	161
Table 4.11 Optimal BE-TS Models for Tri-County Case (KD involvement).....	162
Table 4.12 Comparison of the IVs for Tri-County Case	165
Table 4.13 Consistency of the IVs for Tri-County Case	173
Table 4.14 Sprawl Score of Some Counties in the US.....	176
Table 4.15 Transit Score of Some Counties in the US.....	178
Table 5.1 Information on Konya Case	182
Table 5.2 Crash Frequencies in Konya Case.....	185
Table 5.3 Built Environment Variables.....	189
Table 5.4 Descriptive Statistics of the Single County Case Dataset	190
Table 5.5 Optimal BE-TS Models for the Konya Case (FQ severity)	191
Table 5.6 Optimal BE-TS Models for the Konya Case (FQ involvement)	192
Table 5.7 Optimal BE-TS Models for the Konya Case (KD severity).....	194
Table 5.8 Optimal BE-TS Models for the Konya Case (KD involvement)	194
Table 5.9 Comparison of the IVs for the Konya Case	196
Table 5.10 Consistency of the IVs for the Konya Case	200
Table 5.11 Significant Variables in Each Model.....	203
Table 5.12 Measuring Data Contribution in Single County Frequency Models...	205
Table 5.13 Measuring Data Contribution in Tri-County Frequency Models	206
Table 5.14 Measuring Data Contribution in Single County Density Models	207
Table 5.15 Measuring Data Contribution in Tri-County Density Models	208
Table 5.16 Distribution of Crash Types in Varying Cases.....	211

Table 6.1 Factor Significance and Consistency: VRU Crashes in Konya Case ...	223
Table E.1 Goodness-of-Fit Statistics of Single County Frequency Models	315
Table E.2 Goodness-of-Fit Statistics of Single County Density Models.....	315
Table E.3 Goodness-of-Fit Statistics of Tri-County Frequency Models	316
Table E.4 Goodness-of-Fit Statistics of Tri-County Density Models.....	316



LIST OF FIGURES

FIGURES

Figure 1.1. Research Map.....	6
Figure 2.1. Components of Sustainability	10
Figure 2.2. Annual Total Population at Mid-Year, 1950-2050 (billion)	11
Figure 2.3. % of Population at Mid-Year Residing in Urban Areas, 1950-2050	12
Figure 3.1. Proposed Methodological Process	69
Figure 3.2. Possible Geographic Unit of Analyses and Conversion	71
Figure 3.3. ¼ Mile Buffer Zone Defined for Each Crash.....	72
Figure 3.4. Crash Assignment by Applying a Buffer Zone.....	74
Figure 3.5. (a) Thiessen Polygons Applied to Census Tract Centroid in Travis County (b) 10x10 miles Grid Structure Applied to Florida	77
Figure 3.6. Flowchart of the algorithm.....	79
Figure 3.7. Tabu-search Pseudocode.....	84
Figure 3.8. Tabu-search Pesudocode in this Study.....	86
Figure 3.9. Population Block Groups and Urban Region in Leon County.....	95
Figure 3.10. Population Block Groups and Urban Region in Tri-County Case.....	96
Figure 3.11. Crashes According to Involvement and Severity for the Single County Case	98
Figure 3.12. Crashes According to Involvement and Severity of the Tri-County Case	99
Figure 3.13. Spatial Distribution of Different Types of Crashes in Single County Case	103
Figure 3.14. Spatial Distribution of Different Types of Crashes in Tri-County Case	104
Figure 3.15. Kernel Density of Crashes Classified According to Severity in Single County Case.....	107
Figure 3.16. Kernel Density of Crashes Classified According to Involvement in Single County Case (continued from Figure 3.15).....	108

Figure 3.17. Kernel Density of Crashes Classified According to Severity in Tri-County Case	109
Figure 3.18. Kernel Density of Crashes Classified According to Involvement in Tri-County Case (continued from Figure 3.17)	110
Figure 3.19. Data Sources and Classification	113
Figure 3.20. Experimental Design for the Single County Case	128
Figure 4.1. Tabu-Search Based Setting.....	133
Figure 4.2. Tabu-search Results – Single County - FQ_S3	135
Figure 4.3. Tabu-search Results – Single County - FQ_ALL	138
Figure 5.1. Traffic Analysis Zones and the Provincial Border of Konya	182
Figure 5.2. Crashes According to Involvement and Severity of the Konya Case	184
Figure 5.3. Crash Frequency Distributions in Konya (a) Injury Crashes (b) Fatal Crashes (c) Total Crashes (d) Motor Vehicle Involved Crashes (e) VRU Crashes	186
Figure 5.4. Kernel Density of Crash Typology in Konya (a) Injury Crashes (b) Fatal Crashes (c) Total Crashes (d) Motor Vehicle Involved Crashes (e) VRU Crashes	187
Figure A.1. FQ_S1 Negative Binomial.....	271
Figure A.2. FQ_S2 Negative Binomial.....	271
Figure A.3 FQ_S3 Negative Binomial.....	272
Figure A.4. FQ_ALL Negative Binomial	272
Figure A.5. FQ_MOT Negative Binomial	273
Figure A.6. FQ_VRU Negative Binomial	273
Figure A.7. KD_S1 Gamma.....	274
Figure A.8. KD_S2 Gamma.....	274
Figure A.9. KD_S3 Gamma.....	275
Figure A.10. KD_ALL Gamma	275
Figure A.11 KD_MOT Gamma	276
Figure A.12. KD_VRU Gamma	276
Figure A.13. KD_S1 Lognormal	277

Figure A.14. KD_S2 Lognormal	277
Figure A.15. KD_S3 Lognormal	278
Figure A.16. 16 KD_ALL Lognormal	278
Figure A.17. KD_MOT Lognormal	279
Figure A.18. KD_VRU Lognormal.....	279
Figure A.19. KD_S1 Weibull	280
Figure A.20. KD_S2 Weibull.....	280
Figure A.21. KD_S3 Weibull	281
Figure A.22. KD_ALL Weibull	281
Figure A.23. KD_MOT Weibull	282
Figure A.24. KD_VRU Weibull.....	282
Figure B.1. FQ_S1 Negative Binomial	283
Figure B.2. FQ_S2 Negative Binomial	283
Figure B.3. FQ_S3 Negative Binomial	284
Figure B.4. FQ_ALL Negative Binomial.....	284
Figure B.5. FQ_MOT Negative Binomial.....	285
Figure B.6. FQ_VRU Negative Binomial	285
Figure B.7 KD_S1 Gamma	286
Figure B.8 KD_S2 Gamma	286
Figure B.9. KD_S3 Gamma	287
Figure B.10 KD_ALL Gamma.....	287
Figure B.11. KD_MOT Gamma.....	288
Figure B.12. KD_VRU Gamma	288
Figure B.13. KD_S1 Lognormal	289
Figure B.14. KD_S2 Lognormal	289
Figure B.15. KD_S3 Lognormal	290
Figure B.16. KD_ALL Lognormal.....	290
Figure B.17. KD_MOT Lognormal.....	291
Figure B.18. KD_VRU Lognormal	291
Figure B.19. KD_S1 Weibull	292

Figure B.20. KD_S2 Weibull.....	292
Figure B.21. KD_S3 Weibull.....	293
Figure B.22. KD_ALL Weibull	293
Figure B.23. KD_MOT Weibull	294
Figure B.24. KD_VRU Weibull	294
Figure C.1. FQ_S2 Negative Binomial.....	295
Figure C.2. FQ_ALL Negative Binomial	295
Figure C.3. FQ_MOT Negative Binomial	296
Figure C.4. FQ_VRU Negative Binomial.....	296
Figure C.5. KD_S2 Gamma.....	297
Figure C.6. KD_ALL Gamma	297
Figure C.7. KD_MOT Gamma	298
Figure C.8. KD_VRU Gamma.....	298
Figure D.1. Single County Frequency S2	299
Figure D.2. Single County Frequency ALL.....	300
Figure D.3. Single County Frequency MOT.....	301
Figure D.4. Single County Frequency VRU	302
Figure D.5. Single County Kernel Density S2.....	303
Figure D.6. Single County Kernel Density ALL	304
Figure D.7. Single County Kernel Density MOT	305
Figure D.8. Single County Kernel Density VRU.....	306
Figure D.9. Tri-County Frequency S2	307
Figure D.10. Tri-County Frequency ALL.....	308
Figure D.11. Tri-County Frequency MOT.....	309
Figure D.12. Tri-County Frequency VRU	310
Figure D.13. Tri-County Kernel Density S2.....	311
Figure D.14. Tri-County Kernel Density ALL	312
Figure D.15. Tri-County Kernel Density MOT	313
Figure D.16. Tri-County Kernel Density VRU.....	314

CHAPTER 1

INTRODUCTION

One of the main requirements of human beings is to shelter. According to Maslow's Hierarchy of Needs, sheltering is one of the five main physiological needs that have to be satisfied to pursue higher-order needs (Maslow, 1943). The spaces where people live have utmost importance besides production or planning of spaces where people live is under the profession of urban planning, which is an ongoing process.

There are significant parts of this planning process, such as determining land uses, determining the growth direction, or transportation planning. Those parts could not be isolated according to the nature of the urban planning profession. Thus, they are performed in an intertwined fashion. Therefore, when the term urbanization is used, it refers to the structures and spaces and the links and relations connecting or disconnecting those structures and spaces, so a network of parts working together ideally. Moreover, the policies and the planning principles that produce this network should also be considered under urbanization and the urban planning discipline that has to organize them all. This organization was mostly under control till the boom period at the end of the massive population losses in world wars.

The process of urbanization has increased after the Second World War. By the triggering effect of the increase in private car ownership, this increase extended the cities towards their fringes where cheaper land parcels were available. After that, the urban form (land uses and the built environment) had significantly altered by the impacts that emerged from the changes in transportation. Those impacts are embedded in the change of urban built environment not only as being triggers but also as being triggered by the changes in the built environment itself. As in the case of private car usage, urban areas are scattered around as low-density suburbs since they are accessible (at least reachable) by cars; this also increased car ownership as

being the only alternative in transportation since public transit is not feasible to serve those low-density areas. This excessive (in terms of urban built-up area), scattered (dispersed), homogenous and low-density residential development type is specified as urban sprawl. The increase in car ownership that was the primary driver led to further scattering and increased negative externalities such as increased vehicle miles traveled, traffic accidents, and traffic congestion, as hypothesized by transport scholars.

Among those externalities, special attention should be given to traffic accidents since they are directly related to human life. Every year, 1.35 million people are killed in road traffic crashes worldwide, which corresponds to an average rate of 3,700 people daily. Roadway traffic injuries were estimated to be the eighth leading cause of death for all age segments and the number one for the younger population between the ages 5 and 29 in 2016 (WHO, 2018). Moreover, traffic crashes and fatalities also have substantial economic costs. For example, with 32,999 fatalities in 2010, the total economic cost of roadway crashes in the US was about \$242 billion (about 1.6% of GDP), and fatalities reached up to the level of 34,439 people in 2016 (Blincoe et al., 2015; ITF, 2018).

An increasing rate of roadway crashes happens in urban areas globally (ITF, 2018; WHO, 2018). In 2017 only, 44% of deaths occurred on urban roadways, 40% on rural roads, and 15% on highways/motorways in the US. According to the International Transport Forum (ITF, 2018), rural roadways were the deadliest in the US two decades ago; however, this pattern has changed as urban vehicle miles traveled increased 13%, whereas rural vehicle miles traveled decreased by 3% since 2008 (Highway Traffic Safety Administration, 2019; ITF, 2018). Therefore, further evaluations of the factors and causes of road crashes are needed; thus, many studies have recently focused on these aspects.

As urban crashes take the lead in causing fatalities and injuries, urban built environments require special attention. A critical question that remains to be addressed is whether the urban built environment is a significant contributor to the

traffic safety problem, if so, from what perspectives. The situation is even more critical for developing countries, where both motorization and urban development are happening simultaneously, making it harder to control roadway safety. This requires an examination of the relation between the built environment (BE) and traffic safety (TS). Although this relationship was the focus in different studies, the relationship between the BE and TS has not been conclusive yet (Merlin et al., 2020). Moreover, very few of these studies focused on developing country cases, where data quality and availability are the prime concerns.

Depending on this context, this study's main aim is to analyze the relationship between the built environment and traffic safety. This aim is organized by the following objectives (secondary aims):

- Determining the built environment factors that have significant associations with traffic crashes
- Comparing the BE-TS relationship findings within different case studies and contexts
- Developing a generalizable methodology to evaluate those different cases in various contexts
- Providing an evaluation scheme for the value of data availability, which is a critical issue in developing countries

Moreover, the aim and the related objectives are formulated by the following research questions:

RQ1) Which built environment factors are related to traffic crashes?

RQ2) Are the relationships between the built environment factors and varying types of crashes consistent (e.g., are the factors associated with fatal crashes similar to the aspects related to property damage only [PDO] crashes)?

- RQ3) Do the built environment factors affect varying types of crashes in the same way (e.g., whether the associations are consistently positive or negative)?
- RQ4) Are there any similarities or differences among varying study areas?
- RQ5) Are there any similarities or differences between the findings of this study and the findings in the literature?
- RQ6) Could there be a generalizable relationship pattern among the built environment factors and traffic safety?

Accordingly, the study is organized into chapters as follows:

Chapter 2 provides the background of the study. The conceptualization of the built environment and transportation within the sustainability frame is introduced. Then the associations of the built environment with travel demand and traffic safety are discussed. Moreover, a thorough review of the relevant literature on the relationships between the built environment, travel demand, and travel safety is provided with a focus on the relationship between the built environment and traffic safety.

Chapter 3 presents the methodological framework, including the aims and related research questions. In this framework, dataset preparation is followed by selecting the key built environment variables associated with traffic safety. Different data extraction methods and statistical distributions are introduced for the data preparation stage. Moreover, a tabu-search-based algorithm that consists of iterative solutions of multiple regression models is presented at the subset selection stage. The best sets of built environment variables are used to identify critical factors for the aforementioned relationship. The overall methodology is flexible so that it could be employed for various emphasizes with minor changes. For example, this study focuses on the critical variables that are valid for most of the crashes; but the focus could easily be shifted towards fatal crashes.

Chapter 4 introduces the developed country context. There are two different cases: a small-sized single county case and a tri-county urban agglomeration case, both from

the State of Florida, the US. Other than this spatial variation, there are two crash typologies used; one is based on crash severity (e.g., property damage only crashes). The other is based on involvement (e.g., vulnerable road user involved crashes). Also, the crash representation is alternated for accuracy purposes (i.e., crash count and density-based representation). The results of these two particular cases are discussed, and within a “variable significance-variable consistency” frame, the key built environment factors are determined.

Chapter 5 introduces the developing country context. The case study from this context is selected as Konya, Turkey. The methodology introduced in Chapter 3 (data preparation, subset selection, and evaluation) is also applied to the Konya case. The key built environment factors are identified using the “variable significance-variable consistency” frame. Moreover, the results of the developed and developing country contexts are compared. Finally, a straightforward evaluation scheme is presented for the value of built environment data while structuring the relationship between the built environment and traffic safety.

Chapter 6 concludes the study with a summary of the research. This is followed by a section dedicated to exemplifying how the generalizable frame proposed in this study could be adapted with a different focus. Then, a discussion on the findings within the planning perspective is presented. At last, the limitations of the study and future research possibilities are discussed. The research map is presented in Figure 1.1.

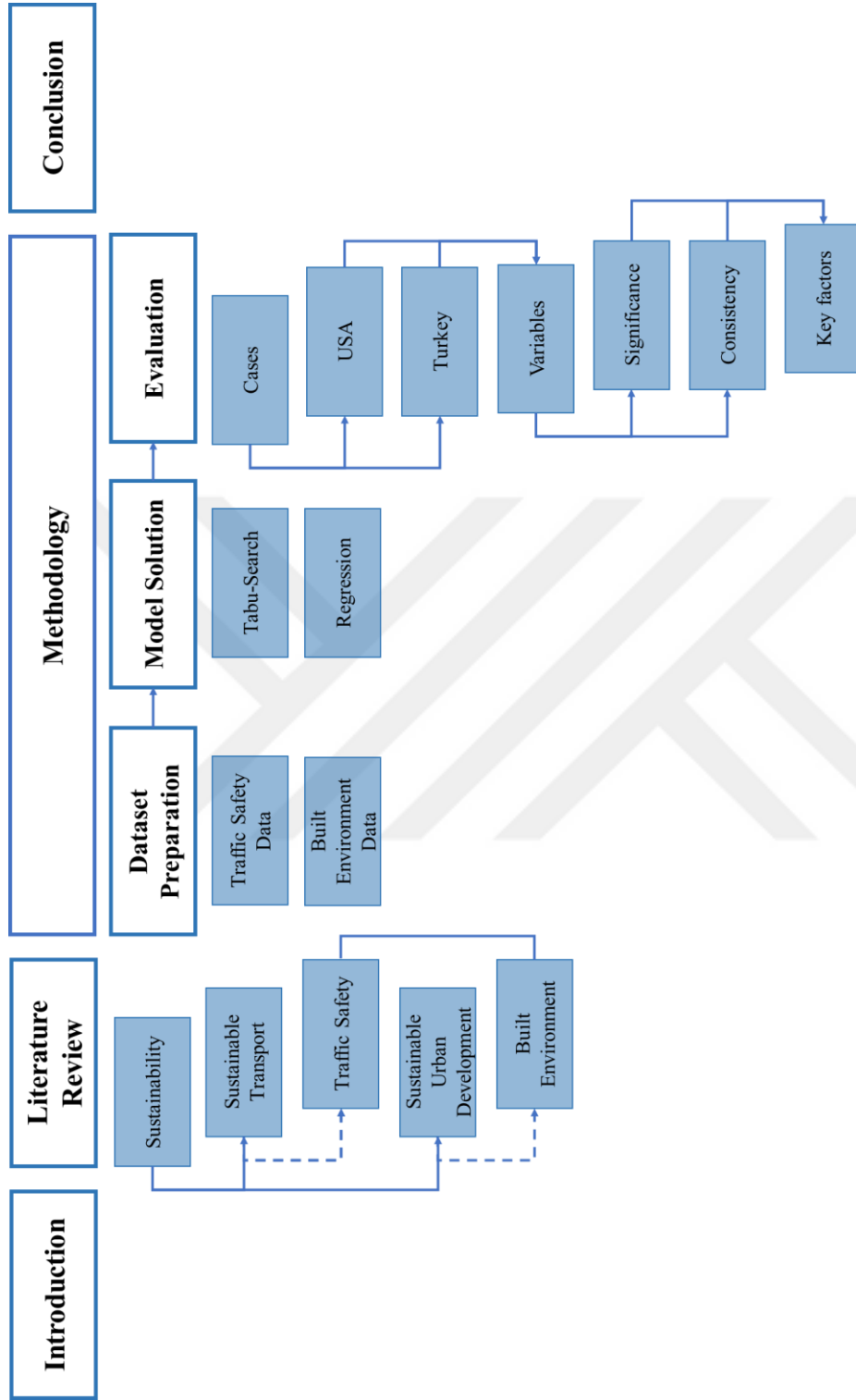


Figure 1.1. Research Map

CHAPTER 2

BACKGROUND: UNDERSTANDING THE RELATIONSHIPS BETWEEN THE BUILT ENVIRONMENT, TRAVEL, AND TRAFFIC SAFETY FROM SUSTAINABLE DEVELOPMENT AND URBAN GROWTH PARADIGMS

This chapter presents a background review regarding sustainability, sustainable development, and particularly the urban development process. This focus on the concept of sustainability is because of the significant emphasis that the sustainable development paradigm places on the built environment and its effects on urban transportation. Consequently, while presenting debates of sustainable development, in order to elaborate on the hypothesized relationship between the built environment and traffic safety, an extended review considering the built environment, travel, and traffic safety is presented, while the literature review's primary attention is paid to the association between the built environment and traffic safety.

2.1 Sustainable Development

The present sustainable development utilization began after the Second World War when developed nations became increasingly aware of the rapid industrialization combined with population growth, leading to substantial negative effects on the environment. These adverse impacts cannot be solved within the developed nations' area of responsibility. “Silent Spring” by Rachel Carson (1962), the “Tragedy of the Commons” by Garrett Hardin (1968), and “Limits to Growth” by Meadows et al. (1972) are prominent examples of publications that took attention and raised concerns about environmental problems and raised consciousness about ecology (Gerike & Koszowski, 2017). Besides, the understanding of development has changed drastically from the 1970s onwards due to the realization of oil dependence (while fossil fuels are limited) and the harm given to the natural environment through greenhouse gas emissions (causing global warming and even changes in climates).

In the first half of the 1970s, the United States supported Israel during the Arab-Israeli War. Thus, during 1973-74 an oil embargo against the United States was imposed by the Arabic states of the Organization of Petroleum Exporting Countries (OPEC) (*Milestones: 1969–1976 - Office of the Historian*, n.d.). This embargo shed light on the world's oil dependence and resulted in a paradigm shift concerning systemic thinking. The world realized that the development and growth motives were unsustainable. However, the foundations of the new paradigm were put forward a decade later.

Sustainability and *sustainable development* terms have been framing -at least conceptually- most of the future considerations after the World Commission on Environment and Development's (known as Brundtland Commission) report in 1987 (WCED, 1987). Sustainable development was defined as “Development that meets current generations' needs without compromising future generations' ability to meet their own needs.” In 1993, the Rio Declaration was signed by 150 countries. After that, sustainability became one of the policy makers' main concerns since it was seen that the world is not limitless, and the resources are scarcer than estimated. This debate included an uncontrolled increase in population, depletion of agricultural lands, energy consumption, and global warming (Jenks et al., 2005).

Albeit the arguments and concerns seemed to be limited by an environmentalist frame when sustainability became popular, there were other salient dimensions (economic and social dimensions) to be discussed. In Figure 2.1, the three major dimensions and their sub-concepts are depicted (Khan, 1995).

Environmental sustainability concerns the earth's natural resources in every aspect. While using natural resources and disposing of the wastes, the earth's carrying capacity should be considered. Carrying capacity is related to human life directly and to other living beings; thus, protecting bio-diversity is an essential consideration of environmental sustainability. Reduced greenhouse gas emissions, reduced share of non-renewable fuels, and reduced urban sprawl are some specific goals under environmental sustainability.

Traditional economic sustainability concerns the market's capacity to develop more efficient resource allocation depending on unlimited resources. The growth will, in turn, create technologies that would be capable of renewing the used resources (Khan, 1995). However, it was understood that natural resources are not limitless. The new paradigm's economic sustainability implies a balanced growth and production system that satisfies today's consumption levels without hampering future needs. An increase in productivity and diffusing the growth, production, and productivity to underdeveloped countries while considering natural constraints are essential elements of economic sustainability. Moreover, along with the production side changes, changes in the consumption habits are required on the demand side.

Social sustainability is mainly about equity, so the distribution of wealth and reducing poverty. Social services such as education and health should be available for everyone. Gender equality, political accountability, and participation are some other critical elements of social sustainability. However, from the welfare economics perspective, these elements cannot be achieved without sustained economic growth and development. Moreover, economic growth alone does not guarantee social sustainability since it would result in more profound disparities. Economic growth without equity regulations often causes social inequality and limits the availability of benefits; besides, prolonged inequality would hinder economic growth in the long term. Therefore, economic and social sustainability must support each other. Likewise, environmental sustainability could not be realized without social and economic sustainability. Poverty reduction and sustainable economic growth are important elements of environmental sustainability. Therefore, being sustainable requires harmonious integration of these three dimensions; social sustainability, environmental sustainability, and economic sustainability, as depicted in Figure 2.1 (Basiago, 1998; Khan, 1995).

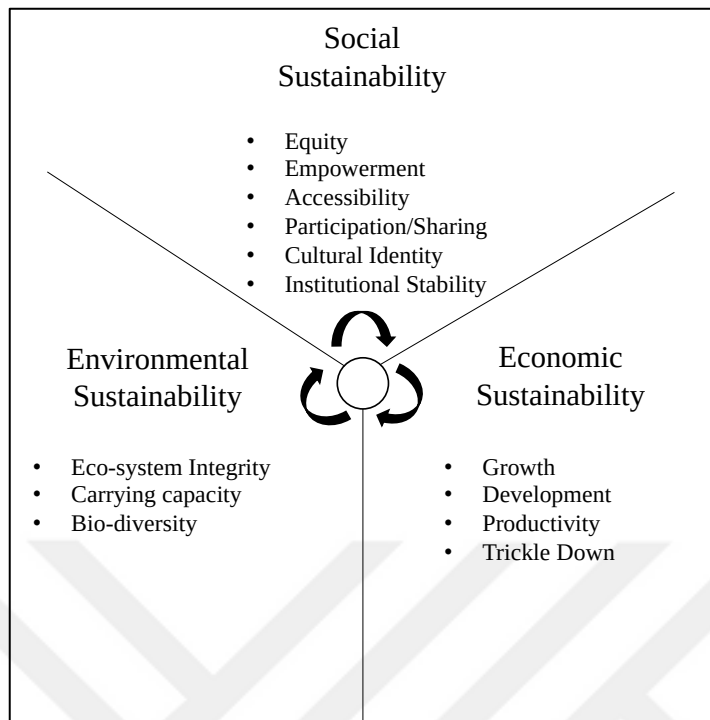


Figure 2.1. Components of Sustainability

2.2 Sustainable Development for Urban Growth: Compactness vs. Sprawl Debate

The alerting situation led to many concerns about mother nature in terms of its carrying capacity. As the long-term impact of industrialization, employment opportunities in urban areas far surpassed rural employment opportunities. This change resulted in mass migrations to the urban areas. Additionally, the advancements in medicine which were mostly available through urban healthcare systems and the end of world-wide wars, resulted in increased life expectancy, so the world population increased drastically from the 1950s onward. As depicted in Figure 2.2 (United Nations-Population Division, 2019), the world population has increased from 2.5 billion in 1950 to 7.8 billion in 2020, and as UN Population Division predicts, it will be 9.8 billion in 2050. Thus, the natural resources will have to satisfy

an additional 2 billion people while the situation is now emerging, considering global warming, climate change, and poverty levels. Currently, most people live in less developed countries, and this domination will be expected to continue. However, the number of people living in the least developed countries will surpass the number of people living in more developed countries (United Nations-Population Division, 2019).

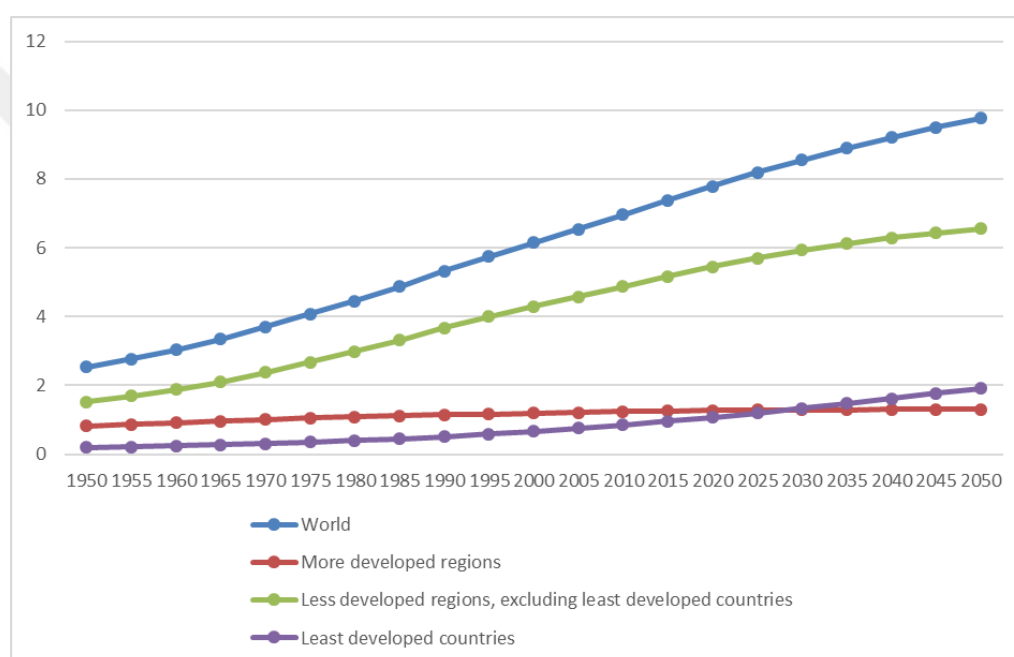


Figure 2.2. Annual Total Population at Mid-Year, 1950-2050 (billion)¹

¹-More developed regions comprise Europe, Northern America, Australia/New Zealand and Japan.

- Less developed regions comprise all regions of Africa, Asia (except Japan), Latin America and the Caribbean plus Melanesia, Micronesia and Polynesia.

- The group of least developed countries, as defined by the United Nations General Assembly in its resolutions (59/209, 59/210, 60/33, 62/97, 64/L.55, 67/L.43, 64/295 and 68/18) included 47 countries in June 2017: 33 in Africa, 9 in Asia, 4 in Oceania and one in Latin America and the Caribbean.

- Other less developed countries comprise the less developed regions excluding the least developed countries.

Moreover, there is another drastic trend depicted in Figure 2.3 (United Nations-Population Division, 2019). The rate of urbanization has an irreversible increasing trend. In 1950, the percentage of the urban population was around 29.6 % in the world. While more developed regions had 54.8 % urban population, the least developed regions had just 7.5 % urban population. In 2020, the overall rate was 56.2 %; developed regions with 79.1 % and least developed regions with 34.6 %. The estimated rates are 68.4%, 86.6%, 52.5% for the world, more developed regions, and the least developed regions. The rate of urbanization is increasing in addition to the increasing population, meaning that the world will accommodate most of its population in cities (United Nations-Population Division, 2019).

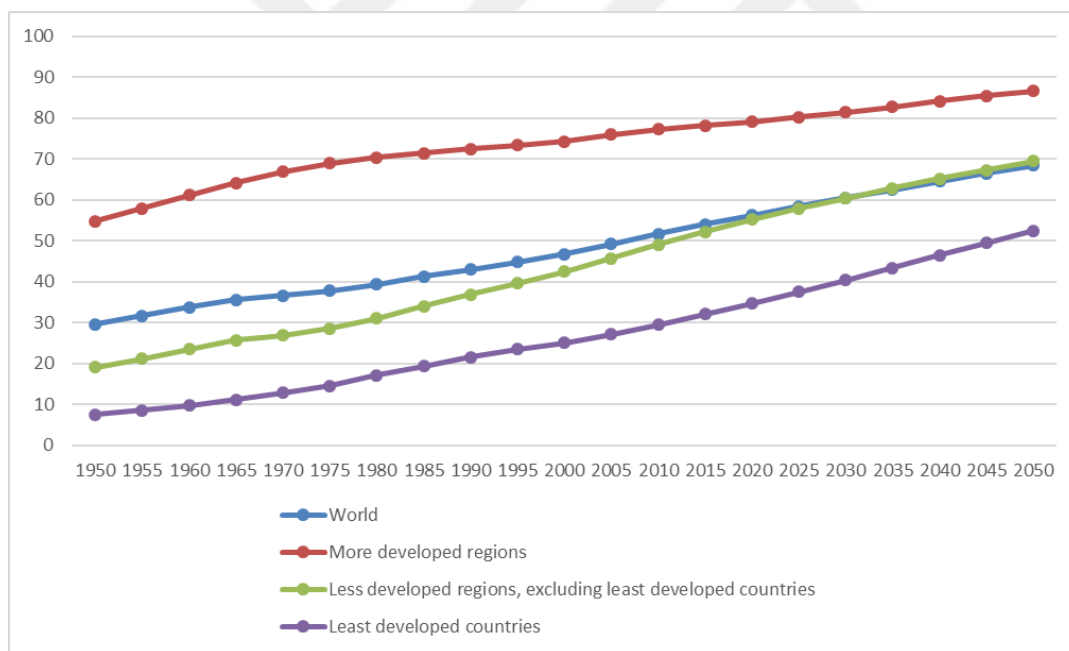


Figure 2.3. % of Population at Mid-Year Residing in Urban Areas, 1950-2050

Therefore, the role of cities in the sustainability realm is salient. According to White (1994, as cited in Jenks, Burton, & Williams, 2005):

“...although the problems are generally known, the complexity of cities, and differences in the urban experience of their inhabitants, lead to a variety of issues that make the search for effective solutions a daunting task. Nevertheless, it is a task that needs to be undertaken; a large part of the response to unsustainable development should come from the cities because that is where the most intense environmental damage is taking place, and it is there that many improvements can effectively be made”

Particular attention should be paid to the growth patterns of those complex structures since they are and will be the places to accommodate most of the world's population. There is a vast literature condensed on analyzing urban growth. The “compactness” characterizing traditional cities was the concretization of ambiguous and abstract terms of sustainability. The new concerns could be reflected in the urban space through this terminology (Çalışkan, 2004). The main elements of compact cities are defined in Çalışkan (2004) as centrality, contiguity, density, diversity, intensity; compact cities are characterized as having high density, having mixed land-uses, having relational factors close to each other (to be within walking distance), and elements to be nested.

However, in the current situation, the cities are far from being compact. For this stance, it is essential to discuss the main direction of urban growth and the resulting urban form. As private cars became affordable, the number of people who own and use automobiles in commuting increased, which led to fundamental changes in the urban development process. Low-density suburbs emerged at the fringes of the densely populated urbanized areas. These low-density extensions had their characteristic, which is total dependence on private cars. This total dependence on

private cars was again reflected in space as the further scattering of low-density residential areas. Like a vicious circle, this process resulted in the production of unsustainable environments named urban sprawl in the literature.

The term urban sprawl is ambiguous and hard to define uniquely. There are varying definitions used in the literature, mostly due to the characteristics of the studied urban structure. Some definitions could be listed as follows:

- “The low-density, residential development beyond a city’s limits (The Heritage Foundation as cited in Gillham and MacLean, 2002)”
- “Transitional period between rural and urban land use (Reason Public Policy Institution as cited in Gillham and MacLean, 2002)”
- “A pattern of growth that has largely occurred in an unplanned, ad hoc fashion (U.S. Environmental Protection Agency as cited in Gillham and MacLean, 2002)”
- “Sprawl is a form of urbanization distinguished by leapfrog patterns of development, commercial strips, low density, separated land uses, automobile dominance, and a minimum of public open space (Ewing, 2008).”

In Table 2.1, characteristics of urban sprawl defined in the literature are listed by revising the classification in Habibi and Asadi (2011).

Table 2.1 Characteristics of Urban Sprawl

Study	Characteristics
(Ewing, 1997)	Low residential density, poor residential accessibility, lack of functional open spaces
(Downs, 1999)	Unlimited outward extension of development, low-density residential and commercial settlements, leapfrog development, dominance of private automobiles, lack of centralized planning or control of land uses, widespread strip commercial development
(Pendall, 1999)	Low density; ribbon on strip, scattered, leapfrog or isolated development
(USHUD, 1999)	Both residential and non-residential low density, dominance of automobiles, unlimited outward expansion
(Brueckner, 2000)	Excessive spatial growth
(Galster et al., 2001)	A combination of low levels of density, continuity, concentration, clustering, centrality, nuclearity, mixed uses, proximity
(Kahn, 2001)	Employment decentralization
(Ludlow, 2006)	Physical pattern of low-density expansion of large urban areas into the surrounding agricultural areas
(Song & Zenou, 2006)	Scattered, and poorly planned low-density development beyond the edge of urbanized areas
(Angel et al., 2007)	Longer distances, excessive using of lands, low density, fragmentation and fingerlike extensions
(Frenkel & Ashkenazi, 2008)	Higher growth rates in the suburbs than in the center, low density, homogenous and scattered development, poor accessibility
(Habibi & Asadi, 2011)	More growth than usual, excessive nature
(Sohn et al., 2012)	Poor accessibility to urban functions

Definitions of urban sprawl could be summarized as excessive (considering the required growth of the urban built-up area with respect to the population growth), scattered (dispersed), homogenous (not having a diversity of land uses), and low-density residential development with the paucity of public open spaces and good accessibility. Alternatively, from the transportation perspective, the sprawling areas could be identified as places where public transportation is not feasible. However, there are viewpoints that favor urban sprawl.

According to Gordon and Richardson (1997), sprawl is an outcome of free-market decisions made by households and firms; thus, it maximizes the overall welfare of society. Downs (1999) argues that sprawling character provides easy access to open spaces, both in one's own home and in the country-side, and relatively short commuting times for most of those who both live and work in the suburbs. Burchell, Shad, Listokin, and Phillips (1998) state that sprawling patterns are usually associated with high-income rates; hence, they have lower crime rates than do large cities. Residents in sprawling small communities have a greater influence on government decisions than do residents in larger cities, being able to pass zoning regulations and other regulations that exclude undesirable development in their neighborhoods. According to Kahn (2008), sprawl provides affordable housing, so it creates an equal opportunity for housing across racial lines. In Table 2.2 (Gillham & MacLean, 2002), the major controversial issues are listed.

Table 2.2 Viewpoints in Sprawl Debate

Issues	Anti-sprawl	Pro-development
Land and open use	Sprawl consumes valuable, limited land resources, including farmland.	There is more than enough land and farmland left to develop.
Endangered habitat	Sprawl fragments habitats, threatening endangered species.	Wildlife is increasing, not decreasing, in suburban areas.
Traffic congestion	Auto dependent sprawl causes traffic congestion.	Dispersal of land-uses may also result in dispersal of traffic and hence less concentration (congestion) of traffic
Energy consumption	Auto dependent sprawl consumes unsustainable amounts of energy.	Auto technology is changing, and oil reserves remain adequate.
Air pollution	Increased auto travel caused by sprawl is contributing to global warming and air pollution.	Global warming is still basically unproven. Air pollution is an urban, not a suburban problem.
Water pollution	Sprawl destroys wetlands and contributes to water pollution from increased runoff.	Environmental restrictions increase housing costs and are unfair to landowners.
Public health	In addition to polluting air and water, sprawl contributes to obesity and increased stress levels.	Risks to public health are considered unproven.
Community	Suburbia is destroying community life and character (in the case of United States).	Suburbs allow plenty of opportunities for community involvement.
Aesthetics	Sprawl is devastating (America's) treasured landscapes.	More people want to live in suburbia than want to live in cities.
Economic	Sprawl costs more than compact development.	Cities are more expensive than suburbs.
Social divides	Sprawl geographically divides races and classes.	Suburbs are becoming more diverse. Opportunities are now equal.
Cities	Sprawl has drained the cities, leaving them problem ridden.	Cities are responsible for their own problems.

The authors favoring urban sprawl mainly assert that the amount of agricultural land that is lost due to sprawl is relatively small compared to the full terrain. National Association of Home Builders (as cited in Gillham and MacLean, 2002) says that the cropland in the United States in 1992 is nearly the same as in 1945; 451 million acres and 461 million acres. Additionally, traffic and energy-related negativities linked with sprawl are seen as the problems of densely settled urban areas. According to the statistics of the Heritage Foundation (as cited in Gillham and MacLean, 2002), vehicle miles traveled (VMT), fuel waste, and travel time are higher in dense urban areas and decline in less dense, more suburban environments. Also, it is argued that air quality and traffic congestion are generally worse in dense, urban areas. Moreover, increasing density could increase infrastructure costs after a certain point (Staley, 1999, as cited in Gillham and MacLean, 2002).

There are also many studies done to analyze the impacts of urban sprawl. Ewing, Schieber and Zegeer (2003), sought the relationship between urban sprawl and traffic fatalities. In the study, a sprawl index is created, and regression analysis is applied for 448 US counties in order to derive the association between urban sprawl and traffic fatalities. According to the study, urban sprawl is directly related to traffic fatalities; counties that are more compact have lower fatality rates, especially for pedestrians. Ewing, Hamidi and Grace (2016) investigated the association between urban sprawl and traffic crash rates. A sprawl index for 994 US counties is constructed. By structural equation modeling, it is found out that higher levels of vehicle crashes are observed in sprawling counties compared to compact counties. Berrigan, Tatalovich, Pickle, Ewing and Ballard-Barbash (2014) examined the possible association between urban sprawl and total- and obesity-related cancer mortality. A significant association is determined in such a way that the cancer mortality rates are higher in less sprawling areas. Ewing, Hamidi, Grace and Dennis (2016) investigated the impact of urban sprawl and compactness on social mobility for commuting zones in the US. It is found out that upward social mobility in compact areas is significantly higher than in sprawling areas.

Other than the health and demographic impacts, urban sprawl's transportation impacts are widely analyzed. Ewing, Pendall and Chen (2003) examined urban sprawl's transportation impacts for the largest US metropolitan areas. After the operationalization of sprawl by factor analysis, those factors are tried to be related to transportation outcomes such as vehicle ownership, mode choice, traffic delay per capita, and traffic fatalities per capita. According to the results, compact areas perform better than the sprawling ones for most transportation outcomes. Trowbridge and McDonald (2008) examined the relationship between urban sprawl and teenage driving. According to the results, in sprawling counties (depending on the sprawl index generated by Ewing, Pendall, et al., (2003), teenage driving is significantly higher than in compact counties. Zhao (2010) analyzed the impacts of urban expansion on transportation in Beijing. According to the findings, trip distance and car usage are increased by urban expansion on the fringe, which increases carbon emissions. García-Palomares (2010) investigated the relationship between urban sprawl and the journey to work in the metropolitan area of Madrid. By the low-density metropolitan expansion in a sprawling fashion, travel distances increase, and the mode choice changes in favor of automobiles. Holcombe and Williams (2010) examined transportation externalities that are alleged results of urban sprawl. As opposed to the findings of Ewing, Pendall, et al. (2003); Ewing, Schieber, et al. (2003) and Hamidi, Ewing, Preuss and Dodds (2015), no significant relationship is found between urban sprawl and transportation externalities: commute times, automobile ownership, fatal auto accidents, air pollution, and highway expenditures. Zolnik (2011) examined the impacts of sprawl on private-vehicle commuting outcomes (duration and length of commutes). According to the findings, sprawl is associated with shorter commuting times. Kim and Brownstone (2013) examined the relationship between residential density and vehicle usage and fuel consumption in the US.

According to the results, the relationship between residential density and vehicle usage and fuel consumption is significant; as density increases, mileage and fuel consumption decrease, but the economic value is not as significant as the relation is.

Hamidi and Ewing (2014) conducted a longitudinal analysis to validate former sprawl measures by comparing 2000 and 2010 compactness scores of the largest urbanized areas in the US. Moreover, compactness scores are associated with transportation outcomes. Compactness is found to be positively related to walking and transit mode choice and negatively related to average drive time. Hamidi and Ewing (2015) questioned one of the stated benefits of sprawl – providing affordable housing (Kahn, 2008). It is argued that this benefit of sprawl is biased because transportation costs are ignored. In this study, housing affordability is reconsidered by taking transportation costs into account for 221 metropolitan statistical areas in the US. According to the results, the total cost (housing and transportation) decreases as the compactness score increases. Hamidi et al. (2015) revisited a former study of Ewing, Pendall and Chen (2002) and refined sprawl indices for the US metropolitan areas. According to the results, compactness is positively related to walking and transit mode shares; and negatively related to average vehicle ownership per household. Lastly, in Ewing, Tian, and Lyons (2018), the relationship between urban sprawl and congestion is investigated in the US cities. In sprawling areas, travel speeds increase, but distances commuted increase, so they offset each other. Besides, compact developments are not very successful at alleviating the congestion problem.

In addition to the discussions here, city form and development densities are reevaluated amid the COVID-19 outbreak. During the pandemic, dense and highly populated megacities reported emerging situations. The most straightforward fact is that the spread of the virus could be amplified in dense areas due to people's proximity to each other. Moreover, it could be argued that private car usage would be safer than public transportation, which is extensively used in those megacities like New York. However, some recent studies put forward contradicting results (Hamidi et al., 2020; Hamidi & Zandiatashbar, 2021). Although compact settlements had the stated disadvantages, it is easier to reach healthcare units or any facility in denser cities. It is argued that sprawling cities had worse conditions due to spatial inequality in terms of accessing health services (Zhao et al., 2020).

Moreover, the unhealthy lifestyles due to car usage (less activity and walking) are correlated to the virus's impacts. Therefore, denser areas are less vulnerable in this stance (Hamidi et al., 2020). Though the relation is not conclusive yet since there would be many other factors affecting the spread of the pandemic, the evidence favors denser settlements, at least in terms of accessibility to health care facilities.

With all the pros and cons, the sprawl issue is a hot debate in the literature since the resulting environments are recently evaluated in terms of sustainability. Sustainability concerning climate change was hitherto debated on the grounds of the reality of climate change, but it is most commonly accepted as the earth has to face. In line with these, recently escalated discussions on the size and form of cities' growth seemed to prove the compact city arguments that are in the opposite direction with urban sprawl on the scale of urban form. Therefore, while having sprawling settlements on hand, the proposals for compact cities and sustainable transportation modes would result in significant changes in how the cities operate. Thus, planning for the cities according to the emerging facts and policies requires a better understanding of the processes and the analysis of the cities in detail. While sprawl versus compact city discussion could be considered as a larger (even in regional) scale debate, the impacts of those on the built environment could be better analyzed at lower scales such as neighborhoods or population blocks (in the U.S. case).

2.2.1 Urban built environment

As given in the previous section, cities' growth pattern (compact or sprawling) is well suited in the sustainability debate. Albeit urban growth patterns could be conceptualized at smaller scales such as neighborhoods, the discussion about the growth patterns (and city form) is generally regional. While considering community levels or doing smaller scale and localized analyses, urban areas are usually referred to as the built environment.

2.2.1.1 Dimensions of the built environment

In order to analyze the built environment, it should be operationalized. Cervero and Kockelman (1997) firstly mentioned Ds of the built environment as density, design, and diversity to define the built environment. These three Ds are expanded to five Ds with the addition of destination accessibility and distance to transit (Ewing & Cervero, 2010). In addition to these five main dimensions/characteristics of the built environment, two more dimensions could be used as built environment characteristics and better formulate the built environment analyses. One is demographic and socioeconomic properties, and the other is travel demand management. Ewing and Cervero (2010) proposed these two dimensions as the sixth and seventh Ds. (Note: these are the built environment dimensions/characteristics put forward by the authors favoring compact cities rather than sprawling ones; however, wide usage of these dimensions/characteristics in the literature enabled them to be the objective basis). Albeit those dimensions/characteristics are not strictly separated, they could be defined as follows:

- **Density:** It is the most commonly used variable describing the built environment (Merlin et al., 2020). It is the quantity of the variable of interest (e.g., population, employment, or dwelling units) per unit of area, which can be gross or net (Ewing & Cervero, 2010). Household and employment densities (P. Chen & Zhou, 2016), population, jobs, and housing densities (Ouyang & Bejleri, 2014), and population, household and employment densities (Lord & Mannering, 2010) are used as built environment indicators.
- **Diversity:** The measures under diversity depend on the number of varying land uses in a specific area and the extent to which they are represented in land area, floor area, or employment. In many studies, diversity is measured by entropy, where lower values indicate single (homogenous) uses of land, and higher values indicate varied (heterogenous/mixed) uses of land (Ewing & Cervero, 2010). In P. Chen and Shen (2016, 2019); Ouyang and Bejleri (2014), land use mix (entropy of different land uses) whereas proportion of

varying land uses such as industrial land use, open land use, and commercial land use proportions are used as indicators in Kim and Yamashita (2002), and different land uses are used to distinguish crash frequencies in Kim, Pant and Yamashita (2010).

- **Design:** It is related to street network features in an area. Some of the measures could be listed as average block size, the ratio of four-way (or three-way) intersections, and the density of intersections (e.g., the number of intersections per square mile) (Ewing & Cervero, 2010). For example, Kim and Yamashita (2002) uses number of all-way stop intersections, signalized intersections, number of three-way intersections, Ouyang and Bejleri (2014) uses street length, number of street segments, and number of intersections, and C.-Y. Yu (2014) uses number of intersections and number of dead-ends as independent variables. Depending on the focus of the study, it can also be measured using sidewalk coverage, average street widths, or the number of pedestrian crossings, other physical elements that distinguish pedestrian-focused environments from auto-focused environments (Ewing & Cervero, 2010). For example, in Milton and Mannering (1998), along with roadway and intersection densities, sidewalk completeness, and bike lane completeness are used as built environment indicators.
- **Distance to transit:** It is the ease of accessing public transportation alternatives from home or work (e.g., bus stop or rail station within a quarter to half a mile of trip origin). It can be measured as the mean of the shortest street routes from the residences or workplaces in an area to the nearest bus stop or rail station. Likewise, it can be measured by transit route density (e.g., the distance between transit stops, or the number of stops per unit area) (Ewing & Cervero, 2010). The number of bus stops in C.-Y. Yu (2014), the number of bus stops and distance to bus stops in Ouyang and Bejleri (2014), and the number of bus stops and the number of subway stations in Kim and Yamashita (2002) are used as the indicators of the built environment.

- ***Destination accessibility:*** It is the measure of the ease of accessibility to trip destinations/attraction points (central business district or other centers). It is often measured at the zonal level in terms of distance or time. If it is measured at the regional level, it is the travel distance or time to the central business district (Ewing & Cervero, 2010). In Khattak, Wang and Zhang (2010), distance to school, distance to supercenter and distance to bridge/tunnel are used, whereas in Ouyang and Bejleri (2014), distance to commercial area is used.
- ***Demographic and socioeconomic properties:*** Those are the indicators related to demographics such as population, age, gender, household size, and socioeconomic statuses such as the number of people employed in a household, median income, or educational attainment. These indicators are not directly part of the built environment but are related to the built environment since they reflect the composition of the population living or working at a geographical unit. Many studies analyzing the built environment considered those variables (at least for controlling demographic and socioeconomic influences). For example, Hadayeghi, Shalaby, and Persaud (2003) used total population, the number of households, full-time employed people, part-time employed people, and the number of vehicles; Zhibin Li, Wang, Liu, Bigham and Ragland (2013) used the percentage of the male population, percentage of under 18 population, percentage of 65 and over age population, and percent of the population below the poverty line.
- ***Travel demand management:*** As discussed previously, supplying more infrastructure to match transportation demand is both not feasible and sustainable. Therefore, the supply-side should be constrained concerning increasing travel demand. The indicators under this dimension could be related to the supply-side such as parking supply (Z. Guo, 2013) or demand-side such as trip frequency or vehicle miles traveled (Ewing & Cervero, 2010).

2.2.2 Urban transportation

Transportation (international, national, and urban), as being one of the most polluting sectors along with energy production, industry, and housing, also takes part in the sustainability discussion. Sustainable transportation is one of the basics of achieving sustainable development (Sustainable Mobility for All, 2017). According to EPA (EPA, n.d.), based on 2010 global emission data, transportation had a significant share with 14% (electricity and heat production had 25% share).

Not only the increase in the international and national transportation (freight and passenger) volumes, urban transportation volumes are also significant, primarily private cars as an unsustainable mode of transport. Moreover, 72% of transportation emissions are related to road transportation (*AR5 Climate Change 2014: Mitigation of Climate Change — IPCC*, n.d.). Increased car ownership (initially was seen as an advantage for congested cities to alleviate traffic problems) became a significant problem. Cars are the most polluting vehicles considering per capita emissions. Moreover, cities dispersed in an uncontrolled manner with the mobility potential of automobiles. Mobility is an essential element that should not be reduced (considering equity and accessibility components of social sustainability); however, the benefits must be considered without neglecting the adverse impacts. Increased mobility by cars would, in turn, yield increased traffic congestion, pollution, and urban sprawl. Considering the environmental and traffic-related adverse impacts of transportation, satisfying the mobility needs of all to achieve social sustainability is challenging but fundamental to achieve sustainable development.

Reducing greenhouse gas emissions of transportation, reducing externalities of transportation (such as congestion, accidents, environmental impacts), using clean vehicles and alternative fuels (such as electric cars), increasing multimodal transportation (to integrate a greater percentage of sustainable modes) are the goals of European Commission to achieve sustainability in transportation (European Commission, n.d.). Moreover, Eurostat provides indicators to operationalize and

measure Sustainable Development Goals; which has transportation-related ones such as (Gerike & Koszowski, 2017; *Indicators - Eurostat*, n.d.):

- “energy consumption of transport relative to GDP”
- “modal split of freight and passenger transport”
- “volume of freight and passenger transport relative to GDP”
- “greenhouse gas emissions from transport”
- “people killed in road accidents”
- “average CO₂-emissions per kilometer from new passenger cars”

Therefore, reducing travel demand for unsustainable modes and reducing road accidents are essential factors for sustainable transportation and sustainable development. Operational indicators of travel demand and traffic safety will be given next.

2.2.2.1 Travel demand

As discussed previously, the increase in car ownership has changed how the cities developed and travel behavior (modal choice in favor of private cars). Sprawling cities had to provide more infrastructure to deal with increasing demand; resulting problems such as air pollution and congestion took considerable attention that states and localities have been trying to implement planning and design principles to reduce car dependence. It was impossible to match automobiles' travel demand by supplying more roads, and curbing the demand to match supply was the plausible alternative (Ewing & Cervero, 2001). Constraining the demand rather than supplying more roads is in line with the sustainability principles, e.g., reducing emissions, modal split in favor of sustainable modes (public transportation, biking, walking), and increasing multimodal transportation to integrate those sustainable modes. Ergo, the focus here is the travel demand of automobile users, and the aim is to define alternatives to reduce this travel demand. According to Ewing and Cervero (2010),

some indicators used to operationalize travel demand (travel) for analyses are vehicle miles traveled (VMT), vehicle hours traveled (VHT), mode choice, trip frequency, and trip length. Moreover, annual average daily traffic (AADT), person miles traveled, average delay, and traffic volume are other indicators used to measure travel demand (Federal Highway Administration, n.d.-a).

2.2.2.2 Traffic safety

Each year, 1.35 million people are killed on roadways around the world, and about 3,700 people are killed in traffic accidents every day. Besides, road traffic injuries are estimated to be the eighth leading cause of death (see Table 2.3) for all age segments and number one for the children and younger adults between the ages 5-29 in 2016 (WHO, 2018).

Table 2.3 Leading causes of death, all ages, 2016 (WHO, 2018)

Rank	Cause	% of total deaths
1	Ischaemic heart disease	16.6
2	Stroke	10.2
3	Chronic obstructive pulmonary disease	5.4
4	Lower respiratory infections	5.2
5	Alzheimer's disease and other dementias	3.5
6	Trachea, bronchus, lung cancers	3.0
7	Diabetes mellitus	2.8
8	Road traffic injuries	2.5
9	Diarrhoeal diseases	2.4
10	Tuberculosis	2.3

Moreover, traffic accidents and fatalities have substantial economic costs. For example, according to Blincoe, Miller, Zaloshnja and Lawrence (2015), the total

economic cost of roadway crashes in the U.S. is about \$242 billion in 2010, 1.6% of GDP (32,999 fatalities, which is 34,439 in 2016). As being one of the primary sources of human deaths, traffic accidents, a higher percentage (53%) of those accidents occur in urban areas globally (WHO, 2018). The U.S. has a pattern similar to the global one, as in 2017, most of the fatalities occurred on urban roads. In 2017, 44% of deaths occurred on urban roads, while 40% on rural roads and 15% on highways/motorways. According to the International Transport Forum (ITF, 2018), rural roads were the deadliest in the U.S. case. However, this pattern has changed as urban vehicle miles traveled increased by 13%, whereas rural vehicle miles traveled decreased by 3% since 2008 (Highway Traffic Safety Administration & Department of Transportation, 2019). Thus, urban road casualties should be considered in particular. For analyzing traffic safety, there are classifications of accident severity. For the United Kingdom case, accidents are reported as “killed,” “serious injury,” or “slight injury” (Department of Transport, 2012); for the United States case, classifications vary, but the common ones are “fatal,” “incapacitating injury,” “nonincapacitating injury,” “possible injury,” and “no injury” (Federal Highway Administration, n.d.-b).

2.2.3 Background on the relationships between travel, built environment, and traffic safety

This section reviews the studies focusing on the relationships between the urban built environment and urban transportation. Traffic safety and travel demand are the two separate urban transportation topics associated with the urban built environment. Moreover, studies examining the relationship between travel demand and traffic safety are also reviewed.

2.2.3.1 Built environment and travel

In this section, the studies investigating the relationship between the built environment and travel demand are presented (see Table 2.4). As the discussion mainly originated from the developed country-literature, most reviewed articles studied developed country cases (most from the United States). Some studies are from developing countries such as China (Guan & Wang, 2019; L. Yu, Xie, & Chan, 2019), Chile (Zegras, 2010), and Vietnam (Chi Nguyen et al., 2020). Among the reviewed articles, most studies directly mention the term “built environment”; however, some other studies embedded built environment in growth scenarios (as discussed in previous sections) (Bartholomew & Ewing, 2008), land use typology (Boarnet, Nesamani, & Smith, 2003), urban form (Bento, Cropper, Mobarak, & Vinha, 2005), or neighborhood characteristics (X. Cao, Mokhtarian, & Handy, 2007). For this review, it is not convenient to directly mention all the variables because there are numerous variations of each dimension's indicators. Thus, if variables falling into a dimension/category exist in a study, an “X” is used and left blank otherwise. For example, population density, employment density, and housing density are different density variables. The critical information provided in Table 2.4 is that whether any density variables are used or not in a study. So, either one, two, or three variables are used; they are reflected by the same symbol “X.” The most common dimension is demographic and socioeconomic, almost all studies employed demographic and socioeconomic variables in their analysis. Diversity and density are the other two common dimensions. Transit/distance to transit is the least employed dimension.

It could be seen from Table 2.4 that all of the studies have variable(s) of the “Demographic and Socioeconomic” category. Moreover, most studies focused on individual/household level analyses; that is to say, they investigated the relationship between link/zonal/regional level-built environment factors and person/household level travel outcomes (VMT, number of trips, mode choice). Some studies structured their analyses at the community level, such as traffic zones (Choi, 2018) or at the

regional level (Bartholomew & Ewing, 2008). From the methodological perspective, varying approaches were employed, e.g., negative binomial regression (Chatman, 2008, 2009; Shay et al., 2006; Shay & Khattak, 2005), structural equation modeling (Bagley & Mokhtarian, 2002; X. Cao et al., 2007; L. Yu et al., 2019), multilevel regression (Park et al., 2018), logistic regression (Boer et al., 2007; Gehrke & Wang, 2020), and geographically weighted regression (Nkeki & Asikhia, 2019; Nowrouzian & Srinivasan, 2013).

The relationship between travel and the built environment was extensively reviewed by Ewing and Cervero (2001, 2010). Their findings supported that each variable was not elastic enough to imply a significant change in the VMT, with the highest absolute weighted average elasticity, which was 0.39. Though the sole elasticities were not high, combined effects would make a difference (Ewing & Cervero, 2010). This finding would be a direction for integrated transportation and urban planning perspective. Moreover, the main finding of Salon (2015) was that the relations between the built environment dimensions and VMT depend on neighborhood type (e.g., central city, urban, suburb, and rural) and the purpose of travel (e.g., non-work VMT, one-way commute VMT). Thus, the patterns should be analyzed within their contexts; and the geographic unit would play an important role in the analysis considering that the majority of the studies are based on the individual/household scale.

Table 2.4 Review on Built Environment and Travel

Study	Case Study Location	Aim	Dependent Variable(s)	Unit of Analysis
(Bagley & Mokhtarian, 2002)	UNITED STATES / CALIFORNIA / San Francisco	Examining the relationship between residential neighborhood type and travel behavior	VMT	Individual
(Bartholomew & Ewing, 2008)	UNITED STATES	Determining how much compact growth scenarios are estimated to reduce travel below current trends	Percentage of difference in VMT between the planning scenario and the trend scenario	Region Scenario (scenarios nested within regions)
(Bento et al., 2005)	UNITED STATES / 114 Urban areas	Examining the effects of urban form and public transit supply on commute mode choices and annual VMTs of households	Mode choice Vehicle choice Driving decision VMT	Urban area
(Boarnet & Wang, 2019)	UNITED STATES / CALIFORNIA / Los Angeles, Orange, Ventura, San Bernardino, Riverside counties	Examining the relationship between urban spatial structure and driving	Household VMT	Household
(Boarnet et al., 2008)	UNITED STATES / OREGON / Portland	Examining the health benefits from urban design characteristics that are related to increased walking	Walking distance	Individual
(Boarnet et al., 2003)	UNITED STATES / OREGON / Portland	Examining whether land use matters more for trip generation or vehicle miles traveled (VMT) for an individual	Number of nonwork trips Total nonwork VMT	Individual
(Boer et al., 2007)	UNITED STATES	Investigating to whether design guidelines stimulate walking	Odds ratio to walk	Household
(X. (Jason) Cao et al., 2009)	UNITED STATES / CALIFORNIA	Investigating the relation between the residential environment and nonwork travel frequencies by auto, transit, and walk/bicycle modes	Frequency of home-based nonwork trip by auto by walking/biking transit	Individual
(X. Cao et al., 2007)	UNITED STATES / CALIFORNIA	Examining whether changes in neighborhood characteristics trigger changes in travel behavior	Change in auto ownership Change in driving behavior Change in accessibility Change in spaciousness	Individual

Table 2.4 (continued)

(Cervero, 2002)	UNITED STATES / MARYLAND / Montgomery county	Examining mode choice around a model that considers the influences of built environments factors and generalized cost related factors and travelers' socioeconomic attributes	Mode choice	Individual
(Cervero & Duncan, 2006)	UNITED STATES / CALIFORNIA	Examining whether jobs-housing balance or locating retail and services closer to residential areas leads greater travel-reduction	Daily work tour VMT Daily work tour VHT Daily shopping services VMT Daily shopping services VHT	Household
(Cervero & Kockelman, 1997)	UNITED STATES / CALIFORNIA / San Francisco	Examining the effects of 3Ds' on trip rates and mode choice	Daily personal VMT Proportion of trips	Individual
(Cervero & Murakami, 2010)	UNITED STATES / 370 Urban areas	Examining the effects of built environments on VMT	VMT per capita	Urban area
(Chatman, 2003)	UNITED STATES	Examining how density and mixed uses of workplace impact personal commercial travel and commuting mode choice	VMT on personal commercial trips	Individual
(Chatman, 2008)	UNITED STATES / CALIFORNIA / Alameda, San Francisco, Santa Clara counties	Examining how density impacts travel	Auto speed Non-work activities VMT	Individual
(Chatman, 2009)	UNITED STATES / CALIFORNIA / San Diego, San Francisco, Oakland, San Jose	Examining the impact of built environment on nonwork travel	Number of work trips by auto by transit by walk/bike	Household
(Chi Nguyen et al., 2020)	VIETNAM / HANOI	Examining the relation between the built environment and travel behavior	Travel mode choice	Individual
(Choi, 2018)	CANADA / CALGARY	Investigating how land use and transportation factors influence household vehicle travel	Household daily auto travel distance Transit travel distance Active (walking, cycling) travel distance	TZ
(Chowdhury & Scott, 2020)	CANADA / HALIFAX	Examining whether the built environment impacts total distance traveled by auto	Worker personal vehicle kilometers traveled (PKT) Non-worker personal vehicle kilometers traveled (PKT)	Individual

Table 2.4 (continued)

(Diao & Ferreira, 2014)	UNITED STATES / MASSACHUSETTS / Boston	Examining the link between household vehicle use and their residential locations	VMT per vehicle VMT per household VMT per capita	Grid
(Gehrke & Wang, 2020)	UNITED STATES / OREGON / Portland	Examining the difference in the relationship between travel and the built environment while measuring the land development model, urban design, and transportation system features at various scales	Mode choice	BG CT Grid
(Greenwald & Boarnet, 2001)	UNITED STATES / OREGON / Portland	Examining nonwork pedestrian travel regarding the built environment as decisive	Number of nonwork trips	Individual
(Guan & Wang, 2019)	CHINA / Beijing	Examining the built environment impacts on travel with a household-based viewpoint	Share of travel by car Travel time by non-motorized modes	Household
(Guzman et al., 2020)	COLOMBIA / BOGOTA	Examining the part of the built environment and sociodemographic characteristics on walking distances	Travel distance	Household
(Khattak & Rodriguez, 2005)	UNITED STATES / NORTH CAROLINA	Investigating whether the residents of neo-traditional neighborhoods replace walking for driving trips or do they make more trips overall	Household trips in a day Trip distance Trip duration	Household
(Kitamura et al., 1997)	UNITED STATES / CALIFORNIA	Examining the effects of land-use and attitudinal features on travel behavior	Total number of person trips Number of transit trips Number of non-motorized trips	Individual
(Nkeki & Asikhia, 2019)	NIGERIA / BENIN	Analyzing the spatial difference in the interactions between travel behavior and built environment	Travel mode choice	Neighborhood
(Nowrouzian & Srinivasan, 2013)	UNITED STATES / FLORIDA / Broward, Palm Beach, Miami-Dade counties	Evaluating the effect of land-use on person miles traveled (PMT)	Person miles traveled (PMT)	Individual
(Park et al., 2018)	UNITED STATES / 8 Metropolitan areas	Considering the role of neighborhood-level built environment characteristics of rail station areas by examining several travel outcomes	Auto trips VMT Transit trips Walk trips	Region Station Area Household

Table 2.4 (continued)

(Rodríguez & Joo, 2004)	UNITED STATES / NORTH CAROLINA / Chapel Hill, Carrboro	Examining the relation between the local physical environment and non-motorized mode choice	Mode choice	Individual
(Shay & Khattak, 2005)	UNITED STATES / NORTH CAROLINA / Chapel Hill, Carrboro	Exploring the distinctions in automobile ownership and use between conventional and neotraditional neighborhoods	Number of trips Travel time Travel distance	Household
(Shay et al., 2006)	UNITED STATES / NORTH CAROLINA / Chapel Hill	Analyzing utilitarian trip-making behavior within a neo-traditional neighborhood and comparing total trips to mode-specific trips	Number of total trips	Individual
(Thao & Ohnmacht, 2019)	SWITZERLAND	Examining the impacts of the built environment on travel behavior	Distance traveled Number of trips	Household
(Vance & Hedel, 2007)	GERMANY	Examining to what extent does urban form influence the dependence on auto travel	Car usage Travel distance	Individual
(L. Yu et al., 2019)	CHINA / SHENZHEN	Exploring the impacts of the built environment on transit travel	Transit mode choice Travel distance Travel time	Neighborhood unit Individual
(Zegras, 2010)	CHILE / SANTIAGO	Analyzing the relationships between the built environment and motor vehicle ownership and use	Motor vehicle choice VKT	Household

Table 2.4 (Continued)

Study	Independent Variable(s)						Method / Modeling Approach
	Demographics	Density	Diversity	Design	Transit	Destination	
(Bagley & Mokhtarian, 2002)	X						Structural Equation Modeling (SEM)
(Bartholomew & Ewing, 2008)	X	X	X				Multilevel Meta-Analysis
(Bento et al., 2005)	X	X	X		X		Multinomial Logit Model OLS Model
(Boarnet & Wang, 2019)	X	X					Two-stage Tobit Model
(Boarnet et al., 2008)	X	X	X	X	X	X	Tobit Regression

Table 2.4 (continued)

(Boarnet et al., 2003)	X	X	X	X	X	X	Ordered Probit Model Ordinary Least Squares Regression
(Boer et al., 2007)		X	X	X			Logistic Regression
(X. (Jason) Cao et al., 2009)	X		X	X		X	Seemingly Unrelated Regression (SURE)
(X. Cao et al., 2007)	X		X			X	Structural Equations Modeling (SEM)
(Cervero, 2002)	X	X	X			X	Logit Models
(Cervero & Duncan, 2006)	X		X			X	Log-linear Regression
(Cervero & Kockelman, 1997)	X	X	X	X	X	X	Multiple Regression Binomial Logit Model Factor Analysis
(Cervero & Murakami, 2010)	X	X	X	X			Structural Equation Modeling (SEM)
(Chatman, 2003)	X	X	X			X	Joint Logit Tobit Model
(Chatman, 2008)	X	X	X	X	X	X	Negative Binomial Regression Tobit Regression
(Chatman, 2009)	X		X	X	X	X	Negative Binomial Regression
(Chi Nguyen et al., 2020)	X	X	X		X		Multinomial Logit Regression
(Choi, 2018)	X	X	X	X	X	X	Segmented Regression
(Chowdhury & Scott, 2020)	X		X	X		X	Ordinary Least Squares Spatial Lag Model
(Diao & Ferreira, 2014)	X	X	X	X	X	X	Factor analysis Ordinary Least Squares (OLS) Regression
(Gehrke & Wang, 2020)		X	X	X			Multinomial Logistic Regression
(Greenwald & Boarnet, 2001)	X	X	X			X	Ordered Probit Model
(Guan & Wang, 2019)	X				X	X	Principal Component Analysis Factor Analysis
(Guzman et al., 2020)	X	X	X	X	X	X	Log-linear Model
(Khattak & Rodriguez, 2005)	X			X			Logit Model Negative Binomial Regression Linear Regression
(Kitamura et al., 1997)	X		X	X		X	Linear Regression
(Nkeki & Asikhia, 2019)	X	X	X	X	X	X	Factor Analysis Global Logistic Regression (GLR) Geographically Weighted Logistic Regression (GWLRL)
(Nowrouzian & Srinivasan, 2013)	X	X	X	X		X	Linear Regression Geographically Weighted Regression (GWR)
(Park et al., 2018)	X	X	X	X	X	X	Multilevel Modeling
(Rodríguez & Joo, 2004)	X	X		X			Logit Model Nested Logit Model Heteroscedastic Extreme Value (HEV) Model
(Shay & Khattak, 2005)	X			X			Negative Binomial Regression OLS Regression
(Shay et al., 2006)	X		X			X	Negative Binomial Regression
(Thao & Ohnmacht, 2019)	X	X	X		X	X	Linear Regression
(Vance & Hedel, 2007)	X	X	X		X	X	Probit Model OLS Regression
(L. Yu et al., 2019)	X	X	X		X	X	Structural Equation Modeling (SEM)
(Zegras, 2010)	X	X		X	X	X	Multinomial Logit Model OLS Regression

2.2.3.2 Travel and traffic safety

This section presents a review of the studies investigating the relationship between travel and traffic safety. Those studies examined this relationship by assuming traffic safety and travel variables as dependent and independent variables, respectively. The former reviewing structures presented in the last two sections are also valid for this section (Table 2.5). The crash categories are TC (total crashes), V (single-vehicle crashes), V-V (multiple-vehicle crashes), MC (motorcycle crashes), P-V (pedestrian-vehicle crashes), and B-V (bicycle-vehicle crashes). Crash severity, crash count, and crash rate properties are also presented in Table 2.5.

Most of the studies analyzed the US cases (M. A. Abdel-Aty & Radwan, 2000; C. Chen & Xie, 2016; Clark & Cushing, 2004; Jovanis & Chang, 1986; Liu & Sharma, 2017; Stimpson et al., 2014; Yeo et al., 2015; Zheng, 2012), while there are studies with cases from European Countries (Romania; Cadar, Boitor, & Dumitrescu, 2017, France; Martin, 2002), Spain; Santamariña-Rubio, Pérez, Olabarria, & Novoa, 2014, and Czech Republic; Šliupas, 2009). The analysis unit varies from the intersection (C. Chen & Xie, 2016) to state (Clark & Cushing, 2004).

The modeling approach also varies from negative binomial regression (Šliupas, 2009) to structural equation modeling (Stimpson, Wilson, Araz, & Pagan, 2014). The number of studies analyzing the relationship between travel and traffic safety is less than the number of studies analyzing the relationship between the built environment and travel and the built environment and traffic safety.

Though there are exceptions (for example; Dong et al., 2014, found a negative relation between DVMT and total crashes and explained this as higher DVMT would be related to higher traffic density which would result in lower speeds and so lower crash rates.), there is a generalizable relation: as the travel amount increases, the number of crashes would increase since travel is stated as an exposure factor in the literature, and the increase in exposure increases the crash risk (Merlin et al., 2020). Ewing and Dumbaugh (2009) put the relation in a different fashion. Compact

developments are dense and require less VMT per capita. The increased density results in congestion and lower speeds. Property damage only accidents would be more in number in those settings, but fatal crashes would be lower than the sprawled settings where the speeds are higher than compact settlements. Moreover, the strength of the relationship between VMT and traffic safety would vary by location (Liu & Sharma, 2017). Therefore, analyzing the impacts of location-specific factors such as the built environment characteristics (as in the previous section) would better explain the variations in crash occurrences.



Table 2.5 Review on Travel and Traffic Safety

Study	Case Study Location	Aim	Unit of Analysis
(M. A. Abdel-Aty & Radwan, 2000)	UNITED STATES / FLORIDA	Modeling the frequency of accident occurrence and involvement	Roadway segment
(Cadaru et al., 2017)	ROMANIA	Examining the impact of traffic volumes on accident occurrence	Road
(C. Chen & Xie, 2016)	UNITED STATES / MASSACHUSETTS	Modeling the effects of annual average daily traffic (AADT) on predicting multiple-vehicle crashes	Intersection
(Clark & Cushing, 2004)	UNITED STATES / 13 States	Analyzing the effect of population density on the motor vehicle mortality rates in rural and urban areas	State
(Jovanis & Chang, 1986)	UNITED STATES / INDIANA	Modeling the relation of accidents to vehicle miles traveled	Roadway segment
(Liu & Sharma, 2017)	UNITED STATES / IOWA	Examining spatial and temporal impacts in traffic crash trend analysis	County
(Martin, 2002)	FRANCE	Examining the relationship between crash rates and hourly traffic volume	Motorway section
(Santamaría-Rubio et al., 2014)	SPAIN / CATALONIA	Examining gender differences in the rate of road traffic injury	Region
(Šenk & Ambros, 2011)	CZECH REPUBLIC	Quantification of the safety performance of roundabouts in terms of accident frequency considering their traffic and geometric characteristics	Roundabout
(Stimpson et al., 2014)	UNITED STATES / 100 Urban areas	Investigating whether an increasing share of mass transit use, compared to vehicle miles traveled on public roads, did associate with decreased motor vehicle fatalities	Urban area
(Yeo et al., 2015)	UNITED STATES / 147 Urbanized areas	Examining the effects of urban sprawl and VMT on traffic fatalities	Urban area
(Zheng, 2012)	UNITED STATES / CALIFORNIA	Examining the relation between traffic conditions and the crash occurrence possibility	Data collection point

Table 2.5 (continued)

Study	Dependent Variable(s)				Independent Variable(s) for Travel	Method / Modeling Approach
	Crash Type	Crash Severity	Crash Count	Crash Rate		
(M. A. Abdel-Aty & Radwan, 2000)	TC		XX		AADT	Negative Binomial Modeling
(Cadar et al., 2017)	TC		XX		AADT	Linear Regression Crash Prediction Models (CPM)
(C. Chen & Xie, 2016)	V-V		XX		AADT	Generalized Additive Models (GAMs) Piecewise Linear Negative Binomial (PLNB) Regression Models
(Clark & Cushing, 2004)	TC	X		XX Crash/pop.	VMT	Linear Regression
(Jovanis & Chang, 1986)	TC		XX		VMT	Poisson Regression
(Liu & Sharma, 2017)	TC	X	XX		VMT	Bayesian Spatio-temporal Models with Poisson and Zero Inflated Poisson Models
(Martin, 2002)	TC		XX	XX Crash/Road length	Hourly traffic flow	Poisson Regression Negative Binomial Regression Logistic Regression
(Santamariña-Rubio et al., 2014)	TC	X	XX		Person hours traveled	Poisson Regression
(Šenk & Ambros, 2011)	TC		XX		AADT	Negative Binomial Regression
(Stimpson et al., 2014)	V-V		XX		Share of annual mass transit miles traveled per capita	Structural Equation Modeling (SEM)
(Yeo et al., 2015)	TC	X	XX		VMT	Path Analysis
(Zheng, 2012)	TC			X Crash occurrence: 1 O/W: 0	Speed, Congestion	Binary Logistic Model

2.2.3.3 Built environment and traffic safety

There is considerable literature examining the spatial relationships of traffic safety. Many studies in the literature either focused on or referred to the relationship between traffic safety and the urban built environment since many accidents happen in urban areas. Those studies could be classified based on three main dimensions:

- **the spatial unit of analysis**, categories are individual spot level (e.g., road intersection or segment), zonal level (e.g., neighborhood, block group, traffic analysis zone or census tract), or regional level (e.g., urban cluster, county or metropolitan area) (Ouyang & Bejleri, 2014; Ziakopoulos & Yannis, 2020);
- **the methodology applied** (e.g., Poisson regression or Bayesian modeling) (Lord & Mannering, 2010; Ziakopoulos & Yannis, 2020), and
- **the built environment characteristics used** (those measures could be single such as density, land uses, or street network configuration or a composite such as urban sprawl) (Merlin et al., 2020).

The literature review organization was adapted from the classification scheme presented in Ziakopoulos and Yannis (2020), in which the spatial unit of analysis is the primary dimension of classification. Thus, the studies were clustered into three major groups: intersection/link level studies, zonal (community) level studies, and regional level studies. The review tables in this section were also based on this classification:

- Table 2.6: intersection/link level studies
- Table 2.7: zone level studies
- Table 2.8: regional level studies

In all of the following review tables, standard fields are used. Since this section is dedicated to reviewing the relationship between the built environment and traffic safety, the primary dependent variable of interest is “traffic crashes.” The

“Dependent Variable(s)” section consisted of “Crash Type,” “Crash Severity,” “Crash Count,” and “Crash Rate” subclasses (this typology is also offered by Ziakopoulos & Yannis, 2020). The “crash type” section defined the crash participants in the reviewed studies. The alternatives are as follows:

- TC refers to total crashes (all types of crashes)
- V refers to single-vehicle crashes
- V-V refers to multiple-vehicle crashes
- MC refers to motorcycle crashes
- P-V refers to pedestrian-vehicle crashes
- B-V refers to bicycle-vehicle crashes

The “crash severity” section provides information about the dependent variable in terms of severity level. If the crash severity (e.g., fatal, incapacitating, nonincapacitating, and property damage only) was estimated, it is represented by “XX.” (e.g., the outcome of the analysis is the crash severity; Barua, El-Basyouny, & Islam, 2014). However, if used as a filter, it is represented by “X” (e.g., the dependent variable is fatal and injury total crashes; Kaygisiz, Senbil, & Yildiz, 2017). The same notation is valid for the “crash count” and “crash rate” sections. If crash counts/frequencies or crash rates were estimated in a study, “XX” is used, whereas “X” represents a filtering process in these categories.

The independent variables of those studies are given under the built environment dimensions/characteristics section within “Demographic and Socioeconomic,” “Density,” “Diversity,” “Design,” “Distance to Transit,” “Destination Accessibility,” and “Travel” categories.

Intersection/link level studies analyze the crash occurrences at the junctions or road segments, which requires a fewer amount of aggregation of crashes than zonal and regional studies. The immediate built environment characteristics of the road segments or intersections are usually analyzed. The most common built environment

dimension that was considered all of the studies except Dai, Taquechel, Steward, and Strasser (2010) is travel, particularly annual average daily traffic, reflecting the traffic volume for a road segment. Since the unit of analysis is related to the roads, one of the most common built environment characteristics is “design,” which includes the geometric features of roads and intersections. For example, Dai et al. (2010), S. Q. Xie, Dong, Wong, Huang, and Xu (2018) considered road width, Alarifi, Abdel-Aty, Lee, and Park (2017), Alarifi, Abdel-Aty, Lee, and Wang (2018), Barua, El-Basyouny, and Islam (2016), and Chiou and Fu, (2013) considered the number of lanes, Chiou and Fu (2013), and Dai and Jaworski (2016) made use of road curvature, and H. Huang, Zhou, Wang, Chang, and Ma (2017), and Lee, Abdel-Aty, and Cai, (2017) considered the number of legs as independent factors in the analyses. The other common built environment characteristic is diversity, the diverse types of land uses in close proximity. Barua, El-Basyouny, and Islam (2014), and Barua et al. (2016), and El-Basyouny and Sayed (2009) used business land uses, Dai and Jaworski (2016) considered the number of stores/restaurants and the number of drinking places in a similar fashion, Kaygisiz, Senbil, and Yildiz (2017) considered distinct land uses such as light industry, mixed, residential, industry and non-urban, and Cai, Abdel-Aty, Lee, Wang, and Wang (2018), F. Guo, Wang, and Abdel-Aty (2010), Lee et al. (2017), and Lee, Abdel-Aty, Huang, and Cai (2019) considered urban-rural distinction in their analyses.

At the zonal level, there are more types than the segment/intersection level. For the United States case, the common spatial units are block groups (BG), traffic analysis zones (TAZ), and census tracts (CT), zip-code tabulation area (ZCTA), and traffic analysis districts (TAD) in the hierarchical order. In this order, block groups consist of census blocks, the smallest unit of census data. The average population of census blocks is 85 people (Lee et al., 2017). There are no studies in this review at this unit. On average, 39 blocks obtain block groups. One-fifth of the reviewed studies employed this geographic unit (M. Abdel-Aty et al., 2013; Clifton & Kreamer-Fults, 2007; Dumbaugh & Li, 2011; Y. Huang et al., 2018; Jermprapai & Srinivasan, 2014;

Marshall & Garrick, 2011; Ouyang & Bejleri, 2014; Saha et al., 2020; P. Xu et al., 2018; C.-Y. Yu & Xu, 2018; Zhai et al., 2019).

USCB defines traffic analysis zones as the geographic units used, especially for transportation statistics generated from the decennial census (USCB, 1994). TAZ is the most common (in almost half of the reviewed studies) used geographic unit (M. Abdel-Aty et al., 2013; Bao et al., 2017; Cai et al., 2016; P. Chen, 2015; Ding, Wang, et al., 2018; Dong et al., 2014, 2015; Gomes et al., 2017; Q. Guo et al., 2017; Hadayeghi et al., 2007, 2010, 2003; Jang et al., 2017; Jiang et al., 2016; Ladrón de Guevara et al., 2004; Lee et al., 2017; Lee, Abdel-Aty, & Jiang, 2015; G. Lovegrove et al., 2010; G. R. Lovegrove & Sayed, 2007; G. Lovegrove & Sayed, 2006; Osama & Sayed, 2017; Pirdavani et al., 2014; Rhee et al., 2016; Wei & Lovegrove, 2013; P. Xu et al., 2017; P. Xu & Huang, 2015; Yasmin & Eluru, 2016). Note that the TAZ level and definition are not unique to the US case. Other countries also use traffic analysis zones in their data creation processes; in this review, Belgium, Brazil, Canada, China, and South Korea have this geographic unit in common.

The Block group is a subdivision of tracts. Census tracts are somewhat homogeneous geographic units in terms of demographics, socioeconomic status, and living conditions, usually having 2,500 to 8,000 inhabitants (USCB, 1994). Similar to block groups, around one-fifth of the studies employed census tract as the unit of analysis (Cai, Abdel-Aty, Lee, et al., 2017; Cottrill & Thakuriah, 2010; LaScala et al., 2000; Loukaitou-Sideris et al., 2007; Morency et al., 2012; Narayanamoorthy et al., 2013; Ukkusuri et al., 2012; Y. Wang & Kockelman, 2013; Wier et al., 2009; K. Xie et al., 2019; C.-Y. Yu, 2014).

Zip-code tabulation area (ZCTA) is not an actual geographic unit but an assemblage of mail delivery routes. This unit is rarely used; among the studies reviewed, only Bao, Liu, Qin, & Zhou (2018) employed ZCT areas.

Traffic analysis zone (TAZ) is a subdivision of traffic analysis districts (TAD). This unit of analysis is also rarely applied. Cai, Abdel-Aty, and Lee (2017); Cai, Abdel-Aty, Lee, et al. (2017), and Cai, Abdel-Aty, Lee, and Huang (2019) used traffic

analysis districts. There are also other geographic units used for case studies from countries other than the United States. For example, for the United Kingdom, census wards and enumeration districts are used as the geographic unit of analysis (Noland & Quddus, 2004a, 2005). For the case of Australia, statistical area level 1 (SA1), statistical area level 2 (SA2), state electoral division (SED), postal areas (POA) are alternative units (Amoh-Gyimah et al., 2017).

In terms of the dependent variables, a higher portion of studies used the crash count of a single crash type as the outcome variable. Also, the severity level of crashes is used as a filtering feature in more than half of the studies. For example, Gomes et al. (2017) modeled the number of injury (fatal and non-fatal) total crashes, and Q. Guo et al. (2017) modeled killed and seriously injured pedestrian-vehicle crashes. The most common built environment dimensions are demographic and socioeconomic, density, and travel dimensions in terms of independent variables. Then, the design characteristics were extensively considered in studies examining the spatial relationships of traffic accidents. The least considered dimensions were the transit service and destination accessibility dimensions.

For the regional level studies listed in Table 2.8, larger geographic units than zonal level units are used. Some regional units used in those studies are county (Ewing, Hamidi, & Grace, 2016; Ewing, Schieber, et al., 2003; Flask & Schneider, 2013; Han et al., 2018; H. Huang et al., 2010; Zhenning Li et al., 2019; Zhibin Li et al., 2013; Noland & Oh, 2004), urban area (Najaf et al., 2018), and metropolitan area (Lee et al., 2018; Lee, Abdel-Aty, Huang, et al., 2019; Renne & Tolford, 2018) in the US case, city (Jang et al., 2017; Moeinaddini et al., 2014) in general, province in Turkish case (Erdogan, 2009), and provincia in Italian case (Lee, Abdel-Aty, de Blasiis, et al., 2019). In the studies reviewed at the regional level, the most common outcome variable is crash count/frequency. In all studies, crash severity is used either as a filter (in most cases) to define the outcome variable or as one of the outcome variables. For example, in Jang et al. (2017), severe total crashes were modeled, whereas, in Ewing, Hamidi, and Grace (2016), both crash frequency and crash severity are considered endogenous variables. The built environment dimensions

used are similar to the zonal level; demographic and socioeconomic, design, and travel dimensions are the most common ones.

According to the literature review presented in Table 2.6-2.8, most of the studies are at the zonal-level (compared to link/segment and regional level studies). One reason for this is the availability of built environment variables at the zonal level, increasing the possibility of providing insight into the impacts of the built environment factors on traffic crashes (Ouyang & Bejleri, 2014; Ukkusuri et al., 2011). For the regional level studies, a higher level of aggregation of the crash and built environment data than the link and zonal levels is used. Thus, the other reason is that crash analysis would be more appropriate for policy-oriented, macro-level, long-term, or indirect impacts at the regional level. Moreover, though the intersection/link level crash analysis is also considerably studied since the accidents usually happen at road segments, it is hard to assign built environment characteristics. At the zonal level, traffic accidents are aggregated, and a degree of aggregation is better for analysis to tackle with randomness. For example, when the crash type is considered, this level is better for pedestrian and cyclist crashes because of the low occurrence (Merlin et al., 2020). This is also valid when the crash severity is considered. Fatal and incapacitating injury crashes are rare in nature, so accidents' aggregation would give better insight.

Albeit the unit of analysis was the primary dimension, the methodological dimension should be mentioned. In Lord and Mannering (2010), statistical methods used in crash frequency data are reviewed, and accordingly, some of the most common approaches are as follows: Poisson model (Khattak et al., 2010); negative binomial/Poisson-gamma model (Dumbaugh & Li, 2011; Gomes et al., 2017; Hadayeghi et al., 2003; Jermprapai & Srinivasan, 2014; Khattak et al., 2010; K. Kim et al., 2006; Zhibin Li et al., 2013; Marshall & Garrick, 2011; Noland & Quddus, 2004a; Ouyang & Bejleri, 2014; Ukkusuri et al., 2012); zero-inflated model (ordered logit model) (Cai et al., 2016; P. Chen & Shen, 2016, 2019; Khattak et al., 2010; Ukkusuri et al., 2012); generalized additive model (P. Chen & Shen, 2016; Tasic et al., 2017), and multivariate models (M. Abdel-Aty et al., 2013; Aguero-Valverde &

Jovanis, 2010). The Poisson regression model is the most straightforward technique to model the crash frequency. Since crash frequency data consists of non-negative integer values, discrete modeling such as Poisson regression is required rather than the methods assuming continuous variables.

Although the Poisson model is the most basic one, it has a particular drawback: it cannot handle under- and over-dispersion, which is very common in crash data. Negative binomial/Poisson-gamma models are extensions to the Poisson model to overcome the overdispersion problem. Though under-dispersed data cannot be modeled using a negative binomial model, it is the most employed to model crash frequencies. Other approaches, such as zero-inflated models to handle a large number of zero crash counts frequently used in crash data analysis, would create theoretical inconsistencies such as having zero long-term means (Lord & Mannering, 2010). The generalized additive models could handle nonlinear relationships and bring a flexible, functional form involving smoothing functions for the explanatory variables. However, having more parameters, the approach is complex both to process and interpret than the previous count models. Multivariate models are useful to model the specific types of crash categories (e.g., property damage only, non-incapacitating injury, or fatal crash counts) jointly (not independent from each other). That is to say, the multivariate approach is adopted to model the counts of more than one type of crash while evaluating the correlation between these different crash types if they are dependent. Those models' drawback is their complexity and requirement of forming a correlation matrix for those crash types that are dependent (Lord & Mannering, 2010).

Moreover, there is a tendency to employ Bayesian techniques in crash-related research, as Bayesian approaches often perform better than their non-Bayesian counterparts (Ziakopoulos & Yannis, 2020). A quarter of the studies in this review applied Bayesian techniques (M. Abdel-Aty et al., 2013; Alarifi et al., 2018; Barua et al., 2014, 2016; Dong et al., 2014; El-Basyouny & Sayed, 2009; F. Guo et al., 2010; Lee et al., 2017; Lee & Abdel-Aty, 2018; Saha et al., 2020; P. Xu et al., 2018; P. Xu & Huang, 2015).

Table 2.6 Review on Built Environment and Traffic Safety - Link Level

Study	Case Study Location	Aim	Dependent Variable(s)			
			Crash Type	Crash Severity	Crash Count	Crash Rate
(Alarifi et al., 2017)	UNITED STATES / FLORIDA / Orange, Broward counties	Examining traffic safety considering corridor-level variables	TC		XX	
(Alarifi et al., 2018)	UNITED STATES / FLORIDA / Orange, Broward counties	Examining the impacts of various neighboring structures on “spatial hierarchical joint crash frequency models”	TC		XX	
(Barua et al., 2014)	CANADA / BRITISH COLUMBIA / Richmond, Vancouver	Examining the addition of spatial correlation in multivariate count-data models for crash severity	TC	XX	XX	
(Barua et al., 2016)	CANADA / BRITISH COLUMBIA / Vancouver	Examining the effects of introducing spatial heterogeneity in multivariate random-parameters models and their influence on various crash severity levels	TC	XX	XX	
(Cai, Abdel-Aty, Lee, et al., 2019)	UNITED STATES / FLORIDA / Orlando	Linking micro-level and macro-level safety analyses based on spatial correlation	TC		XX	
(Cai et al., 2018)	UNITED STATES / FLORIDA / Orange, Seminole, Osceola counties	Analyzing the zonal effects on traffic safety for segments and intersections	TC		XX	
(Chiou & Fu, 2013)	TAIWAN	Examining crash frequency and crash severity simultaneously	TC	XX	XX	XX
(Dai & Jaworski, 2016)	UNITED STATES / GEORGIA / DeKalb county	Examining the impacts of the built environment on pedestrian crashes	P-V		X	
(Dai et al., 2010)	UNITED STATES / GEORGIA / Georgia State University campus	Examining the relation between specific built environment features and pedestrian crashes	P-V			
(El-Basyouny & Sayed, 2009)	CANADA / BRITISH COLUMBIA / Vancouver	Examining the incorporation of spatial effects in models of accident prediction	TC		XX	
(Flahaut, 2004)	BELGIUM	Analyzing the impact of road features and local spatial environment on the roadway safety	TC	XX	XX	
(Guerra et al., 2019)	UNITED STATES / PHILADELPHIA	Examining the relationships among neighborhood socio-demographic factors, urban form, roadway features, and traffic crashes	TC; P-V	XX	XX	
(F. Guo et al., 2010)	UNITED STATES / FLORIDA / Hillsborough, Orange counties	Examining the safety impacts of risk factors (e.g., geometric design characteristics, traffic control, and traffic flow features)	TC	XX	XX	

Table 2.6 (continued)

(H. Huang et al., 2017)	UNITED STATES / FLORIDA / Hillsborough county	Examining motor vehicle, bicycle, and pedestrian crash occurrences at urban intersections	TC; V-V; P-V; B-V		XX	
(Kaygisiz et al., 2017)	TURKEY / ESKISEHIR	Examining the effects of the built environment on traffic crashes in the developing country context	TC; V-V; P-V	X	XX	
(J. K. Kim et al., 2010)	UNITED STATES / NORTH CAROLINA	Analyzing injury severity of pedestrian-vehicle crashes	P-V	XX	X	
(Lee et al., 2017)	UNITED STATES / FLORIDA	Modeling intersection level crashes with diverse macro-level geographic unit data	TC; P-V; B-V	X	XX	
(Miranda-Moreno et al., 2011)	CANADA / QUEBEC / Montreal	Analyzing the influence of built environment on pedestrian activity and pedestrian-vehicle crash occurrence	P-V	X	XX	
(Morency et al., 2012)	CANADA / QUEBEC / Montreal	Examining the impacts of traffic volume and road geometry on traffic injuries concerning social inequality	V-V; V-M; P-V; B-V	X	XX	
(Moudon et al., 2008)	UNITED STATES / WASHINGTON / King county	Examining the risk of occurrence of a pedestrian-motor vehicle crash at a specific point on a state route	P-V		XX	
(Quistberg et al., 2015)	UNITED STATES / WASHINGTON / Seattle	Evaluating the risk of pedestrian crashes at intersections and mid-blocks	P-V		XX	
(L. Wang et al., 2019)	UNITED STATES / FLORIDA	Examining real-time crash risk concerning expressway ramps	TC			X
(K. Xie et al., 2013)	CHINA / SHANGHAI	Examining corridor-level signalized intersection safety	TC		XX	
(K. Xie et al., 2014)	CHINA / SHANGHAI	Modeling crash frequency considering the hierarchy and spatial correlation of close intersections	TC		XX	
(S. Q. Xie et al., 2018)	CHINA / HONG KONG	Determining the potential factors contributing to the pedestrian crash occurrences at signalized intersections	P-V		XX	
(C.-Y. Y. Yu, 2015)	UNITED STATES / TEXAS / Austin	Analyzing multi-level (road environment and census tract) built environments' concerning severe injuries for pedestrians	P-V	XX		X
(C.-Y. Yu & Zhu, 2016)	UNITED STATES / TEXAS / Austin	Examining school siting and surrounding built environments' impact on motorist and pedestrian crash rates around public schools	V-V; P-V			XX

Table 2.6 (Continued)

Study	Independent Variable(s)							Unit of Analysis	Method / Modeling Approach
	Demographics	Density	Diversity	Design	Transit	Destination	Travel		
(Alarifi et al., 2017)		X		X			X	Corridor; Road segments; Intersections	Multilevel Poisson-Lognormal Joint Model (i) with Corridor and Sub-Corridor Random Effects (ii) with Corridor and Sub-Corridor Random Parameters
(Alarifi et al., 2018)		X		X			X	Road segments; Intersections	13 Bayesian Hierarchical Poisson-Lognormal Joint Spatial Models with Adjacency-Based, Adjacency-Route, Distance-Order, and Distance-Based Spatial Weight Features
(Barua et al., 2014)		X	X	X			X	Urban road segments	Full Bayesian Poisson Lognormal Univariate and Multivariate Random Parameters Models (i) with Heterogenous Effects (ii) with CAR Priors for Spatial Heterogeneity (iii) with both
(Barua et al., 2016)		X	X	X			X	Urban road segments	Full Bayesian Poisson Lognormal Multivariate Random Parameters Models (i) with Heterogenous Effects (ii) with CAR Priors for Spatial Heterogeneity (iii) with both
(Cai, Abdel-Aty, Lee, et al., 2019)	X	X	X	X		X	X	Road segments; Intersections	Bayesian Spatial Crash Frequency Model (i) a Non-integrated Model for the Macro-Level, (ii) a Non-integrated Model for the Micro-Level, and (iii) an Integrated Model for both Levels

Table 2.6 (continued)

(Cai et al., 2018)	X	X	X	X		X	X	Road segments; Intersections	Poisson-Lognormal Models: (i) Fixed Parameter Univariate Model; (ii) Grouped Random Parameter Univariate Spatial Model; (iii) Grouped Random Parameter Univariate Spatial Model with Zonal Factors; (iv) Grouped Random Parameter Multivariate Spatial Model with Zonal Factors
(Chiou & Fu, 2013)			X	X			X	Highway segments	Multinomial Generalized Poisson (MGP) Model with or without Error Components; Nested Generalized Poisson Models
(Dai & Jaworski, 2016)	X	X	X	X	X		X	Road segments	Network Kernel Density Estimation (NKDE); Bivariate Analysis and Negative Binomial Models
(Dai et al., 2010)				X	X			Road segments; Intersections	Network-Based Kernel Density Estimation (NKDE); Network-Based K-Function
(El-Basyouny & Sayed, 2009)			X	X			X	Urban road segments	Full Bayesian Multivariate Poisson Lognormal with and without CAR Prior; Full Bayesian Multiple Membership Model; Full Bayesian Extended Multiple Membership Model
(Flahaut, 2004)				X		X	X	Rural & Highway segments	Logistic Regression with and without Spatial Autocorrelation
(Guerra et al., 2019)	X	X	X	X			X	Road segment	Multilevel Negative Binomial Models
(F. Guo et al., 2010)			X	X			X	Intersections	Fixed Effects Bayesian Poisson Regression; Fixed and Mixed Effects Bayesian Negative Binomial Regression; Spatial CAR Prior Extended Poisson and Negative Binomial Models

Table 2.6 (continued)

(H. Huang et al., 2017)	X	X		X	X		X	Intersections	Poisson Regression (Univariate, Multivariate Lognormal and Spatial Random Effects Models)
(Kaygisiz et al., 2017)			X	X	X		X	Road segments	Binary Logit Models; Count Data Regression Models (Poisson Regression, Zero Inflated Poisson Regression, Negative Binomial Regression, Zero Inflated Negative Binomial Regression)
(J. K. Kim et al., 2010)	X		X	X				Individual based	Mixed Logit Model
(Lee et al., 2017)	X	X	X	X	X		X	Intersections	Mixed effects Negative Binomial Models with Micro-level Variables, with Micro-level and Macro-level Variables and with Micro-level and Macro-level Variables with Random-Effects
(Miranda-Moreno et al., 2011)	X		X	X	X		X	Intersections	Negative Binomial Regression; Generalized Negative Binomial Regression; Latent Class Negative Binomial Regression
(Morency et al., 2012)	X	X		X	X		X	Intersections	Multilevel Poisson Regression
(Moudon et al., 2008)	X	X	X	X	X		X	Points	Binomial Logit Model
(Quistberg et al., 2015)	X	X	X	X				Intersections and mid-block locations	Multilevel Mixed Effects Poisson Models
(L. Wang et al., 2019)	X	X	X	X			X	Ramps	Bayesian Logistic Regression
(K. Xie et al., 2013)			X	X			X	Intersections; Urban segments	Hierarchical Bayesian Negative Binomial Regression (Basic, Random Parameter, Hierarchical)

Table 2.6 (continued)

(K. Xie et al., 2014)			X	X			X	Intersections; Urban segments	Bayesian Negative Binomial Regression (Basic, Random Effect, Random Parameter, Hierarchical, Hierarchical CAR)
(S. Q. Xie et al., 2018)				X	X		X	Intersections	Poisson Lognormal Model; Bayesian Measurement Errors Model
(C.-Y. Y. Yu, 2015)	X	X	X	X	X		X	Road segment	Multilevel Binomial Logistic Model
(C.-Y. Yu & Zhu, 2016)	X	X	X	X	X		X	School parcel location	Log-linear Regression

Table 2.7 Review on Built Environment and Traffic Safety - Zonal Level

Study	Case Study Location	Aim	Dependent Variable(s)			
			Crash Type	Crash Severity	Crash Count	Crash Rate
(M. Abdel-Aty et al., 2013)	UNITED STATES / FLORIDA / Hillsborough and Pinellas counties	Examining the impact of variables on crashes at different geographic units	TC; P-V	XX	XX	
(Amoh-Gyimah et al., 2017)	AUSTRALIA / MELBOURNE	Examining the effects of differences in spatial units on unobserved heterogeneity	TC	X	XX	
(Bao et al., 2018)	UNITED STATES / NEW YORK / Manhattan, the Bronx, Brooklyn and Queens	Examining the contribution of the extraction of variables from taxi GPS data to spatially aggregated crashes in urban areas	TC	XX	XX	
(Bao et al., 2017)	UNITED STATES / LOS ANGELES	Examining how to consolidate human activity information in spatial analysis of crashes in urban areas	TC; V-V; P-V		XX	
(Cai, Abdel-Aty, & Lee, 2017)	UNITED STATES / FLORIDA	Examining exogenous factors that affect pedestrian and bicyclist crashes at the macroscopic level	TC; P-V; B-V		XX	
(Cai, Abdel-Aty, Lee, et al., 2017)	UNITED STATES / FLORIDA	Identifying the best zonal system for macroscopic crash modeling (considering census tracts, state-wide traffic analysis zones, and traffic analysis districts)	TC; P-V; B-V	XX		
(Cai, Abdel-Aty, Lee, et al., 2019)	UNITED STATES / FLORIDA / Orlando	Examining the association between the crash counts of macro-level and micro-level based on spatial interaction	TC		XX	

Table 2.7 (continued)

(Cai, Abdel-Aty, Sun, et al., 2019)	UNITED STATES / FLORIDA	Developing a deep learning approach for predicting crashes with the high-resolution data	TC		XX	
(Cai et al., 2016)	UNITED STATES / FLORIDA	Developing dual-state models for pedestrian and bicycle crash frequency analysis	P-V; B-V		XX	
(Chen, 2015)	UNITED STATES / WASHINGTON / Seattle	Analyzing the relationship between built environment factors and motor vehicles involved bicycle crashes	B-V		XX	
(Chen & Shen, 2016)	UNITED STATES / WASHINGTON / Seattle	Examining the effects of built environment factors on automobile-involved bicycle crashes considering possible spatial dependence among crash locations	B-V	XX	XX	
(Chen & Shen, 2019)	UNITED STATES / FLORIDA / Orlando, WASHINGTON / Seattle, MINNESOTA / Minneapolis	Examining the associations between the built environment features and cyclist injury severity and the spatial variation of these associations among cities	B-V	XX	XX	
(Cho et al., 2009)	UNITED STATES / WASHINGTON D.C. / Montgomery county	Analyzing how perceived and actual crash risk are associated with each other considering the built environmental characteristics	P-V; B-V	X	XX	
(Clifton & Kreamer-Fulfs, 2007)	UNITED STATES / MARYLAND / Baltimore City	Examining the relation between the physical and social attributes and the pedestrian-vehicle crashes around the schools	P-V	XX	XX	
(Clifton et al., 2009)	UNITED STATES / MARYLAND / Baltimore City	Examining the effects of personal and environmental characteristics on injury severity of pedestrian-vehicle crashes	P-V	XX		
(Cottrill & Thakuriah, 2010)	UNITED STATES / ILLINOIS / Chicago	Examining the relationship between pedestrian-vehicle crashes and the aspects of areas with high minority populations and lower-income	P-V	X	XX	XX
(Ding, Chen, et al., 2018)	UNITED STATES / WASHINGTON / Seattle	Analyzing the comparative importance of built environment characteristics on automobile-involved pedestrian crashes by considering their mutual interaction and their non-linear effects	P-V		XX	
(Dong et al., 2015)	UNITED STATES / FLORIDA / Hillsborough county	Evaluating the spatial proximity effects by using complex and multi-dimensional spatial data in crash prediction	TC		XX	

Table 2.7 (continued)

(Dong et al., 2014)	UNITED STATES / FLORIDA / Hillsborough county	Evaluating crash prediction models at traffic analysis zones level with different spatial-proximity structures	TC		XX	
(Dumbaugh & Li, 2011)	UNITED STATES / TEXAS / Bexar county / San Antonio	Analyzing whether urban crashes are the products of random error or whether characteristics of the built environment may influence it	V-V; P-V; B-V		XX	
(Gomes et al., 2017)	BRAZIL / FORTALEZA	Comparing local spatial models and global non-spatial crash prediction models	TC	X	XX	
(Guerra et al., 2019)	UNITED STATES / PHILADELPHIA	Examining the relation between neighborhood characteristics (socio-demographics, urban form, roadway) and traffic safety outcomes (traffic collisions, injuries, and fatalities)	TC; P-V	XX	XX	
(Q. Guo et al., 2017)	CHINA / HONG KONG	Examining the impacts of road network patterns on pedestrian safety	P-V	X	XX	
(Hadayeghi et al., 2003)	CANADA / ONTARIO / Toronto	Predicting the number of accidents in planning zones as a function of zonal features	TC	X	XX	
(Hadayeghi et al., 2007)	CANADA / ONTARIO / Toronto	Developing zonal-level crash prediction models	TC	X	XX	
(Hadayeghi et al., 2010)	CANADA / ONTARIO / Toronto	Examining the spatial variations in the association between the number of zonal crashes and possible transportation planning predictors	TC	X	XX	
(Huang, Wang, & Patton, 2018)	UNITED STATES / MICHIGAN / Wayne, Oakland and Macomb counties	Examining the relationship between the built environment and vehicle crashes	TC			XX
(Jang et al., 2017)	UNITED STATES / CALIFORNIA / San Diego county	Examining the relationship between the urban built environment and traffic crash occurrences	TC	X	XX	
(Jermprapai & Srinivasan, 2014)	UNITED STATES / FLORIDA	Developing macroscopic or planning-level pedestrian safety models	P-V	X	XX	
(Jiang et al., 2016)	UNITED STATES / FLORIDA / Orange, Seminole, Osceola counties	Analyzing the feasibility of employing a random forest model for hot zone identification at macro-level	TC; P-V; B-V	X	XX	
(Ladrón de Guevara et al., 2004)	UNITED STATES / ARIZONA / Tucson	Forecasting crashes at the planning level	TC	X	XX	
(LaScala et al., 2000)	UNITED STATES / CALIFORNIA / San Francisco	Investigating geographic correlates of pedestrian injury crashes through spatial analysis	P-V	X	X	XX

Table 2.7 (continued)

(LaScala et al., 2004)	UNITED STATES / CALIFORNIA	Analyzing annual child pedestrian injury rates with a focus on the contribution of schools to injury risk	P-V; B-V	X	XX	
(Lee & Abdel-Aty, 2018)	UNITED STATES / FLORIDA	Analyzing bicycle safety at macro-level	B-V		XX	
(Lee et al., 2014)	UNITED STATES	Developing a zone system for traffic safety analysis at macro-level	TC	X	XX	
(Lee, Abdel-Aty, & Jiang, 2015)	UNITED STATES	Developing multivariate crash models at the macroscopic level for crashes by different transportation modes	V-V; P-V; B-V		XX	
(Lee, Abdel-Aty, Choi, et al., 2015)	UNITED STATES / FLORIDA	Determining hot zones for pedestrian safety	V-V; P-V; B-V		XX	
(Loukaitou-Sideris et al., 2007)	UNITED STATES / LOS ANGELES	Examining the impacts of socio-demographic, urban form, land use, density, and traffic characteristics on pedestrian crash rates	P-V		X	XX
(Lovegrove & Sayed, 2006)	CANADA / BRITISH COLUMBIA / Greater Vancouver Regional District	Developing a safety planning decision-support tool to facilitate a proactive approach to community planning	TC	X	XX	
(Lovegrove & Sayed, 2007)	CANADA / BRITISH COLUMBIA / Greater Vancouver Regional District	Analyzing the use of macrolevel collision prediction models in a reactive safety application	TC	X	XX	
(Lovegrove & Sayed, 2006)	CANADA / BRITISH COLUMBIA / Greater Vancouver Regional District	Employing macro-level collision prediction models in road safety planning applications	TC	X	XX	
(Lovegrove, Lim, & Sayed, 2010)	CANADA / BRITISH COLUMBIA / Greater Vancouver Regional District	Developing and testing macro-level collision prediction models in a regional road safety planning application	TC	X	XX	
(Marshall & Garrick, 2011)	UNITED STATES / CALIFORNIA	Examining the effect of street and street network features on various crash types	TC	X	XX	
(Morency et al., 2012)	CANADA / QUEBEC / Montreal	Examining the impacts of traffic volume and road geometry on traffic injuries concerning social inequality	V-V; V-M; P-V; B-V	X	XX	
(Narayanamoorthy et al., 2013)	UNITED STATES / NEW YORK / Manhattan	Analyzing the traffic crash-related counts of bicyclists and pedestrians by injury severity	P-V; B-V	X	XX	
(Nashad et al., 2016)	UNITED STATES / FLORIDA	Macrolevel modeling of pedestrian and bicycle crashes	P-V; B-V		XX	

Table 2.7 (continued)

(Noland & Quddus, 2004a)	UNITED KINGDOM / ALL WARDS	Examining the relations between area-wide factors with traffic fatalities, serious injuries, and slight injuries	TC	X	XX	
(Noland & Quddus, 2005)	UNITED KINGDOM / LONDON DISTRICTS	Examining the impacts of congestion on traffic safety	TC; P-V	X	XX	
(Osama & Sayed, 2017)	CANADA / VANCOUVER	Examining the association between pedestrian-vehicle crashes and several zonal characteristics	P-V		XX	
(Ouyang & Bejleri, 2014)	UNITED STATES / FLORIDA / Miami-Dade County	Examining the impacts of built environment on traffic crashes	TC; P-V; B-V	X	XX	
(Pirdavani et al., 2014)	BELGIUM / FLANDERS	Analyzing fatal and injury crashes spatially by employing geographically weighted regression	TC	X	XX	
(Rhee et al., 2016)	SOUTH KOREA / SEOUL	Examining the distribution of traffic crash frequency	TC	X	XX	
(Rothman et al., 2014)	CANADA / TORONTO	Analyzing the influence of the built environment on walking to school and child pedestrian collisions	P-V			XX
(Saha et al., 2020)	UNITED STATES / NORTH CAROLINA / Mecklenburg county	Examining the association between built environment variables and crash frequency, considering environmental mediating factors	TC	X	XX	
(Ukkusuri et al., 2012)	UNITED STATES / NEW YORK	Analyzing the relation of land use and road design on pedestrian safety, and the impacts of the spatial aggregation level on pedestrian crash frequency	P-V	X	XX	
(Y. Wang & Kockelman, 2013)	UNITED STATES / TEXAS / Travis county / Austin	Spatial analysis of pedestrian crash counts across neighborhoods	P-V	X	XX	
(Wei & Lovegrove, 2013)	CANADA / BRITISH COLUMBIA / Central Okanagan Regional District	Assessing the safety of cyclists by community-based, macro-level collision prediction models	B-V		XX	
(Wier et al., 2009)	UNITED STATES / CALIFORNIA / San Francisco	Developing an area-level regression model for vehicle-pedestrian injury crashes	P-V	X	XX	
(Xie, Ozbay, & Yang, 2019)	UNITED STATES / NEW YORK / Manhattan	Analyzing the spatial autocorrelation of neighboring sites and the correlation across different crash types	TC	X	XX	
(P. Xu & Huang, 2015)	UNITED STATES / FLORIDA / Hillsborough county	Examining spatial heterogeneity in regional safety modeling	TC	X	XX	

Table 2.7 (continued)

(P. Xu et al., 2018)	UNITED STATES / FLORIDA / Hillsborough county	Addressing the modifiable areal unit problem (MAUP) in safety analysis	TC		XX	
(P. Xu et al., 2017)	UNITED STATES / FLORIDA / Hillsborough county	Examining the spatially differing relationships between crash frequency and risk factors	TC	X	XX	
(Yasmin & Eluru, 2016)	CANADA / MONTREAL, TORONTO	Examining factors affecting bicycle crashes at the Traffic Analysis Zone (TAZ) level	B-V		XX	
(Yu & Xu, 2018)	UNITED STATES / TEXAS / Austin	Investigating the influence of the built environment on crashes with different injury severity levels	TC	X	XX	
(Yu, 2014)	UNITED STATES / TEXAS / Austin	Examining the influence of income, ethnicity, and built environmental features on the proportions of workers who walk/bike and on pedestrian/cyclist crash rates	P-V; B-V			XX
(Zhai et al., 2019)	UNITED STATES / FLORIDA / Hillsborough county	Examining the effects of zonal configurations on macro-level traffic safety analysis for crashes of various severity levels	TC	X	XX	

Table 2.7 (Continued)

Study	Independent Variable(s)							Unit of Analysis	Method / Modeling Approach
	Demographics	Density	Diversity	Design	Transit	Destination	Travel		
(M. Abdel-Aty et al., 2013)	X	X				X	X	TAZ; CT; BG	Bayesian Multivariate Poisson Lognormal Regression
(Amoh-Gyimah et al., 2017)	X	X	X		X		X	SA1; SA2; TAZ; SED; POA; GRID	Random Parameter Negative Binomial Model; Semi-parametric Poisson GWR
(Bao et al., 2018)	X	X	X	X		X	X	ZIP Code Tabulation Areas (ZCTA)	Geographically Weighted Poisson Regression (GWPR); Latent Dirichlet Allocation
(Bao et al., 2017)	X	X	X	X			X	TAZ	Geographically Weighted Regression (GWR)
(Cai, Abdel-Aty, & Lee, 2017)	X	X		X	X	X	X	Traffic Analysis Districts (TAD)	Bayesian Negative Binomial Regression; Bayesian Logit Regression Model; Bayesian Joint Model; Elasticity Analysis
(Cai, Abdel-Aty, Lee, et al., 2017)	X	X		X		X	X	CT; TAZ; TAD	Poisson-lognormal Models; with and without Consideration of Spatial Autocorrelation

Table 2.7 (continued)

(Cai, Abdel-Aty, Lee, et al., 2019)	X	X	X	X		X	X	TAD	Bayesian Spatial Crash Frequency Model (i) Non-integrated Models for the Macro-and Micro-level, (ii) an Integrated Model for both Macro and Micro-levels
(Cai, Abdel-Aty, Sun, et al., 2019)		X	X	X			X	GRID	Deep Learning Architecture of a Convolutional Neural Network (CNN)
(Cai et al., 2016)	X	X	X	X	X		X	TAZ	Negative Binomial Spatial and Non-spatial Models (Basic, Zero-inflated, and Hurdle)
(Chen, 2015)	X	X	X	X	x		X	TAZ	Poisson Lognormal Random Effects Model Using Hierarchical Bayesian Estimation
(Chen & Shen, 2016)	X	X	X	X				TAZ (crash-based analysis, but the BE factors of associated TAZ are used)	Generalized Ordered Logit Model and a Generalized Additive Model
(Chen & Shen, 2019)	X	X	X	X	X		X	BG (crash-based analysis, but the BE factors of associated BG are used)	Generalized Ordered Logit Model
(Cho et al., 2009)		X	X	X	X	X		Community Analysis Zones (CAZs)	Exploratory Factor Analysis; Path Analysis
(Clifton & Kreamer-Fults, 2007)	X	X	X	X	X			BG	Multivariate Regression
(Clifton et al., 2009)	X	X	X	X	X			BG (crash-based analysis, but the BE factors of associated BG are used)	Generalized Ordered Probit Model
(Cottrill & Thakuriah, 2010)	X	X		X	X	X	X	Environmental Justice Areas (EJ); CT	Poisson Regression with Heterogeneity
(Ding, Chen, et al., 2018)		X	X	X	X		X	TAZ	Multiple Additive Poisson Regression Trees (MAPRT) Model
(Dong et al., 2015)	X		X	X			X	TAZ	Support Vector Machine with Correlation-based Feature Selector; Bayesian Multivariate Poisson Lognormal with CAR Prior
(Dong et al., 2014)	X	X		X			X	TAZ	Bayesian Multivariate Poisson Lognormal with CAR Prior Regression for Boundary and Non-boundary Area Models
(Dumbaugh & Li, 2011)	X	X	X	X			X	BG	Negative Binomial Regression
(Gomes et al., 2017)	X	X	X	X				TAZ	Negative Binomial Regression; Poisson GWR; Negative Binomial GWR
(Guerra et al., 2019)	X	X	X	X			X	CT	Multilevel Negative Binomial Models

Table 2.7 (continued)

(Q. Guo et al., 2017)	X	X	X				X	TAZ	Poisson Lognormal Regression; Bayesian Poisson Lognormal with CAR Prior Regression with Contiguity, Geometry-Centroid Distance and Road Network Connectivity
(Hadayeghi et al., 2003)	X	X		X			X	TAZ	GWR; Negative Binomial Regression
(Hadayeghi et al., 2007)	X	X	X	X			X	TAZ	Generalized Linear Modeling (GLM)
(Hadayeghi et al., 2010)	X	X	X	X			X	TAZ	Geographically Weighted Poisson Regression (GWPR); Negative Binomial Regression; Poisson Regression
(Huang, Wang, & Patton, 2018)	X	X	X	X				BG	Geographically Weighted Regression (GWR)
(Jang et al., 2017)	X	X	X	X			X	TAZ	Multi-level Poisson Regression
(Jermprapai & Srinivasan, 2014)	X	X	X	X	X	X		BG	Negative Binomial Regression
(Jiang et al., 2016)	X	X	X	X			X	TAZ	Random Forest Models (CART trees); Wilcoxon Tests
(Ladrón de Guevara et al., 2004)	X	X	X	X	X		X	TAZ	Negative Binomial Regression; Simultaneous Equation Estimation
(LaScala et al., 2000)	X		X	X			X	CT	Spatial Autocorrelation Regression Log-linear Model
(LaScala et al., 2004)	X	X		X			X	Geographic units	Linear Regression Models
(Lee & Abdel-Aty, 2018)	X	X	X	X	X	X	X	ZIP	Bayesian Poisson Lognormal CAR Models
(Lee et al., 2014)	X	X	X	X			X	TSAZ; TAZ	Bayesian Multivariate Poisson Lognormal Regression
(Lee, Abdel-Aty, & Jiang, 2015)	X	X	X	X			X	TAZ	Univariate and Multivariate Bayesian Poisson Lognormal with CAR Prior Regression
(Lee, Abdel-Aty, Choi, et al., 2015)	X	X	X	X		X	X	ZIP	Bayesian Poisson Lognormal Simultaneous Equations Spatial Error Model
(Loukaitou-Sideris et al., 2007)	X	X	X				X	CT	OLS regression
(Lovegrove & Sayed, 2006)	X	X		X		X	X	Neighborhood - TAZ	Groups of Macrolevel Crash Prediction Models using GLMs (Negative Binomial Error Structure)
(Lovegrove & Sayed, 2007)	X	X		X		X	X	Neighborhood - TAZ	Groups of Macrolevel Crash Prediction Models using GLMs
(Lovegrove & Sayed, 2006)	X	X		X			X	Neighborhood - TAZ	Groups of Macrolevel Crash Prediction Models using GLMs; Before-after analysis
(Lovegrove, Lim, & Sayed, 2010)	X	X		X			X	TAZ	Groups of Collision Prediction GLMs; Modified T-Tests

Table 2.7 (continued)

(Marshall & Garrick, 2011)	X	X	X	X		X	X	BG	Negative Binomial Regression
(Morency et al., 2012)	X	X		X	X		X	CT	Multilevel Poisson Regression
(Narayanamoorthy et al., 2013)	X	X	X	X	X		X	CT	Customized Generalized Ordered-Response Spatial Multivariate Count Model
(Nashad et al., 2016)	X		X	X	X	X	X	statewide TAZ	Negative Binomial Regression (Copula-based)
(Noland & Quddus, 2004a)	X	X	X	X			X	Census Ward	Negative Binomial Regression
(Noland & Quddus, 2005)	X	X		X	X	X		Enumeration District	Negative Binomial Regression; ANOVA
(Osama & Sayed, 2017)		X	X	X			X	TAZ	Negative Binomial Generalized Linear Regression; Full Bayes with Spatial Effects
(Ouyang & Beijleri, 2014)		X	X	X	X	X		BG	Negative Binomial Regression
(Pirdavani et al., 2014)	X	X		X			X	TAZ	Moran's I; Geographically Weighted GLM; Negative Binomial Regression
(Rhee et al., 2016)	X		X	X	X		X	TAZ	OLS Regression; Spatial-lag Regression; Spatial Error Regression; Poisson GWR
(Rothman et al., 2014)	X	X	X	X				School attendance boundaries	Negative Binomial Regression
(Saha et al., 2020)	X		X	X	X		X	BG	Bayesian CAR Modeling
(Ukkusuri et al., 2012)	X		X	X	X			CT ZIP	Negative Binomial Regression; Negative Binomial Regression with Heterogeneity in Dispersion Parameter; Zero-Inflated Negative Binomial Regression
(Y. Wang & Kockelman, 2013)	X	X	X	X	X	X	X	CT; Thiessen polygons	Multivariate Poisson Lognormal Regression with and without CAR Priors
(Wei & Lovegrove, 2013)	X	X		X	X		X	TAZ	Negative Binomial Macrolevel Crash Prediction Models
(Wier et al., 2009)	X		X	X	X		X	CT	Log-linear Multivariate OLS Regression Model
(Xie, Ozbay, & Yang, 2019)	X		X	X	X		X	CT	Poisson-Gamma (PG) Model; Univariate Conditional Autoregressive (UCAR) Model; Multivariate Conditional Autoregressive (MVCAR) Model
(P. Xu & Huang, 2015)	X	X		X			X	TAZ	Negative Binomial Regression; Bayesian Negative Binomial Model with CAR Prior; Random Parameter Negative Binomial Model; Semi-Parametric Poisson GWR
(P. Xu et al., 2018)	X	X		X			X	BG; CT; TAZ; ZIP	Bayesian Negative Binomial Model with CAR Prior
(P. Xu et al., 2017)	X	X		X	X		X	TAZ	Bayesian Spatially Varying Coefficients Model

Table 2.7 (continued)

(Yasmin & Eluru, 2016)	X		X	X	X	X	X	TAZ	Poisson Regression; Negative Binomial Regression (Basic and Latent Segmentation)
(Yu & Xu, 2018)	X	X	X	X	X	X	X	BG; CT; TAZ	Negative Binomial Regression; Geographically Weighted Negative Binomial Regression (GWNBR)
(Yu, 2014)	X	X	X	X	X		X	CT	Path Models
(Zhai et al., 2019)	X	X		X			X	BG; TAZ; CT; ZIP	Bayesian Poisson-Lognormal Models with Multivariate CAR Priors

Table 2.8 Review on Built Environment and Traffic Safety - Regional Level

Study	Case Study Location	Aim	Dependent Variable(s)			
			Crash Type	Crash Severity	Crash Count	Crash Rate
(Bishai et al., 2006)	41 COUNTRIES	Investigating the reasons for the increase in traffic fatalities with GDP per capita in lower-income countries and decrease with GDP per capita in higher-income countries	TC	XX	XX	
(Erdogan, 2009)	TURKEY	Analyzing the provincial differences in traffic accidents and roadway mortality	TC	XX		XX
(Ewing, Hamidi, & Grace, 2016)	UNITED STATES / ALL STATES	Evaluating the relationship between urban sprawl and the rates of traffic crashes	TC	XX	XX	
(Ewing, Schieber, et al., 2003)	UNITED STATES / 448 US Counties	Analyzing the relation between urban sprawl and traffic fatalities	TC; P-V	X	X	XX
(Flask & Schneider, 2013)	UNITED STATES / OHIO	Examining the multi-level spatial correlation in motorcycle crashes	MC	X	XX	
(Han et al., 2018)	UNITED STATES / FLORIDA	Analyzing the variations in the impact of road-level factors on crash frequency for diverse regions	TC	X	XX	
(H. Huang et al., 2010)	UNITED STATES / FLORIDA	Estimating county-level crash risk differences by controlling for exposure variables, i.e., daily vehicle miles traveled (DVMT) and population	TC	X	XX	X
(Jang et al., 2017)	UNITED STATES / CALIFORNIA / San Diego county	Examining the relation between urban built environment and traffic crashes	TC	X	XX	
(LaScala et al., 2001)	UNITED STATES / CALIFORNIA	Examining the relationship between neighborhood characteristics (including alcohol availability and alcohol consumption patterns) and pedestrian injury crashes	P-V	X	X	XX

Table 2.8 (continued)

(Lee et al., 2017)	UNITED STATES / FLORIDA	Modeling intersection crashes	TC; P-V; B-V	X	XX	
(Lee et al., 2018)	UNITED STATES	Generating trip and safety models in one structure at the same time	P-V; B-V	X	XX	
(Lee, Abdel-Aty, de Blasiis, et al., 2019)	ITALY, UNITED STATES	Examining the international transferability of safety performance functions (SPFs) for the USA and for Italy	TC; P-V; B-V	X	XX	
(Lee, Abdel-Aty, Huang, et al., 2019)	UNITED STATES / US Metropolitan Areas	Combining the walking exposure analysis and pedestrian crash analysis	P-V	X	XX	
(Zhenning Li et al., 2019)	UNITED STATES / IDAHO	Distinguishing unobserved heterogeneity generated from multiple sources for impaired-driving crashes	TC	X	XX	
(Zhibin Li et al., 2013)	UNITED STATES / CALIFORNIA	Examining the spatial heterogeneity in the relationship between crash counts and explanatory variables across counties	TC	X	XX	
(Moeinaddini et al., 2014)	20 CITIES WORLDWIDE	Analyzing the association between street network factors with traffic crashes and fatalities at city level	TC	X	XX	
(Najaf et al., 2018)	UNITED STATES / 100 Major Urban Areas	Examining the impacts of urban form on traffic safety at city level	TC; V-V; P-V	X	XX	
(Noland & Oh, 2004)	UNITED STATES / ILLINOIS	Examining whether several changes in road network infrastructure and geometric design can be related to changes in road fatalities and accidents	TC	X	XX	
(Noland & Quddus, 2004b)	UNITED KINGDOM / 11 STATISTICAL REGIONS	Analyzing pedestrian and bicycle casualties at the regional level	P-V; B-V	X	XX	
(Renne & Tolford, 2018)	UNITED STATES / LOUISIANA / Baton Rouge, New Orleans	Analyzing possible changes in the growth of vehicle miles traveled (VMT) and related traffic safety outcomes	TC	X		XX

Table 2.8 (Continued)

Study	Independent Variable(s)							Unit of Analysis	Method / Modeling Approach
	Demographics	Density	Diversity	Design	Transit	Destination	Travel		
(Bishai et al., 2006)	X			X				Country	OLS; Fixed Effects Model; Random Effects Model

Table 2.8 (continued)

(Erdogan, 2009)	X			X				Provinces	Moran's I and Geary's c Values, Z and G statistics; Geographically Weighted Regression (GWR)
(Ewing, Hamidi, & Grace, 2016)	X	X	X				X	County	Structural Equation Modeling (SEM)
(Ewing, Schieber, et al., 2003)	X	X						County	Principal Component Analysis; Regression Analysis
(Flask & Schneider, 2013)	X			X				County; Township	Bayesian Negative Binomial Regression with Mixed Effects
(Han et al., 2018)				X			X	County	Bayesian Hierarchical Random Parameter Model; Bayesian Hierarchical Random Intercept Model; Bayesian Poisson Lognormal Model
(H. Huang et al., 2010)	X	X		X		X	X	County	Bayesian Spatial CAR Priors Regression
(Jang et al., 2017)	X	X	X	X			X	City	Multi-level Poisson Regression
(LaScala et al., 2001)	X	X	X				X	Communities	Spatial Autocorrelation Regression Log-linear Model
(Lee et al., 2017)	X	X	X	X	X		X	County; County Division	Mixed Effects Negative Binomial Models with: (i) Micro-level Variables, (ii) Micro-Macro-level Variables (iii) Micro- and Macro-level Variables with Random-Effects
(Lee et al., 2018)	X	X	X	x	X	X	X	Metropolitan areas	Bayesian Integrated and Non-integrated Bivariate Models
(Lee, Abdel-Aty, de Blasiis, et al., 2019)	X	X	X		X		X	County; Provincia	Negative Binomial Regression; Calibration factors; Transferability Indices
(Lee, Abdel-Aty, Huang, et al., 2019)	X	X	X				X	Metropolitan areas	Multiple Linear Regression Model Integrated in a Poisson Lognormal Model
(Zhenning Li et al., 2019)	X			X			X	County	Hierarchical Bayesian Random Parameters Models
(Zhibin Li et al., 2013)	X	X		X			X	County	Negative Binomial Regression; Geographically Weighted Poisson Regression GWPR
(Moeinaddini et al., 2014)		X		X	X			City	Gamma-distributed GLM
(Najaf et al., 2018)	X	X	X	X	X	X	X	Urban Areas	Factor Analysis (FA); Structural Equation Modeling (SEM)
(Noland & Oh, 2004)	X			X				County	Negative Binomial Panel Regression
(Noland & Quddus, 2004b)	X			X				Statistical regions	Fixed Effects Negative Binomial Regression
(Renne & Tolford, 2018)	X	X		X	X		X	Metropolitan Area	Structural Equation Modeling (SEM)

2.3 Summary

According to this review, most studies examined the relationship between traffic safety and the urban built environment. Definition of the built environment is not crystal clear, and the variables defining the built environment vary from one study to another. Even in the broadest built environment definition (by using D-variables), related variables may change. Therefore, selecting the built environment variables is a paramount concern, which is one of this study's aims.

In terms of the unit of analysis, most studies considered zonal level variables related to data availability and aims of those studies. The majority of the studies examined cases from developed countries (especially from the US), mainly due to the mentioned data availability issue, compared to the developing countries. Thus, studying developed country cases would be desirable due to the data concerns and the availability of comparable results. On the other hand, there is more to do for developing countries. The contribution of developing country cases would be valuable in terms of data availability. Data collection and sharing are expensive processes. With limited resources, it is essential to know which data is more critical and should be collected. Thus, an evaluation scheme would be developed using developed and developing country cases. For this purpose, the examination of the relation between the built environment (BE) and traffic safety (TS) would be informative.

In terms of traffic safety variables (the dependent variables), several studies analyzed alternative types of crashes, but most of the studies analyzed a single crash type, e.g., total crashes. A few studies directly estimated severity level in their analyses; severity was usually employed as a filter to select the crash type, e.g., fatal vulnerable road user involved crashes. Therefore, examining the crashes by severity would be a viable alternative.

To conclude, due to the mentioned aspects throughout the literature review chapter, this study focuses on determining the essential built environment factors associated

with traffic safety (defined by varying crash typologies) at a macroscopic level. The crashes are categorized according to severity and involvement. In addition to the factor significance determined by regression models, factor consistency is considered important. Thus, an iterative algorithm is proposed to trace factor behavior throughout the solution. The cases are mainly from the developed country context (i.e., the United States). In determining the value of information, a Turkish case is used as a developing country context.

In the next chapter, the methodology of this study is presented in detail.



CHAPTER 3

METHODOLOGY

This section presents the methodology applied in this study in detail. Firstly, the aim and research objectives are presented. Then the methodological frame is introduced. Finally, the case study areas are presented.

3.1 Aim of the Study and Research Questions

In the previous chapters, three salient and widely studied relationships were presented:

- The relationship between the built environment and travel (BE-T)
- The relationship between travel and traffic safety (T-TS)
- The relationship between the built environment and traffic safety (BE-TS)

The relationship between travel and traffic safety is the most straightforward. Travel increases the exposure for all types of road users. High exposure bears a higher risk of traffic crashes, even in the safest urban environment and modes. The other relationships are not as straightforward as the BE-T. There are mixed results regarding the impacts of the built environment on travel (demand) and traffic safety. The built environment impacts the travel demand, and both travel demand and built environment impact traffic safety. Therefore, this study focuses on the relationship between the built environment and traffic safety (BE-TS) since the travel variables could also be evaluated under the built environment dimensions, and more variables could be considered together.

Depending on this context, the main aim of this study is to analyze the relationship between the built environment and traffic safety. The following objectives are organized according to this aim:

The first objective is trying to figure out which built environment factors have significant associations with traffic crashes. The second objective is to contribute to the literature regarding the findings of the stated relationship. Thirdly, while analyzing the relationship between the built environment and traffic safety, this study seeks to develop a generalizable methodology to evaluate different relationships in various contexts. The final objective is to provide an evaluation scheme for the value of information (data) available in some cases but not in others (i.e., developed country cases vs. developing country cases).

Depending on the presented objectives of this study, the following research questions are formulated to reach a conclusive frame:

- RQ1) Which built environment factors are related to traffic crashes?
- RQ2) Are the relationships between the built environment factors and varying types of crashes consistent (e.g., are the factors associated with fatal crashes similar to the aspects related to property damage only [PDO] crashes)?
- RQ3) Do the built environment factors affect varying types of crashes in the same way (e.g., whether the associations are consistently positive or negative)?
- RQ4) Are there any similarities or differences among varying study areas?
- RQ5) Are there any similarities or differences between the findings of this study and the findings in the literature?
- RQ6) Could there be a generalizable relationship pattern among the built environment factors and traffic safety?

3.2 Methodology

As stated in the aims, while analyzing the relationship between the built environment and traffic safety, this study tries to propose a generalizable methodology as follows:

- Extraction and classification of geospatial data
- Reducing the number of variables/dimensions through statistical analyses to reach a core set of variables/dimensions
- Applying a meta-heuristics method (tabu-search) to come up with the best subset of variables/dimensions explaining the variability in dependent-variables
- Evaluation of the results and comparing them with the findings of the similar studies

The general scheme of this methodological process is given in Figure 3.1.

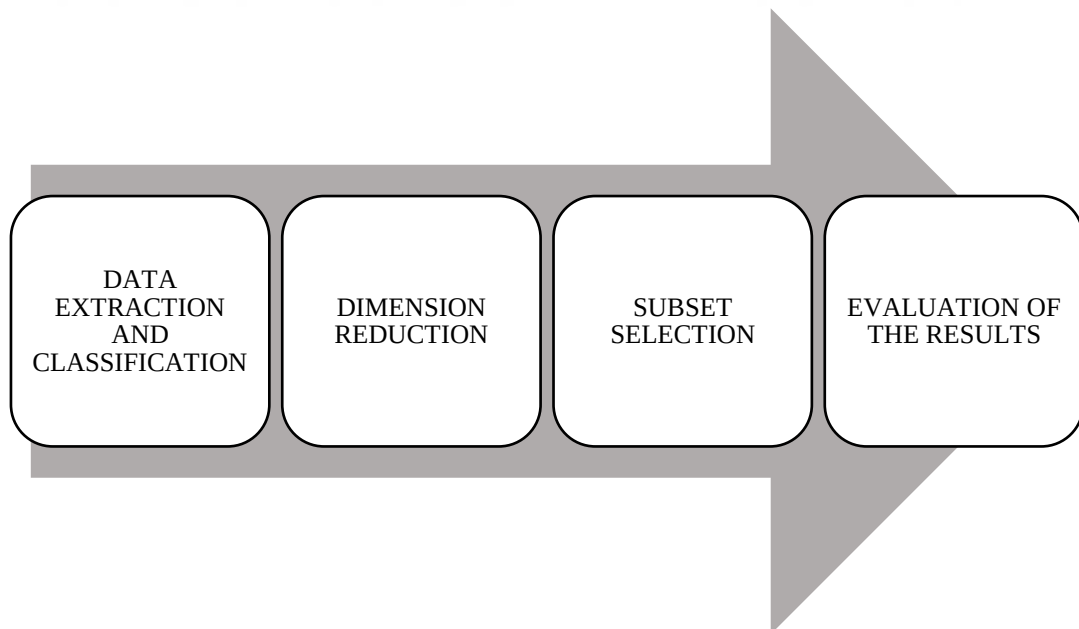


Figure 3.1. Proposed Methodological Process

3.2.1 Extraction and classification of geospatial data

Merriam-Webster dictionary defines geospatial as: “consisting of, derived from, or relating to data that is directly linked to specific geographical locations” (“Geospatial | Definition of Geospatial by Merriam-Webster,” n.d.). Thus, geospatial data has a geographic component, providing information about the location such as coordinates, address, postal code, city, and country (“What is the Difference Between GIS and Geospatial? - GIS Lounge,” n.d.).

Both crash and built environment data are geospatial. While doing a spatial analysis for crashes, it is essential to know where those accidents happen. For example, crashes usually occur at intersections or road segments, and the determination of crash “hot spots” on a given road requires spatial information about the accidents. Having spatial information is also valid for the built environment features/characteristics. Since most of the built environment data is geospatial, this data is prepared with spatial references or geocoded later on and presented concerning the geographic unit such as census blocks, block groups, tracts, counties, or states in the US, and neighborhoods, districts, provinces, or regions in Turkey.

The spatial references are required in geospatial data analysis. This requirement is twofold first; as mentioned above, the data must be related to a geographic unit; thus, the first requirement is related to the data generation. The second part is associated with the analysis. In spatial studies, such as examining the relationship between the built environment and traffic safety, it is central to decide on the unit of analysis, e.g., household, road segment, zone, or region. For example, a block group's office space density can be calculated by dividing the total office space in that block group by this block group's surface area. The resulting density profile is then assigned to the respective block group and its spatial reference. If all the data is available in the same unit of analysis, and this is the desired unit of analysis, it is convenient to proceed based on this analysis unit.

However, if all the data is not available in the same unit of analysis or the desired unit of analysis is not available, some conversion/assignment is required. This conversion would be an aggregation or a disaggregation method according to the spatial features of the data. For example, in terms of aggregation, the individual crashes aggregated to intersections for intersection level study. In Figure 3.2, possible level differences and conversion types are exemplified.

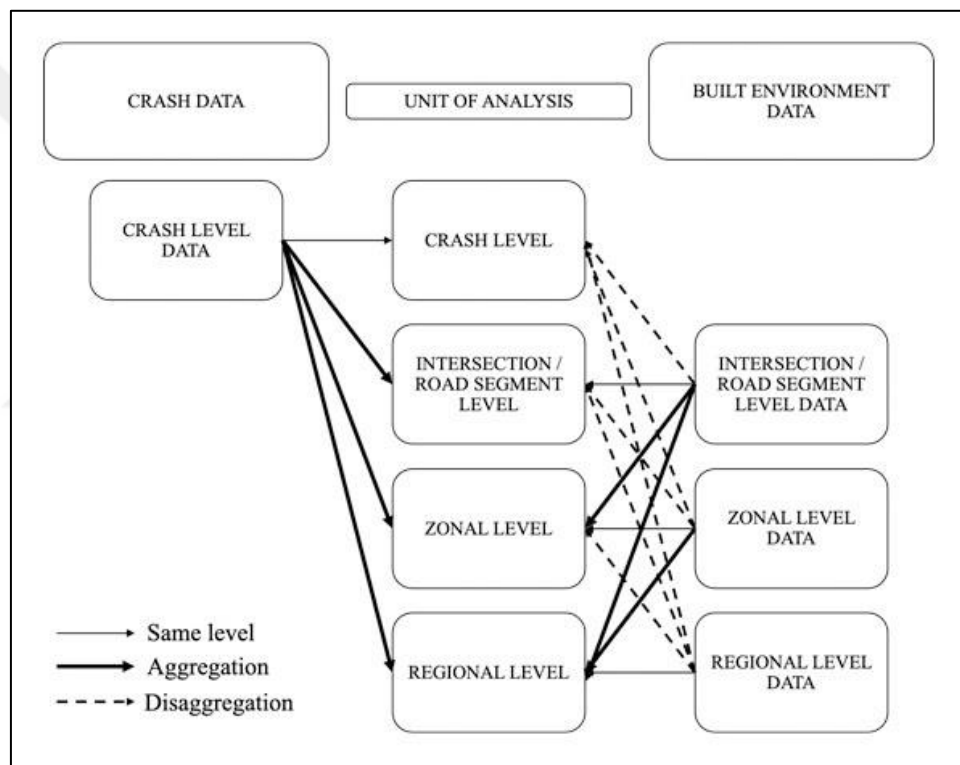


Figure 3.2. Possible Geographic Unit of Analyses and Conversion

For example, for an intersection/road segment-based analysis, the built environment features around the intersections/road segments should be assigned to the respective intersections/road segments. This assignment could be done by applying a buffer zone around each intersection/road segment or directly assigning the neighboring

units (area, neighborhood, block group). For example, Chen and Shen (2016) assigned the related TAZ's employment density and the land use mix within 50 m buffers of crash sites. Chen and Shen (2019) assigned population and employment densities of the related block groups and the length of bike lanes within a buffer of 50 feet to the crashes, and Clifton et al. (2009) used a quarter-mile buffer to assign block group properties to crashes (see Figure 3.3). Moreover, for any study, except crash-based ones, crash data should be extracted systematically, e.g., counting the number of crashes within a buffer zone of an intersection would be a method for an intersection level analysis. Two main crash data extraction techniques, crash frequency calculation and crash density estimation techniques, and their assignments to spatial units, will be discussed next.



Figure 3.3. 1/4 Mile Buffer Zone Defined for Each Crash (Clifton et al. [2009])

3.2.1.1 Crash frequency

“Substantive safety” is the term used for the actual or expected traffic safety. It could be quantified and operationalized by i) crash frequency, ii) crash rate, iii) crash type (e.g., vehicle, pedestrian, motorcycle), and iv) crash severity (e.g., fatal, incapacitating, nonincapacitating, and property damage only). Crash frequency is the count of crashes at a given intersection or road segment for a specified period of time. The crash rate is the normalized frequency to take exposure into account, where measures of exposure could be vehicle miles traveled (VMT), average daily traffic (ADT), or annual average daily traffic (AADT) (Federal Highway Administration, n.d.-b).

Crash frequency is a simple metric to operate with, while there are other methods to quantify crashes, such as determining property damage only equivalent to each more severe crash. Additionally, if the aim is to differentiate severity levels or crash types, frequency is proper. According to the review section presented in Chapter 2, crash frequency is the most common traffic safety indicator.

After determining the indicator to use, this indicator should be assigned to the geographic unit of analysis. For example, suppose the unit of analysis is the road segment. In that case, the number of crashes that happened at each road segment during a specific time period is summed and assigned to the respective segment. Most crashes happen at roads, so it is easier to assign crashes to segments, but an assignment method should be developed for areal units. This assignment method is required because the boundaries of most geographic units are roads. It is straightforward to assign the crashes within the geographic units and calculate the crash frequencies (i.e., by summing up), but the boundaries' crashes should also be assigned. The most common method is proposing a buffer zone around the geographic unit and assigning the crashes according to the buffer zones. There are two decisions to make for this method: i) the depth of the buffer zone and ii) whether to assign an occurrence to a single or several geographic units.

For the first decision, there is no valid for all value used in all studies. 200-ft is a shared buffer depth used by Dumbaugh and Li (2011), Dumbaugh and Rae (2009), Dumbaugh, Rae and Wunneberger (2011), Ouyang and Bejleri (2014) for block groups and Zhai, Huang, Xu and Sze (2019) for TAZs. Other studies employed different buffer depths, e.g., Huang, Wang, and Patton (2018) used a 50 m buffer around each block group, and Cai, Abdel-Aty, Lee, and Huang (2019) used a 250 ft buffer around traffic analysis districts (TAD). An example is given in Figure 3.4, in which a 200 ft buffer is applied to a block group border in Leon County, Florida. The crashes within the buffer zone are assigned to this block group, whereas the crashes out of the buffer zone (encircled with red) are not assigned.



Figure 3.4. Crash Assignment by Applying a Buffer Zone (produced by the author)

For the second decision, either the crashes falling into conflicting buffer zones are counted for each conflicting zonal unit (Dumbaugh et al., 2011; Y. Huang et al., 2018; Ouyang & Bejleri, 2014), or each crash is assigned to a single zonal unit (Saha et al., 2020). Though all the studies considered that adjacent areal units influence crashes, the latter approach was mainly proposed to eliminate overcounting crashes.

This study's approach is to apply a 125m buffer around block groups and assign each crash to all block groups within the buffer distance. Having a buffer distance that is higher than the interstate road width and counting the crashes for multiple block groups are selected to keep all the block groups nearby responsible for each crash. These assumptions would inflate the crashes at the city centers due to the smaller block group sizes; however, those parts of the cities are more problematic in terms of the activity concentration and the number of crashes. Thus, the mentioned emphasis is kept on purpose.

3.2.1.2 Crash density

Other than the crash frequency/rate, crash density is a measure of crashes that could be employed in crash analysis. Crash density is the crash count during a period per unit area and is mostly used to determine crash “hotspots.” Identification of concentrations or hotspots is salient in the detection of high-risk locations at which some incidents such as crimes, diseases, and crashes frequently occur; i.e., “hotspot analysis” has been widely employed in crime-focused studies for directing the security forces and focusing police patrol routes around the specific locations determined in these analyses, in similar fashion addressing issues related to public health (Deepika & Saradha, 2014; Hart & Zandbergen, 2014; Kloog et al., 2009).

In the crash frequency case, the crashes were assigned to respective geographic units. A similar methodology could be applied to the crash density case. After the crash assignment, the geographic unit area could be used to calculate crash densities per unit area. There is another method to employ; creation of particular boundaries,

calculating densities to those boundaries and then assigning to the geographic units. There are two standard-boundary creation techniques: creating Thiessen Polygons (Voronoi diagram/tessellations) and the other is creating grid cells. According to Bolstad (2012 cited in Quistberg et al., 2017), the Voronoi diagram is a spatial interpolation technique that allocates spatial areas around a given set of points proportionally, such that all possible locations in a cell is nearer to that cell than the other cells. For example, Wang and Kockelman (2013) built Thiessen Polygons around each census tract centroids (see Figure 3.5-a).

The grid cell structure is another method to distribute and analyze crashes. The grid system can also be used as a geographic unit of analysis (other than the zonal and regional units described in the previous chapter). Various studies have been conducted based on grids. For example, Kim, Made Brunner, Yamashita, Brunner and Yamashita (2006) developed a uniform grid structure consisting of 0.1 mile-squared grids to explore the impact of built environment characteristics such as land use and demographics on crashes. Kim, Pant, and Yamashita (2010) used a similar grid structure to examine the impact of land use and demographics on crashes and accessibility. Cai, Abdel-Aty, Lee, and Eluru (2017) developed various grids sizes from 1 mile-square to 100 miles-square and compared them with zonal systems (census tract, traffic analysis zones, and traffic analysis districts) for crash prediction (see Figure 3.5-b).

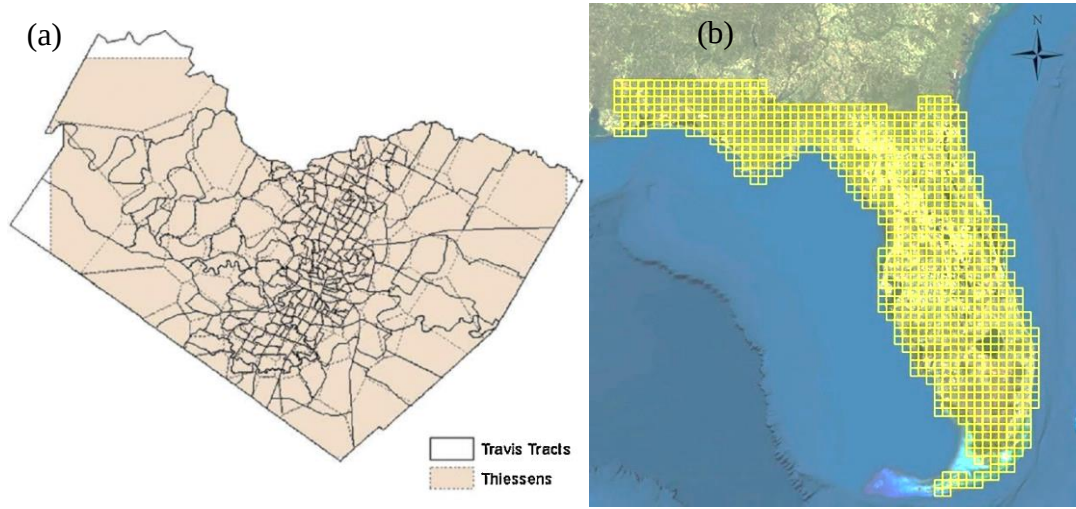


Figure 3.5. (a) Thiessen Polygons Applied to Census Tract Centroid in Travis County (Y. Wang & Kockelman, 2013) (b) 10x10 miles Grid Structure Applied to Florida (Cai, Abdel-Aty, Lee, et al., 2017)

For density calculation purposes, the grid structure is preferred to create a uniform base. This base is used in Kernel Density Estimation (KDE), which is a non-parametric method to estimate the probability density function (pdf) of a random variable, which is incidence points (crash points in this study). Thus, spatially-distributed point data is used to create density surfaces using KDE (Brunsdon, 1995). KDE analysis creates a density network having high and low-density spots (Ulak et al., 2017).

The purpose of implementing KDE is to estimate the shape of the density function of an independent and identically distributed sample from an unknown distribution. The kernel density estimator of this density function is given in equation (1) as follows:

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right) \quad (1)$$

Where $\hat{f}_h(x)$ is density estimation at position x , n is the number of observations, h is the bandwidth or smoothing parameter, $K(\cdot)$ is the kernel function, x_i is the position of the i^{th} observation. There are various kernel functions, such as uniform and normal, for estimating kernel densities. However, the kernel function's choice does not significantly affect the density estimation compared to the effect of bandwidth selection. The density function's shape is essentially defined by the bandwidth selection rather than the kernel function (Ulak, 2018).

For an urban network, bandwidths between 50 meters to 300 meters are considered appropriate (Steenberghen et al., 2010; Z. Xie & Yan, 2013, as cited in Ulak et al. 2017). For rural regions, a bandwidth of 1000 meters is appropriate (Blazquez & Celis, 2013). In this study, a bandwidth of 300 meters was applied, which was cited as appropriate in a former study with a similar context. In this former study, spatial analysis of aging population involved crashes were done for case studies from Escambia, Leon, and Bay counties, Florida, the US (Ulak et al., 2017). The study also mentioned that one-tenth of the bandwidth is appropriate for the cell width (Okabe and Sugihara, 2012, as cited in Ulak et al., 2017). Therefore, the cell width is selected as 30 meters (one-tenth of 300 meters). After the density surface is created, density values could be assigned to the desired spatial unit, such as block groups. Moreover, as in the case of crash frequency (count) assignment, to generate emphasized crash estimates, total kernel density figures are used in the regression models as dependent variables.

After the preparation of independent and dependent variables, dimension reduction, subset selection, and evaluation stages follow, depicted in Figure 3.6 and described in the following sections.

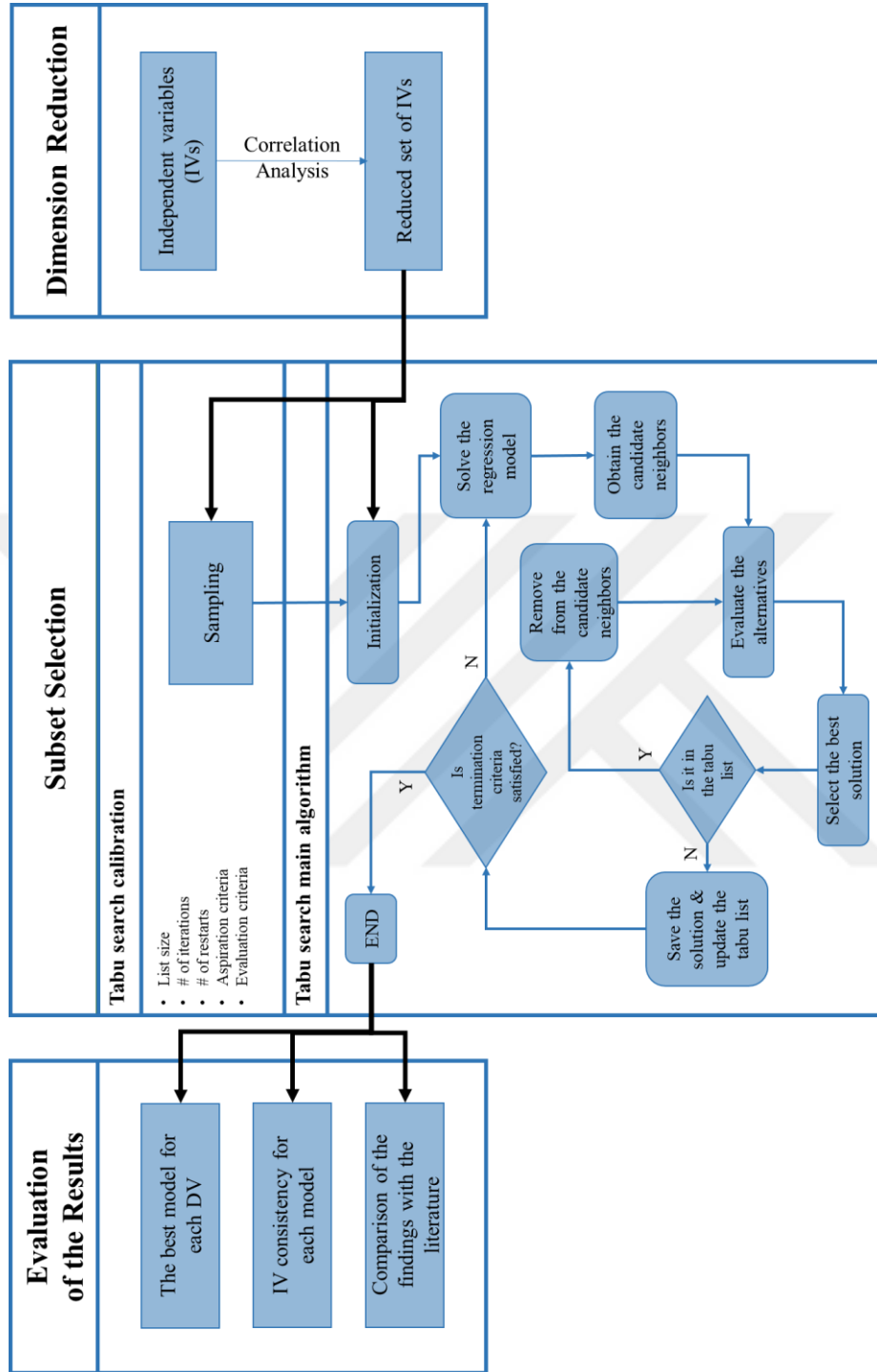


Figure 3.6. Flowchart of the algorithm

3.2.2 Dimension reduction

In statistical analyses, each variable in the model constitutes a dimension in the solution space. For the studies considering many variables, thus many dimensions, there would be problematic instances due to those variables' relations. For example, in a regression analysis, the main aims are to develop the best relational model whose independent variables can explain the maximum variability in the dependent variable and estimate the outcomes for unobserved instances with minimum error. If the independent variables are correlated, their contributions are disputable. In some cases, variables become redundant since these variables are linear combinations of other variables, known as singularity; and in some other cases, they are strongly correlated, known as multicollinearity (Tabachnick & Fidell, 2007).

Tabachnick and Fidel (2007) define correlation as “the measure of the size and direction of the linear relationship between the two variables,” so it is used to quantify the association between variables. Suppose there is a positive correlation between two variables; an increase in one variable's values is associated with an increase in the other variable's values. However, if there is a negative correlation between two variables, an increase in one variable associate with a decrease in the other variable. The mostly used correlation coefficient, Pearson product-moment correlation coefficient, can be calculated by the following equation (Tabachnick & Fidell, 2007):

$$r = \frac{N \sum XY - (\sum X)(\sum Y)}{\sqrt{[N \sum X^2 - (\sum X)^2][N \sum Y^2 - (\sum Y)^2]}} \quad (2)$$

Where N is the number of observations, X is the vector of one variable, Y is the other variable vector. The value of r is in the interval of [-1,1], where 1 is associated with perfect positive correlation, and -1 is associated with the perfect negative correlation, while 0 represents no correlation.

If the correlation between variables is too strong, the problem of multicollinearity occurs. According to Dohoo, Ducrot, Fourichon, Donald, and Hurnik (1997) and Tabachnick and Fidel (2007), for the cases with absolute correlation coefficients over 0.9, multicollinearity is almost undisputed to be a problem. However, there is no definite minimum level to consider multicollinearity. For example, L. Wang, Abdel-Aty, Lee and Shi (2019) eliminated one of the two variables that have an absolute correlation coefficient higher than 0.3; Amoh-Gyimah, Saberi, and Sarvi (2016), and Wen, Sun, Zeng, Zhang and Yuan (2018) set the threshold to 0.6. Gao, Kamphuis, Helbich, and Ettema (2020) used 0.8 as the maximum absolute correlation level to have both variables in the model. Since there is no consensus on the minimum value, the researchers determine their own levels considering the severity of the multicollinearity problem and the variables that have very important in the study.

In this study, more than 200 variables were considered. Since the number of variables was too much, and some of the variables were derived as linear combinations of each other (having singularity problem), the number of variables was reduced. One method of dimension reduction is correlation analysis, as discussed above. Most of the available built environment variables are complementary and correlated. For example, the number of people employed in an area is related to the employment opportunities in the same area. Auto ownership is closely related to the amount of income dedicated to car maintenance. The strongly correlated variables were eliminated in this study by setting the absolute maximum value of tolerable correlation as 0.75.

3.2.3 Subset selection

After creating datasets, the proposed methodology focused on selecting the best subset of independent variables. There is a large amount of data about the built environment and its main dimensions, depending on the geographic unit of analysis and the case area. It is very intriguing to make a model using all available variables. However, this creates a variety of data analysis problems (Broadhurst et al., 1997):

- Some variables may be entirely unrelated to the model's goals and may blur any significant relationships between other variables.
- To obtain consistent estimates of the parameters, each variable's number of observations should be significantly more than the number of variables.
- Variables can be associated, in which case the duplicate information is unnecessary.
- The signal-to-noise ratio of certain variables may be so low that their incorporation into the model is questionable, especially if other “cleaner” correlated variables are present.
- When model parameters are optimized using iterative methods, more parameters can result in a more complex error surface to optimize. This complexity can affect the model's overall convergence time.

Thus, using a subset would increase the model's accuracy by eliminating the redundant variables, describing the dataset parsimoniously, and estimating regression coefficients with smaller standard errors (which is more serious when correlated variables exist) (Miller, 1984).

This study aims to develop the best subset of a comprehensive set of built environment variables while examining their relationships with traffic accidents. The evaluation criteria would depend on the model; for example, the best subset in a multiple linear regression model would minimize errors (Chung et al., 2020) or keep the minimum set of explanatory variables without compromising the model's explanatory power (Örkcü, 2013). Moreover, the built environment variables' associations with the traffic safety variables should be evaluated in terms of consistency or dependability. For example, it is important to observe whether a variable has a consistently positive or negative relationship with traffic crashes or its relation is affected by other variables to vary in direction. Therefore, rather than solving a regression model once with all variables, an iterative process that considers a subset of variables each time would be preferred.

Subset selection is a combinatorial optimization problem. Stepwise regression, forward selection, and backward elimination are widely used iterative greedy methods (Broadhurst et al., 1997). However, since those algorithms are greedy, they proceed with the highest increases or decreases, which would stack at a local minimum or maximum. Thus, other heuristic algorithms, such as genetic algorithms (Örkcü, 2013) or Tabu search, are viable alternatives (Bae et al., 2016), yielding better solutions and permitting the tracing of the behavior of the variables throughout the solution process.

3.2.3.1 Tabu-search based subset selection

In this study, the tabu-search algorithm proposed in Fouskakis and Draper (2002) is utilized to determine the best set of independent variables that optimize various multiple regression models' objective value. Tabu search (TS), initially proposed by Glover (1986), is a meta-heuristic approach for solving optimization problems. Tabu search is a stepwise local search procedure during which some moves are enlisted in a tabu-list, so being restricted/prohibited as the name “tabu” refers. Thus, rather than applying a tabu move, the algorithm would proceed to an inferior solution until a termination criterion is satisfied. This tabu search property could prevent the algorithm from being trapped in a local optimum, which is the main distinction from the greedy and improvement type searching algorithms (Drezner et al., 1999).

The TS procedure needs some definitions and parameters. These are the evaluation criteria of the solutions (e.g., the objective function value), the initial solution, the termination criteria (e.g., the number of iterations), the aspiration criterion (e.g., comparison of the solution to the best objective function value) and the tabu-list size. Accordingly, the generalized pseudocode of the TS algorithm is given in Figure 3.7.

Begin with an initial solution and an empty tabu-list, and set the solution as the best solution (set the solution value as the aspiration value).

While (any termination criterion is not met) **Do**:

1. Obtain the neighbors of the current solution.
2. Select the best neighbor that is not in the tabu-list or satisfies the aspiration criterion (if defined). Set this neighbor as the current solution.
3. Update the tabu-list, the aspiration value (if changed), and the best solution (if changed).

Return the best solution.

Figure 3.7. Tabu-search Pseudocode

Tabu search algorithm has been employed as a solution procedure in numerous optimization problems such as crew scheduling (Sargut et al., 2017), vehicle routing problem (H. Kim et al., 2009), traffic forecasting (Hong et al., 2006). Though not commonly utilized, the selection of regression variables is one of those problems for which the tabu-search algorithm could be employed. In Drezner et al. (1999) and Hage, Hage, Ghnatios, and Hamade (2018), tabu-search is used to determine the optimal set of independent variables of multiple linear regression models by minimizing the significance level; in Drezner et al. (1999), the results are compared to stepwise regression results, and in Hage et al. (2018) the results are compared to the benchmarked ones. In Pacheco, Casado, and Núñez (2009), tabu-search is utilized to select a subset of variables for logistic regression models, and the results are compared to the classical methods (stepwise, forward selection, and backward elimination). In Ghnatios, Hage, and Hage (2019), tabu-search is employed to optimize the least-squares regression kernel (functions of independent variables).

As given in Fouskakis and Draper (2002), the algorithm begins with an initial solution and proceeds stepwise with the evaluation of the neighbors. The moves

taken are kept in a limited tabu-list (of size s) for a number of iterations unless the aspiration criterion (yielding a better solution than the best solution obtained so far) is satisfied. Accordingly, a move kept in the list leaves the list either it satisfies the aspiration criterion or the list is full (the first in-first out rule applies). The algorithm consists of three main stages: preliminary, intensification, and diversification. All the moves described above are applied in the preliminary search stage until a termination criterion (either time limit or the number of iterations) is satisfied. The intensification stage starts with the best solution obtained in the preliminary search stage with an empty tabu-list. In this stage, the same preliminary search is applied, and if a better solution is obtained, it restarts. The number of restarts is defined by the user. Finally, the diversification stage starts with a tabu-list consisting of the most frequently used s moves and a random state. Thus, this stage gives the chance to search in the infrequently visited solution space. After a termination criterion for this stage is satisfied, the overall algorithm's best solution is returned. The pseudocode of the tabu search algorithm employed in this study is given in Figure 3.8.

Initialization:

Begin with an initial solution and an empty tabu-list, set the solution as the best solution, and so the aspiration value.

Body:

While (the maximum number of restarts for the whole algorithm is not reached) **Do**:

Preliminary Search:

While (the maximum number of iterations for the preliminary search stage is not reached) **Do**:

1. Evaluate the objective values of the neighbors of the current solution and keep them in the available neighbors set.
2. Select the neighbor with the best objective value from the available neighbors set.
 - 2.1. If the move is not in the tabu-list, set it as the current move;
 - 2.1.1. If the current solution value is better than the aspiration value, update the aspiration value as the current solution value;
 - 2.1.2. Else, do not update the aspiration value.
 - 2.2. Else (current move is in the tabu-list),
 - 2.2.1. If the current solution is better than the aspiration value, remove the move from the tabu-list, and update the aspiration value.
 - 2.2.2. Else (the current solution is not better than the aspiration value), remove the neighbor from the available neighbors set, and go to step 2.
 - 2.3. Add the selected move to the end of the tabu-list.
3. Go to Intensification with the best solution found.

Intensification:

4. **While** (maximum number of restarts for the intensification is not reached) **Do**:
 - 4.1. Empty the tabu-list.
 - 4.2. Repeat Preliminary Search (1-3) Until a better solution value than the aspiration value is found OR the maximum number of iterations is reached
5. Go to Diversification.

Diversification:

6. Empty the tabu-list.
7. Add the most frequent s moves to the tabu-list.
8. **Begin** with a random state.
9. Repeat Preliminary Search (1-3) Until the maximum number of iterations for the diversification is reached.

End:

Return the best solution.

Figure 3.8. Tabu-search Pseudocode in this Study

In heuristic algorithms such as tabu search, parameter selection is essential since selecting different parameters can yield very different results. Thus, the tabu-search algorithm requires parameters to be determined in advance. If benchmarked parameters do not exist, they should be calibrated by applying the algorithm on the same or a sample dataset. Several combinations of parameters should be tested to determine the best setting for the dataset. According to the proposed algorithm, the following parameters should be calibrated, the number of neighbors, the list size, the termination criteria (the number of iterations), the number of restarts, and the objective function.

3.2.3.2 Fitting statistical distributions

In this study, there are too many explanatory variables available, and the set of variables used vary for each study in the literature. This study aims to select the best set of variables to explain the relationship between the built environment and traffic crashes at a specific time and place and try to discuss the generalizability of the results. This problem is embedded in the proposed tabu search algorithm, an iterative tool for solving optimization problems. Ergo, the problem is solving multiple regression models many times in which various statistical distributions could be fitted to come up with the best set of explanatory variables for each setting.

Among the alternatives, negative binomial modeling was chosen. Poisson and negative binomial modeling are widely used in traffic safety data analysis, while tabu search implementation is the novel part.

Since crash frequencies on a specific analysis unit (e.g., a highway segment, population block group, census tract, or a city) take non-negative integer values during a specific period of time. Thus, count-data modeling techniques (e.g., Poisson and negative binomial regression) are solid alternatives (Milton & Mannering, 1998). The following equation (3) defines the Poisson regression model:

$$P(y_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \quad (3)$$

where, $P(y_i)$ is the probability of having exactly y_i (a nonnegative integer) crashes at unit i (population block group in this study), and μ_i is the expected crash frequency at unit i (i.e., $\mu_i = E(y_i)$). The expected crash frequency at the unit i could be represented by the equation (4), where X_i is a vector of explanatory variables (built environment variables) and β is a vector of parameters to be estimated (noting that the variance is equal to the expected value in Poisson regression):

$$\mu_i = e^{\beta X_i} \quad (4)$$

Since accident frequency data are overdispersed by having a variance that exceeds the mean, the negative binomial model is more appropriate than the Poisson model, in which variance is assumed to be equal to the expected value. The following equation (5) defines the negative binomial model:

$$P(y_i) = \frac{\Gamma(y_i + \frac{1}{\alpha})}{\Gamma(y_i + 1)\Gamma(\frac{1}{\alpha})} \left[\frac{1}{1 + \alpha\mu_i} \right]^{\frac{1}{\alpha}} \left[\frac{\alpha\mu_i}{1 + \alpha\mu_i} \right]^{y_i} \quad (5)$$

Where Γ is the gamma function, μ_i is the mean value of unit i , and α is the dispersion parameter ($\alpha \geq 0$). Thus, as the variance formula (6) shows, the variance could be higher than the mean to capture dispersion (when α converges to 0, the model approximates to the Poisson model) (Miaou, 1994).

$$Var(y_i) = \mu_i + \alpha \mu_i^2 \quad (6)$$

Poisson and Poisson gamma/negative binomial distributions are applied to discrete data with no negative values. Since crash frequency is discrete, these distributions are proper; however, due to the overdispersion of crash data and its property of having positive outliers, Poisson gamma/negative binomial distribution is more suitable (“Statistical Distributions,” n.d.).

For the continuous nonnegative data case, Poisson and negative binomial distributions are not appropriate. For continuous, nonnegative, right-skewed data, Gamma, Lognormal and Weibull distributions are appropriate (“Statistical Distributions,” n.d.). The properties of Gamma, Lognormal and Weibull distributions are given in Table 3.1 (Siswadi & Quesenberry, 1982). These were the distributions applied to fit crash density data in this study.

Table 3.1 Applied Statistical Distributions and Their Parameters

Distribution	Probability Density Function (PDF)	Parameters
Gamma	$\frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$	k, θ
Lognormal	$\frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}$	μ, σ
Weibull	$\frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}$	k, λ

As Gamma, Lognormal and Weibull distributions are appropriate for nonnegative continuous crash data, these three distributions are employed for the regression fits for crash densities.

3.2.4 Evaluation of the results

The results are evaluated according to the research questions formulated in this chapter. First of all, for each case and each setting, the model fit evaluations are done. The goodness of fit statistics, given in Table 3.2, are used (Cox & Snell, 1998; de-Graft Acquah, 2010; Nagelkerke, 1991). For Akaike information criterion and Bayesian information criterion-based comparisons, smaller values are better, whereas, for R^2 -based comparisons, higher values are better. Though the fit statistics are provided, the model selection is based on the Akaike information criterion (the models that minimize AIC value are selected).

Table 3.2 Fit Statistics for Regression Models

Statistic Name	Formula	
Akaike Information Criterion (AIC)	$AIC = -2 \ln(L) + 2p$	L: log-likelihood estimate p: number of parameters
Bayesian Information Criterion (BIC)	$BIC = -2 \ln(L) + p \ln(n)$	L: log-likelihood estimate p: number of parameters n: number of observations
Likelihood-ratio based R^2 ("Cox & Snell R^2 ")	$R_{Cox \& Snell}^2 = 1 - \left[\frac{-2L_0}{-2L_1} \right]^{2/n}$	L_0 : likelihood of the null model L_1 : likelihood of the full model n: number of observations
Corrected- R^2 ("Nagelkerke R^2 " / "Cragg & Uhler's")	$R_{Nagelkerke}^2 = \frac{1 - \left[\frac{-2L_0}{-2L_1} \right]^{2/n}}{1 - (-2L_0)^{2/n}}$	L_0 : likelihood of the null model L_1 : likelihood of the full model n: number of observations

The best regression models (with minimum AIC value) obtained by the tabu search procedure are then evaluated at the variable level. The proposed tabu-based subset selection algorithm yields the best subset of independent variables (for each BE-TS

model), and the consistency of each variable could be traced iteration by iteration. During the tabu search procedure, the number of times a variable is in the model is an indicator of consistency. Not only being in the model but also the significance level is important. Suppose a variable consistently selected to be in the model with the desired significance level with the same relation type (positive or negative). In that case, this variable could be considered consistent in terms of the specific built environment (BE) and traffic safety (TS) relationship. On the contrary, if a variable is not selected consistently or its significance level is not high, that variable would not be consistent, according to this evaluation frame. For each BE-TS model, each BE variable is traced accordingly.

There are thresholds to be determined; for example, the minimum percentage of the iterations in which a variable is basic (selected to be in the model at a specific iteration) could be considered consistent. In this study, considering the preliminary search (since the algorithm starts with a random set, approximately 10% of the iterations are required to reach a good solution) and diversification (since this stage starts with restricted moves, again approximately 10% of the iterations are required to reach a good solution) stages, 80% was selected as a threshold value (note these percentages are based on the observations specific to this study). Since there is no predetermined value for that type of argument, the process is flexible and could be reconfigured for different settings. Moreover, for consistency, the variables should have a dependable relation (positive or negative) with the dependent variables. Therefore, of the basic variables in more than 80% of the iterations, the ones with the same relation type (positive or negative) in at least 90% of the iterations are assumed to be consistent. This assumption is similar to the minimum significance level assumed in the regression models of this study (p-value should be less than 0.1).

One other aspect of consistency is the support from the literature. The results of each case are also compared to the literature findings to evaluate the similarities and dissimilarities obtained.

The key BE variables are determined in terms of the generalizability of the results and the evaluation of the selected subsets. This determination is done by comparing each model's significant and consistent variables within each case and then comparing the results case-wise. Then the model performance of each BE-TS relationship is evaluated when the key BE variables are used. The same procedures are done for both frequency and density estimates of TS variables. For each TS-BE model, when the TS variables are formulated as kernel densities, Gamma regression, Lognormal regression, and Weibull regression models are solved, and the results are compared according to the fit statistics presented in Table 3.2 and the additional fit statistics presented in Table 3.3 (D'Agostino & Stephens, 1986). According to the comparisons, one distribution is selected to be employed in further models, such as measuring the key BE variables' performance and modeling other case studies.

Table 3.3 Additional Goodness of Fit Statistics

Statistic	Formula	Parameters
Root mean squared error*	$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - E_i)^2}{n}}$	O_i : value of the i^{th} observation E_i : expected value for the i^{th} observation n : sample size (number of observations)
Mean absolute error*	$MAE = \frac{\sum_{i=1}^n O_i - E_i }{n}$	O_i : value of the i^{th} observation E_i : expected value for the i^{th} observation n : sample size (number of observations)
Mean absolute percentage error*	$MAPE = \frac{\sum_{i=1}^n \frac{ O_i - E_i }{O_i}}{n}$	O_i : value of the i^{th} observation E_i : expected value for the i^{th} observation n : sample size (number of observations)
Normalized root mean squared error* (by standard deviation)	$NRMSE_{sd} = \frac{RMSE}{\sigma}$	$RMSE$: root mean squared error σ : standard deviation of the observations
*Lower values indicate better model fit		

3.2.5 The frame of the methodology

The general frame of the methodology introduced in this study is summarized in Table 3.4. The first three of the objectives are related to the review chapter. The following two objectives are related to the methodology introduced in this chapter, and the others are related to the case study applications.

Table 3.4 Research Frame (adapted from Wang, 2010)

Research Objectives	Research Method	Chapter #
Identifying the factors to define traffic safety (TS)	Review of TS indicators used in previous studies	Chapter 2
Identifying various built environment (BE) factors associated with traffic safety (TS)	Review of BE factors used in previous studies	Chapter 2
Identifying various mathematical models to analyze the relationship between the built environment and traffic safety	Review of relevant literature for mathematical models	Chapter 2
Preparing the data for analysis	Classification of the TS and BE data. Application of GIS operations to assign BE and TS data to respective analysis zones.	Chapter 3
Reducing the variable dimensions for analysis	Correlations analysis	Chapter 3
Developing the BE-TS relationship for developed county cases	Tabu-search based subset selection	Chapter 4
Evaluating the practical implications of the models	Comparison of the results with the literature and discussion of the findings	Chapter 4
Determining the key BE variables	Significance and consistency-based comparison	Chapter 4
Investigating the BE-TS relationship for a developing country case	Tabu-search based subset selection	Chapter 5
Comparing the findings of developing county case to developed county cases	Regression models and R^2 -corrected based scheme	Chapter 5

3.3 Study Area

The proposed methodology is applied to case studies in the State of Florida, the United States. The case studies' selection from a developed country is twofold: the first is related to the data (both availability and reliability of data). Since this study's focus is not collecting primary data but rather researching secondary data, data availability and reliability are significant concerns. The second is related to the accumulated literature. According to the comprehensive literature review presented in the previous chapter, both the theories about urban growth patterns are structured depending on the developed country context (especially in the U.S. case), and the majority of the case studies analyzing the relationship between the built environment and traffic safety were developed country cases. Thus, to create a generalizable methodology and compare the findings with the ones in the literature, the United States cases are selected.

The first case study area is Tallahassee, Florida's capital city, located in Leon County in the northwestern part of Florida. According to the U.S. Census 2010, the total population of Leon county is 275,487 (see Table 3.5).

Table 3.5 Information on the Case Study Areas

	SINGLE COUNTY	TRI-COUNTY			
	Leon	Broward	Miami- Dade	Palm Beach	Overall
Population in 2010	275,487	1,748,066	2,496,435	1,320,134	5,564,635
Population in 2019	293,582	1,952,778	2,716,940	1,496,770	6,166,488
Total area (mi²)	667	1,210	1,898	1,970	5,077
Population density in 2010 (person per mi²)	413	1,445	1,315	670	1096
Number of population block groups	177	939	1,593	884	3,416

The study area (with county borders of Florida), the urban areas in Leon county, and block group borders are given in Figure 3.9.

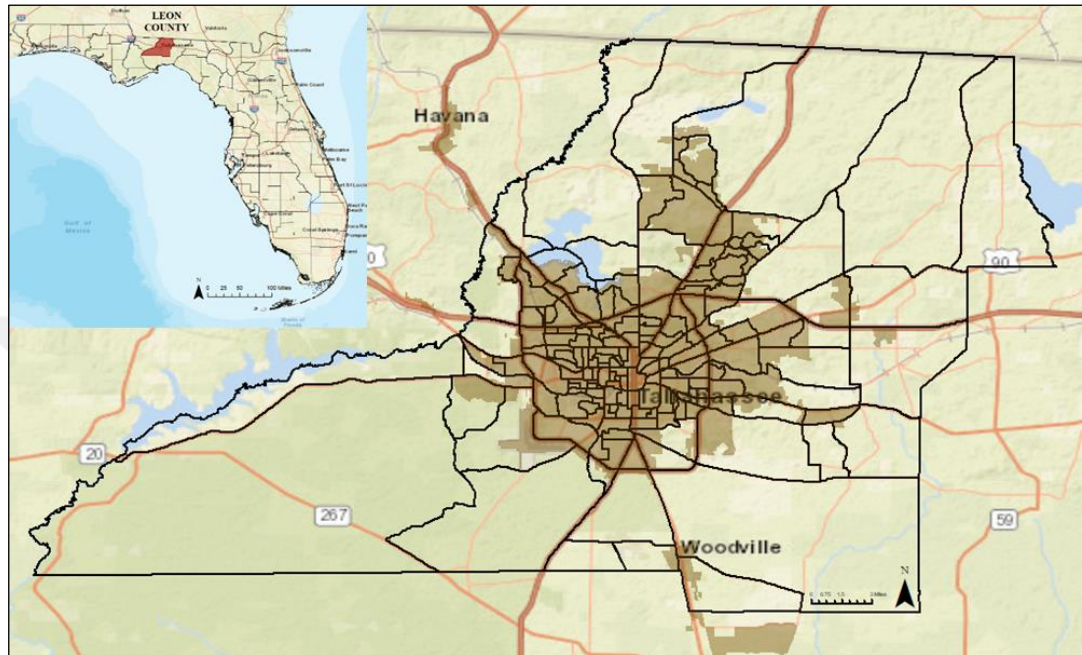


Figure 3.9. Population Block Groups and Urban Region in Leon County

The second case study area consists of the three most populous counties of the State of Florida. Palm Beach, Miami-Dade, and Broward counties are considered together as an extended urban agglomeration in this case (see Figure 3.10).

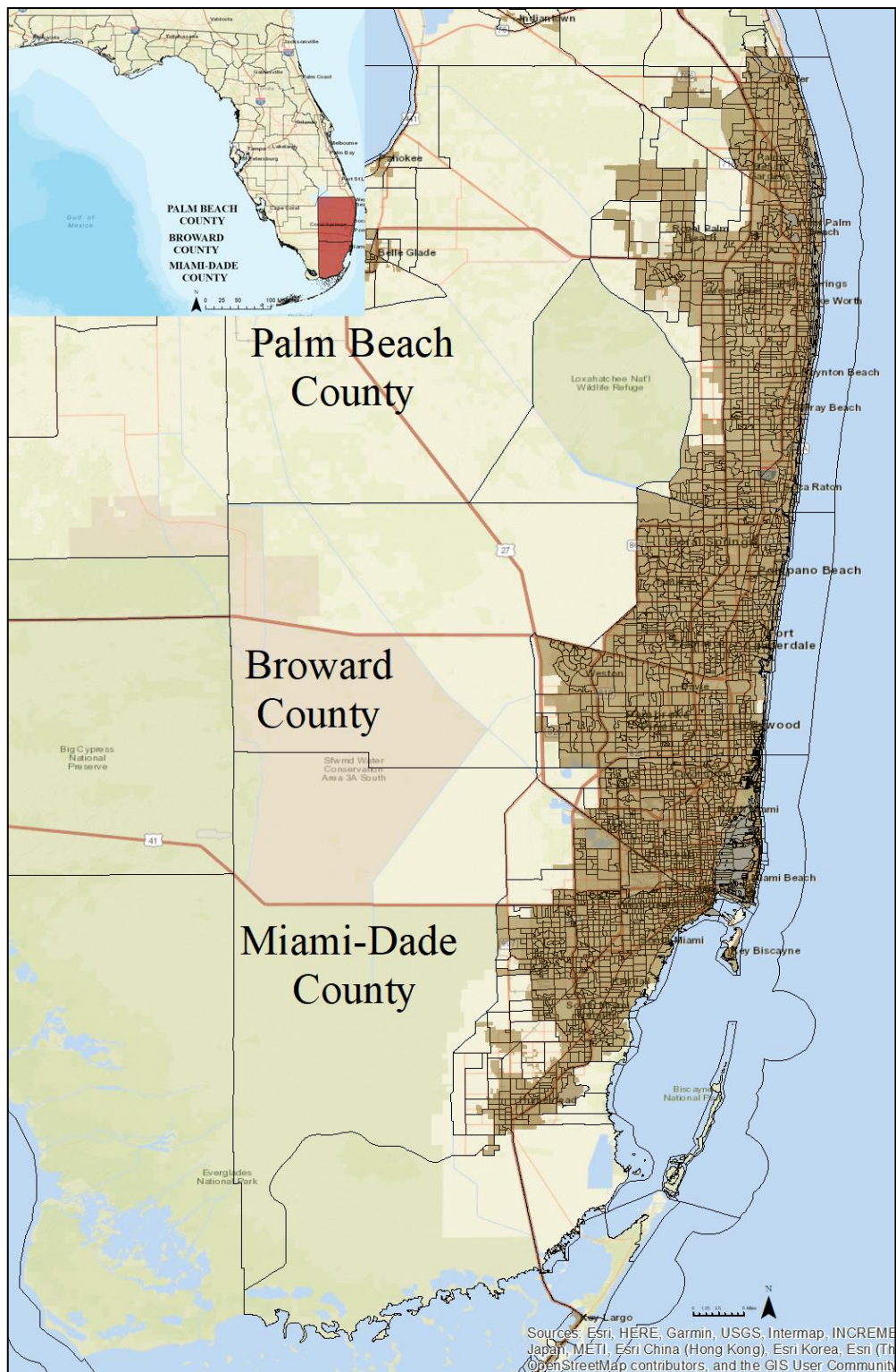


Figure 3.10. Population Block Groups and Urban Region in Tri-County Case

The cumulative population of these three counties in 2010 is 5,564,635 (see Table 3.5). This tri-county structure is the largest urban agglomeration in the State of Florida.

The selection of these two particular cases is related to one of the aims of this study as coming up with a generalizable methodology and research questions of determining the common factors that explain the relationship between the built environment and traffic safety. Accordingly, the selected cases have different structures. Leon county is relatively small and has a single urban center. As being the capital city and having three international universities, Tallahassee inhabits many students and public employees. In contrast, the tri-county case is a large and extended urban structure with multiple urban centers in each county. These are coastal counties and attractive touristic centers. Thus, analyzing these two distinct structures would give insight into the commonalities and differences in varying settings.

3.4 Data Extraction and Classification for the Study Area

In this section, the extraction and preparation of the datasets are presented. Traffic safety variables are the dependent variables, and the built environment variables are the independent variables of this study. The preparation processes for both dependent and independent variable sets are explained in detail.

3.4.1 Traffic safety dataset

The dataset mainly consisted of the crash records for Leon County, Broward County, Miami-Dade County, and Palm Beach County between 2009 and 2011. It was gathered from the Florida Department of Transportation (FDOT, 2015). In Figure 3.11a and Figure 3.11b, crashes according to the crash typologies presented for the single county case.

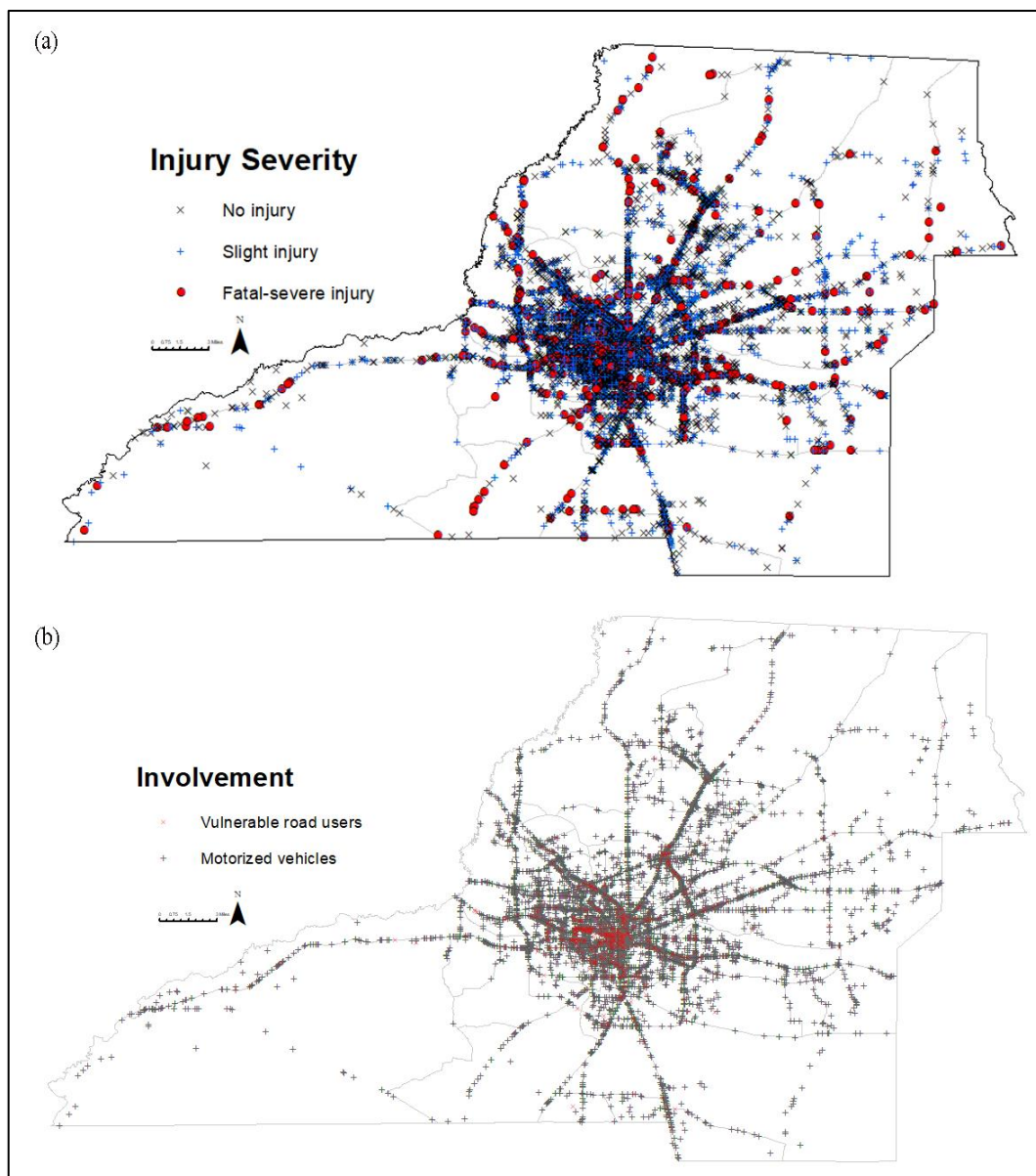


Figure 3.11. Crashes According to Involvement and Severity for the Single County Case

Similarly, in Figure 3.12a and Figure 3.12b, crashes according to the crash typologies presented for the tri-county case are presented.

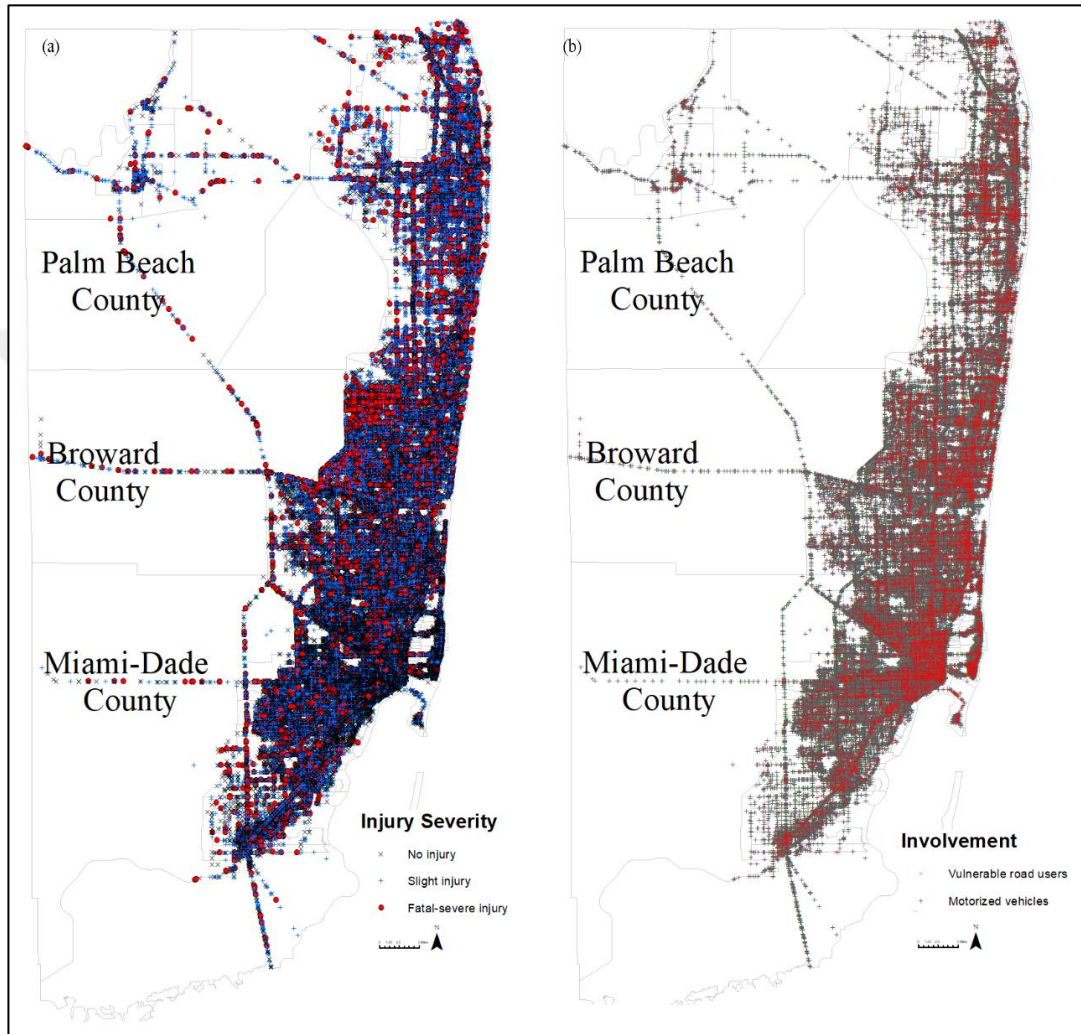


Figure 3.12. Crashes According to Involvement and Severity of the Tri-County Case

According to the crash distributions presented in Figure 3.11a and Figure 3.12a, it could be observed that vulnerable road user involved crashes are densified around the city centers relative to the motor vehicle crashes. According to the second

typology presented in Figure 3.11b and Figure 3.12b, it could be observed that, though the number of fatal crashes is quite smaller than the number of slight injury crashes and no injury crashes in general, fatal crashes are reasonably equally distributed around the peripheral parts in both the single county and tri-county cases. The increased speeds could explain this at the higher-order roads with a higher share of the peripheral parts.

The next section will elaborate on the crash frequency calculations and the spatial distribution of the crashes.

3.4.1.1 Crash frequency

The crash data were classified according to severity and involvement. In Table 3.6, crash typology and respective crash frequencies are given for the single county and tri-county cases, respectively. The total crash counts for each category are presented under the “original” column. This is the cumulative number of crashes distributed to counties as defined in the dataset originally (for the 2009-2011 period).

Table 3.6 Crash Frequencies in the Study Areas

CRASH TYPE	SINGLE COUNTY			TRI-COUNTY		
	Crash count			Crash count		
	original	assigned	assigned / original	original	assigned	assigned / original
S1-Property damage only (PDO) crashes	7,706	17,106	2.22	167,944	422,934	2.56
S2-Possible injury or non-incapacitating injury crashes	5,891	13,474	2.29	92,992	232,721	2.58
S3-Incapacitating injury or fatal crashes	601	1,270	2.11	12,423	30,400	2.51
ALL- Motorized and non-motorized	14,198	31,870	2.24	273,359	686,294	2.56
MOT-Motor-vehicle crashes only	13,695	30,767	2.25	262,313	658,868	2.56
VRU-Vulnerable road user crashes	503	1,103	2.19	11,046	27,426	2.58
	Average: 2.22			Average: 2.56		

Under the “severity typology”, there are three severity levels: S1, S2, and S3 stand for crashes with no injury, crashes with possible or non-incapacitating injury, and crashes with incapacitating injury or fatality within 30 days.

The “involvement” typology is about the parts influenced by the crashes. It all stands for all types of crashes. The other two classes are motorized vehicle crashes and vulnerable road user crashes (involving pedestrian and bicycle user crashes). As expected, the highest number of crashes are of no injury cases for the first typology, and the lowest number of crashes are of the highest severity level. For the second typology, very few of the accidents are about vulnerable road users. Additionally, these patterns are valid for both cases.

In Table 3.6, the numbers presented under the “assigned” column are the cumulative number of crashes assigned to population block groups according to the assignment method introduced in this chapter. The crash counts for the 2009-2011 period are assigned to the census block groups within a 125m buffer zone, and each crash is assigned to all block groups within the buffer distance. Since each crash is assigned to more than one population block group, if it is within the buffer distance of more than one block group, the number of crashes under the “assigned” heading is always higher than the “original” number of crashes.

For this assignment, a normalization technique is not applied since the dependent variable is assumed to be the crash count for a specific geographic unit (census block group in this case). As dependent variable crash rate (e.g., crash count per year per road segment, crash count per daily vehicle miles traveled) is also common, but in this study, the assigned crash counts are used as in some former studies (Ouyang & Bejleri, 2014; Y. Wang & Kockelman, 2013). The assigned crash counts are used as dependent variables because the normalizing variables such as population, area, or traffic volume are used as explanatory variables.

One other aspect related to the dependent variables and the assignment technique is that the multiple counting of each crash exaggerates the counts due to the mentioned buffer zone assignment technique. However, this is aimed in this study because the

surrounding block groups within a reasonable distance would be responsible for each crash. This reservation is particularly valid for the smaller block groups usually situated around the cities' central parts. Built environment features such as density or employment density of a neighboring geographic unit would influence the exposure by generating more activity.

Concerning the assignment rates, the average crash assignment in the single county case is less than the average crash assignment in the tri-county case, 2.22 and 2.56, respectively (note that the ratios and the average assignment rates are presented to discuss the behavior of the buffer zone assignment technique). This is due to the number and sizes of block groups. When compared (Figure 3.9 vs. Figure 3.10), there are more small-sized block groups in the tri-county case, so on average, each crash is assigned to more block groups by using the same buffer distance. This ratio is the smallest (2.11 and 2.51 for the single county and tri-county cases, respectively) for the fatal crashes (S3), which would be due to the spatial distribution of the crashes presented at the block group level in Figure 3.11.

According to the distribution of crashes, fatal crash frequencies are higher at peripheral block groups relatively. This would be related to the increased speeds at the periphery (compared to the congested center) and the block groups' size, which tend to increase from center to periphery, covering more surface area with primary roads. The highest rate (2.29 and 2.58 for the single county and tri-county cases) belongs to the second severity level (possible injury or non-incapacitating injury). This is probably due to the intensification at and around the city center, where the block group sizes are smaller.

The spatial distributions -crash frequencies (crash counts for 2009-2011 period) of block groups- of all crash types of both the single county and tri-county cases are given in Figure 3.13 and Figure 3.14, respectively. It could be observed that while there are similarities in the spatial distribution of crash types, there are also apparent differences in some types in both cases.

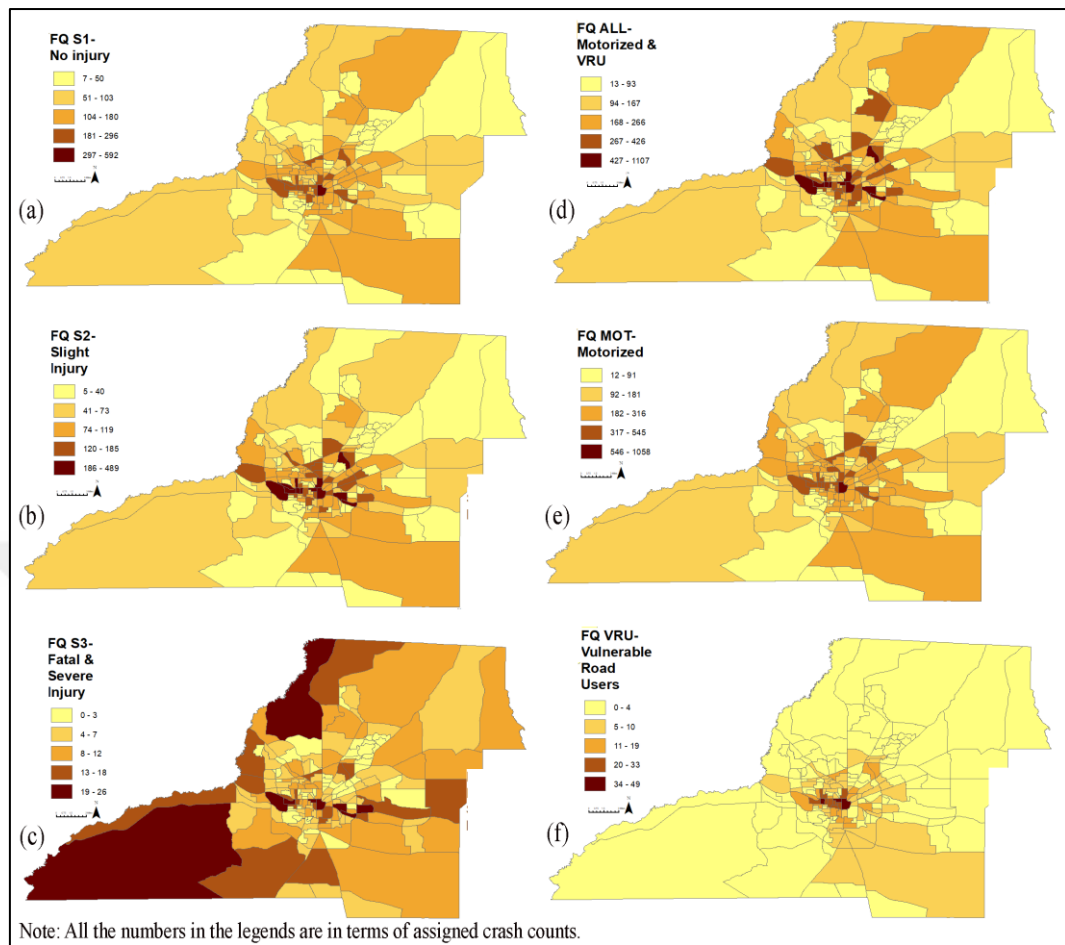


Figure 3.13. Spatial Distribution of Different Types of Crashes in Single County Case

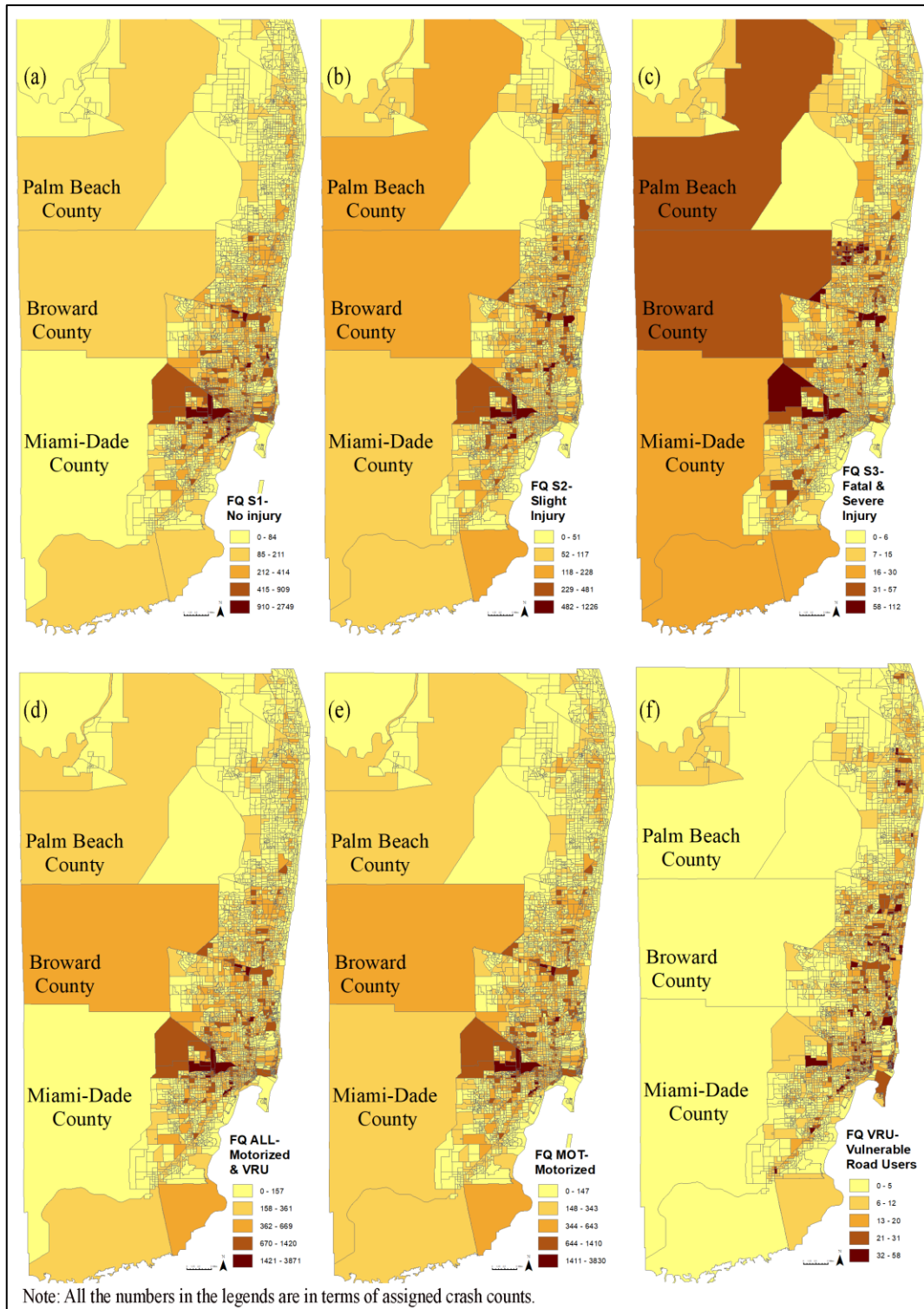


Figure 3.14. Spatial Distribution of Different Types of Crashes in Tri-County Case

The similarity in the distribution of overall crashes (Figure 3.13a and Figure 3.14a) with the motor vehicle crashes (Figure 3.13b and Figure 3.14b) and no-injury crashes (Figure 3.13d and Figure 3.14d) is obvious. The reason is that their frequencies are close. As usual, most of the crashes are vehicle crashes, 13,695 of 14,198 crashes (30,767 of 31,870 assigned crashes) in the single county case, and 262,313 of 273,359 crashes (658,868 of 686,294 assigned crashes) in the tri-county case. Moreover, more than half are PDO crashes, 17,106 and 422,934 assigned crashes in single and tri-county cases, respectively.

Vulnerable road user (pedestrian and bicycle user) involved crashes also happen but less frequently. They are clustered around the city center, as expected, since pedestrian movement is maximum at city centers (Figure 3.13c and Figure 3.14c). On the other hand, fatal crashes usually happen at the block groups at the fringe and close to the rural areas (Figure 3.13f and Figure 3.14f). This is also compatible with the emerging theory by (Ewing & Dumbaugh, 2009), which states that city centers (especially compact ones) have lesser vehicle miles compared to the auto-dependent suburban (especially sprawling) developments. Though fewer vehicle miles are observed, and lower speed limits are applied, the number of crashes is higher in the city centers since there are more cars, pedestrians, and conflict zones. However, among those crashes, PDOs have the highest share. When both speeds and vehicle miles increase (risk and exposure factors, respectively), the crashes would be fatal.

3.4.1.2 Crash density

The crash data that were classified according to severity (property damage only, non-incapacitating injury, fatal or incapacitating injury crashes) and involvement (total, motor vehicle involved, and vulnerable road user involved crashes) was also used to compute crash densities (as of Kernel Density) by the Kernel Density Estimation method presented in this chapter.

The points in the original case represented crashes. These points were used to calculate kernel density figures and transformed into a smoothed raster image, consisting of cells of a specific size (30 meters x 30 meters). As defined in the respective section, the cell width was set to one-tenth of the bandwidth, which is assumed to be 300 meters. These figures are also defined and used in a similar study (Ulak et al., 2017). Therefore, the points representing the crashes (with geographic coordinates) were used to calculate the density figure for each cell by kernel density estimation in ArcGIS (ESRI, 2011).

The calculated raster images are represented in the figures Figure 3.15-Figure 3.18. In Figure 3.15, kernel density estimates for the single county case crashes classified according to severity types are presented. In Figure 3.16, kernel density estimates for the single county case crashes classified according to involvement types are depicted. In Figure 3.17, kernel density estimates for the tri-county case crashes classified according to severity types are presented. In Figure 3.18, kernel density estimates for the tri-county case crashes classified according to involvement types are given.

Finally, for the modeling purpose, those crash density figures for each cell were assigned to population block groups by spatially joining the cells intersecting block group polygons (note that the final crash kernel density figure of a population block group is the sum of the density figures of the cells that are intersecting that block group).

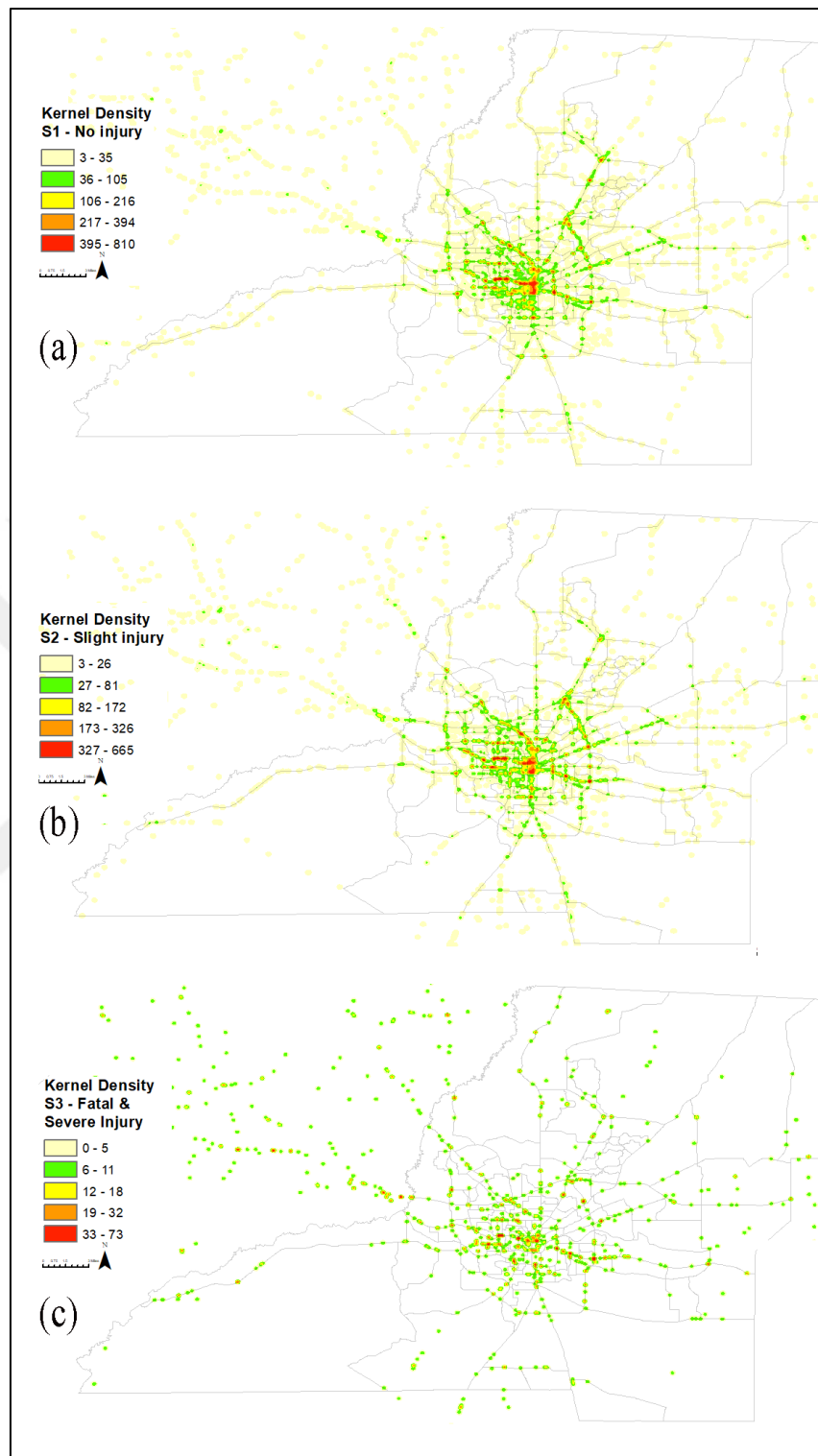


Figure 3.15. Kernel Density of Crashes Classified According to Severity in Single County Case

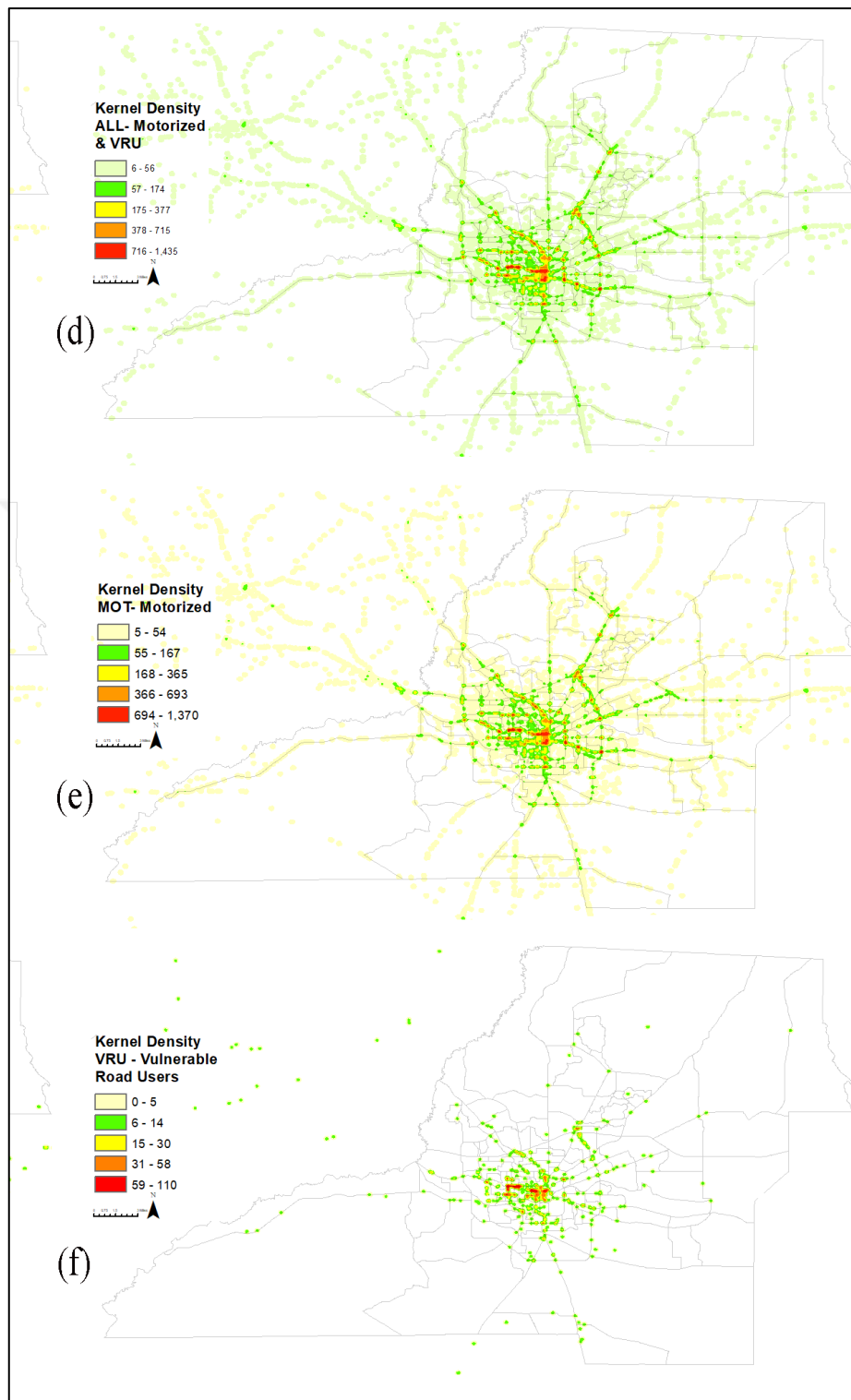


Figure 3.16. Kernel Density of Crashes Classified According to Involvement in Single County Case (continued from Figure 3.15)

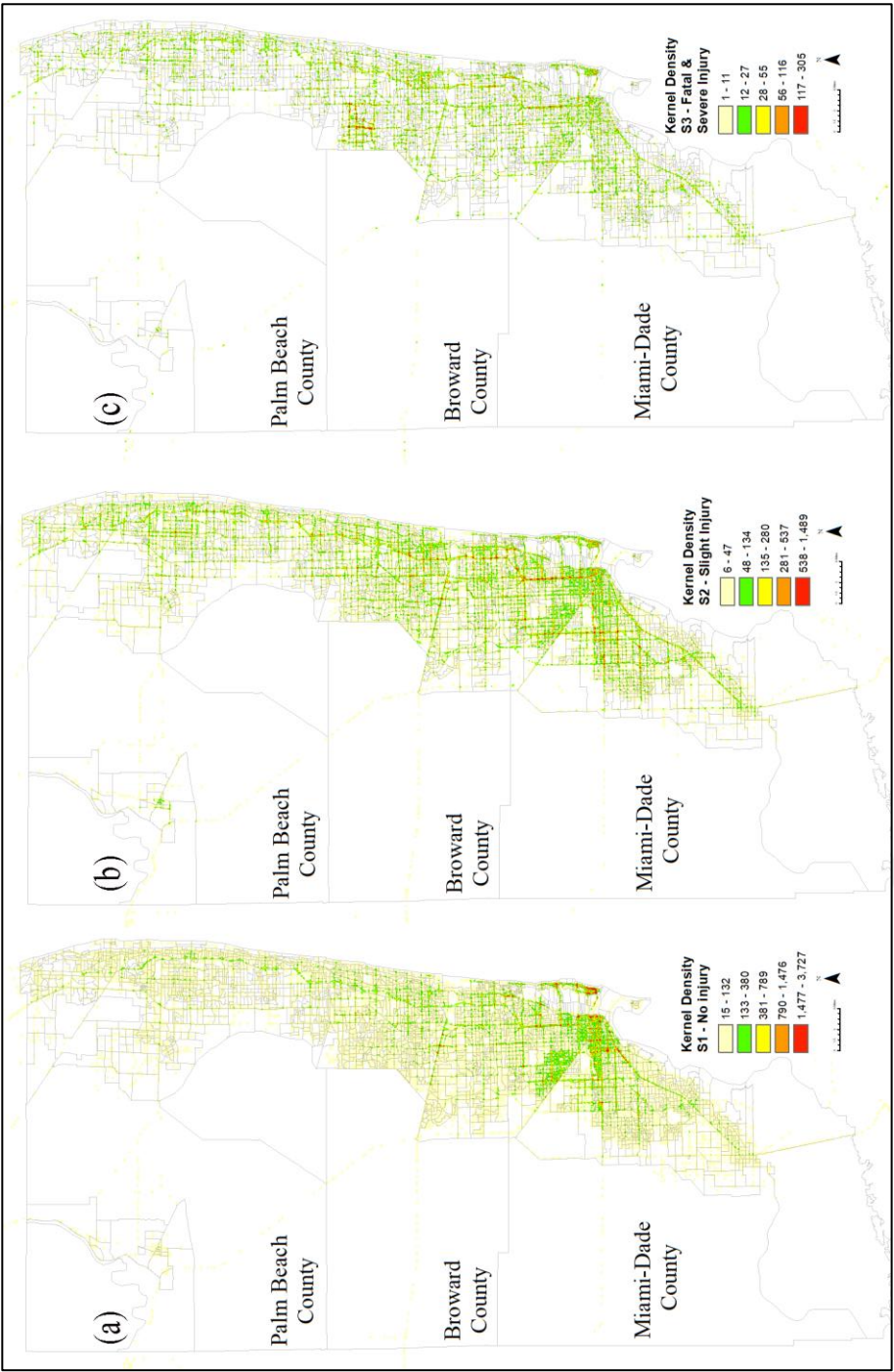


Figure 3.17. Kernel Density of Crashes Classified According to Severity in Tri-County Case

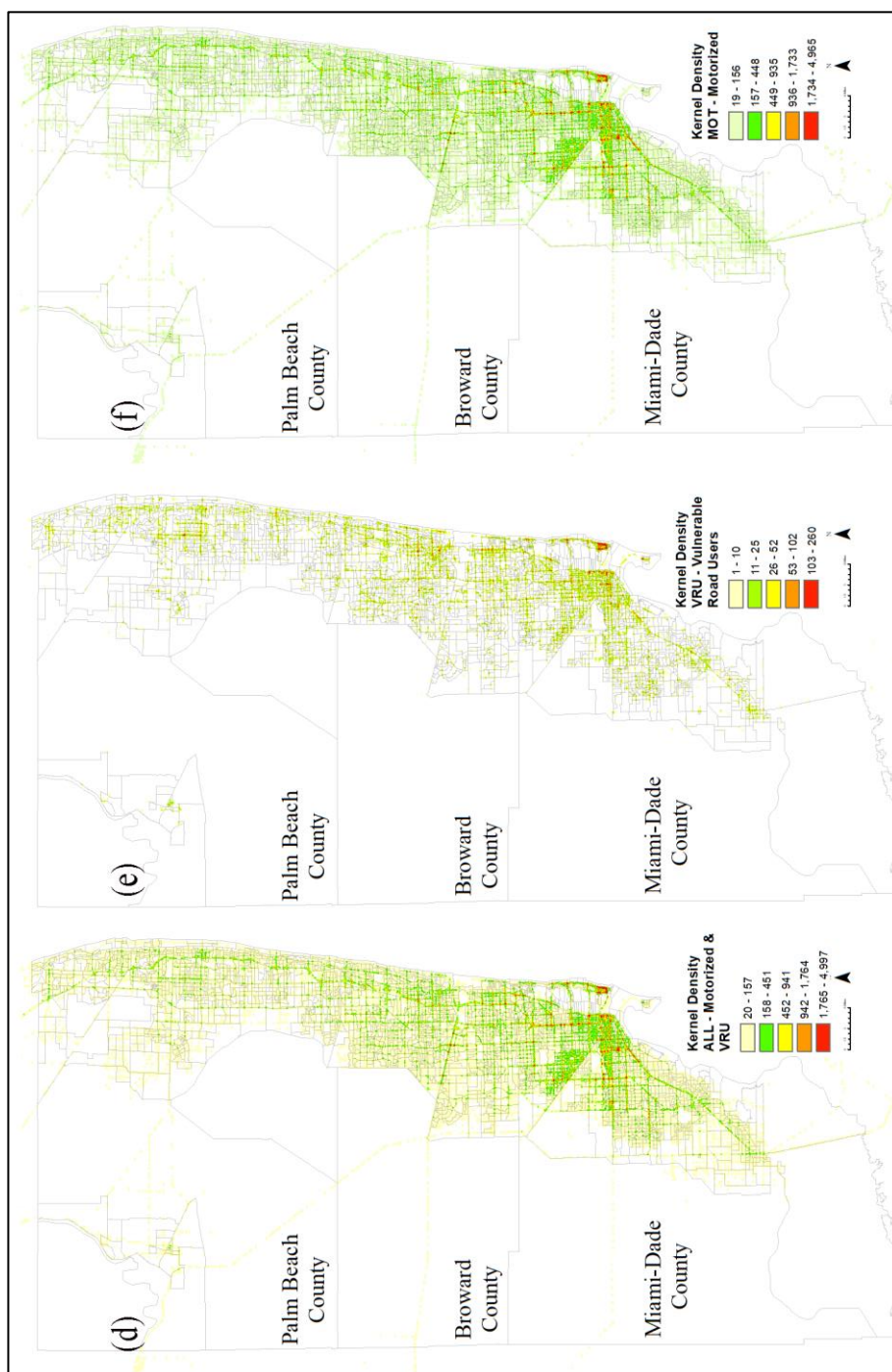


Figure 3.18. Kernel Density of Crashes Classified According to Involvement in Tri-County Case (continued from Figure 3.17)

3.4.2 Built environment dataset

Other than the crash dataset, there are several other datasets used in this study. These datasets were used to construct the independent variables dataset.

Traffic volume (annual average daily traffic) and street network data were gathered from the Florida Department of Transportation. A similar assignment procedure for crashes was applied to traffic volume assignments. The averages of 2013, 2014, and 2015 traffic volumes were assigned to block groups. Moreover, the total roadway length for each block group was calculated by summing the respective roadway assignments, where the roads were classified into three major categories in the U.S. Tiger Roads data: “Primary Road,” “Secondary Road,” and “Local neighborhood road, rural road, city street.” In road length and road density calculations, this classification is employed. For “Primary Road,” “Secondary Road,” and “Local neighborhood road, rural road, city street,” buffer distances of 48 meters, 24 meters, and 12 meters were used, respectively (note: these distances are the standard road widths for each category). For each road category, total road length and road surface density for each population block group were calculated. Additionally, a similar procedure was applied to calculate the total sidewalk length for each population block group using the Florida Department of Transportation's sidewalk data.

Built environment data were mainly gathered from the smart location database (“Smart Location Mapping | Smart Growth | US EPA,” n.d.) and the housing and transportation affordability index database (“Housing + Transportation Affordability Index | Center for Neighborhood Technology,” n.d.). The smart location database's organizational structure was used in this study to gather all the explanatory variables gathered from different sources. We classified variables under the five main built environment characteristics, density, design, diversity, destination accessibility, transit, and two additional dimensions, demographic and socioeconomic properties, and travel. The definitions of the D-variables are not strict, and those vary from one study to another. In this study, density, diversity, design, and destination accessibility categories are used (Ewing & Cervero, 2010; Ouyang & Bejleri, 2014); however,

under the transit category, transit usage in general and transit mode choice for commuting are provided rather than distance to transit to consider the direct ridership rather than the proximity of transit service. Under the transit category, in addition to the mentioned variables, the number of transit stops was used. The GIS data of the transit stops was gathered from the Florida Department of Transportation (“Florida Transit Data Exchange - FTIS: Florida Transit Information System,” n.d.). Dai and Jaworski (2016) mention half a mile as buffer distance for transit stops since this distance was mentioned as a “reasonable walking distance” for leisure and commuting (Boone, Buckley, Grove, & Sister, 2009; Cho et al., 2009 cited in Dai & Jaworski, 2016). In this study, bus stops were assigned to block groups by a radius of 415 m (5 minutes of walking distance), and metro stops were assigned to block groups by a radius of 1250 m (15 minutes of walking distance).

Moreover, the 2010 U.S. census block group data having the demographic, economic, and land use information (U.S. Census Bureau, 2011) was considered to provide insight to the analysis. Travel mode choice (except public transportation) and traffic volume variables were considered under the travel dimension, which would have direct relationships with crash occurrences. The increase in any travel mode choice and traffic volume would increase the exposure in the respective mode and the whole system, respectively, which would increase crashes (Merlin et al., 2020).

The primary datasets and their classifications are given in Figure 3.19. The Census, HTI, and SLD variables were already at the population block group level, so they were assigned to population block groups directly. Crash data, traffic volume, road length, and density variables were processed according to the procedures detailed above to assign to the population block groups.

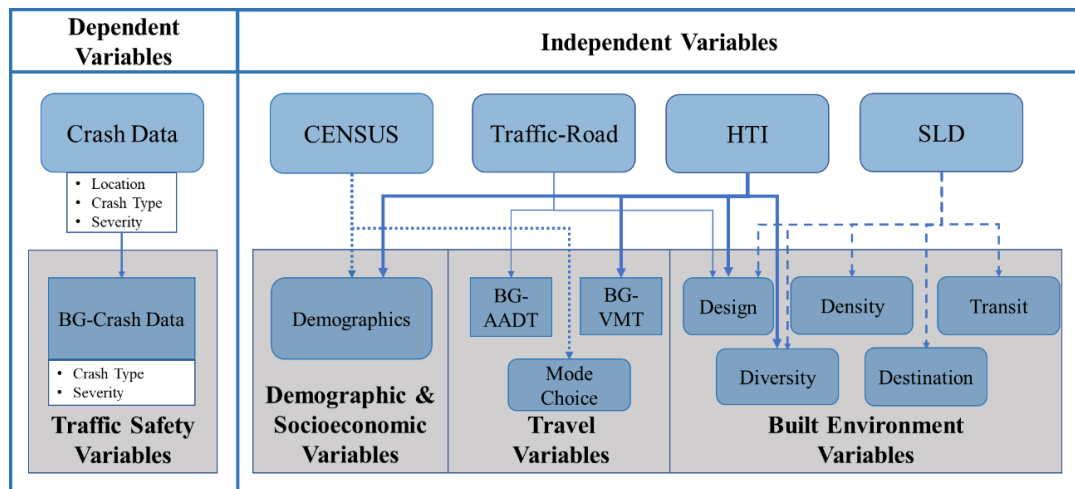


Figure 3.19. Data Sources and Classification

In Table 3.7 – Table 3.13, the built environment data is presented according to the built environment dimension. In Table 3.7, demographic and socioeconomic variables, in Table 3.8 density variables, in Table 3.9 diversity variables, in Table 3.10 design variables, in Table 3.11 transit variables, in Table 3.12 destination accessibility variables, and Table 3.13, travel variables are presented. Most of these variables were directly assigned to block groups, and the others were processed. The (*) sign near the variables indicated that the respective variable was calculated by using the data source presented in the final column the respective table. There are 34 demographic and socioeconomic variables, 12 density variables, 14 diversity variables, 19 design variables, 8 accessibility variables, 6 transit variables, and 7 travel variables, so 100 variables in total. These variables are the reduced set of 354 variables from the mentioned sources. A qualitative pre-processing was applied to these variables. The repetitive variables, similar variables, and variables with missing observations were eliminated to develop this set of 100 variables.

Table 3.7 Demographic and Socioeconomic Variables

Variable Code	Variable Explanation	Source
D0A*	Urban indicator (0 or 1) (1 Urban: pop. density > 500 persons per sq. mile – 0 Rural: pop. density < 500 persons per sq. mile)	Census
D0B	Total population (x1000 persons)	Census
D0B_W_P	White % of population (0-100)	Census
D0B_B_P	Black % of population (0-100)	Census
D0B_H_P	Hispanic % of population (0-100)	Census
D0B_U5_P	% of population under the age of 5 (0-100)	Census
D0B_U18_P	% of population under the age of 18 (0-100)	Census
D0B_O65_P	% of population above the age of 65 (0-100)	Census
D0B_O85_P	% of population above the age of 85 (0-100)	Census
D0B_M_P*	Male % of population (0-100)	Census
D0C_HHS	Average household size (persons)	Census
D0C_HU	Number of housing units (x1000 housing units)	Census
D0C_OC_P	% of occupied housing units (0-100)	Census
D0C_B90_P*	% of housing units built before 1990 (0-100)	Census
D0C_CO2_HH	Annual CO ₂ emission per household (tonnes)	HTI
D0C_CO2_A	Annual CO ₂ emission per acre (tonnes)	HTI
D0C_A_HH	Number of automobiles per household (automobiles)	HTI
D0C_A0_P	% of households with zero automobiles (0-100)	SLD
D0C_A1_P	% of households with one automobile (0-100)	SLD
D0C_A2P_P	% of households with two or more automobiles (0-100)	SLD
D0D_L9	% of population 25 years and more with no schooling completed (0-100)	Census
D0D_CE	% of population 25 years and more with at least bachelor's degree (0-100)	Census
D0E	Total number of workers (x1000 workers)	Census
D0E_HI	Median household income (x\$1000)	Census
D0E_LWH	% of low wage (less than \$1250/month) workers (home location) (0-100)	Census
D0E_LWW	% of low wage (less than \$1250/month) workers (work location) (0-100)	Census
D0E_BP	% of households with income below poverty level (0-100)	Census
D0E_H_P	% of income spent for housing (0-100)	HTI
D0E_T_P	% of income spent for travel (0-100)	HTI
D0E_A_P	% of income spent for auto ownership (0-100)	HTI
D0E_H_C	Housing cost (annual) (x\$1000)	HTI
D0E_T_C	Transportation cost (annual) (x\$1000)	HTI
D0E_A_C	Auto ownership cost (annual) (x\$1000)	HTI
D0E_R	Median gross monthly rent (x\$1000)	HTI

Table 3.8 Density Variables

Variable Code	Variable Explanation	Source
D1A_LA	Land area (x1000 acres)	Census
D1A_BS	Average block size (acres)	Census
D1A	Gross residential density on unprotected land (total housing units/land area)	SLD
D1B	Gross population density on unprotected land (total population/land area)	SLD
D1C	Gross employment density on unprotected land (total number of jobs/land area)	SLD
D1C_5_ENT	Gross entertainment employment density (number of entertainment jobs/land area)	SLD
D1C_5_IND	Gross industrial employment density (number of industrial jobs/land area)	SLD
D1C_5_OFF	Gross office employment density (number of office jobs/land area)	SLD
D1C_5_RET	Gross retail employment density (number of retail jobs/land area)	SLD
D1C_5_SVC	Gross service employment density (number of service jobs/land area)	SLD
D1D	Gross activity density on unprotected land [(total employment + total housing units)/land area]	SLD
D1E	Compact Neighborhood Score (0-10) (10: the most compact)	HTI

Table 3.9 Diversity Variables

Variable Code	Variable Explanation	Source
D2A_EP_HH_MX	Employment and household entropy	SLD
D2A_JP_HH	Jobs per household (total employment/number of households)	SLD
D2A_R	Ranking all of the block groups by D2A_EP_HH_MX (employment and household entropy) (1-20)	SLD
D2A_WRK_EMP	Workers per job (number of workers/total employment)	SLD
D2B_E5_MX	Employment entropy (five-tier employment entropy)	SLD
D2B_R	Ranking all of the block groups by their D2B_E5_MX value (employment mix) (1-20)	SLD
D2C_TRIP_EQ	Trip productions and trip attractions equilibrium index (0-1)	SLD
D2C_TRIP_MX	Employment and household trip entropy (based on vehicle trip production and trip attractions)	SLD
D2C_WRK_MX	Workers per job equilibrium index (0-1)	SLD
D2D	Total employment (x1000 jobs)	SLD
D2D_ONDX	Job access score (1-10)	HTI
D2D_GRVT	Employment access index (0-100)	HTI
D2D_NDX	Employment mix index (0-100)	HTI
D2E	Household intensity (sum of the “number of households/squared distance” values of the block groups)	HTI

Table 3.10 Design Variables

Variable Code	Variable Explanation	Source
D3A	Total road network density (miles per square mile)	SLD
D3A_AO	Network density of auto-oriented links per square mile (miles per square mile)	SLD
D3A_MM	Network density of multi-modal links per square mile (miles per square mile)	SLD
D3A_PO	Network density of pedestrian-oriented miles of multi-modal links per square mile (miles per square mile)	SLD
D3B_G	Gross intersection density (total number of intersections/area)	HTI
D3B_W	Street intersection density (auto oriented intersections eliminated) (number of street intersections/area)	SLD
D3B_AO	Intersection density of auto-oriented intersections per square mile (number of auto-oriented intersections/area)	SLD
D3B_MM3	Intersection density of multi-modal intersections having three legs per square mile (number of 3-leg intersections/area)	SLD
D3B_MM4	Intersection density of multi-modal intersections having four or more legs per square mile (number of 4-leg intersections/area)	SLD
D3B_PO3	Intersection density of pedestrian-oriented intersections having three legs per square mile (number of pedestrian-oriented 3-leg intersections/area)	SLD
D3B_PO4	Intersection density of pedestrian-oriented intersections having four or more legs per square mile (number of pedestrian-oriented 4-leg intersections/area)	SLD
D3B_R	Street intersection density ranked (1-20)	SLD
D3D_R1_P*	% of primary roads (0-100)	FDOT
D3D_R2_P*	% of secondary roads (0-100)	FDOT
D3D_R3_P*	% of local roads (0-100)	FDOT
D3D_R1_D*	Density of primary roads (miles per square mile)	FDOT
D3D_R2_D*	Density of secondary roads (miles per square mile)	FDOT
D3D_R3_D*	Density of local roads (miles per square mile)	FDOT
D3E*	Total sidewalk length (miles)	FDOT

Table 3.11 Transit Variables

Variable Code	Variable Explanation	Source
D4A	Proximity to transit stops (from population-weighted centroid of census block group to nearest transit stop) (meters)	SLD
D4A_R	Ranking all of the block groups by their D4A value (1-20)	SLD
D4B_025	Proportion of census block group employment within ¼ mile of fixed-guideway transit stop (0-1)	SLD
D4B_050	Proportion of census block group employment within ½ mile of fixed-guideway transit stop (0-1)	SLD
D4C_W_P	Transit ridership % of workers	HTI
D4C_T_HH	Transit trips per household (annual) (x1000 trips)	HTI
D4C_C_HH	Transit cost per household (annual) (x\$1000)	HTI
D4D*	Number of bus stops (stops)	FDOT

Table 3.12 Destination Accessibility Variables

Variable Code	Variable Explanation	Source
D5A_E	Working-age population within 45 minutes auto travel time (x1000 people)	SLD
D5A_R	Jobs within 45 minutes auto travel time (x1000 jobs)	SLD
D5B	Walking index (1-20)	SLD
D5C_EI	Regional centrality index – accessibility of working age population (1-100)	SLD
D5C_R	Accessibility to regional destinations	SLD
D5C_RI	Regional centrality index – job accessibility by transit (1-100)	SLD

Table 3.13 Travel Variables

Variable Code	Variable Explanation	Source
D6A_M_P*	% of motorized modes for commuting (excluding transit) (0-100)	Census
D6A_NM_P*	% of non-motorized modes for commuting (0-100)	Census
D6B_AA*	Annual average daily traffic (average of 3 years) for all modes (x1000 vehicles)	FDOT
D6B_AT*	Annual average daily traffic (average of 3 years) for trucks (x1000 trucks)	FDOT
D6C_T	Annual vehicle miles traveled (VMT) per household (x1000 miles)	HTI
D6C_LNT*	Natural Logarithm of VMT per household	HTI
D6C_C	Annual vehicle miles traveled (VMT) cost per household (x\$1000)	HTI

3.4.3 Dimension reduction for the built environment dataset

The previously introduced built environment dataset was reduced to case-specific datasets by employing the dimension reduction procedure defined earlier in this chapter. Therefore, for each specific case, the full dataset (100 variables) was reduced. For reducing the dataset to eliminate severe multicollinearity problems, correlation analysis was used. In this analysis, the threshold values were chosen as -0.75 and 0.75. Moreover, the literature review is used to decide on the variables to be selected. This analysis was done in three main stages. In the first stage, variables under each built environment dimension were analyzed. Correlation tables were created, and the variables that had strong correlation coefficients were filtered. At the end of this stage, for each built environment dimension, the full set of variables were reduced to a set of variables that did not have correlation problems. In the second stage, all the reduced dimensions were evaluated together by the same process. At the end of the second stage, the reduced set of variables that would be used in the subset selection procedure was obtained. These two stages were applied to both the single county case dataset and the tri-county dataset. At the end of the three stages, for each case, two distinct datasets were obtained. Those datasets are used in the subset selection procedure separately.

The resulting datasets are presented in the following tables. For the single county case, the reduced set of variables are given in Table 3.14, whereas, for the tri-county case, the reduced set of variables are given in Table 3.15.

Table 3.14 Descriptive Statistics of the Single County Case Dataset

Category	Variable Name	Min	Max	Mean	Variance
Traffic safety (Dependent Variables)	FQ_S1	7	592	96.64	5390.12
	FQ_S2	5	489	76.12	3990.54
	FQ_S3	0	26	7.18	29.86
	FQ_ALL	13	1107	180.06	19332.95
	FQ_MOT	12	1058	173.82	17596.51
	FQ_VRU	0	49	6.23	64.75
	KERNEL_S1_T	2.51	568.09	47.38	2627.05
	KERNEL_S2_T	2.19	459.17	36.18	1771.44
	KERNEL_S3_T	0.00	23.36	3.46	10.06
	KERNEL_ALL_T	5.49	1061.05	87.96	9170.45
	KERNEL_MOT_T	5.37	1013.52	84.81	8355.14
	KERNEL_VRU_T	0.00	46.50	2.97	23.53
Demographic & socioeconomic	D0A	0	1	0.79	0.17
	D0B	0	5	1.56	0.57
	D0B_B_P	1.13	98.11	31.51	668.13
	D0B_H_P	0.60	30.83	5.64	15.48
	D0B_U5_P	0.00	16.36	5.63	7.39
	D0B_O65_P	0.00	59.06	10.20	55.42
	D0B_M_P	6.74	70.94	47.61	29.64
	D0C_HHS	0	3	2.31	0.35
	D0C_OC_P	0.00	100.00	87.80	188.88
	D0C_B90_P	0.00	100.00	63.36	593.43
	D0C_A0_P	0.00	49.14	6.81	84.20
	D0C_A1_P	0.00	88.17	36.80	289.07
	D0D_L9	0.00	23.88	2.95	17.91
	D0D_CE	0.00	87.81	38.70	441.40
	D0E	0	2	0.64	0.11
	D0E_HI	0	137	49.83	957.49
	D0E_LWH	15.36	52.88	28.09	73.37
	D0E_LWW	0.00	85.00	34.59	313.94
	D0E_R	0	2	0.81	0.20
Density	D1A_LA	0.04	81.25	2.41	60.78
	D1B	0.03	37.43	4.65	25.18
	D1C_5_ENT	0.00	3.21	0.31	0.36
	D1C_5_IND	0.00	2.51	0.13	0.10
	D1C_5_OFF	0.00	43.11	0.70	11.58
	D1C_5_RET	0.00	7.73	0.30	0.71
Diversity	D2A_JP_HH	0.00	6792.00	47.97	266193.27
	D2A_WRK_EMP	0.00	240.00	8.76	478.35
	D2B_E5_MX	0.00	0.97	0.53	0.10
	D2C_TRIP_EQ	0.00	0.98	0.36	0.09
	D2C_WRK_MX	0.00	0.99	0.30	0.09
	D2D_ONDX	0.00	9.10	5.96	5.14
	D2D_NDX	0	91	84.77	174.37

Table 3.14 (continued)

Design	D3A_AO	0.00	7.18	0.95	1.44
	D3A_MM	0.00	11.26	2.07	4.01
	D3B_G	0	436	115.60	7466.30
	D3B_MM3	0.00	73.36	11.41	159.26
	D3B_MM4	0.00	47.70	4.29	50.72
	D3B_PO3	0.7	149.7	39.04	855.43
	D3B_PO4	0.00	156.21	7.98	263.60
	D3D_R1_P	0.00	39.61	3.25	69.27
	D3D_R2_P	0.00	47.27	17.49	140.86
	D3D_R2_D	0.00	39.38	8.09	50.76
	D3E	0	38	6.36	39.25
Transit	D4C_W_P	0	7	1.99	3.95
	D4D	0	80	17.13	269.38
Destination	D5A_R	0.03	74.72	43.80	210.52
	D5B	1.33	12.33	7.01	4.98
Travel	D6A_M_P	0	100	89.89	239.66
	D6B_AA	0	605	117.96	9220.77
	D6C_T	0	25	20.00	17.92

Table 3.15 Descriptive Statistics of the Tri-County Case Dataset

Category	Variable Name	Min	Max	Mean	Variance
Traffic safety (Dependent Variables)	FQ_S1	0.0	2749	123.81	25674.23
	FQ_S2	0.0	1226	68.13	5289.15
	FQ_S3	0.0	112	8.90	106.16
	FQ_ALL	0.0	3871	200.91	52132.90
	FQ_MOT	0.0	3830	192.88	50135.92
	FQ_VRU	0.0	58	8.03	60.57
	KERNEL_S1_T	0.0	1959.48	54.22	6974.07
	KERNEL_S2_T	0.0	739.37	29.87	1559.10
	KERNEL_S3_T	0.0	65.48	3.84	28.13
	KERNEL_ALL_T	0.0	2654.86	88.48	14855.78
	KERNEL_MOT_T	0.0	2622.81	84.90	14331.98
	KERNEL_VRU_T	0.0	38.12	3.42	13.92
Demographic & socioeconomic	D0A	0.0	1	0.97	0.03
	D0B	0.0	15	1.63	0.81
	D0B_B_P	0.0	98.99	19.93	685.77
	D0B_H_P	0.0	100.00	39.62	899.58
	D0B_U5_P	0.0	33.33	5.49	6.21
	D0B_O65_P	0.0	99.58	17.79	259.45
	D0B_M_P	0.0	100.00	48.32	31.48
	D0C_HHS	0.0	6	2.67	0.52
	D0C_OC_P	0.0	100.00	86.54	170.19
	D0C_B90_P	0.0	100.00	74.46	890.94
	D0C_CO2_HH	0.0	13.20	7.49	7.31
	D0C_A0_P	0.0	100.00	8.60	129.29
	D0C_A1_P	0.0	100.00	39.64	321.25
	D0D_L9	0.0	80.75	8.23	92.43

Table 3.15 (continued)

	D0D_L9	0.0	80.75	8.23	92.43
	D0D_CE	0.0	100.00	29.03	351.40
	D0E_HI	0.0	250	56.08	1079.89
	D0E_LWH	0.0	100.00	24.33	28.44
	D0E_LWW	0.0	100.00	32.42	247.84
	D0E_BP	0.0	100.00	14.33	166.02
	D0E_R	0.0	4	1.04	0.46
Density	D1A_LA	0.0	573.45	0.95	287.73
	D1A_BS	0.0	1792	13.48	1602.63
	D1B	0.0	406.28	13.94	260.24
	D1C	0.0	527.85	4.23	249.53
	D1C_5_ENT	0.0	88.96	0.74	13.09
	D1C_5_IND	0.0	183.11	0.49	11.79
	D1C_5_OFF	0.0	254.01	0.75	33.28
	D1C_5_RET	0.0	22.33	0.48	1.63
	D1C_5_SVC	0.0	118.39	0.89	17.07
Diversity	D2A_JP_HH	0.0	9540.00	14.73	65714.43
	D2A_WRK_EMP	0.0	708.00	13.62	1642.34
	D2B_E5_MX	0.0	0.99	0.50	0.11
	D2C_TRIP_EQ	0.0	1.00	0.36	0.10
	D2C_WRK_MX	0.0	1.00	0.28	0.09
	D2D	0.0	57	0.62	4.48
	D2D_ONDX	0.0	9.90	6.75	4.31
	D2D_NDX	0.0	94	88.53	81.16
Design	D3A	0.0	70.01	21.63	53.42
	D3A_AO	0.0	30.68	1.77	8.86
	D3A_MM	0.0	37.53	2.64	8.48
	D3B_G	0.0	1251	215.52	15655.02
	D3B_MM3	0.0	356.96	11.92	389.69
	D3B_MM4	0.0	162.19	8.34	160.87
	D3B_PO3	0.0	1761.5	79.79	5165.17
	D3B_PO4	0.0	730.23	28.49	1187.00
	D3D_R1_P	0.0	79.89	3.90	87.81
	D3D_R2_P	0.0	100.00	12.69	216.63
	D3E	0.0	21	2.59	8.03
Transit	D4B_025	0.0	1.00	0.02	0.01
	D4C_W_P	0.0	40	7.03	35.25
	D4D	0.0	199	20.72	263.37
Destination	D5A_R	0.0	296.40	149.79	3497.59
	D5B	1.0	19.83	12.02	15.05
Travel	D6A_M_P	0.0	100	86.36	223.23
	D6A_NM_P	0.0	88	2.68	39.54
	D6B_AA	0.0	2218	141.18	25840.80
	D6C_T	0.0	27	18.99	13.01

In these tables, descriptive statistics of both independent variables of each built environment dimension (e.g., design and diversity) and the dependent variables representing traffic safety (e.g., frequency of severe crashes and kernel density of vulnerable road user crashes) are presented.

In both tables (for both single county and tri-county cases), it could be observed that all the dependent variables have variance figures that are far larger than their mean values, which means that the data of each dependent variable is overdispersed. Therefore, the selection of the statistical distributions (negative binomial) for the count models and (Gamma, Lognormal, Weibull) for the density models are appropriate.

The variables presented are the remaining ones after the dimension reduction process; thus, they are used in the subset selection process. One of this study's objectives is to determine the factors best to explain the traffic safety variables, and the variables presented in Table 3.14 and Table 3.15 are the ones from which the subsets are selected. There are more variables in Table 3.15 (57 variables) than in Table 3.14 (50 variables), so the set of alternative variables for the tri-county case is larger than the set of alternative variables for the single county case. In terms of combinatorial optimization, that means the total number of combinations is far higher. Since there are 7 more variables, there are 127 times more (2^7-1) alternatives in the tri-county case. Since even 50 variables case is not easy to handle (2^{50}) by complete enumeration, it is not feasible to enumerate 2^{57} alternatives. In those combinatorial optimization problems, heuristic algorithms are quite effective, and tabu-search, as a meta-heuristics algorithm, is employed in this study.

3.4.4 Subset selection-tabu-search calibration

As mentioned in the previous sections, to apply a tabu-search algorithm, its parameters should be determined, and its criterion should be decided. In this section, the calibration of the tabu-search process is described in detail.

For the calibration part, two instances from the single county case (instances 1 and 2) and two instances from the tri-county case were randomly selected (instances 3 and 4). For the tabu-search algorithm proposed in this study, the parameters -the number of neighbors, list size, the number of restarts, the number of iterations, and the model evaluation criterion- were preoptimized. For the number of neighbors {15, 30, total number of independent variables}, for the list size {5, 10, 15, 20}, for the number of iterations {20, 50, 75}, and for the number of restarts {3, 5, 10, 15} were the candidates. Those parameters were tested with regression settings of 4 different dependent variables, so in total, there were 576 different settings. The results of those preliminary experiments are summarized in Table 3.16.

Table 3.16 Results of Preliminary Experiments

	# of iterations	# of optimal results	# of restarts	# of optimal results	# of neighbors	# of optimal results	List size	# of optimal results
Instance 1	20	32	3	39	15	28	5	36
	50	48	5	39	30	64	10	40
	75	32	10	39	50	64	15	40
			15	39			20	40
Instance 2	20	0	3	3	15	0	5	12
	50	4	5	3	30	4	10	0
	75	4	10	3	50	8	15	0
			15	3			20	0
Instance 3	20	32	3	41	15	36	5	44
	50	48	5	41	30	64	10	40
	75	36	10	41	58	64	15	40
			15	41			20	40
Instance 4	20	16	3	16	15	0	5	16
	50	16	5	16	30	0	10	16
	75	16	10	16	58	64	15	16
			15	16			20	16

For each instance, the best solution was selected as a benchmark score, and the result obtained at each combination was compared to that benchmarked result. If the best solution was obtained at that specific combination, that combination was assigned 1 and 0 otherwise. Then, for each parameter, the sums of those assigned values are presented in Table 3.16, as the numbers given in the “# of optimal results” column. For example, in instance 1, 32 of all combinations with 20 iterations reached the benchmarked result, whereas the number 48 of the combinations with 50 iterations reached the benchmarked result. For each parameter, the total number of iterations that reached the benchmarked result is presented under the respective column—the higher the number, the better the parameter for the process. For the same number of optimal results, the smaller parameter was selected. For example, all alternatives for the “number of restarts” parameter yielded the same number of optimal results, so 3 as the smallest alternative is selected. According to the results presented in Table 3.16, the following values were assigned to the parameters:

- **Number of neighbors:** It is set to be the total number of independent variables.
- **List size (s):** It is set to 5.
- **Termination criteria:** For each stage (preliminary, intensification, and diversification) the number of iterations is set to 50 (Due to the number of restarts at the intensification stage, the total number of iterations may vary).
- **Number of restarts:** The number of restarts of the intensification stage is set to 3, while the number of restarts of the whole algorithm is set to 1.

For selecting the objective function value, preliminary testing was done. An example is presented in Table 3.17. The same setting was solved by optimizing six different objective functions previously introduced in Table 3.2, i.e., likelihood-based R^2 , R^2 -corrected (Negalkerke R^2), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Log-likelihood, and residual deviance. For the objective functions of likelihood-based R^2 and R^2 -corrected (Negalkerke R^2), the

algorithm tried to maximize the objective; whereas, for the objective functions of AIC, BIC, Log-likelihood, and residual deviance, the algorithm tried to minimize the objective (note that the pseudo- R^2 estimates used in this study are preferred in generalized linear models).

In Table 3.17, the independent variables in the example are presented. The models with the same set of dependent and independent variables are solved for each objective function separately. The results according to each objective function are presented in Table 3.17. For each solution, there are two columns. The first column is named “Basic” and the second is named “Sig. level”. The “Basic” column depicts whether a variable was selected to be in the model or not. If there is a “✓” sign, it means that that variable was in the model at the optimum iteration of the respective model; otherwise, the variable was not basic. The “Sig. level” column depicts the significance level of the respective variable at that optimum iteration. For example, for the tabu-search based subset selection model with BIC as the objective function, DOB variable was in the optimum subset with the highest significance level.

In this preliminary testing, it was observed that the AIC as the objective function returns the best results so that the number of significant variables is optimized. That is to say, at the optimum iteration, the number of independent variables used is lower when AIC and BIC were the objective function values. In the results obtained by employing objective functions other than AIC and BIC, more variables were selected; however, not all of these independent variables were significant. The independent variables selected by AIC and BIC functions turn out to be significant in almost all cases. When AIC and BIC values are compared on this stance, the variables selected by BIC value are more parsimonious in number than the variables selected by AIC. In the BIC case, almost all of the variables have the highest significance level ($p\text{-value} < 0.01$). However, by AIC value, variables with lesser significance levels could be captured while mostly avoiding insignificant variables. Thus tabu-search in this study employs AIC as the objective function to evaluate regression models.

Table 3.17 Test Results for Objective Function

VARIABLE	OBJECTIVE FUNCTION											
	R ² -Likelihood ratio based		R ² -Corrected		AIC		BIC		Loglikelihood		Residual deviance	
	Basic	Sig. level	Basic	Sig. level	Basic	Sig. level	Basic	Sig. level	Basic	Sig. level	Basic	Sig. level
Intercept	✓	***	✓	***	✓	***	✓	***	✓	***	✓	***
D0A	✓	**	✓	**	✓	***	✓	***	✓	**	✓	*
D0B	✓	.	✓	.	✓	**			✓	.	✓	***
D0B_B_P	✓		✓						✓		✓	
D0B_H_P	✓		✓		✓				✓			
D0B_U5_P	✓		✓						✓			
D0B_O65_P	✓		✓						✓			
D0B_M_P	✓		✓						✓			
D0C_HHS	✓		✓						✓			
D0C_OC_P	✓		✓						✓		✓	
D0C_B90_P	✓	*	✓	*	✓	*			✓	*		
D0C_A0_P	✓		✓						✓			
D0C_A1_P	✓		✓						✓		✓	
D0D_L9	✓		✓						✓			
D0D_CE	✓		✓						✓		✓	
D0E	✓		✓						✓			
D0E_HI	✓		✓						✓			
D0E_LWH	✓		✓						✓			
D0E_LWW	✓	*	✓	*	✓	***	✓	***	✓	*		
D0E_R	✓	***	✓	***	✓	***	✓	***	✓	***	✓	*
D1A_LA	✓	***	✓	***	✓	**			✓	***	✓	***
D1B	✓		✓						✓			
D1C_5_ENT	✓		✓						✓			
D1C_5_IND	✓	*	✓	*	✓	*			✓	*	✓	
D1C_5_OFF	✓		✓						✓		✓	
D1C_5_RET	✓		✓						✓			
D2A_JP_HH	✓		✓						✓		✓	
D2A_WRK_EMP	✓	**	✓	**	✓	***	✓	***	✓	**	✓	.
D2B_E5_MX	✓	*	✓	*	✓	***	✓	**	✓	*	✓	*
D2C_TRIP_EQ	✓		✓						✓		✓	
D2C_WRK_MX	✓	**	✓	**	✓	**	✓	*	✓	**	✓	*
D2D_ONDX	✓	**	✓	**	✓	**			✓	**	✓	**
D2D_NDX	✓	.	✓	.	✓	**			✓	.	✓	.
D3A_AO	✓		✓						✓			
D3A_MM	✓		✓						✓			
D3B_G	✓		✓						✓			
D3B_MM3	✓		✓						✓			
D3B_MM4	✓		✓						✓			
D3B_PO3	✓		✓						✓			
D3B_PO4	✓		✓						✓			
D3D_R1_P	✓	*	✓	*	✓	**			✓	*		
D3D_R2_P	✓		✓						✓			
D3D_R2_D	✓		✓						✓			
D3E	✓		✓						✓		✓	
D4C_W_P	✓		✓						✓			
D4D	✓	**	✓	**	✓	***	✓	***	✓	**		
D5A_R	✓	**	✓	**	✓	***			✓	**	✓	***
D5B	✓		✓						✓		✓	
D6A_M_P	✓	**	✓	**	✓	***	✓	**	✓	**		
D6B_AA	✓	***	✓	***	✓	***	✓	***	✓	***	✓	***
D6C_T	✓		✓						✓			
Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1												
Variable: basic "✓" non-basic ""												

3.5 Summary

In this chapter, the methodological framework was presented, including the aims and related research questions. In this framework, dataset preparation is followed by creating a reduced set of variables and then selecting the key built environment (BE) variables associated with traffic safety (TS). Different data extraction methods and statistical distributions are recommended depending on the traffic safety variables (frequency or kernel density). Though the extraction methods vary, the main procedure to determine critical factors is the same: tabu-search-based subset selection (statistical distribution to fit also varies as mentioned).

In addition to the methodological framework, this chapter has introduced the dataset to be employed in the following chapter. “Smart Location Database,” “2010 US Census,” and “Housing and Transportation Affordability Index” data are the main data sources. The case studies have been selected from Florida, the US. Two particular cases, one as a single urban structure (Leon County) and the other as an urban agglomeration (Broward, Miami-Dade, and Palm Beach Counties), have been selected. The crash data for those counties have been classified into two typologies: crash severity and involvement.

According to the methodology and dataset introduced in this chapter, built environment-traffic safety models are analyzed for single and tri-county cases. The experimental design for the single county case (the same design is also valid for the tri-county case) is presented in Figure 3.20. There are two cases (single and tri-county), two types of crash representation (frequency and density), four different distributions (negative binomial for frequency estimation and Gamma, Lognormal and Weibull for density estimation), and two crash typologies (severity and involvement) each having three sub-categories. Therefore, there are 48 different BE-TS formulations. The results will be presented and discussed in the next chapter.

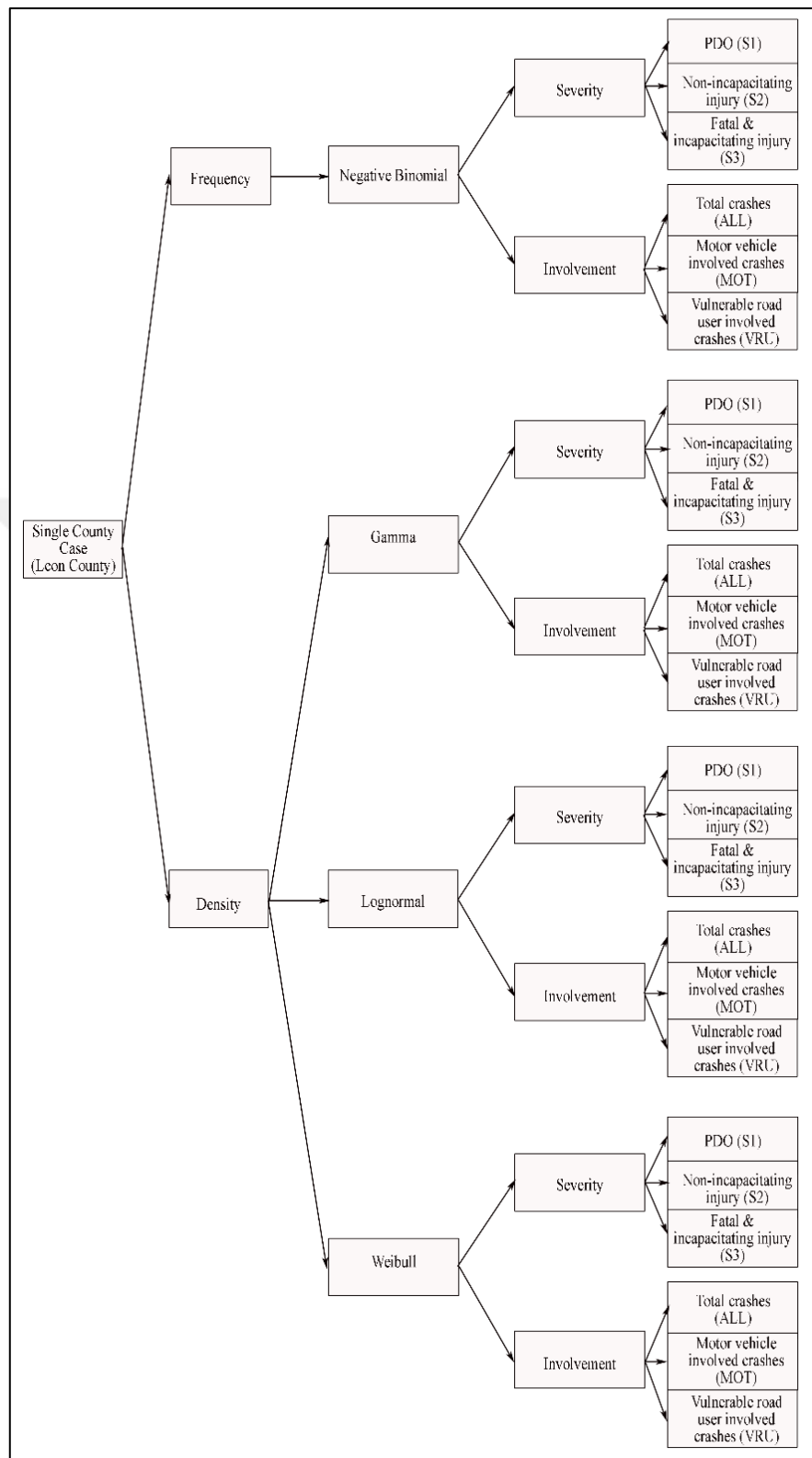


Figure 3.20. Experimental Design for the Single County Case

Finally, the software programs used for applying the proposed methodology are as follows. Data preparation processes are performed by MS Excel, ArcGIS (version 10.7) (ESRI, 2011), SPSS (version 26) (IBM Corp, 2019), and R software (R Core Team, 2020). The overall algorithm is coded in R software (R Core Team, 2020). For tabu search process *tabuSearch* package (Domijan, 2018), for generalized linear regression model solutions, *MASS* (Ripley et al., 2020; Venables & Ripley, 2002) and *GAMLSS* packages (Rigby & Stasinopoulos, 2005) are used. All the computations are performed on a personal computer with Intel(R) Core i5-2320 3.00GHz processor and 8 GB ram.



CHAPTER 4

RESULTS FROM ACCIDENT FREQUENCY AND DENSITY MODELS, DEVELOPED COUNTRY CONTEXT

This chapter presents the results of the tabu-search based subset selection methodology applied to the case studies introduced in the previous chapter.

4.1 Background

The relationship between the built environment and traffic safety has been investigated. This relationship was analyzed in two distinct settings, one was a small-sized single county case, and the other was a tri-county urban agglomeration. Other than this setting, there were two crash typologies used; one considers a crash classification based on the severity, and the other considers the occupants (involving parts) of those crashes. Moreover, each crash type was estimated in two forms, frequency (crash counts) and kernel density (total density estimates).

The effect of the urban built environment on the crash frequency and crash density has been examined using the accident frequency models described in Chapter 3. In the following sections, the tabu-search-based algorithm results will be presented in detail on two sample models.

Following this introductory part, the model estimation results and findings are presented in this chapter. The organization of the settings is given in Table 4.1 (in a total of 48 settings). This chapter is organized as follows: first, the results from the single county case are presented. The results of the tri-county case follow this. Then, the predictive performances of these models are tested. Finally, a detailed discussion of the findings is presented. Note that, for the frequency method, negative binomial distribution was used, so the results presented in this chapter are based on negative binomial regression. However, for the density method, three different statistical

distributions were used. Gamma distribution had a better fit, so only the Gamma regression results are presented in this chapter (Note that all the remaining results are presented in Appendix A and Appendix B for the single county case and tri-county case, respectively).

Table 4.1 Organization of the Settings

Case	DV-extraction method	Distribution	Crash typology	
Single county	Frequency	Negative Binomial	Severity*	Involvement**
	Density	Gamma	Severity	Involvement
		Lognormal	Severity	Involvement
		Weibull	Severity	Involvement
Tri-county	Frequency	Negative Binomial	Severity	Involvement
	Density	Gamma	Severity	Involvement
		Lognormal	Severity	Involvement
		Weibull	Severity	Involvement

*Severity levels are: S1-PDO, S2-non-incapacitating injury, S3-fatal or incapacitating injury

**Involvement types are: ALL, MOT (motorized vehicles), VRU (vulnerable road users)

4.2 Tabu-Search Based Solution Algorithm: Overview

Before introducing the results of the tabu-search-based subset selection method, the general scheme of this method is described in this section. As introduced in the methodology section, tabu-search is an iterative technique. Through several iterations, the algorithm searches for the optimum result. The overall algorithm proposed in this study is presented in Figure 4.1. In this study, tabu-search was used to determine the best set of built environment factors that has the minimum AIC value when modeled with traffic safety variables.

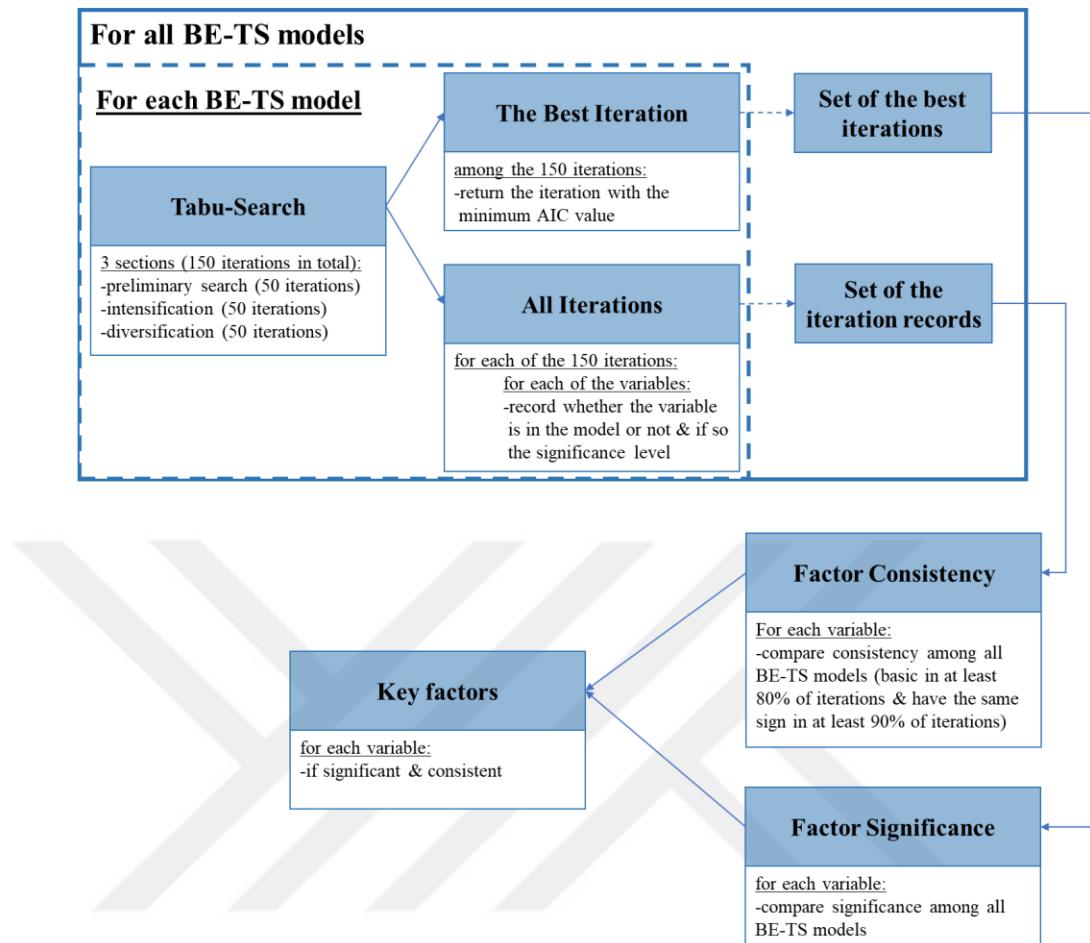


Figure 4.1. Tabu-Search Based Setting

The proposed BE-TS model settings are presented both in Figure 3.20 and Table 4.1. For each setting, built environment-traffic safety relationship model (e.g., built environment variables and vulnerable road user involved crashes in single county case), the tabu-search process was applied. There were three sections, preliminary search, intensification, and diversification, 50 iterations each. The algorithm stopped after solving 150 iterations (note that the number of iterations at the intensification stage, so the total number of iterations would vary if the algorithm requires restarts). Then the best iteration (with the minimum AIC value) of those 150 iterations was

returned. The variables that were basic at that best iteration and the significance levels of those variables were recorded. Moreover, all 150 iterations were traced for each variable. For each variable, the number of iterations (among 150 iterations) in which that specific variable was basic (selected to be in the model) and its association with the dependent variable (positive or negative) were recorded. This process was applied to each BE-TS model in each case. The best iterations were collected to generate the set of best iterations that will be used to discuss factor significance. The information about all iterations was collected to generate the set of iteration records that will be used to discuss factor consistency. Factor significance and factor consistency will be used to identify the key built environment factors that will be discussed later in this chapter.

4.3 Tabu-Search Based Solution Algorithm: Examples

In Figure 4.2, tabu-search results for the single county fatal or incapacitating injury crashes case (S3) are given. There are 150 iterations in total (50 iterations for each section: preliminary search, intensification, and diversification). Each major iteration (in which all neighboring iterations are solved) is presented in the body of the graph. Each iteration was a recurrent solution of multiple regression models with the subset of variables determined at that iteration. The algorithm starts with a random set of independent variables and proceeds by selecting the best possible combination of independent variables to minimize the Akaike Information Criterion (AIC) value. For each iteration, the AIC value and the number of variables are presented at the bottom part. These two values are about the model behavior during iterations.

Moreover, BE variable behavior was also traced. For each variable, the number of iterations in which that specific variable was basic is presented by the chart at the figure's righthand side. The significance value of each variable in the model at a specific iteration is represented via color code, e.g., the highest significance level of 99% ($p < 0.01$) is depicted by red, 95% significance level, and 90% significance levels are depicted by orange and yellow respectively.

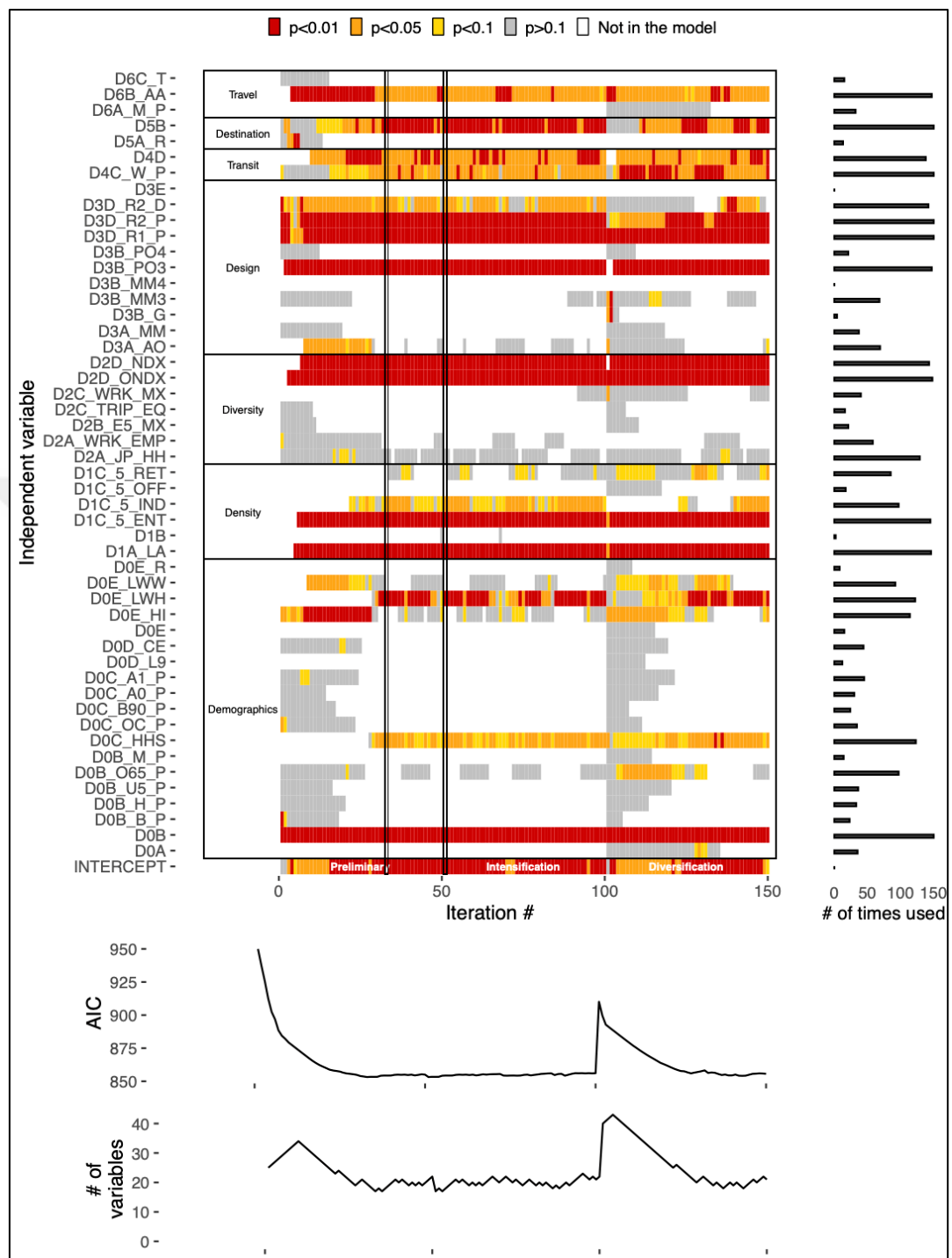


Figure 4.2. Tabu-search Results – Single County - FQ_S3

As mentioned, the number of times each variable was selected to be in the subset is given on the same figure's right-hand side. It is possible to observe which variables are preferred mostly and what are their significance levels while they are in the model. As an example, for this case, the most preferred variables are: annual average daily traffic (D6B_AA), walking index (D5B), number of bus stops (D4D), transit ridership percentage of workers (D4C_W_P), the density of secondary roads (D3D_R2_D), percentage of secondary roads (D3D_R2_P), percentage of primary roads (D3D_R1_P), intersection density in terms of pedestrian-oriented intersections having three legs per square mile (D3B_PO3), employment mix index (D2D_NDX), job access score (D2D_ONDX), gross entertainment employment density (D1C_5_ENT), land area in acres (D1A_LA), average household size (D0C_HHS), total population (D0B). These are the most frequently selected variables, but they do not have the same level of significance overall.

Among them, percentage of secondary roads (D3D_R2_P), percentage of primary roads (D3D_R1_P), intersection density in terms of pedestrian-oriented intersections having three legs per square mile (D3B_PO3), employment mix index (D2D_NDX), job access score (D2D_ONDX), gross entertainment employment density (D1C_5_ENT), land area in acres (D1A_LA), and total population (D0B) are the ones that consistently had high significance levels ($p < 0.01$). The others usually had lower significance values. Although those variables' significance levels are usually high or low when they are in the model, their level would change from one iteration to another. For example, the walking index (D5B) was always in the selected subset, but its significance level fluctuates from one iteration to the other from 90% to 99%. The variables that were consistently in the model and having stable significance levels (preferably higher than 99%) would be the built environment's critical factors.

The optimal objective value (minimum AIC value) obtained by this procedure for the model estimating S3 crashes was 853.18 (note that the optimum iteration is depicted with black lines; and if there were multiple iterations with the minimum AIC value, they are all represented on the graph). This value was obtained at the 33rd and 51st iterations with the same subset of independent variables. The subsets of the

BE variables in the model at those optimum iterations and their significance levels are given in the next section.

In Figure 4.3, tabu-search results of the single county case for the ALL (motorized and vulnerable road user) crashes are given. Each major iteration (in which all neighboring iterations were solved) is presented in the graph.

For this case, the most preferred variables were annual average daily traffic (D6B_AA), percentage of motorized modes for commuting (D6A_M_P), jobs within 45 minutes auto travel time (D5A_R), number of bus stops (D4D), percentage of primary roads (D3D_R1_P), job access score (D2D_ONDX), employment entropy (D2B_E5_MX), household workers per job (D2A_WRK_EMP), gross industrial employment density (D1C_5_IND), land area in acres (D1A_LA), median gross monthly rent (D0E_R), percentage of low wage workers (work location) (D0E_LWW), percentage of low wage workers (home location) (D0E_LWH), percentage of housing units built before 1990 (D0C_B90_P), Hispanic % of the population (D0B_H_P), total population (D0B), urban indicator (D0A).

The listed variables were frequently selected to be in the subset and had high significance levels ($p < 0.01$). Other variables were frequently used, such as household workers per job equilibrium index (D2C_WRK_MX) and percentage of population 25 years and more with at least bachelor's degree (D0D_CE); however, the significance levels of those variables were lower than the first set of variables.

The optimal objective value (minimum AIC value) obtained by this procedure for the model estimating ALL crashes was 1798.52. This value was obtained at the 27th and 51st iterations with the same subset of independent variables. The subsets of variables in the model at those optimum iterations and their significance levels are given in the next section with the results of other cases.

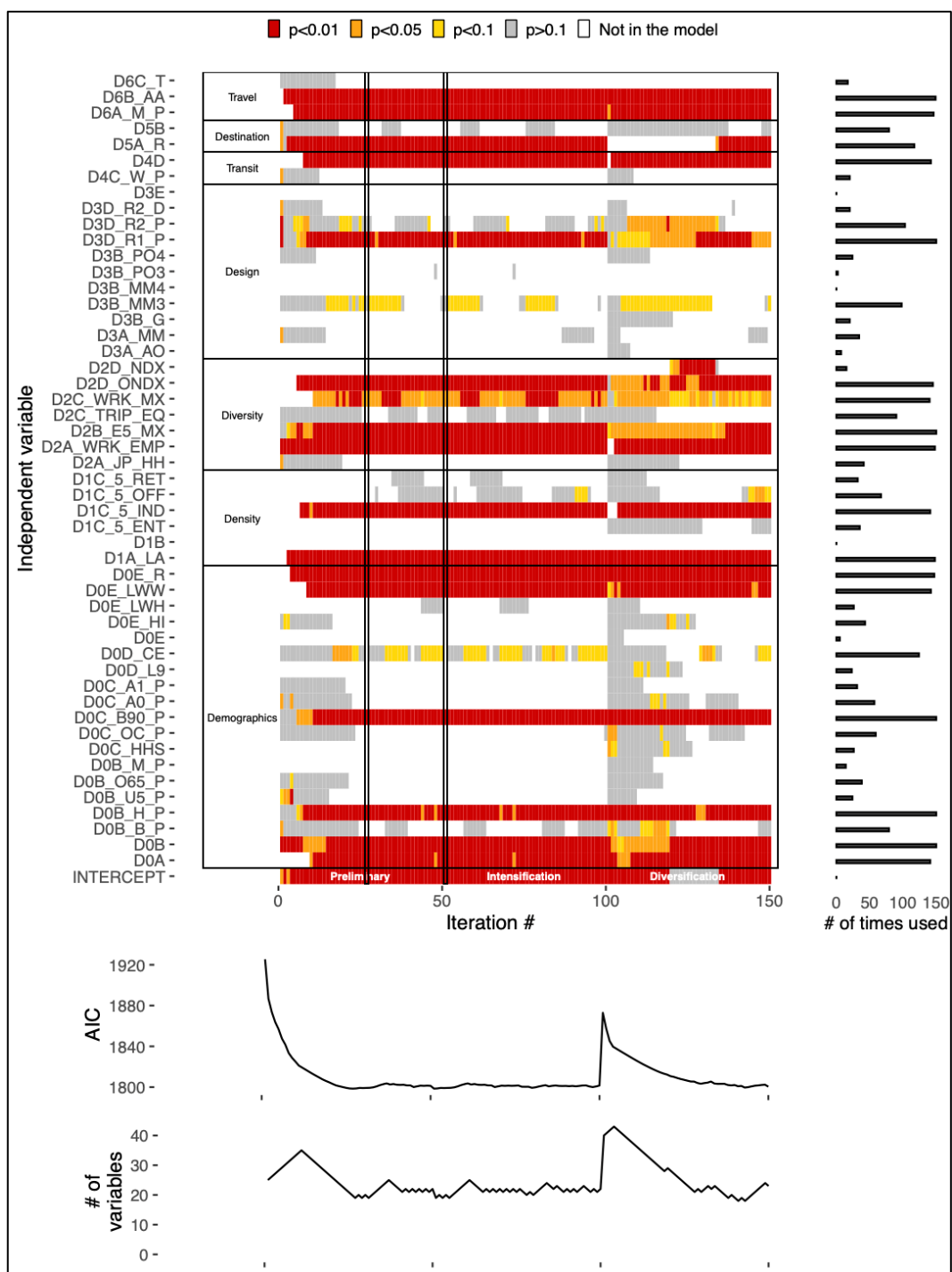


Figure 4.3. Tabu-search Results – Single County - FQ_ALL

4.4 Results of the Single County Case

In this section, the results of the single county case (Leon County) are presented. The presentation order is as follows: firstly, the results of the crash frequency models are presented. For each model, one of the crash types (e.g., frequency of fatal or incapacitating injury – FQ_S3) was used as the dependent variable. The independent variable set was the same for all models. For these crash frequency models, the tabu-based subset selection algorithm employed negative binomial regression at each iteration. Akaike Information Criterion (AIC) was used as the objective function, and the algorithms tried to minimize the AIC value. At the end of the tabu-based subset selection, the algorithm returns the subset of independent variables with the best (minimum) AIC value obtained. Secondly, the results of the crash density models are given. Three different statistical distributions were used in those models to fit the data: Gamma, Lognormal, and Weibull (only the Gamma regression results are presented in this chapter). The same procedure with the frequency models was applied. AIC values were minimized for the models, and the subsets of independent variables obtained at the optimum iteration are presented.

4.4.1 Single county-crash frequency modeling results

Table 4.2 shows the results of crash frequency models for severity types, and in Table 4.3, the results of crash frequency models for involvement (occupancy) are presented. In terms of BE variables, for each crash category, there were significant variables for different models.

Five variables were significant for all six TS variables (crash categories): total population (D0B), land area (D1A_LA), industrial employment density (D1C_5_IND), number of bus stops (D4D), and annual average daily traffic (D6B_AA). Among those five variables, industrial employment density had negative relationships with all crash types, whereas the other four variables had positive relationships with all crash types.

Table 4.2 Optimal BE-TS Models for Single County Case (FQ severity)

IV \ DV	FQ_S1			FQ_S2			FQ_S3		
	Beta		SE.	Beta		SE.	Beta		SE.
Intercept	3.65	***	0.31	(2.99)		2.06	(9.11)	***	2.72
D0A	(0.27)	***	0.09	(0.21)	**	0.09	---		---
D0B	0.13	***	0.04	0.13	***	0.05	0.20	***	0.05
D0B_H_P	0.02	**	0.01	0.03	***	0.01	---		---
D0C_HHS	---		---	---		---	(0.19)	**	0.09
D0C_B90_P	0.00	**	0.00	0.00	***	0.00	---		---
D0C_A0_P	---		---	---		---	---		---
D0D_CE	0.00	*	0.00	---		---	---		---
D0E_LWH	---		---	---		---	0.02	***	0.01
D0E_LWW	0.00	***	0.01	0.00	**	0.00	---		---
D0E_R	0.24	***	0.06	0.18	***	0.06	---		---
D1A_LA	0.01	***	0.00	0.01	***	0.00	0.02	***	0.00
D1B	---		---	---		---	---		---
D1C_5_ENT	---		---	0.11	*	0.06	0.32	***	0.07
D1C_5_IND	(0.25)	***	0.09	(0.31)	***	0.09	(0.27)	**	0.13
D1C_5_RET	---		---	---		---	---		---
D2A_JP_HH	---		---	(0.01)	**	0.01	(0.01)		0.01
D2A_WRK_EMP	(0.01)	***	0.00	(0.01)	***	0.00	---		---
D2B_E5_MX	0.36	***	0.10	0.26	***	0.10	---		---
D2C_WRK_MX	0.29	***	0.10	---		---	---		---
D2D_ONDX	(0.08)	***	0.03	(0.05)	**	0.02	(0.21)	***	0.03
D2D_NDX	---		---	0.08	***	0.02	0.13	***	0.03
D3B_MM3	0.00	*	0.00	---		---	---		---
D3B_PO3	---		---	---		---	(0.01)	***	0.00
D3D_R1_P	0.01	**	0.00	0.01	***	0.00	0.01	***	0.00
D3D_R2_P	---		---	0.01	***	0.00	0.03	***	0.01
D3D_R2_D	---		---	---		---	(0.02)	**	0.01
D3E	---		---	---		---	---		---
D4C_W_P	---		---	---		---	(0.07)	**	0.03
D4D	0.01	***	0.00	0.01	***	0.00	0.01	**	0.00
D5A_R	0.01	***	0.00	---		---	---		---
D5B	---		---	---		---	0.09	***	0.03
D6A_M_P	(0.01)	***	0.00	(0.01)	***	0.00	---		---
D6B_AA	0.00	***	0.00	0.00	***	0.00	0.00	**	0.00
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$									
Goodness of Fit									
AIC	1605.35			1513.28			853.18		
R²- Corrected (Nagelkerke)	0.83			0.85			0.68		

Table 4.3 Optimal BE-TS Models for Single County Case (FQ involvement)

IV \ DV	FQ_ALL			FQ_MOT			FQ_VRU		
	Beta		SE	Beta		SE	Beta		SE
Intercept	4.09	***	0.30	4.03	***	0.30	0.52		0.35
D0A	(0.24)	***	0.08	(0.24)	***	0.09	(0.30)	**	0.15
D0B	0.15	***	0.04	0.13	***	0.04	0.20	***	0.07
D0B_H_P	0.02	***	0.01	0.02	***	0.01	0.03	**	0.01
D0C_HHS	---		---	---		---	---		---
D0C_B90_P	0.00	***	0.00	0.00	***	0.00	0.01	***	0.00
D0C_A0_P	---		---	---		---	0.01	***	0.00
D0D_CE	---		---	0.00		0.00	---		---
D0E_LWH	---		---	---		---	0.02	***	0.01
D0E_LWW	0.00	***	0.00	0.01	***	0.00	---		---
D0E_R	0.21	***	0.06	0.21	***	0.06	---		---
D1A_LA	0.01	***	0.00	0.01	***	0.00	0.01	**	0.01
D1B	---		---	---		---	0.03	***	0.01
D1C_5_ENT	---		---	---		---	0.12	*	0.06
D1C_5_IND	(0.32)	***	0.09	(0.31)	***	0.09	(0.46)	***	0.12
D1C_5_RET	---		---	---		---	0.07	**	0.03
D2A_JP_HH	---		---	---		---	---		---
D2A_WRK_EMP	(0.01)	***	0.00	(0.01)	***	0.00	(0.01)		0.00
D2B_E5_MX	0.32	***	0.10	0.31	***	0.10	0.37	**	0.16
D2C_WRK_MX	0.24	**	0.10	0.24	**	0.10	0.50	***	0.14
D2D_ONDX	(0.09)	***	0.03	(0.09)	***	0.03	---		---
D2D_NDX	---		---	---		---	---		---
D3B_MM3	0.00		0.00	0.00		0.00	0.01	***	0.00
D3B_PO3	---		---	---		---	---		---
D3D_R1_P	0.01	***	0.00	0.01	***	0.00	---		---
D3D_R2_P	0.00		0.00	0.00	*	0.00	---		---
D3D_R2_D	---		---	---		---	---		---
D3E	---		---	---		---	0.01	*	0.01
D4C_W_P	---		---	---		---	---		---
D4D	0.01	***	0.00	0.01	***	0.00	0.01	**	0.00
D5A_R	0.02	***	0.00	0.02	***	0.00	---		---
D5B	---		---	---		---	---		---
D6A_M_P	(0.01)	***	0.00	(0.01)	***	0.00	(0.02)	***	0.00
D6B_AA	0.00	***	0.00	0.00	***	0.00	0.00	***	0.00
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$									
Goodness of Fit									
AIC	1798.52			1793.07			729.05		
R²- Corrected (Nagelkerke)	0.84			0.84			0.82		

Urban indicator (D0A), the percentage of the Hispanic population (D0B_H_P), the percentage of the housing units built before 1990 (D0C_B90_P), employment entropy (D2B_E5_MX), and percentage of motorized modes for commuting (D6A_M_P) had significant associations with all crash types except fatal or incapacitating injury crashes. Among those variables, urban indicator and the percentage of motorized modes for commuting had negative associations, while the others had positive associations.

Job access score (D2D_ONDX) and the percentage of primary roads (D3D_R1_P) were the other variables that have significant associations with most crash types (all except the crashes involving vulnerable road users). Job access score had negative associations with crashes, but the percentage of primary roads had positive associations.

The percentage of low wage workers living in the block group (D0E_LWH), median gross monthly rent (D0E_R), and household workers per job (D2A_WRK_EMP) were found to have significant associations with crash types other than the fatal or incapacitating injury crashes and the crashes that involve vulnerable road users. Household workers per job variable had a negative association, whereas the percentage of low wage workers and median gross monthly rent had positive associations. Moreover, household workers per job equilibrium index (D2C_WRK_MX) had positive associations with PDO crashes, ALL crashes, motor vehicle, and vulnerable road user involved crashes.

4.4.2 Single county-crash density modeling results

In Table 4.4 and Table 4.5, the results of crash density models with severity types and the results of crash density models with involvement (occupancy) are presented, respectively.

Table 4.4 Optimal BE-TS Models for Single County Case (KD severity)

IV \ DV	KD_S1			KD_S2			KD_S3		
	Beta		SE	Beta		SE	Beta		SE
Intercept	2.95	***	0.38	2.32	***	0.37	0.01		0.48
D0A	(0.22)	**	0.09	(0.28)	***	0.09	---		---
D0B	0.18	***	0.05	0.14	***	0.05	---		---
D0B_B_P	---		---	0.00	**	0.00	---		---
D0B_H_P	---		---	0.02	***	0.01	---		---
D0B_U5_P	(0.02)	**	0.01	(0.02)	*	0.01	---		---
D0C_B90_P	---		---	0.00	***	0.00	---		---
D0C_A0_P	0.01	*	0.00	---		---	---		---
D0D_L9	(0.01)		0.01	---		---	---		---
D0E_HI	---		---	---		---	---		---
D0E_LWH	0.01	**	0.01	---		---	0.04	***	0.01
D0E_LWW	0.01	***	0.00	0.00	***	0.00	---		---
D0E_R	0.22	***	0.07	0.16	**	0.06	---		---
D1A_LA	0.01	***	0.00	0.02	***	0.00	0.02		0.02
D1B	(0.03)	***	0.01	---		---	---		---
D1C_5_IND	(0.24)	***	0.09	(0.32)	***	0.09	---		---
D2A_WRK_EMP	(0.01)	***	0.00	(0.01)	***	0.00	---		---
D2B_E5_MX	0.31	***	0.11	0.42	***	0.10	---		---
D2C_WRK_MX	0.24	**	0.11	---		---	---		---
D2D_ONDX	(0.10)	***	0.03	(0.10)	***	0.03	(0.16)	**	0.07
D3B_PO3	---		---	(0.00)	*	0.00	(0.02)	***	0.01
D3D_R1_P	0.01	**	0.00	0.01	***	0.00	---		---
D4D	0.02	***	0.00	0.01	***	0.00	---		---
D5A_R	0.02	***	0.01	0.02	***	0.01	---		---
D5B	---		---	---		---	0.14	**	0.06
D6A_M_P	(0.01)	***	0.00	(0.01)	**	0.00	---		---
D6B_AA	0.00	***	0.00	0.00	***	0.00	0.00	***	0.00
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$									
Goodness of Fit									
AIC	1367.38			1250.17			681.55		
R²- Corrected (Nagelkerke)	0.84			0.87			0.26		

Table 4.5 Optimal BE-TS Models for Single County Case (KD involvement)

IV \ DV	KD_ALL			KD_MOT			KD_VRU		
	Beta		SE	Beta		SE	Beta		SE
Intercept	3.44	***	0.36	3.36	***	0.37	3.32	**	1.33
D0A	(0.28)	***	0.09	(0.28)	***	0.09	---		---
D0B	0.18	***	0.05	0.17	***	0.05	0.33	*	0.19
D0B_B_P	---		---	---		---	---		---
D0B_H_P	---		---	---		---	---		---
D0B_U5_P	(0.02)	*	0.01	(0.02)	*	0.01	---		---
D0C_B90_P	0.00	*	0.00	0.00	*	0.00	---		---
D0C_A0_P	---		---	---		---	---		---
D0D_L9	---		---	---		---	---		---
D0E_HI	---		---	---		---	(0.02)	***	0.00
D0E_LWH	0.01	**	0.00	0.01	**	0.00	---		---
D0E_LWW	0.01	***	0.00	0.01	***	0.00	---		---
D0E_R	0.17	***	0.06	0.17	***	0.06	---		---
D1A_LA	0.02	***	0.00	0.02	***	0.00	---		---
D1B	(0.02)	**	0.01	(0.02)	**	0.01	---		---
D1C_5_IND	(0.26)	***	0.08	(0.26)	***	0.08	---		---
D2A_WRK_EMP	(0.01)	***	0.00	(0.01)	***	0.00	(0.01)	*	0.01
D2B_E5_MX	0.31	***	0.10	0.32	***	0.10	---		---
D2C_WRK_MX	0.22	**	0.10	0.22	**	0.10	---		---
D2D_ONDX	(0.11)	***	0.03	(0.12)	***	0.03	---		---
D3B_PO3	---		---	---		---	---		---
D3D_R1_P	0.01	***	0.00	0.01	***	0.00	---		---
D4D	0.02	***	0.00	0.02	***	0.00	---		---
D5A_R	0.02	***	0.01	0.02	***	0.01	---		---
D5B	---		---	---		---	---		---
D6A_M_P	(0.01)	***	0.00	(0.01)	***	0.00	(0.03)	**	0.01
D6B_AA	0.00	***	0.00	0.00	***	0.00	0.01	***	0.00
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$									
Goodness of Fit									
AIC	1549.47			1542.01			325.25		
R²- Corrected (Nagelkerke)	0.86			0.86			0.31		

According to the results, fewer BE variables were found to be significant for the density models than the frequency models. According to the crash density models, when Table t533 and Table t534 are considered together, annual average daily traffic (D6B_AA) was positively associated with all crash types as expected. Other than the traffic volume, total population (D0B), household workers per job (D2A_WRK_EMP), job access score (D2D_ONDX), and the percentage of motorized modes for commuting (D6A_M_P) were the common BE variables found to be significantly associated with all TS variables (crash types). The total population was positively associated with all crash types except fatal or incapacitating injury crashes. Household workers per job and percentage of motorized modes for commuting were also associated with the same crash types but negatively. On the other hand, job access score was negatively associated with all the crash types except the crashes involving vulnerable road users.

The other common BE variables and their associations are as follows: urban indicator (D0A), the percentage of population under the age of 5 (D0B_U5_P), and gross industrial employment density (D1C_5_IND) were negatively associated with PDO crashes, non-incapacitating injury crashes, ALL crashes and motor vehicle involved crashes. Whereas the percentage of low wage workers working in the block group (D0E_LWW), median gross monthly rent (D0E_R), land area (D1A_LA), employment entropy (D2B_E5_MX), the percentage of primary roads (D3D_R1_P), the number of bus stops (D4D) and the number of jobs within 45 minutes auto travel time (D5A_R) were positively associated with the same crash types. Moreover, the percentage of low-wage workers living in the block group (D0E_LWH) was positively associated with PDO crashes, fatal or incapacitating injury crashes, ALL crashes, and motorized vehicle involved crashes.

4.4.3 Discussion of the single county crash modeling outcomes

In this section, the results of crash frequency and crash density models are discussed. In Table 4.6, for each crash category (PDO as S1, non-incapacitating injury as S2,

fatal or incapacitating injury as S3, total crashes as ALL, motor vehicle involved crashes as MOT, and vulnerable road user involved crashes as VRU) both frequency (crash counts for 2009-2011 period) and density (total kernel density of crashes) modeling results are depicted.

Table 4.6 Comparison of the IVs for Single County Case

IV	90% 99%	+	-	DV	S1		S2		S3		ALL		MOT		VRU	
					NB	GA	NB	GA	NB	GA	NB	GA	NB	GA	NB	GA
D0A																
D0B																
D0C_B90_P																
D0E_LWW																
D0E_R																
D1A_LA																
D1C_5_IND																
D2A_WRK_EMP																
D2B_E5_MX																
D2D_ONDX																
D3D_R1_P																
D4D																
D5A_R																
D6A_M_P																
D6B_AA																

For each crash category, the first column is dedicated to frequency modeling (negative binomial regression), and the second column is dedicated to density modeling (gamma regression). The table's rows are dedicated to the BE variables for which a significant association (higher than 90%) was found at least in one model.

The color code represents both the significance levels (90% and 99%) and the direction of association (positive or negative). For example, the urban indicator (D0A) was found to be negatively associated with PDO crashes with a significance level of 99% in the frequency model (NB) and a significance level of 90% in the density model (GA); whereas, the total population (D0B) was found to be positively associated with the same crash type with 99% significance for both models. While presenting the results, the most common variables that were significant in at least four of the crash types were selected.

According to the results, urban identification was negatively associated with PDO crashes, non-incapacitating injury crashes, ALL crashes, and motor vehicle involved crashes significantly. In the former studies, “distance to the nearest urban area” was found to be negatively associated with total crashes (Cai, Abdel-Aty, Lee, et al., 2017) and with vulnerable road user crashes (Cai, Abdel-Aty, Lee, et al., 2017, 2019), so, urbanized areas are prone to increased crash frequencies. Similarly, urban areas were linked positively to total (Jermprapai & Srinivasan, 2014) and bike crash frequencies (Siddiqui et al., 2012). Whereas fatal crashes were generally linked to rural areas since driver speeds are higher in rural areas, and as speeds increase, the severity of the crashes increase (Jermprapai & Srinivasan, 2014; Zhibin Li et al., 2013; Noland & Quddus, 2004a). Therefore, the findings in this study contradicted the ones in the literature. Since population and traffic densities are higher in the urban areas, it is expected to have increased crash rates in the urban areas. However, the block groups' size gets larger while approaching the fringe, so urban block groups are smaller than rural and suburban block groups. One explanation for the negative association with urban identification could be this size-effect. Though the crash densities increase at the urban core, the total number of crashes could be higher at the larger block groups, supported by the positive and significant association of the land area with all crash types except vulnerable road user involved crashes. VRU crashes would be more likely to happen in urban areas where block groups are smaller than the rural block groups, and there are more attraction points.

Under the demographic and socioeconomic properties dimension, the total population was found to be positively associated with all crash types, as expected. The total population was considered a surrogate for exposure and was positively related to total crashes (Amoh-Gyimah et al., 2017; H. Huang et al., 2016; K. Kim et al., 2006; P. Xu et al., 2017). As exposure increases, the likelihood of crashes increases. Furthermore, total population and pedestrian crashes had a positive association since a higher population would generate more mobility and exposure (K. Kim et al., 2006; Nashad et al., 2016; Noland & Quddus, 2004b; Quistberg et al., 2015; Siddiqui & Abdel-Aty, 2012). Therefore, as the population increases, all crash types' frequency (and density) likely to increase.

The studies that considered the relationship between household income and traffic crashes found negative associations with total crashes (Amoh-Gyimah et al., 2017; Noland & Quddus, 2004a; Pirdavani et al., 2013; Rhee et al., 2016; K. Xie et al., 2019; C. Xu et al., 2017). A similar negative relation was found between income and pedestrian crashes (Bao et al., 2017; Cottrill & Thakuriah, 2010; Siddiqui et al., 2012). The percentage of low-wage workers was also found to be significantly positive with PDO, non-incapacitating injury crashes, all and motor vehicle involved crashes. On the other hand, median gross rent as a surrogate of income had a different picture with the same crash types. Here, the findings are the complete opposite: as the amount of rent paid increases (so the wealth indirectly), the crash frequency (and density) increases. One explanation for this result is the possible increase in car ownership. As wealth increases, car ownership and usage would increase, so the exposure. Thus, wealth could also be negatively associated with traffic safety.

Industrial employment density was negatively associated with PDO, non-incapacitating injury, total, motor vehicle involved, and vulnerable road user involved crashes. The findings in the literature about industrial employment are mixed. Some studies found a negative association between vulnerable road user-involved crashes and industrial employment (Cai et al., 2016; P. Chen & Zhou, 2016; Ding, Chen, et al., 2018). The negative association was explained by the low likelihood of pedestrian activities in those sites for those crashes. However, some

studies found a positive association (Miranda-Moreno et al., 2011; Ukkusuri et al., 2012) for the vulnerable road user involved crashes by increased traffic volumes in those sites. In this study, the relationship was negative, supporting the commonsense that those areas experience lower levels of vulnerable road user involved crashes. For the other crash types, the findings are also mixed. Some studies found negative associations (Kaygisiz et al., 2017; Pulugurtha et al., 2013), and some found a positive association (Chiou & Fu, 2013). The negative associations of industrial employment density with all crash types in this study were mostly due to the monotonous structure of the industrial districts compared to office and retailing employment. This is also supported by the employment entropy, which was positively associated with PDO, non-incapacitating injury, total, and motorized vehicle crashes. That is to say, as the heterogeneity of employment increases, the likelihood of crashes increases due to the exposure, and as the employment becomes homogenous, the likelihood decreases.

The household workers per job variable was found to have negative associations with PDO, non-incapacitating, all, and motor vehicle involved crashes. As household workers per job increases, employment density decreases (not a work zone). Those block groups would have residential characteristics that are also homogenous. Job access score was also found to be negatively associated with PDO, non-incapacitating, fatal crashes, all, and motor vehicle involved crashes. Higher values of job access score indicate higher access to a variety of jobs. Therefore, the negative relationship would be due to both better accessibility and alternative jobs.

According to the results, the percentage of the primary roads was positively associated with PDO, non-incapacitating injury, fatal, all, and motor vehicle involved crashes. Most of the former studies came up with similar results that primary roads were positively associated with crash risk and frequency. The road density of major-roads was positively associated with total crashes (Bao et al., 2017, 2018; C.-Y. Yu & Xu, 2018). Road density is also positively associated with pedestrian crashes (Osama & Sayed, 2017; Ukkusuri et al., 2011; Y. Wang & Kockelman, 2013; Wier et al., 2009).

According to the single county modeling results, the number of bus stops was positively associated with all types of crashes in the frequency models and all crashes except vulnerable road user crashes and fatal crashes. The findings in the literature support the positive associations. The number of bus stops found to be positively related not only to the total crashes (Cai, Abdel-Aty, & Lee, 2017; Ouyang & Bejleri, 2014), but also the motor vehicle involved crashes (Kaygisiz et al., 2017; Ouyang & Bejleri, 2014), and pedestrian crashes (P. Chen & Zhou, 2016; Ouyang & Bejleri, 2014; Ukkusuri et al., 2012; Y. Wang & Kockelman, 2013). Bus stops are the conflict points for all types of road users and vehicles. Vehicles may have conflicts with buses (Cai, Abdel-Aty, & Lee, 2017), or the buses block pedestrians' sight while crossing (P. Chen & Zhou, 2016). Thus, increased conflict points increase the likelihood of crashes.

The percentage of motorized modes (excluding transit) was negatively associated with property damage only crashes, non-incapacitating injury crashes, total crashes, and motor vehicle involved crashes. It was surprising to have that association since it was expected to observe higher rates of crashes of any type as motor vehicle usage increases. Those block groups with higher car ownership rates and relatively higher income levels were located at the urban periphery. Those properties, especially income level, were associated with lower crash rates in former studies (Amoh-Gyimah et al., 2017; Bao et al., 2017; Pirdavani et al., 2013). Therefore, rather than the exposure impact by car ownership, the deprivation level would surpass the positive impact for the deprived communities who use transit and non-motorized modes primarily. One other indirect impact would be due to the positive association between education level and income. Income is positively associated with education level, and education level is negatively associated with crash rates, so there would be a latent impact of income through the level of education. Additionally, pedestrian and bicyclist activities are expected to be lower at the block groups close to the fringe than the block groups close to the city center. Moreover, on the other side, in a recent study, it was found that non-motorized modes and transit positively associated with crashes in Florida (Lee, Abdel-Aty, de Blasiis, et al., 2019).

Finally, as an indicator of traffic volume and exposure, annual average daily traffic (AADT) was positively associated with all types of crashes. The result is evident since increasing traffic volume would result in more exposure and increased likelihood of crashes, which complies with the former studies (Cai, Abdel-Aty, Lee, et al., 2019; H. Huang et al., 2016; Loukaitou-Sideris et al., 2007; J. Wang & Huang, 2016; Zeng & Huang, 2014).

4.4.4 Consistency of the single county crash modeling results

The consistency of the selected BE variables for the single county case will be discussed in this section. In the previous sections, the best models (the best iteration in the tabu-search based solution procedure for each model) and the BE variables at the optimum solutions were presented. It is also essential to determine the consistency so the BE variables' consistency concerning the TS variables. Since the algorithm is an iterative process with approximately 7500 trials, the selected variables often vary. Moreover, the relation (positive, negative, or none) between a BE variable and a TS variable would also vary due to the model's variables at a specific iteration. A variable would have a positive association at an iteration and a negative association at another iteration. This variability would imply that the other variables would play a role in the association of this BE variable with the TS variable.

A scheme was developed to detect the consistency levels of the BE variables, which is depicted in Table 4.7. In this scheme, the tabu search iterations were used. For each BE variable during each BE-TS model solution process, four statistics were kept. The number of times the variable was in the subset, the number of times that specific variable was negative, the number of times the variable was positive, and the number of times was zero when it is in the model.

Then the BE variables were classified into five categories depending on two dimensions. The first dimension was about the percentage of iterations that variable

was in the subset (basic). That dimension had three sub-categories: less than 80%, 80% or more, and more than 90%.

The second dimension was about the type of association of the BE variable with the TS variable. That dimension also had two sub-categories: positive (+) and negative (-). For this dimension, 90% was the threshold value. To be evaluated as positive (or negative), a BE variable should have a positive (or negative) association in at least 90% of the iterations it was selected to be in the model.

Therefore, four categories were defined and depicted depending on these two dimensions. As the fifth category, the variables that appeared in less than 80% of the iterations and the variables that appeared in at least 80% of the iterations with no association were considered together and identified as inconsistent. This category was represented by white color. As an example, the urban indicator (D0A) consistently had a negative relationship with PDO crashes (S1), total crashes (ALL), and motor vehicle involved crashes (MOT) for both frequency and density models. The association's existence and type were less consistent for non-incapacitating crashes (S2) for the same modeling frames. The urban indicator variable was inconsistent for the fatal or incapacitating injury crashes (S3), and vulnerable road user involved crashes (VRU).

Table 4.7 Consistency of the IVs for Single County Case

IV	DV	S1		S2		S3		ALL		MOT		VRU	
		NB	GA	NB	GA	NB	GA	NB	GA	NB	GA	NB	GA
D0A													
D0B													
D0B_B_P													
D0B_H_P													
D0B_U5_P													
D0C_HHS													
D0C_OC_P													
D0C_B90_P													
D0C_A0_P													
D0D_CE													
D0E_HI													
D0E_LWH													
D0E_LWW													
D0E_R													
D1A_LA													
D1C_5_ENT													
D1C_5_IND													
D2A_JP_HH													
D2A_WRK_EMP													
D2B_E5_MX													
D2C_WRK_MX													
D2D_ONDX													
D2D_NDX													
D3A_MM													
D3B_PO3													
D3D_R1_P													
D3D_R2_P													
D3D_R2_D													
D4C_W_P													
D4D													
D5A_R													
D5B													
D6A_M_P													
D6B_AA													

One of the main aims of this study was to determine the generalizable built environment factors affecting traffic safety. Thus, the BE variables that were considered consistent concerning at least four TS variables were considered consistent in this scheme. According to this consistency scheme, the following variables were found consistent in their associations with TS variables:

- Urban indicator (D0A)
- Total population (D0B)
- Hispanic % of population (D0B_H_P)
- % of housing units built before 1990 (D0C_B90_P)
- % of low wage (less than \$1250/month) workers (work location) (D0E_LWW)
- Median gross monthly rent (D0E_R)
- Land area in acres (D1A_LA)
- Gross industrial employment density (D1C_5_IND)
- Household workers per job (D2A_WRK_EMP)
- Employment entropy (D2B_E5_MX)
- Job access score (D2D_ONDX)
- % of primary roads (D3D_R1_P)
- Number of bus stops (D4D)
- % of motorized modes for commuting (excluding transit) (D6A_M_P)
- Annual average daily traffic (average of 3 years) for all modes (D6B_AA)

Finally, the BE variables were filtered depending on the significance and consistency. The BE variables' significance levels presented in the previous section

and the consistency evaluations presented in this section were amalgamated. Thus, the variables that were both significant at the optimum iterations and consistent during the process were determined by comparison. Among the variables presented as consistent, the Hispanic percentage of population (D0B_H_P) was not significant in most BE-TS models for the single county case. Among the significant variables in Table 4.6, jobs within 45 minutes auto travel time (D5A_R) was not consistent. Therefore, the remaining variables were evaluated as the key factors (both significant and consistent) for the single county case.

- Urban indicator (D0A)
- Total population (D0B)
- % of housing units built before 1990 (D0C_B90_P)
- % of low wage (less than \$1250/month) workers (work location) (D0E_LWW)
- Median gross monthly rent (D0E_R)
- Land area in acres (D1A_LA)
- Gross industrial employment density (D1C_5_IND)
- Household workers per job (D2A_WRK_EMP)
- Employment entropy (D2B_E5_MX)
- Job access score (D2D_ONDX)
- % of primary roads (D3D_R1_P)
- Number of bus stops (D4D)
- % of motorized modes for commuting (excluding transit) (D6A_M_P)
- Annual average daily traffic (average of 3 years) for all modes (D6B_AA)

4.5 Results of the Tri-County Case

This section presents the tri-county (Broward, Miami-Dade, and Palm Beach counties). BE-TS modeling results. Firstly, the crash frequency modeling results, then the crash density modeling results are presented.

4.5.1 Tri-county-crash frequency modeling results

This section presents the tri-county crash frequency modeling results Table 4.8 shows the results of crash frequency models concerning severity types, and in Table 4.9, the results of crash frequency models concerning involvement are presented.

According to the results, among the BE variables, total population (D0B), gross entertainment employment density (D1C_5_ENT), gross retail employment density (D1C_5_RET), the percentage of secondary roads (D3D_R2_P), total sidewalk length (D3E), transit ridership percentage of workers (D4C_W_P), the number of bus stops (D4D), walking index (D5B), the percentage of non-motorized modes for commuting (D6A_NM_P) and annual average daily traffic (D6B_AA) were found to be positively related to all crash types. The proportion of total employment within a quarter-mile of a fixed-guideway transit stop (D4B_025) and gross population density (D1B) were negatively related to all crash types.

Three BE variables had mixed relations with crash types though being significantly related to all crash types. Hispanic percentage of the population (D0B_H_P) and the number of jobs within 45 minutes auto travel time (D5A_R) were found to be negatively related to fatal or incapacitating injury crashes and vulnerable road user involved crashes; while positively related to other crash types. Additionally, the percentage of primary roads (D3D_R1_P) was positively associated with all crash types except vulnerable road user involved crashes.

Table 4.8 Optimal BE-TS Models for Tri-County Case (FQ severity)

IV \ DV	FQ_S1			FQ_S2			FQ_S3		
	Beta		SE	Beta		SE	Beta		SE
Intercept	1.51	***	0.16	0.97	***	0.20	0.23		0.33
D0A	(0.27)	***	0.07	---	---	---	(0.19)	**	0.08
D0B	0.14	***	0.01	0.13	***	0.01	0.19	***	0.02
D0B_B_P	---		---	0.00	**	0.00	(0.00)	***	0.00
D0B_H_P	0.01	***	0.00	0.00	***	0.00	(0.01)	***	0.00
D0B_U5_P	(0.02)	***	0.01	---		---	---		---
D0B_O65_P	(0.01)	***	0.00	---		---	---		---
D0B_M_P	---		---	0.00		0.00	0.01	**	0.00
D0C_HHS	---		---	---		---	0.10	***	0.03
D0C_OC_P	0.00	***	0.00	0.01	***	0.00	0.01	***	0.00
D0C_B90_P	0.00	**	0.00	0.00	*	0.00	0.00		0.00
D0C_CO2_HH	0.02	***	0.01	0.01		0.01	0.02	**	0.01
D0C_A0_P	0.00	***	0.00	---		---	---		---
D0C_A1_P	---		---	0.00	**	0.00	(0.00)	**	0.00
D0D_L9	---		---	(0.00)	*	0.00	---		---
D0D_CE	0.00	***	0.00	(0.00)	***	0.00	(0.01)	***	0.00
D0E_HI	---		---	(0.00)	**	0.00	---		---
D0E_LWH	---		---	0.00	*	0.00	---		---
D0E_LWW	---		---	0.00		0.00	0.00	***	0.00
D0E_R	---		---	0.03	*	0.02	---		---
D1A_LA	0.00	***	0.00	0.00	***	0.00	0.00	***	0.00
D1A_BS	---		---	---		---	---		---
D1B	(0.01)	***	0.00	(0.01)	***	0.00	(0.01)	***	0.00
D1C_5_ENT	0.02	***	0.00	0.02	***	0.00	0.01	***	0.00
D1C_5_IND	---		---	---		---	(0.02)	**	0.01
D1C_5_RET	0.02	*	0.01	0.02	*	0.01	0.02	*	0.01
D1C_5_SVC	---		---	---		---	(0.01)	*	0.00
D2A_JP_HH	---		---	0.00	*	0.00	---		---
D2A_WRK_EMP	0.00		0.00	---		---	---		---
D2B_E5_MX	0.18	***	0.04	0.15	***	0.04	---		---
D2C_TRIP_EQ	---		---	0.07	*	0.04	0.18	***	0.04
D2C_WRK_MX	0.15	***	0.04	0.14	***	0.04	---		---
D2D_ONDX	(0.05)	***	0.01	---		---	(0.03)	**	0.01
D2D_NDX	---		---	---		---	0.01	**	0.00
D3A	(0.01)	***	0.00	(0.01)	***	0.00	(0.01)	***	0.00
D3A_AO	---		---	0.02	***	0.00	0.02	***	0.01
D3A_MM	---		---	0.01	**	0.01	0.03	***	0.01
D3B_G	---		---	---		---	(0.00)	***	0.00
D3B_MM3	---		---	(0.00)	**	0.00	(0.00)	***	0.00
D3B_MM4	---		---	(0.00)	*	0.00	(0.00)	*	0.00
D3B_PO3	(0.00)	*	0.00	(0.00)		0.00	---		---
D3B_PO4	0.00	**	0.00	---		---	---		---
D3D_R1_P	0.01	***	0.00	0.01	***	0.00	0.01	***	0.00
D3D_R2_P	0.01	***	0.00	0.01	***	0.00	0.01	***	0.00
D3E	0.02	***	0.00	0.03	***	0.00	0.01	**	0.00
D4B_025	(0.96)	***	0.14	(0.37)	***	0.12	(0.52)	***	0.17
D4C_W_P	0.03	***	0.00	0.01	**	0.00	0.02	***	0.01
D4D	0.01	***	0.00	0.01	***	0.00	0.01	***	0.00
D5A_R	0.01	***	0.00	0.00	***	0.00	(0.00)	***	0.00
D5B	0.05	***	0.01	0.03	***	0.00	0.04	***	0.01
D6A_M_P	---		---	0.00	*	0.00	0.00	**	0.00
D6A_NM_P	0.01	***	0.00	0.01	***	0.00	0.01	***	0.00

Table 4.8 (continued)

D6B_AA	0.00	***	0.00	0.00	***	0.00	0.00	***	0.00
D6C_T	---		---	---		---	(0.06)	***	0.01
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$									
Goodness of Fit									
AIC	35048.37			31526.06			19411.84		
R²- Corrected (Nagelkerke)	0.73			0.68			0.53		

Table 4.9 Optimal BE-TS Models for Tri-County Case (FQ involvement)

IV \ DV	FQ_ALL			FQ_MOT			FQ_VRU		
	Beta		SE	Beta		SE	Beta		SE
Intercept	2.26	***	0.15	2.11	***	0.14	(1.04)	***	0.33
D0A	(0.18)	***	0.06	(0.17)	***	0.06	(0.14)	*	0.07
D0B	0.14	***	0.01	0.14	***	0.01	0.12	***	0.01
D0B_B_P	---		---	0.00		0.00	(0.00)	**	0.00
D0B_H_P	0.00	***	0.00	0.01	***	0.00	(0.00)	***	0.00
D0B_U5_P	(0.01)	*	0.01	(0.01)	**	0.01	0.01		0.01
D0B_O65_P	(0.00)	***	0.00	(0.00)	**	0.00	(0.00)	*	0.00
D0B_M_P	---		---	---		---	0.02	***	0.00
D0C_HHS	---		---	---		---	0.05	**	0.03
D0C_OC_P	0.01	***	0.00	0.01	***	0.00	---		---
D0C_B90_P	0.00	*	0.00	0.00	*	0.00	0.00	***	0.00
D0C_CO2_HH	0.02	***	0.01	0.02	***	0.01	0.01	*	0.01
D0C_A0_P	0.00	**	0.00	0.00	**	0.00	0.00	***	0.00
D0C_A1_P	---		---	---		---	---		---
D0D_L9	---		---	---		---	---		---
D0D_CE	---		---	---		---	(0.00)	***	0.00
D0E_HI	(0.00)		0.00	---		---	(0.00)	**	0.00
D0E_LWH	---		---	---		---	0.01	***	0.00
D0E_LWW	---		---	---		---	0.00		0.00
D0E_R	0.03	*	0.02	0.03	*	0.02	0.04	*	0.02
D1A_LA	0.00	***	0.00	0.00	***	0.00	---		---
D1A_BS	---		---	---		---	(0.00)	*	0.00
D1B	(0.01)	***	0.00	(0.01)	***	0.00	(0.01)	***	0.00
D1C_5_ENT	0.02	***	0.00	0.02	***	0.00	0.02	***	0.00
D1C_5_IND	---		---	---		---	---		---
D1C_5_RET	0.02	**	0.01	0.02	**	0.01	0.03	***	0.01
D1C_5_SVC	---		---	---		---	---		---
D2A_JP_HH	0.00		0.00	0.00		0.00	---		---
D2A_WRK_EMP	0.00	*	0.00	0.00	*	0.00	---		---
D2B_E5_MX	0.18	***	0.04	0.17	***	0.04	0.27	***	0.05
D2C_TRIP_EQ	0.08	*	0.04	0.07		0.04	0.11	**	0.04
D2C_WRK_MX	0.13	***	0.04	0.14	***	0.04	---		---
D2D_ONDX	(0.03)	***	0.01	(0.03)	***	0.01	0.03	***	0.01
D2D_NDX	---		---	---		---	0.01	**	0.00
D3A	(0.01)	***	0.00	(0.01)	***	0.00	---		---
D3A_AO	0.01	***	0.00	0.02	***	0.00	---		---

Table 4.9 (continued)

D3A_MM	---		---	---		---	---		---
D3B_G	---		---	(0.00)	*	0.00	---		---
D3B_MM3	---		---	---		---	---		---
D3B_MM4	---		---	---		---	---		---
D3B_PO3	(0.00)	**	0.00	(0.00)		0.00	(0.00)	***	0.00
D3B_PO4	0.00	*	0.00	0.00	*	0.00	---		---
D3D_R1_P	0.01	***	0.00	0.01	***	0.00	(0.01)	***	0.00
D3D_R2_P	0.00	***	0.00	0.00	***	0.00	0.01	***	0.00
D3E	0.03	***	0.00	0.03	***	0.00	0.03	***	0.00
D4B_025	(0.78)	***	0.12	(0.78)	***	0.12	(0.80)	***	0.12
D4C_W_P	0.03	***	0.00	0.03	***	0.00	0.03	***	0.00
D4D	0.01	***	0.00	0.01	***	0.00	0.01	***	0.00
D5A_R	0.00	***	0.00	0.00	***	0.00	(0.00)	***	0.00
D5B	0.03	***	0.01	0.03	***	0.01	0.04	***	0.01
D6A_M_P	---		---	---		---	---		---
D6A_NM_P	0.01	***	0.00	0.01	***	0.00	0.01	***	0.00
D6B_AA	0.00	***	0.00	0.00	***	0.00	0.00	***	0.00
D6C_T	---		---	---		---	(0.04)	***	0.01
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$									
Goodness of Fit									
AIC	38436.07			38200.96			18364.52		
R²- Corrected (Nagelkerke)	0.72			0.72			0.58		

The percentage of occupied housing units (D0C_OC_P) and total land area (D1A_LA) were positively associated with all crash types other than vulnerable road user involved crashes. The percentage of housing units built before 1990 (D0C_B90_P) and employment entropy (D2B_E5_MX) were positively associated with all crash types except fatal and incapacitating injury crashes. Annual CO2 emission per household (D0C_CO2_HH) was found to be positively associated with all crashes except non-incapacitating injury crashes. In contrast, the urban indicator (D0A) and job access score (D2D_ONDX) were negatively associated with the same crashes. Also, the total road network density (D3A) had negative associations with all crashes other than those involving vulnerable road users.

Some other BE variables that commonly have significant associations with TS variables, and their associations are as follows: the percentage of households with zero automobiles (D0C_A0_P) was positively related to PDO crashes, ALL crashes, moto vehicle involved crashes, and vulnerable road user involved crashes. Median

gross monthly rent (D0E_R) was positively associated with non-incapacitating injury crashes, ALL crashes, moto vehicle involved crashes, and vulnerable road user involved crashes.

Trip productions and trip attractions equilibrium index (D2C_TRIP_EQ) was positively related to non-incapacitating and incapacitating injury crashes, ALL crashes, and vulnerable road user involved crashes. Employment and household entropy (D2C_TRIP_MX) was found to be positively associated with PDO and non-incapacitating injury crashes, ALL, and motor vehicle involved crashes. Network density in terms of facility miles of auto-oriented links per square mile (D3A_AO) had positive associations with non-incapacitating injury and incapacitating injury crashes, ALL and motor vehicle involved crashes. The percentage of the population above the age of 65 (D0B_O65_P) had negative relations with PDO crashes, ALL, motor vehicle involved, and vulnerable road user involved crashes. Furthermore, the percentage of the population with at least a bachelor's degree (D0D_CE) had mixed associations. It was positively associated with PDO crashes but negatively associated with non-incapacitating and incapacitating injury crashes and vulnerable road user involved crashes.

4.5.2 Tri-county-crash density modeling results

In this part, the results of the tri-county case in terms of crash density models are presented. Table 4.10 results from crash density models for severity types, and in Table 4.11, the results of crash density models for involvement are presented. When considered together, the following results are obtained.

Table 4.10 Optimal BE-TS Models for Tri-County Case (KD severity)

IV \ DV	KD_S1			KD_S2			KD_S3		
	Beta		SE	Beta		SE	Beta		SE
Intercept	0.64	**	0.31	0.10		0.22	(0.33)		0.49
D0A	(0.49)	***	0.08	(0.26)	***	0.07	(0.610)	***	0.15
D0B	0.21	***	0.02	0.19	***	0.02	0.25	***	0.03
D0B_B_P	---		---	0.00	**	0.00	(0.00)	*	0.00
D0B_H_P	0.01	***	0.00	0.00	***	0.00	(0.01)	***	0.00
D0B_U5_P	(0.02)	***	0.01	---		---	---		---
D0B_O65_P	(0.01)	***	0.00	---		---	---		---
D0B_M_P	0.01	*	0.00	0.01	***	0.00	0.02	***	0.01
D0C_HHS	0.06	**	0.03	---		---	---		---
D0C_OC_P	---		---	0.01	***	0.00	0.01	***	0.00
D0C_B90_P	0.00	***	0.00	0.00	**	0.00	---		---
D0C_CO2_HH	0.02	***	0.01	0.01	*	0.01	0.03	*	0.01
D0C_A0_P	0.01	***	0.00	---		---	---		---
D0C_A1_P	---		---	0.00	**	0.00	---		---
D0D_CE	0.00	***	0.00	(0.00)	***	0.00	(0.01)	***	0.00
D0E_HI	---		---	(0.00)	**	0.00	---		---
D0E_LWH	0.01	***	0.00	0.01	***	0.00	---		---
D0E_LWW	---		---	0.00		0.00	0.00	**	0.00
D0E_BP	0.00		0.00	---		---	---		---
D0E_R	---		---	0.03		---	---		---
D1A_LA	0.00	***	0.00	0.00	***	0.00	0.00	**	0.00
D1B	(0.01)	***	0.00	(0.01)	***	0.00	(0.01)	***	0.00
D1C_5_ENT	0.02	***	0.00	0.02	***	0.00	0.03	***	0.01
D1C_5_IND	---		---	---		---	(0.02)	**	0.01
D1C_5_RET	0.03	***	0.01	0.02	**	0.01	---		---
D2B_E5_MX	0.16	***	0.05	0.18	***	0.05	---		---
D2C_TRIP_EQ	---		---	---		---	0.21	**	0.09
D2C_WRK_MX	0.11	**	0.05	0.15	***	0.04	---		---
D2D	0.01	*	0.01	0.02	**	0.01	---		---
D2D_ONDX	(0.04)	***	0.01	---		---	---		---
D2D_NDX	(0.00)		0.00	0		0	---		---
D3A	(0.01)	***	0.00	(0.01)	**	0.00	---		---
D3A_AO	0.02	***	0.01	0.04	***	0.01	0.03	***	0.01
D3A_MM	0.01	***	0.01	0.01	***	0.01	0.02		0.01
D3B_G	(0.00)	***	0.00	(0.00)	***	0.00	(0.00)	***	0.00
D3B_PO3	---		---	(0.00)	*	0.00	(0.00)	**	0.00
D3B_PO4	0.00	***	0.00	0.00	**	0.00	---		---
D3D_R1_P	0.02	***	0.00	0.01	***	0.00	0.01	*	0.00
D3D_R2_P	0.00	***	0.00	0.00	***	0.00	0.01	***	0.00
D3E	0.03	***	0.00	0.04	***	0.00	0.02	**	0.01
D4B_025	(0.90)	***	0.15	(0.33)	***	0.13	---		---
D4C_W_P	0.03	***	0.00	---		---	---		---
D4D	0.01	***	0.00	0.01	***	0.00	0.01	***	0.00
D5A_R	0.01	***	0.00	0.00	***	0.00	(0.00)	***	0.00
D5B	0.06	***	0.01	0.04	***	0.01	0.05	***	0.01
D6A_NM_P	0.01	***	0.00	0.01	***	0.00	0.01	*	0.00
D6B_AA	0.00	***	0.00	0.00	***	0.00	0.00	***	0.00
D6C_T	---		---	---		---	(0.05)	***	0.01
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$									
Goodness of Fit									
AIC	29281.29			25990.17			12582.69		
R ² - Corrected (Nagelkerke)	0.73			0.68			0.31		

Table 4.11 Optimal BE-TS Models for Tri-County Case (KD involvement)

IV \ DV	KD_ALL			KD_MOT			KD_VRU		
	Beta		SE	Beta		SE	Beta		SE
Intercept	1.28	***	0.25	1.25	***	0.27	(2.26)	***	0.45
D0A	(0.41)	***	0.06	(0.40)	***	0.07	(0.32)	**	0.14
D0B	0.20	***	0.01	0.20	***	0.01	0.20	***	0.03
D0B_B_P	---		---	---		---	---		---
D0B_H_P	0.01	***	0.00	0.01	***	0.00	(0.00)	***	0.00
D0B_U5_P	(0.02)	**	0.01	(0.02)	***	0.01	0.02	**	0.01
D0B_O65_P	(0.00)	***	0.00	(0.00)	***	0.00	0		0
D0B_M_P	0.01	***	0.00	0.01	**	0.00	0.04	***	0.01
D0C_HHS	0.05	**	0.02	0.04	*	0.02	---		---
D0C_OC_P	0.01	***	0.00	0.01	***	0.00	---		---
D0C_B90_P	0.00	***	0.00	0.00	**	0.00	0.00	**	0.00
D0C_CO2_HH	0.02	***	0.01	0.02	***	0.01	---		---
D0C_A0_P	0.00	***	0.00	0.00	***	0.00	---		---
D0C_A1_P	0.00	*	0.00	0.00	*	0.00	---		---
D0D_CE	---		---	---		---	(0.00)	**	0.00
D0E_HI	---		---	---		---	---		---
D0E_LWH	0.01	***	0.00	0.01	***	0.00	0.01	**	0.01
D0E_LWW	---		---	---		---	---		---
D0E_BP	---		---	---		---	0.00	*	0.00
D0E_R	0.03	*	0.02	0.03	*	0.02	0.05		0.04
D1A_LA	0.00	***	0.00	0.00	***	0.00	---		---
D1B	(0.01)	***	0.00	(0.01)	***	0.00	(0.01)	***	0.00
D1C_5_ENT	0.03	***	0.00	0.02	***	0.00	0.03	***	0.01
D1C_5_IND	---		---	---		---	---		---
D1C_5_RET	0.03	***	0.01	0.03	***	0.01	0.05	**	0.02
D2B_E5_MX	0.19	***	0.04	0.18	***	0.04	0.32	***	0.09
D2C_TRIP_EQ	---		---	---		---	0.16	*	0.09
D2C_WRK_MX	0.13	***	0.04	0.14	***	0.04	---		---
D2D	0.01	*	0.01	0.01	*	0.01	---		---
D2D_ONDX	(0.02)	***	0.01	(0.03)	***	0.01	---		---
D2D_NDX	(0.00)		0.00	(0.00)		0.00	---		---
D3A	(0.01)	***	0.00	(0.01)	***	0.00	---		---
D3A_AO	0.02	***	0.01	0.03	***	0.01	---		---
D3A_MM	0.01	***	0.00	0.02	***	0.00	0.02	*	0.01
D3B_G	(0.00)	***	0.00	(0.00)	***	0.00	---		---
D3B_PO3	(0.00)	*	0.00	---		---	---		---
D3B_PO4	0.00	***	0.00	0.00	***	0.00	---		---
D3D_R1_P	0.01	***	0.00	0.01	***	0.00	(0.01)	***	0.00
D3D_R2_P	0.00	***	0.00	0.00	***	0.00	0.00	**	0.00
D3E	0.03	***	0.00	0.03	***	0.00	0.04	***	0.01
D4B_025	(0.68)	***	0.12	(0.69)	***	0.13	(0.82)	***	0.28
D4C_W_P	0.02	***	0.00	0.02	***	0.00	0.02	**	0.01
D4D	0.01	***	0.00	0.01	***	0.00	0.01	***	0.00
D5A_R	0.00	***	0.00	0.00	***	0.00	---		---
D5B	0.04	***	0.01	0.04	***	0.01	0.04	***	0.01
D6A_NM_P	0.01	***	0.00	0.01	***	0.00	0.01	***	0.00
D6B_AA	0.00	***	0.00	0.00	***	0.00	0.00	***	0.00
D6C_T	---		---	---		---	(0.02)	*	0.01
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$									
Goodness of Fit									
AIC	32363.52			32197.77			12281.33		
R ² - Corrected (Nagelkerke)	0.76			0.75			0.27		

Among the 57 BE variables tried in tabu search based subset selection algorithm, total population (D0B), male percentage of the population (D0B_M_P), gross entertainment employment density (D1C_5_ENT), the percentage of secondary roads (D3D_R2_P), total sidewalk length (D3E), the number of bus stops (D4D), walking index (D5B), the percentage non-motorized modes for commuting (D6A_NM_P) and annual average daily traffic (D6B_AA) were found to be positively associated with all the six TS variables (crash types). Urban indicator (D0A) and gross population density (D1B), on the other hand, were negatively related to all TS variables. The Hispanic percentage of the population (D0B_H_P) and the percentage of primary roads (D3D_R1_P) had mixed associations. Both BE variables were positively associated with PDO and non-incapacitating injury crashes, all and motor vehicle involved crashes, also negatively associated with vulnerable road user crashes. However, while the Hispanic percentage was negatively related to fatal or incapacitating injury crashes, the percentage of primary roads had a positive relation.

The percentage of housing units built before 1990 (D0C_B90_P), the percentage of low wage workers living in the block group (D0E_LWH), retail employment density (D1C_5_RET), employment entropy (D2B_E5_MX), and the network density in terms of facility miles of multi-modal links (D3A_MM) were the BE variables having positive associations with all TS variables except fatal or incapacitating injury crashes. Similarly, annual CO2 emission per household (D0C_CO2_HH), land area (D1A_LA), and network density in terms of facility miles of auto-oriented links (D3A_AO) were positively related to all TS variables other than crashes involving vulnerable road users. Gross intersection density (D3B_G) was also associated with the same set of TS variables but negatively.

Another negatively related BE variable was employment proportion within a quarter-mile of a fixed-guideway transit stop (D4B_025). This relation was not significant, just with fatal or incapacitating injury crashes. The other common BE variable was the number of jobs within 45 minutes of auto travel time (D5A_R). It had mixed associations, positive with PDO and non-incapacitating injury crashes, ALL and

motor vehicle involved crashes, but negative with fatal or incapacitating injury crashes.

Household workers per job equilibrium index (D2C_WRK_MX), total employment (D2D), and intersection density in terms of pedestrian-oriented intersections having four or more legs (D3B_PO4) were positively related to PDO and non-incapacitating injury crashes, ALL and motorized vehicle involved crashes.

Transit ridership percentage of workers (D4C_W_P) was found to be positively associated with PDO crashes, ALL, motor vehicle, and vulnerable road user involved crashes. The percentage of occupied housing units (D0C_OC_P) was positively related to non-incapacitating and incapacitating injury, ALL, and motor vehicle involved crashes. Total road network density (D3A) had negative associations with PDO and non-incapacitating injury crashes, ALL, and motorized vehicle involved crashes.

Finally, two BE variables -the percentage of the population with at least a bachelor's degree (D0D_CE) and the percentage of population under the age of 5 (D0B_U5_P)- had mixed relations. As in the crash frequency models, the percentage of the population with at least a bachelor's degree was positively associated with PDO crashes and negatively associated with non-incapacitating and incapacitating injury crashes, and vulnerable road user involved crashes. The percentage of the population under the age of 5 was positively related to vulnerable road user involved crashes but negatively related to PDO, ALL, and motorized vehicle involved crashes.

4.5.3 Discussion of the tri-county crash modeling outcomes

In this section, the results of crash frequency and crash density models for the tri-county case are discussed. In Table 4.12, for each crash category (PDO as S1, non-incapacitating injury as S2, fatal or incapacitating injury as S3, total crashes as ALL, motor vehicle involved crashes as MOT, and vulnerable road user involved crashes as VRU), both frequency and density modeling results are depicted. While

presenting the results, the most common variables that were significant in at least four of the crash types were selected as in the single county case.

Table 4.12 Comparison of the IVs for Tri-County Case

IV	90% 99%	+ -	DV	S1		S2		S3		ALL		MOT		VRU	
				NB	GA	NB	GA	NB	GA	NB	GA	NB	GA	NB	GA
D0A															
D0B															
D0B_H_P															
D0C_OC_P															
D0C_B90_P															
D0C_CO2_HH															
D0D_CE															
D1A_LA															
D1B															
D1C_5_ENT															
D1C_5_RET															
D2B_E5_MX															
D2C_WRK_MX															
D3A															
D3A_AO															
D3D_R1_P															
D3D_R2_P															
D3E															
D4B_025															
D4C_W_P															
D4D															
D5A_R															
D5B															
D6A_NM_P															
D6B_AA															

According to the results, urban identification was negatively associated with PDO, fatal or incapacitating injury crashes, ALL crashes, the motor vehicle involved, and vulnerable road user crashes significantly. These negative associations were similar to those in the single county case, so contradicting the former studies. Additionally, the negative association with the vulnerable user involved crashes in the tri-county case also contradicts the former studies' positive associations (Cai et al., 2016; Cai, Abdel-Aty, Lee, et al., 2017; Siddiqui et al., 2012). These associations would be related to the smaller block group sizes in the urban areas and slower driving speeds in the urban areas.

The total population was positively associated with all crash types as in the single county case. In this case, all the associations were at the highest significance level. Total population is related to activity and exposure, so positive associations were expected.

The Hispanic percentage of the population found to be positively associated with PDO, non-incapacitating injury, all and motorized vehicle involved crashes, while negatively associated with fatal or incapacitating injury and vulnerable road user involved crashes. The percentage of the Hispanic population was usually positively associated with total crashes (Jiang et al., 2016) and pedestrian crashes (Lee et al., 2018; Lee, Abdel-Aty, & Jiang, 2015; Ukkusuri et al., 2011; C.-Y. Yu, 2014). The positive associations were explained by the generally lower income levels of Hispanic populations. As mentioned in the single county case, income level is usually negatively associated with crashes (Amoh-Gyimah et al., 2017; Noland & Quddus, 2004a; Pirdavani et al., 2013; Rhee et al., 2016; K. Xie et al., 2019; C. Xu et al., 2017). However, the average percentage of the population block groups' Hispanic population was almost 40%; thus, Hispanic populations were not the minority populations; however, they would be among the deprived segments with the black people. The uncommon negative associations with fatal crashes could be explained by again income and deprivation. The fatal crashes usually happen at the block groups with a higher percentage of primary roads, which could be related to car usage and ownership. Therefore, the higher income levels and fatal crashes were

negatively associated with the Hispanic percentage of the population. The negative association of vulnerable road users could also be related to the Hispanic populations' lower car ownership. Moreover, as a touristic center, there are many entertainment and commercial facilities at the shore and urban core where the vulnerable road user involved crashes more likely to happen. Those parts accommodate higher-income populations; thus, a negative association with Hispanic percentage was expected. One last factor is the spatial concentration of population segments within the region. Hispanic populations are mostly living in Miami-Dade County compared to the other two counties. Therefore, the vulnerable road user involved crashes occurring in the other two counties have negative associations with the Hispanic population percentage because the Hispanic population percentage is lower in these two counties. These would explain the very uncommon negative association.

The percentage of the occupied housing units could be a surrogate of human activity. Those block groups high in occupancy rate would be preferred sites by respective segments of the population. Therefore, the positive associations found are plausible.

The old housing stock (the percentage of housing units built before 1990) could be associated with deprived communities, having negative associations with income and education (the percentage of the population with at least a bachelor's degree). Therefore, there was a positive association of the old housing stock with PDO crashes. On the other hand, the percentage of the population with at least a bachelor's degree was negatively associated with non-incapacitating, fatal, and vulnerable road user involved crashes. This could be explained not only by income but also education level. The safety status could be worse for the deprived areas with lower income and educational level; moreover, people with higher education level are more probably obey the traffic rules and drive safer (Bao et al., 2017; H. Huang et al., 2010; Zhenning Li et al., 2019; Narayanamoorthy et al., 2013; Ukkusuri et al., 2011). However, the positive association between the education level and PDO crashes is not common. That would be related to increased car ownership and usage. Likewise, the annual CO₂ emission per household is a surrogate of car ownership and exposure. Therefore, as the emission amount per household increases, the crash occurrences

would also increase. Thus, PDO, all, and motor vehicle involved crashes were found to have positive associations with CO₂ emissions, significantly in both frequency and density models.

According to the results, the population density was negatively associated with all crash types significantly in both frequency and density models. The results were contradicting the former studies in which population density was found to be positively related to total crashes (Cai, Abdel-Aty, & Lee, 2017; Cai, Abdel-Aty, Lee, et al., 2017; Khondakar et al., 2010) and pedestrian crashes (Cai et al., 2016; Cai, Abdel-Aty, Lee, et al., 2017; Loukaitou-Sideris et al., 2007; Siddiqui et al., 2012) and bike crashes (Ouyang & Bejleri, 2014; Siddiqui et al., 2012). The population density was accepted as a surrogate to exposure, so the number of crashes was higher. However, the findings supported the negative associations in other studies that found PDO crashes, total crashes, and pedestrian and bicycle crashes negatively associated with residential density (M. Abdel-Aty et al., 2013; Osama & Sayed, 2017). The negative association in the vulnerable road user crashes could be explained by “safety in numbers,” where the drivers could better perceive people and their movements when they are higher in number, so they drive more carefully. The findings of the land area also supported the results of the other types of crashes. The land area was positively associated with PDO, non-incapacitating, fatal, all, and motor vehicle involved crashes. This is similar to the results of the single county case. As land area increases, the total number of crashes increases but the population density decreases. One other explanation could be that the compactness (density being one of the primary factors of compactness) is negatively related to vehicle miles traveled since origins and destinations are closer in a compact block group (Ewing, Hamidi, & Grace, 2016). As vehicle miles traveled (VMT) is most likely less in a compact block group, the crash rate would be lower.

As opposed to population density, entertainment employment density was positively associated with all crash types. Retail employment density was positively associated with PDO, non-incapacitating injury, all, motor vehicle involved, and vulnerable road user involved crashes (with lower significance levels). These findings supported

the findings of former studies (Bao et al., 2017; Hadayeghi et al., 2003; Khondakar et al., 2010; Noland & Quddus, 2004a). Increased employment density means high activities and exposure, so the findings are not surprising. The employment entropy was also corroborated, which was found to have positive associations with all crash types except fatal or incapacitating injury crashes. As in the single county case, heterogeneity of employment increases the likelihood of crashes increases due to the exposure.

The density of auto-oriented links, the percentage of the primary roads, and the percentage of secondary roads had significant implications about the design dimension. The density of roads is a surrogate of exposure. Increased road density increases the likelihood of crashes. The density of auto-oriented links (mainly freeways and arterials with a minimum speed of 41 mph) had a positive association with non-incapacitating injury, fatal, all and motor vehicle involved crashes, as found in the former studies (Bao et al., 2017, 2018; C.-Y. Yu & Xu, 2018). Also, the percentage of the primary roads was positively associated with PDO, non-incapacitating injury, fatal, all, and motor vehicle involved crashes as in the single county case, while having a negative association with vulnerable road user crashes, since pedestrians and bicyclists are not so active on those roads. But, for the secondary roads, vulnerable road users were also positively associated with crash occurrences.

Albeit being safely designed, sidewalks would have negative associations with pedestrian crashes (Y. Wang & Kockelman, 2013), in most of the former studies, sidewalk length was found to be positively related to total crashes (Cai, Abdel-Aty, & Lee, 2017; Cai, Abdel-Aty, Lee, et al., 2017; C.-Y. Yu & Zhu, 2016), and pedestrian crashes (P. Chen & Zhou, 2016; Nashad et al., 2016; C.-Y. Yu, 2014; C.-Y. Yu & Zhu, 2016). Sidewalks could be the surrogates of pedestrian activity and so exposure, so the conflict points and crash risk increase as sidewalks' length increases. It was not surprising to come up with significant positive associations with all crash types.

According to the tri-county modeling results, the number of bus stops was positively associated with all types of crashes in the frequency and density models and all crash types. As mentioned in the single county case, the findings of the former studies support the positive associations (Cai, Abdel-Aty, & Lee, 2017; P. Chen & Zhou, 2016; Kaygisiz et al., 2017; Ouyang & Bejleri, 2014; Ukkusuri et al., 2012; Y. Wang & Kockelman, 2013). The number of bus stops is related to increases in transit mode, walking mode, and increased conflict points for both vulnerable road users and motorized vehicles.

The jobs within 45 minutes of auto travel time positively associated with property damage only crashes, non-incapacitating injury crashes, total, and motor vehicle involved crashes, whereas negatively associated with fatal and incapacitating and vulnerable road user involved crashes. Fatal crashes usually happen on primary roads at the periphery, and employment opportunities are usually clustered at the city centers where it is less likely to have fatal crashes. However, the negative association between the vulnerable road user involved crashes and job locations requires another association. One reason for this would be the metro rail serving the city center. Jobs are concentrated at the urban core of Miami-Dade County, and the core is served by the metro system, which is safer for vulnerable road users. The metro service also explains the negative association of employment proportion within a quarter-mile of a fixed-guideway stop and all types of crashes. This variable was valid for the block groups served by the metro service. Those block groups had lower crash rates compared to the non-served block groups.

The walkability index is a surrogate for exposure since it could be a measure of the attractiveness of the walking mode. As exposure increases, the likelihood of crashes increases. Moreover, it was found to be positively related to all crash types. Total crashes were positively associated with walking (Cai, Abdel-Aty, Lee, et al., 2017; Dong et al., 2016; Lee, Abdel-Aty, de Blasiis, et al., 2019). Also, pedestrian crashes' frequency has a positive association with walking (P. Chen & Zhou, 2016; Lee et al., 2017; Lee, Abdel-Aty, Huang, et al., 2019). The transit ridership percentage of workers and the percentage of non-motorized modes had positive relationships with

all types of crashes. In parallel to these findings, transit usage was positively associated with pedestrian crashes in former studies (Jermprapai & Srinivasan, 2014; Lee et al., 2017, 2018; Narayanamoorthy et al., 2013; Nashad et al., 2016; Ukkusuri et al., 2012). A similar relation was found between transit usage and total crashes (M. Abdel-Aty et al., 2013; Cai, Abdel-Aty, & Lee, 2017; Cai, Abdel-Aty, Lee, et al., 2017; Lee et al., 2017; Lee, Abdel-Aty, de Blasiis, et al., 2019). One explanation for that is related to household income: for economically deprived people, public transportation, walking, and cycling are the affordable alternatives.

Lastly, the annual average daily traffic was positively associated with all types of crashes. This result was consistent for all settings and cases and supported the former study findings both for total crashes (Cai, Abdel-Aty, Lee, et al., 2019; Guerra et al., 2019; H. Huang et al., 2016; Loukaitou-Sideris et al., 2007; Rhee et al., 2016; J. Wang & Huang, 2016; C. Xu et al., 2017; P. Xu & Huang, 2015; C.-Y. Yu & Xu, 2018; Zeng & Huang, 2014) and vulnerable road user involved crashes (Cai, Abdel-Aty, & Lee, 2017; Loukaitou-Sideris et al., 2007; Miranda-Moreno et al., 2011; Wier et al., 2009). As the traffic volume increases, the likelihood of all types of crashes increases. Traffic volume is the most definitive factor in traffic crashes. Even in the safely designed built environments, the crash occurrences would increase with increased traffic volume, so increased exposure.

4.5.4 Consistency of the tri-county crash modeling results

The consistency scheme presented for the single county case was also applied to the tri-county case. The results are depicted in Table 4.13. The variables that were consistent in most of the tabu-search based BE-TS modeling processes were as follows:

- Total population (D0B)
- Hispanic % of the population (D0B_H_P)
- % of occupied housing units (D0C_OC_P)

- % of housing units built before 1990 (D0C_B90_P)
- % of population 25 years and more with at least bachelor's degree (D0D_CE)
- Land area in acres (D1A_LA)
- Gross population density (people/acre) on unprotected land (D1B)
- Gross entertainment employment density (D1C_5_ENT)
- Gross retail employment density (D1C_5_RET)
- Employment entropy (D2B_E5_MX)
- Household workers per job equilibrium index (D2C_WRK_MX)
- Network density in terms of facility miles of auto-oriented links per square mile (D3A_AO)
- % of primary roads (D3D_R1_P)
- % of secondary roads (D3D_R2_P)
- Total sidewalk length (D3E)
- Proportion of CBG employment within ¼ mile of fixed-guideway transit stop (D4B_025)
- Transit ridership percentage of workers (D4C_W_P)
- Number of bus stops (D4D)
- Jobs within 45 minutes auto travel time (D5A_R)
- Walking index (D5B)
- % of non-motorized modes for commuting (D6A_NM_P)
- Annual average daily traffic (average of 3 years) for all modes (D6B_AA)

Table 4.13 Consistency of the IVs for Tri-County Case

IV	DV	S1		S2		S3		ALL		MOT		VRU	
		NB	GA	NB	GA	NB	GA	NB	GA	NB	GA	NB	GA
D0A													
D0B													
D0B_B_P													
D0B_H_P													
D0B_U5_P													
D0B_O65_P													
D0B_M_P													
D0C_OC_P													
D0C_B90_P													
D0C_CO2_HH													
D0C_A0_P													
D0D_CE													
D0E_HI													
D0E_LWH													
D0E_LWW													
D0E_R													
D1A_LA													
D1B													
D1C_5_ENT													
D1C_5_IND													
D1C_5_OFF													
D1C_5_RET													
D2B_E5_MX													
D2C_TRIP_EQ													
D2C_WRK_MX													
D2D													
D2D_ONDX													
D3A													
D3A_AO													
D3A_MM													
D3B_G													
D3B_MM3													
D3B_PO3													
D3B_PO4													
D3D_R1_P													
D3D_R2_P													
D3E													
D4B_025													
D4C_W_P													
D4D													
D5A_R													
D5B													
D6A_M_P													
D6A_NM_P													
D6B_AA													

When those variables were evaluated with the significant variables presented in the previous section, the urban indicator (D0A), annual CO2 emission per household (D0C_CO2_HH), and total road network density (D3A) were significant variables but were not consistent. However, the variables presented as consistent were all significant. Therefore, the BE variables listed above were both significant and consistent.

4.6 The Key Built Environment Variables and Urban Development

Finally, significant and consistent BE variables for both single and tri-county cases were compared to detect the shared variables. The following variables were determined as the key BE variables:

- Total population (D0B)
- % of housing units built before 1990 (D0C_B90_P)
- Land area in acres (D1A_LA)
- Gross employment density²
 - Gross industrial employment density (D1C_5_IND)
 - Gross entertainment employment density (D1C_5_ENT)
 - Gross retail employment density (D1C_5_RET)
- Employment entropy (D2B_E5_MX)
- % of primary roads (D3D_R1_P)
- Number of bus stops (D4D)

² Industrial employment was significant for the single county case, whereas, entertainment and retail employment densities were significant for tri-county case. The type of employment would be distinguished according to the case study area.

- Mode choice³
 - % of motorized modes for commuting (excluding transit) (D6A_M_P)
 - % of non-motorized modes for commuting (D6A_NM_P)
- Annual average daily traffic (average of 3 years) for all modes (D6B_AA)

The percentage of housing units built before 1990, the number of bus stops, and mode choice would be related to TS by the populations' economic status. Deprived communities are captive transit and non-motorized mode users in general. On the other hand, wealth and higher income are usually associated with lower crash rates. Total population, employment density, percent of primary roads, and traffic volume are surrogates of exposure. The increases in those variables increase the likelihood of crashes of any type. Total population, employment density, employment entropy, number of bus stops, and mode choice indicate movement and activity, especially for vulnerable road users. Likewise, those variables also indicate higher motorized vehicle trip volumes and entropy. Thus, these variables were identified as the key built environment variables to understand traffic safety variability from a macro-level perspective. From a dimension-wise perspective, all the D-categories except distance to transit were represented by the selected key variables. For the demographic and socioeconomic dimension total population and the percentage of housing units built before 1990; for the density dimension land area and the sectoral employment density; for the diversity dimension employment entropy; for the design dimension, the percentage of the primary roads, for transit dimension, the number of bus stops, and for the travel dimension mode choice and the traffic volume were the representatives. In order to measure the performance of these key BE variables, an

³ Mode choice would be the general definition. In one case motorized mode composition was significant and in the other non-motorized composition was significant.

evaluation scheme is proposed in the next chapter. Here follows an evaluation of the findings within the compactness vs. sprawl debate.

As introduced in the background section, urban development could be discussed under sustainability and sustainable development contexts. Urban sprawl was seen as an unsustainable urban development form in many studies (Ewing & Cervero, 2010; Ewing & Hamidi, 2014) yet still being discussed since some scholars found compact development contributions relatively small (Stevens, 2017). Nevertheless, the compact development was associated with lower combined transportation and housing expenses, higher alternative transportation modes and higher walking rates, and so healthier lives in former studies (Ewing & Hamidi, 2014). It was mentioned that the sprawl vs. compactness debate is a regional level discussion in essence, but the implications and impacts could be quantified and discussed at lower levels. This study's main aim is to discuss the association between the built environment and traffic crashes. The urban development paradigm frames the built environment, so how the city grows has close connections with the built environment. The traffic crashes are the resulting negative externalities of both the development and travel behavior. Therefore, the implications of the key BE variables identified in this chapter should be evaluated within the sprawl vs. compactness debate frame.

In order to discuss the findings within the sprawl-compactness frame, the general situation of the case studies is presented in Table 4.14 (Ewing & Hamidi, 2014).

Table 4.14 Sprawl Score of Some Counties in the US

	Density score	Land use mix score	Activity centering score	Street connectivity score	Composite (total) score
New York	654.01	144.57	400.25	230.33	425.15
Leon	102.05	106.83	149.96	99.11	118.31
Miami-Dade	137.38	132.85	131.33	156.48	149.93
Broward	120.61	133.24	95.43	148.86	131.01
Palm Beach	107.77	125.08	107.06	118.32	118.40

These scores were calculated by considering four factors as “density score,” “land use mix score,” “activity centering score,” and “street connectivity score” (Ewing & Hamidi, 2014). Under each factor, the related built environment variables were considered. For example, to calculate “land use mix score,” job population balance, the mix of jobs within 1-mile of block groups, and walk score of each census tract’s center were used. After calculating each main factor, a composite score for each metropolitan area and each county was computed. The composite sprawl index scores given in the table were evaluated according to the sprawl-compactness continuum by having the smallest value as the highest sprawl level. The threshold was determined to be 100. Higher values were on the compactness side, and lower values were on the sprawling side (Ewing & Hamidi, 2014). According to the values presented, New York County has the highest compactness score of 425.15 in the US. Miami-Dade County was among the counties having high compactness scores with a score of 149.93, and the other three counties considered in this study were above the average (being fairly on the compactness side).

Along with the sprawl index scores, transit index scores are presented in Table 4.15 (Center for Neighborhood Technology, n.d.-a). These index scores are between 0-10 and consider four main dimensions: “transit trips per week within a half-mile,” “transit routes within a half-mile,” “jobs accessible in 30-minute trip”, and “percentage of commuters who use transit.” As seen from the table, the most compact US county, New York, also had the highest transit score. The four counties in this study had transit scores in parallel to their compactness scores. Miami-Dade had the highest transit score of 6.5, and Leon and Palm Beach Counties had a transit score of 3.3. Compared to the New York case, only Miami-Dade had a moderate transit score (with the contribution of the metro system); other counties seemed to be very poor on transit service and usage rates. In Leon County (single county case), only 1.7% of commuters use public transportation. The transit commuters’ rate of 5.51% of Miami-Dade is also insignificant compared to 63.74% in New York.

Table 4.15 Transit Score of Some Counties in the US

	Transit trips	Transit routes	Jobs accessible	% of commuters using transit	Transit score
New York	22,918	41	2,599,093	63.74	9.9
Leon	773	4	54,342	1.70	3.3
Miami-Dade	1,841	6	169,478	5.51	6.5
Broward	979	NA	116,570	2.94	5.0
Palm Beach	457	2	42,790	2.01	3.3

So, the compactness and transit service seem to corroborate each other in New York, which would be the desirable case. In comparison, the above-average compactness scores of Leon, Palm Beach, and Broward Counties seem to fall apart by the transit usage. Though Miami-Dade was one of the compact counties in the US, the transit score is moderate, and the % of commuters using transit is far less than the rate in the best case; thus, there is room for improvement in transit service.

These figures are important while evaluating the built environment variables. Among the key built environment variables, the total population, the percentage of the old housing stock, and the land area would not be directly structured on the sprawl-compactness continuum. Nevertheless, population density (total population divided by the land area) should be higher in compact block groups. The population density was not among the key built environment factors, but it was a key built environment factor for the tri-county case. The population density was negatively associated with all crash types, a supporting variable for the compactness.

The employment density figures were significant for both the single county and tri-county cases. Employment density should also be higher in compact settlements, but the key aspect is that there should be a balance among land uses. The higher densities of employment would attract trips from other census block groups and increase the likelihood of crashes. As in the tri-county case, retail employment and entertainment employment densities were positively associated with all crash types. The urban

center and the shore side had higher employment and retail employment and higher crash rates as expected. In the compactness continuum, the balance of employment and residential units are favored in general. For trip attraction centers to decrease the likelihood of crashes, integrated transportation plans are required. From the sustainability (and compactness) perspective, reduced car dependency increased non-motorized, and transit mode shares are favorable and equitable. For example, the number of bus stops is a key built environment factor associated with all crash types positively. This association would be reversed by improving the transit service and safety. The rate of transit commuters was very low in both cases, and in some of the former studies, transit usage or the number of bus stops were associated with increased crash rates (Cai, Abdel-Aty, & Lee, 2017; Cai, Abdel-Aty, Lee, et al., 2017; Nashad et al., 2016; Ouyang & Bejleri, 2014; Ukkusuri et al., 2012). Higher quality, safer, equitable, and integrated (with non-motorized modes) transit service should be provided to reverse the positive association between traffic crashes and the number of bus stops and non-motorized mode percentage.

Employment entropy was positively related to all crash types except the fatal crashes. Entropy represents the mixture of employment types. The highest value of one (1) represents the evenest distribution of employment, and the lowest value of zero (0) represents the homogeneity as the domination of one type of employment. In terms of compactness, it is desired to have a balanced employment structure, but it was found to be positively associated with traffic crashes. This negative association would be due to accessibility issues. For example, the job access score (higher values represent higher access to a variety of jobs) had negative associations with traffic crashes. Therefore, if evenly distributed employment opportunities are also accessible, they would lower crash rates from the compactness point of view.

The percentage of the primary roads was positively associated with all crash types except the vulnerable road user involved crashes. Since those roads are generally highways and arterials where pedestrian and cyclist activities are minimal, the results were not surprising. However, those roads are surrogates of volume. In the compactness continuum, compact settlements would be walkable and well served by

local roads; thus, having lower percentages of primary roads. Since primary roads were positively associated with crashes, compactness would be negatively related to crashes regarding the primary road percentage.

Finally, the traffic volume would be lower in compact settlements. Having mixed land uses nearby that requires less travel and less automobile are the properties of compact settlements in the definition. As traffic volume had a positive association with all crashes in all cases and compact settlements require less traffic volume, compactness and exposure would be negatively associated.



CHAPTER 5

RESULTS FROM ACCIDENTS FREQUENCY AND DENSITY MODELS, DEVELOPING COUNTRY CONTEXT

This chapter presents the results of the tabu-search based subset selection methodology applied to the case study from the developing country context.

5.1 Introduction

In the previous chapter, the built environment – traffic safety (BE-TS) relationship was investigated in two case studies in the US. Some generalizable relationships were evaluated, and the key factors were determined depending on these two developed country cases.

As mentioned in the review section, most former studies also considered developed country cases—the lack of developing country cases mostly due to data concerns. A similar situation is valid for Turkey as a developing country. There are very few open-access transportation data, especially at the neighborhood level. For this study, one of the main aims was providing an evaluation scheme for the value of information (data) available in some cases but not in others (i.e., developed country cases vs. developing country cases).

Depending on this context, in this chapter, firstly, a case study from Turkey will be introduced. Then, the BE-TS relationship will be evaluated by using the tabu-search based methodology. Lastly, the resulting models will be compared with the two developed country cases introduced in the previous chapter.

5.2 Case Study Area, Konya

Konya is the largest with its land area (38,257 square kilometers) and is the 7th most populated (2,232,374 as of 2019) city in Turkey (see Table 5.1), located in the Central Anatolia Region (Turkish Statistics Institute, n.d.). Figure 5.1 shows the provincial border and the traffic analysis zones of Konya (Boğaziçi Proje, 2020).

Table 5.1 Information on Konya Case

Total Population in 2019	2,232,374
Total area (km²)	38,257
Population in TAZ 2020	1,934,503
Total TAZ area considered (km²)	1,081
Number of TAZs considered	237

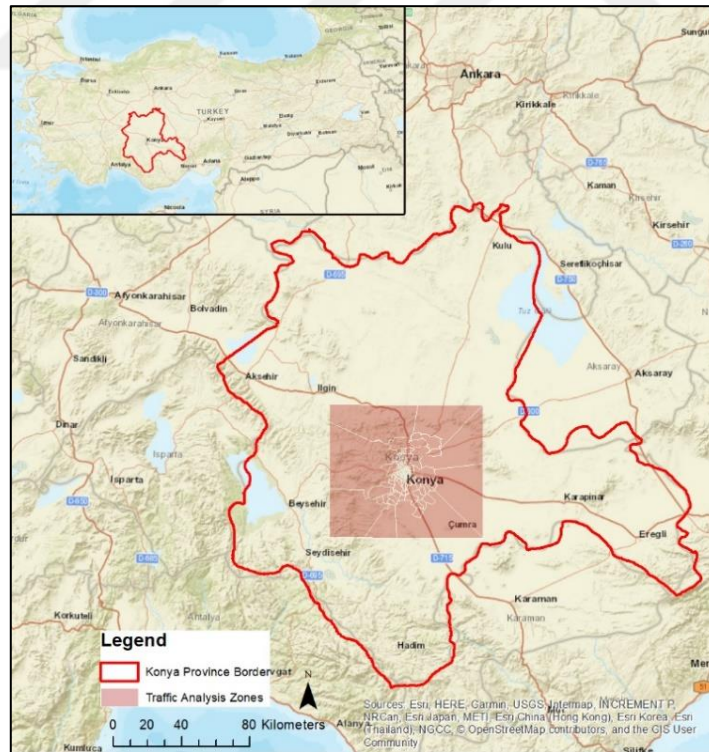


Figure 5.1. Traffic Analysis Zones and the Provincial Border of Konya

5.3 Data Extraction and Classification for Konya

In this section, the extraction and preparation of the built environment and traffic safety dataset are presented.

5.3.1 Traffic safety dataset

The crash dataset mainly consisted of the records for Konya provincial region between 2017 and 2019. It was gathered from the Traffic Department of the General Directorate of Security (Traffic Department, n.d.). The data consisted of fatal and injury crashes. There were neither property damage only (PDO) crash records nor a distinction between injury severity. Therefore, the classification presented in the previous chapters was revised as:

- S2: injury crashes
- S3: fatal crashes
- ALL: total crashes (motor vehicle involved and vulnerable road user involved)
- MOT: motor vehicle involved crashes
- VRU: vulnerable road user involved crashes

In Figure 5.2a and Figure 5.2b, crashes according to the crash typologies presented for the Konya case.

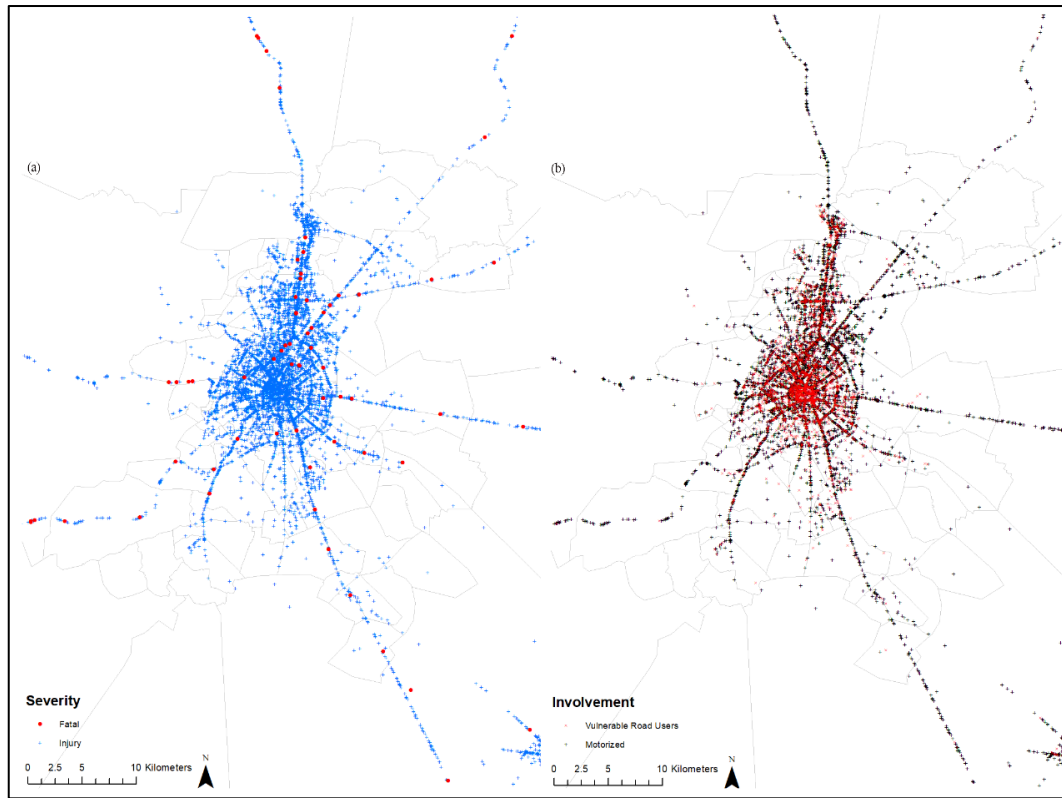


Figure 5.2. Crashes According to Involvement and Severity of the Konya Case

According to the crash distributions presented in Figure 5.2a, it is observed that the number of fatal crashes is smaller than the number of injury crashes, and fatal crashes are around the peripheral parts since the increased speeds are associated with fatal crashes. According to the second typology presented in Figure 5.2b, it is observed that as in the developed country cases, vulnerable road user involved crashes were densified around the city centers relative to the motor vehicle crashes.

The crash data were classified according to severity and involvement as mentioned. In Table 5.2, crash typology and respective crash frequencies are given for the Konya case. Each crash was assigned to the TAZs that were within 125m buffer distance. As in the previous crash assignment procedure, there were more crashes assigned to the traffic analysis zones than the original number of crashes due to the multiple

assignments. The average assignment rate of each crash was 1.77. This figure was far less than the average assignment rates of 2.22 and 2.56 in the single and tri-county cases. This difference would be due to the spatial distribution of the crashes and the zonal definitions. Traffic analysis zones are usually designed to capture the zones' transport specifics rather than just population estimation of the block groups or neighborhoods. Thus, these zones would better distinguish areas in terms of traffic-related issues. The other reason would be related to having larger divisions or having in zone crashes. As the zonal area gets larger in area, more crashes would be covered by each zone, also; if crashes happen at local roads same zones would cover more crashes.

Table 5.2 Crash Frequencies in Konya Case

CRASH TYPE	Crash count		
	original	assigned	assigned / original
S2-Injury crashes	10951	19824	1.81
S3-Fatal crashes	84	127	1.51
ALL- Motorized and non-motorized	11035	19917	1.80
MOT-Motor-vehicle crashes only	7454	12524	1.68
VRU-Vulnerable road user crashes	3581	7393	2.06
	Average: 1.77		

Nevertheless, the pattern is the same as vulnerable road user involved crashes had the highest and fatal crashes had the lowest average. Vulnerable road user involved crashes usually happen around the city center where the zones are smaller in area, and pedestrian activity is more. Whereas fatal crashes most likely occur at primary roads where the zones are larger in size, so the buffer distance around the fatal crashes usually covered by those large zones.

The distribution of the crash types in terms of frequency and kernel density are presented in Figure 5.3 and Figure 5.4, respectively.

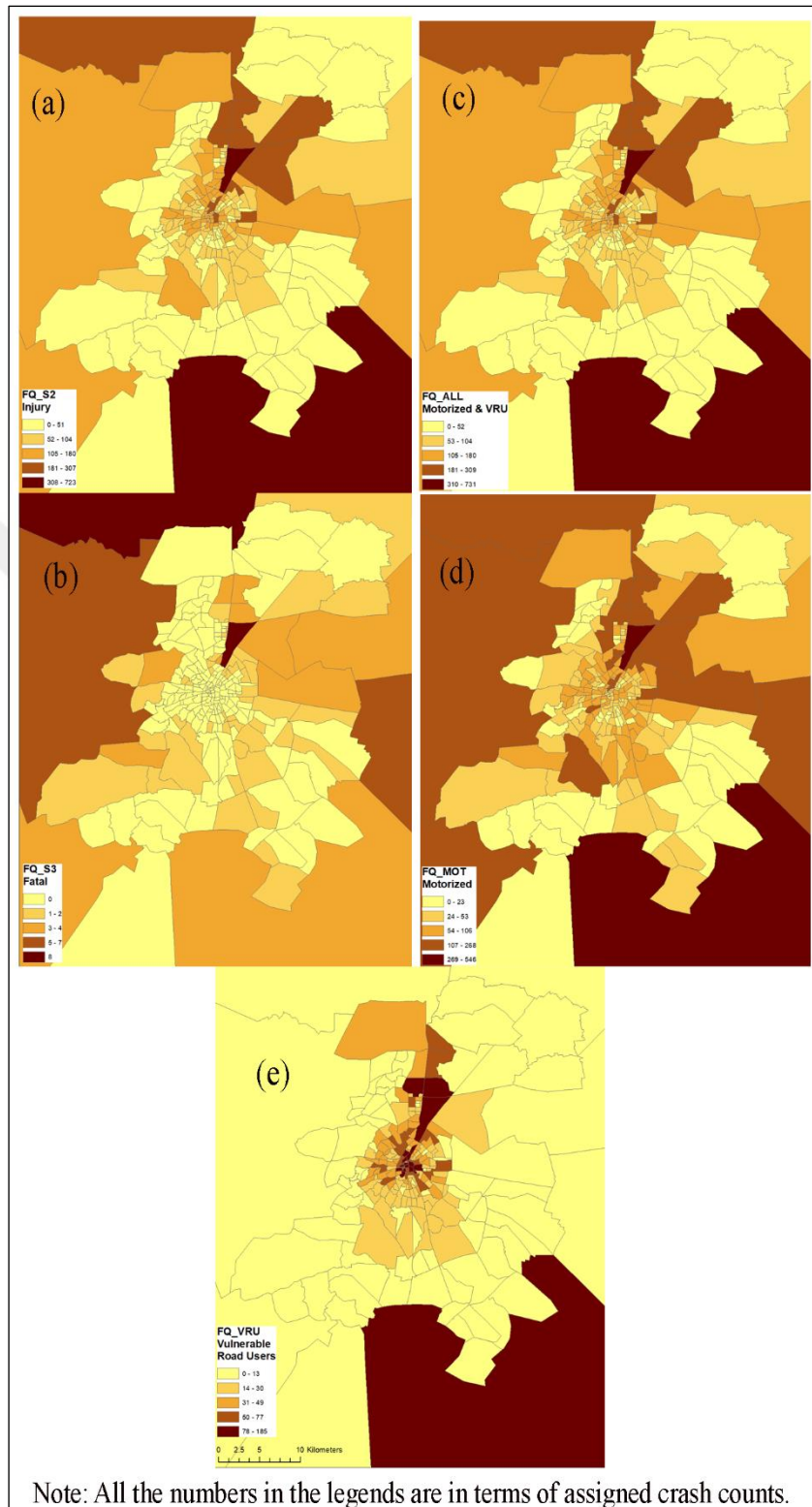


Figure 5.3. Crash Frequency Distributions in Konya (a) Injury Crashes (b) Fatal Crashes (c) Total Crashes (d) Motor Vehicle Involved Crashes (e) VRU Crashes

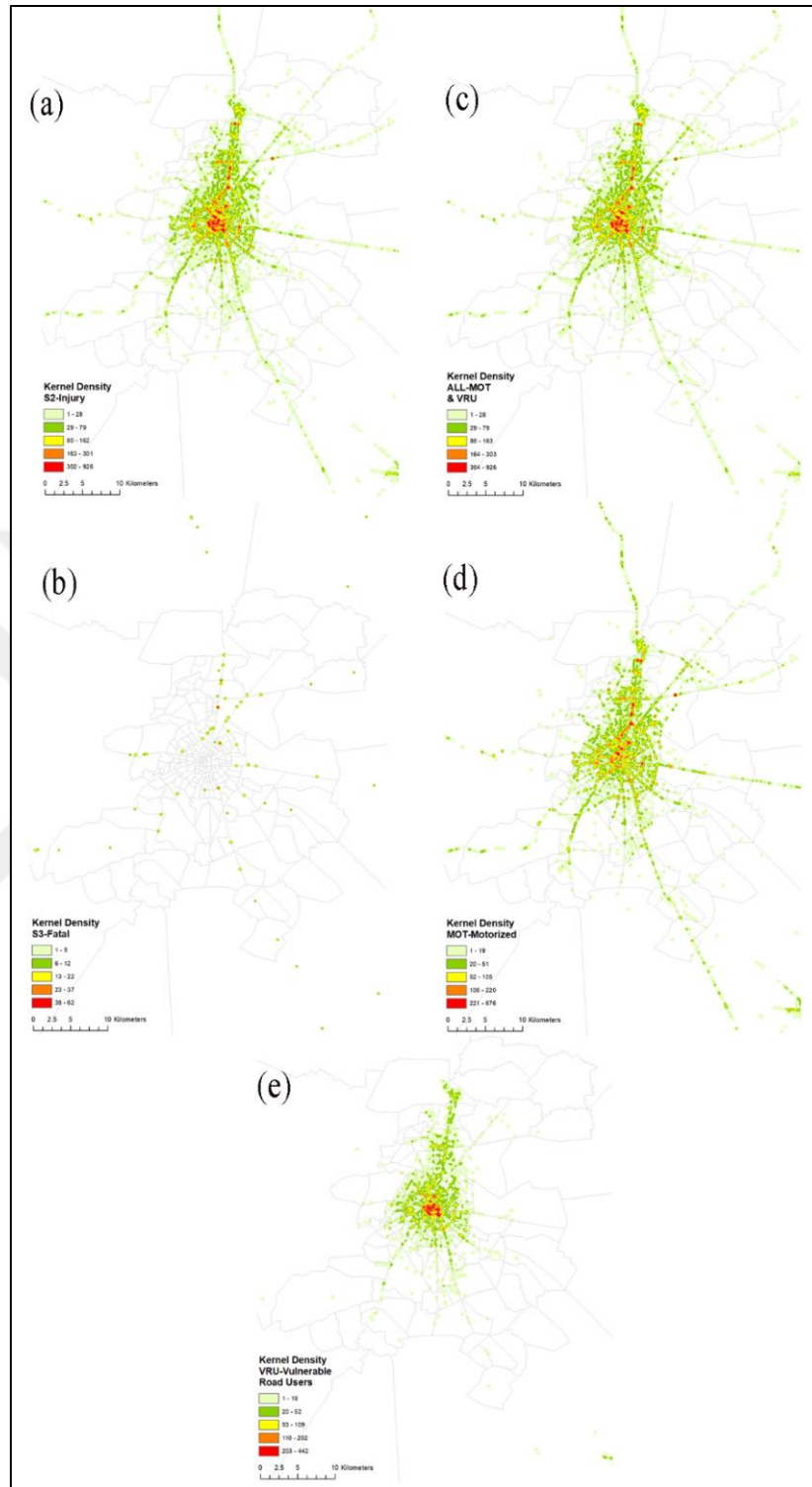


Figure 5.4. Kernel Density of Crash Typology in Konya (a) Injury Crashes (b) Fatal Crashes (c) Total Crashes (d) Motor Vehicle Involved Crashes (e) VRU Crashes

5.3.2 Built environment dataset

Most of the built environment data was gathered from Boğaziçi Proje (2020). This dataset mostly consisted of raw data such as the number of persons and the number of workers designated to each traffic analysis zone. Compared to the built environment data presented in the developed country cases, the dataset was limited, and the built environment variables did not match exactly. Thus, some variables were used as surrogates of the original variables. For example, annual average daily traffic (AADT) was not available in the Konya case, so the total trip count was used as a surrogate of traffic volume. Since road widths were not known for the road network density variable, total road length was used as a substitute value to road surface area (road data was gathered from Open Street Map by Bbbike extract service) (BBBike, n.d.).

Moreover, since each TAZ's household sizes were not known, the household-based variables were calculated as person-based variables. For example, in the developed country cases, D0C_A_HH was the number of automobiles per household, whereas in the Konya case, the same variable code was used to represent the number of households per person. The only exception to this conversion was the mean household income. This data was available in the Konya data; however, it was the median income in the original case. Thus, mean income was used rather than the median income. Table 5.3 presents the complete set of BE variables, their explanations, and calculation methods. It should be mentioned that the variable codes were not changed for the surrogate variables to compare the results with developed county case results.

Table 5.3 Built Environment Variables

BE Dimension	Variable Code	Variable Explanation
Demographic & socioeconomic	D0A*	Urban indicator (0 or 1) (1 Urban: pop. density > 150 persons per km ² – 0 Rural: pop. density < 150 persons per km ²)
	D0B	Total population (x1000 persons)
	D0C_A_HH***	Number of automobiles per person (total number of automobiles/total population)
	D0E	Total number of workers (x1000 persons)
	D0E_HI**	Mean household income (x1000 TL)
Density	D1A_LA*	Land area (km ²)
	D1B*	Gross population density (persons per km ²) (total pop/TAZ area)
	D1C*	Gross employment density (employment per km ²) (employment/TAZ area)
	D1D***	Gross activity density (sum per km ²) [(employment + population)/TAZ area]
Diversity	D2A_JP_HH*	Jobs per person (total employment/total population)
	D2A_WRK_EMP*	Workers per job (workers/employment)
	D2C_TRIP_EQ***	Trip productions and trip attractions equilibrium index $\frac{1}{e} \left \left(\frac{\text{home-based trips produced}}{\text{home-based trips attracted}} \right) - 1 \right $
	D2C_WRK_MX***	Household workers per job equilibrium index $\frac{1}{e} \left \left(\frac{\text{workers}}{\text{total employment}} \right) - 1 \right $
	D2C_WRK_MX_LN***	Natural logarithm of D2C_TRIP_MX
	D2D	Total employment (x1000)
Design	D3A*	Road network density (km per km ²) (total road length/TAZ area)
Travel	D6B_AA***	Total trips (x1000 trips) (produced # of trips + attracted # of trips)
	D6B_AA_LN***	Natural logarithm of D6B_AA_LN
* calculated ** used as a surrogate of the original variable *** calculated and used as a surrogate variable		

When the dimension reduction technique (reducing the variables till the absolute value of each the correlation coefficient is less than 0.75) was applied to the built environment variables presented in Table 5.3, the total number of workers (D0E), mean household income (D0E_HI), gross activity density (D1D), workers per job (D2A_WRK_EMP), and the natural logarithm of total trips (D6B_AA_LN) were reduced. Table 5.4 presents the descriptive statistics for the reduced set of BE variables.

Table 5.4 Descriptive Statistics of the Single County Case Dataset

Category	Variable Name	Min	Max	Mean	Variance
Traffic safety (Dependent Variables)	FQ_S2	0	723	83.65	5150.60
	FQ_S3	0	8	0.39	0.86
	FQ_ALL	0	731	84.04	5221.33
	FQ_MOT	0	546	52.84	2632.87
	FQ_VRU	0	185	31.19	844.31
	KERNEL_S2	0	494.09	44.98	2301.59
	KERNEL_S3	0	9.38	0.41	1.01
	KERNEL_ALL	0	498.55	45.22	2331.34
	KERNEL_MOT	0	367.13	29.01	1261.24
	KERNEL_VRU	0	127.88	15.86	258.90
Demographic & socioeconomic	D0A	0	1	0.74	1.92
	D0B	0	67.11	8.16	77.78
	D0C_A_HH	0	0.37	0.22	0.01
Density	D1A_LA	0.04	73.11	4.56	99.47
	D1B	0	42.85	10.83	103.47
	D1C	0	42.21	3.47	48.51
Diversity	D2A_JP_HH	0	103.78	1.52	90.62
	D2C_TRIP_EQ	0	0.98	0.15	0.06
	D2C_WRK_MX	1.01	42.64x10 ⁷	18.49x10 ⁵	76.72x10 ¹³
	D2C_WRK_MX_LN	0	8.63	1.39	2.07
	D2D	0	58.59	2.59	44.04
Design	D3A	0.16	56.87	21.56	143.91
Travel	D6B_AA	3.06	157.36	24.85	534.57

5.4 BE-TS Modeling Results

For the Konya case, the tabu-search based subset selection algorithm was applied. Since there are fewer variables (13 BE variables), 50 iterations for each section (preliminary search, intensification, and diversification) would almost turn out to be a complete enumeration algorithm so that the results would be the optimum ones. The related tabu-search graphs for each TS variable and model (frequency and kernel density) are presented in Appendix C. This section presents the modeling results of both frequency and kernel density modeling results for the Konya case.

First, the Konya case results when crash frequencies were modeled by negative binomial regression are presented. The results consisted of the best subsets of variables obtained in the tabu-search based selection algorithm while minimizing the Akaike Information Criterion (AIC). In Table 5.5 and Table 5.6, the frequency modeling results for injury severity and involvement are presented, respectively.

Table 5.5 Optimal BE-TS Models for the Konya Case (FQ severity)

IV \ DV	FQ_S2			FQ_S3		
	Beta		SE.	Beta		SE.
Intercept	2.86	***	0.14	(0.88)	***	0.20
D0A	---		---	(1.46)	***	0.27
D0B	0.02	**	0.01	0.05	***	0.01
D0C_A_HH	0.78	*	0.44	---		---
D1A_LA	---		---	---		---
D1B	---		---	---		---
D1C	---		---	---		---
D2A_JP_HH	(0.02)	***	0.00	(0.03)	**	0.01
D2C_TRIP_EQ	0.28		0.18	---		---
D2C_WRK_MX	(0.00)		0.00	---		---
D2C_WRK_MX_LN	(0.03)	**	0.01	---		---
D2D	0.03	**	0.01	0.06	***	0.01
D3A	0.03	***	0.00	---		---
D6B_AA	0.01	***	0.00	---		---
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$						
Goodness of Fit						
AIC	2393.16			328.38		
R²- Corrected	0.56			0.30		

Table 5.6 Optimal BE-TS Models for the Konya Case (FQ involvement)

IV \ DV	FQ_ALL			FQ_MOT			FQ_VRU		
	Beta		SE.	Beta		SE.	Beta		SE.
Intercept	2.88	***	0.14	2.72	***	0.13	1.20	***	0.16
D0A	---		---	---		---	0.42	***	0.14
D0B	0.02	**	0.01	0.02	**	0.01	---		---
D0C_A_HH	0.76	*	0.44	1.08	**	0.45	---		---
D1A_LA	---		---	---		---	(0.02)	***	0.01
D1B	---		---	(0.01)	*	0.01	---		---
D1C	---		---	(0.01)	**	0.01	---		---
D2A_JP_HH	(0.02)	***	0.00	(0.02)	***	0.01	(0.02)	***	0.01
D2C_TRIP_EQ	0.29		0.18	---		---	0.76	***	0.19
D2C_WRK_MX	(0.00)		0.00	(0.00)		0.00	(0.00)		0.00
D2C_WRK_MX_LN	(0.03)	**	0.01	(0.04)	***	0.01	---		---
D2D	0.03	**	0.01	0.04	***	0.01	---		---
D3A	0.03	***	0.00	0.03	***	0.00	0.05	***	0.00
D6B_AA	0.01	***	0.00	0.01	***	0.00	0.02	***	0.00
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$									
Goodness of Fit									
AIC	2395.59			2188.36			1896.06		
R²- Corrected (Nagelkerke)	0.56			0.54			0.63		

The significant variables for the injury crash model (FQ_S2) were total population (D0B), the number of automobiles per person (D0C_A_HH), total employment (D2D), road network density (D3A), and total trips (D6B_AA) with the positive association; and jobs per person (D2A_JP_HH) and the natural logarithm of the employment entropy (D2C_WRK_MX_LN) with the negative association.

The significant variables for the fatal crash model (FQ_S3) were total population (D0B and total employment (D2D) associated positively; jobs per person (D2A_JP_HH) and the urban indicator (D0A) associated negatively.

The significant variables for the total crash model (FQ_ALL) were total population (D0B), the number of automobiles per person (D0C_A_HH), total employment (D2D), road network density (D3A), and total trips (D6B_AA) associated positively;

jobs per person (D2A_JP_HH) and the natural logarithm of the employment entropy (D2C_WRK_MX_LN) associated negatively.

The significant variables for the motor vehicle involved crash model (FQ_MOT) were total population (D0B), the number of automobiles per person (D0C_A_HH), total employment (D2D), road network density (D3A), and total trips (D6B_AA) with positive relationships. Population density (D1B), employment density (D1C), jobs per person (D2A_JP_HH), the natural logarithm of the employment entropy (D2C_WRK_MX_LN) had a negative association.

The significant variables for the vulnerable road user involved crash model (FQ_VRU) were the urban indicator (D0A), trip productions and attractions equilibrium index (D2C_TRIP_EQ), road network density (D3A), and total trips (D6B_AA) with positive relation; land area (D1A_LA) and jobs per person (D2A_JP_HH) with negative relation.

Secondly, the results of the Konya case when kernel densities were modeled by gamma regression are presented. In Table 5.7 and Table 5.8, the density modeling results for injury severity and involvement are presented, respectively. Since the Gamma model for the density of fatal crashes (KD_S3) did not converge, that model was removed from the table.

Table 5.7 Optimal BE-TS Models for the Konya Case (KD severity)

IV \ DV	KD_S2		
	Beta		SE.
Intercept	2.51	***	0.17
D1A_LA	---		---
D2A_JP_HH	(0.01)	*	0.01
D2C_WRK_MX	(0.00)	***	0.00
D3A	0.02	***	0.01
D6B_AA	0.03	***	0.00
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$			
Goodness of Fit			
AIC	2130.33		
R²- Corrected (Nagelkerke)	0.36		

Table 5.8 Optimal BE-TS Models for the Konya Case (KD involvement)

IV \ DV	KD_ALL			KD_MOT			KD_VRU		
	Beta		SE.	Beta		SE.	Beta		SE.
Intercept	2.53	***	0.17	2.53	***	0.11	0.72	**	0.30
D1A_LA	---		---	---		---	(0.02)		0.01
D2A_JP_HH	(0.01)	*	0.01	(0.01)	*	0.01	---		---
D2C_WRK_MX	(0.00)	***	0.00	(0.00)	***	0.00	(0.00)	***	0.00
D3A	0.02	***	0.01	---		---	0.05	***	0.01
D6B_AA	0.03	***	0.00	0.03	***	0.00	0.03	***	0.00
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$									
Goodness of Fit									
AIC	2133.46			1917.92			1404.86		
R²- Corrected (Nagelkerke)	0.36			0.35			0.26		

The significant variables for the injury crash model (KD_S2) were road network density (D3A) and total trips (D6B_AA) with positive relation; jobs per person (D2A_JP_HH), and employment entropy (D2C_WRK_MX) with negative relation.

The significant variables for the total crash model (KD_ALL) were road network density (D3A) and total trips (D6B_AA) associated positively; jobs per person (D2A_JP_HH), and employment entropy (D2C_WRK_MX) associated negatively.

The significant variables for the motor vehicle involved crash model (KD_MOT) were total trips (D6B_AA) with positive relation; jobs per person (D2A_JP_HH), and employment entropy (D2C_WRK_MX) with negative relation.

The significant variables for the vulnerable road user involved crash model (KD_VRU) were road network density (D3A) and total trips (D6B_AA) with positive relation; land area (D1A_LA) and employment entropy (D2C_WRK_MX) with negative relation.

In Table 5.9, the results of crash frequency and crash density models are presented together. As in the developed county cases, for each crash category (injury crashes as S2, fatal crashes as S3, total crashes as ALL, motor vehicle involved crashes as MOT, and vulnerable road user involved crashes as VRU), both frequency and density modeling results were depicted.

For each TS variable, the first column depicts frequency modeling (negative binomial regression) outcomes, and the second column depicts kernel density modeling (gamma regression) outcomes.

The rows are assigned to the BE variables for which a significant association with TS variables (higher than 90%) was found. The color code represents both the significance levels (90% and 99%) and the direction of association (positive or negative). For example, jobs per person (D2A_JP_HH) was found to be negatively associated with injury crashes with a significance level of 99% in the frequency model (NB) and a significance level of 90% in the density model (GA); whereas, the road density (D3A) was found to be positively associated with the same crash type

with a significance level of 99% in both models. Since the Gamma regression model for the fatal crashes (S3) did not converge, the related column (S3-GA) was colored in grey.

Table 5.9 Comparison of the IVs for the Konya Case

IV	DV	S2		S3		ALL		MOT		VRU	
		NB	GA	NB	GA	NB	GA	NB	GA	NB	GA
D0A											
D0B											
D0C_A_HH											
D1A_LA											
D1B											
D1C											
D2A_JP_HH											
D2C_TRIP_EQ											
D2C_WRK_MX											
D2C_WRK_MX_LN											
D2D											
D3A											
D6B_AA											

According to the results, only the urban indicator (D0A) had mixed effects. It was positively associated with vulnerable road user involved crashes (with 99% significance) and negatively associated with fatal crashes (with 99% significance). Having fewer fatal crashes in the urban areas complied with the finding in the previous chapter for the tri-county case. Moreover, this finding was supported by the literature (Jermprapai & Srinivasan, 2014; Lee et al., 2017; Zhibin Li et al., 2013; Noland & Quddus, 2004a). Rural areas were associated with higher speeds, and

urban areas were associated with relatively lower speeds and traffic controls, so having lower fatal crashes in the urban areas was expected.

The positive relation between the urban areas and higher vulnerable road user crashes was not complying with the findings in the developed country cases in this study. The crash frequency models found negative associations. However, some of the former studies supported the positive association found in the Konya case (Cai et al., 2016; Siddiqui et al., 2012). One explanation for that positive association was that higher pedestrian and cyclist activity would be expected at the urban core. Moreover, there would be differences in obeying traffic rules in the two countries. Moreover, Turkey has higher transit and walking mode rates than the US, so it would not be surprising to have higher vulnerable road user crashes in urban areas in Turkey.

Jobs per person variable was found to be negatively associated with all crash types. In the single county case, jobs per household had a significantly negative association with only non-incapacitating injuries. There were no significant associations, either negative or positive, with other crash types, mostly because of having many other variables that would have associations. In the literature, the jobs-housing imbalance was positively associated with traffic crashes (C. Xu et al., 2017). The mismatch between jobs and housing would increase mobility and exposure, whereas balanced zones would be less likely to increase traffic collisions on this stance. On the other hand, mixed uses would increase the collisions by the increase in activities. In the former studies, positive associations with total crashes and mixed uses (C.-Y. Yu & Xu, 2018) and pedestrian crashes and commercial uses (C.-Y. Y. Yu, 2015) were observed.

The relation was not salient in the developed country cases presented in this study, but the economic sectors' density was among the key BE variables. Industrial employment has negative associations for all crash types for the single county case. For the Konya case, jobs per person value was the highest for industrial areas structured at the periphery. Thus, the negative associations were similar to the single county case. The pedestrian activity would be low at those sites, so the negative

association between the vulnerable road user crashes and jobs per person was expected. Moreover, though there would be high truck volume around those sites, the traffic would be more controlled so that the crash rates would be lower.

Employment density had positive associations with all crash types except the vulnerable road user involved crashes (not significant with VRU crashes). One possible reason for the contrast with jobs per person variable would be that employment density was also higher in the urban areas. Commercial and office jobs constitute more share at the urban core. Thus, the positive association in employment density with most crash types would be about the intensity of the employments other than the industry. For a better analysis, the employment density would be differentiated according to sectors as in the developed country case. In the tri-county case, all crashes had positive associations with retail and entertainment employment densities. In contrast, industrial employment density was negatively associated with all crash types in the single county case. This contrast was due to the employment structure and spatial distribution. Retail and entertainment sectors are mainly concentrated at the urban core. However, the light industry is usually located at or close to the periphery and the heavy industry at the cities' outer parts. In the single county case, there were patches of light industry located at the periphery. These parts were associated with higher motor vehicle usage. As motor vehicle usage was negatively associated with all crash types, fewer crashes were observed in those parts of the city.

Total network density had significant positive associations with injury, total, motor vehicle involved, and vulnerable road user involved crashes. This finding could be supported by some of the former studies that the road network density increases the likelihood of crash occurrences (X. Wang et al., 2014; P. Xu et al., 2014). In the single county case, total network density was not significant. In the tri-county case, total network density had negative associations with all crash types except vulnerable road user involved crashes. In a former study, the crash risk was negatively associated with freeway density but positively associated with the principal and

minor arterial densities. This result was attributed to the fact that freeways are generally better designed and have access control (H. Huang et al., 2010).

Moreover, a negative association between local roads and pedestrian crashes (Osama & Sayed, 2017; Y. Wang & Kockelman, 2013) and fatal crashes (C.-Y. Yu & Xu, 2018) were found. Therefore, results could be better evaluated if the road typology were present. Moreover, this density measure gives the same weight to the unit length of a local street and a highway. Therefore, the urban zones were given emphasis.

Total trips also had significant positive associations with injury, total, motor vehicle involved, and vulnerable road user involved crashes. The single county and the tri-county cases had the same results. This finding was almost the most robust one. As traffic volume increases, the exposure increases, so the likelihood of crashes (Cai, Abdel-Aty, Lee, et al., 2019; H. Huang et al., 2016; Loukaitou-Sideris et al., 2007; J. Wang & Huang, 2016; Zeng & Huang, 2014).

Among the BE variables modeled in the Konya case, the most significant variables were jobs per person (D2_JP_HH), road network density (D3A), and total trips (D6B_AA). For each variable used to model the Konya case's BE-TS relationship, the consistency measures were prepared according to the consistency scheme presented in the previous chapter. To determine factor consistency, tabu-search iterations were traced. Throughout the algorithm, the number of times the variable was in the selected subset, the number of times that specific variable was negative, positive, and zero were counted. After that, the BE variables' consistency was determined by the number of times that variable was in the model and the relation of that BE variable with the respective TS variable. To be considered consistent, the variable should be selected at least 80% of the iterations, and it should have the same type of relation (positive or negative) with the TS variable at least 90% of the iterations in which that BE variable was in the model. Table 5.10 presents these consistency measures. In terms of the presented consistency scheme jobs per person (D2_JP_HH), household workers per job equilibrium index (D2C_WRK_MX), the

natural logarithm of D2C_WRK_MX (D2C_WRK_MX_LN), road network density (D3A), and total trips (D6B_AA). The Konya case's key BE variables were determined by intersecting the significant variables and consistent variables sets. Therefore, the most significant and consistent variables in the Konya case were:

- Jobs per person (D2_JP_HH)
- Road network density (D3A)
- Total trips (D6B_AA)

Table 5.10 Consistency of the IVs for the Konya Case

		S2		ALL		MOT		VRU	
		NB	GA	NB	GA	NB	GA	NB	GA
IV	DV								
D0A									
D0B									
D0C_A_HH									
D1A_LA									
D1B									
D1C									
D2A_JP_HH									
D2C_TRIP_EQ									
D2C_WRK_MX									
D2C_WRK_MX_LN									
D2D									
D3A									
D6B_AA									

Since the BE variable dataset was limited in the Konya case, the key BE variables determined in the Konya case differed; however, they were not unfamiliar. The total trips variable in the Konya case was the surrogate of traffic volume in the single and tri-county cases. Traffic volume would be the most important explanatory variable of traffic safety, as expected. Traffic volume is a direct measure of exposure, and as exposure increases, the likelihood of crashes increases. The total trips variable represented the travel category of the BE. Total road network density as an exposure variable would substitute the percentage of primary roads variable of the single and tri-county cases and represented the BE's design category. Finally, the jobs per person variable, representing the BE's diversity category, would substitute the employment density and entropy variables in the single and tri-county cases. Ergo, those key BE variables in the Konya case had conjugates in the key BE variables set of the developed country context.

5.5 Model Comparison

This study aimed to analyze the BE-TS relationship. According to this aim, one of the study objectives was to determine the key built environment factors to explain this relationship. Also, to accomplish this aim, a generalizable methodology was proposed. Accordingly, by employing the proposed methodology, the set of key built environment variables to explain the BE-TS relationship within the developed country context were determined and presented in the previous chapter.

For generalizability of the findings, the developing county case results should be compared to the developed country cases; however, there are not enough studies in the literature considering developing country cases. The case studies' paucity is mostly due to the data concerns, lacking sufficient and reliable (consistent) data in developing countries. Similar concerns are valid for the Turkish case. Since data collection and dissemination are expensive, transportation officials should determine the critical information to collect and produce with limited resources.

Among the BE variables, the most significant and consistent ones were determined in the previous section for the developed country cases. In this chapter, the proposed methodology was applied for a case study from Turkey with limited data. However, it is essential to evaluate the possible contribution of the data that was not available. Therefore, an evaluation scheme is proposed in this section to measure the contribution of data and evaluate the performance of the key BE variables determined in developed country cases.

In this scheme, the developed country cases are reevaluated by solving the models with both the Konya case's significant variables and the key BE variables determined in the previous chapter.

For each TS variable in the first evaluation part, the significant variables in the Konya case were obtained (Table 5.11). For example, the single county PDO crash frequencies were estimated by the D0B, D0C_A_HH, D2A_JP_HH, D2A_WRK_EMP, D2D, D3A, and D6B_AA variables of the single county dataset, where those variables were the significant variables in the Konya PDO crash frequency model.

If a variable's inclusion contributes to explaining the variability in the dependent variable, it would have value. In specific, to evaluate the value of information, the goodness-of-fit statistics were compared. As the coefficient of determination, R^2 , is related to the explained variability in the dependent variable in a linear regression model, it is a proper measure to evaluate a model's performance (*Pseudo R-Squared Measures*, n.d.). Since the models are generalized linear models in this study, the pseudo R^2 measures introduced in the methodology chapter are used. (Cox & Snell R^2 and Nagelkerke R^2 , mentioned as “Likelihood-Ratio Based” and “Corrected,” respectively). These measures reflect how better the resulting model compared to the intercept-only model. Since the dependent variables and the set of all independent variables are the same, the measures are comparable. Thus, after solving each model for each case with the Konya case's significant variables, the goodness-of-fit statistics were gathered. Each model was then solved with the key BE variables

determined in the developed country cases (see Table 5.11). In that case, the same BE variable set was used for every setting.

Therefore, there were three sets of solutions for each setting: (i) the set consisted of the solutions with the significant variables of the Konya case, (ii) the set consisted of the solutions with the key BE variables, and (iii) the one consisted of the original solutions with the entire dataset. The goodness-of-fit statistics of the first set were used as the base, and they were compared to the second and third sets' goodness-of-fit statistics.

Table 5.11 Significant Variables in Each Model

Konya Case	BE variables in frequency models (NB)			
	S2	ALL	MOT	VRU
	D0B	D0B	D0B	D1A_LA
	D0C_A_HH	D0C_A_HH	D0C_A_HH	D2A_JP_HH
	D2A_JP_HH	D2A_JP_HH	D1B	D2C_TRIP_EQ
	D2A_WRK_EMP	D2A_WRK_EMP	D1C	D3A
	D2D	D2D	D2A_JP_HH	D6B_AA
	D3A	D3A	D2A_WRK_EMP	
	D6B_AA	D6B_AA	D2D	
			D3A	
			D6B_AA	
	BE variables in density models (GA)			
	S2	ALL	MOT	VRU
	D2C_WRK_MX	D2C_WRK_MX	D2A_JP_HH	D2C_WRK_MX
	D3A	D3A	D2C_WRK_MX	D3A
	D6B_AA	D6B_AA	D6B_AA	D6B_AA
Single & Tri County Cases	Key BE variables for both frequency and density models (NB & GA)			
	D0B	Total population		
	D0C_B90_P	% of housing units built before 1990		
	D1A_LA	Land area in acres		
	D1C_5_IND	Gross industrial employment density		
	D1C_5_ENT	Gross entertainment employment density		
	D1C_5_RET	Gross retail employment density		
	D2B_E5_MX	Employment entropy		
	D3D_R1_P	% of primary roads		
	D4D	Number of bus stops		
	D6A_M_P	% of motorized modes for commuting (excluding transit)		
	D6A_NM_P	% of non-motorized modes for commuting		
	D6B_AA	Annual average daily traffic for all modes		

In Table 5.12, the crash frequency model results of the single county case are presented. The first column is dedicated to the goodness-of-fit statistics for the models solved with limited data (“SC-KcV” stands for single county statistics with the Konya case variables). The second column is dedicated to the goodness-of-fit statistics for the models solved with key BE variables (“SC-KV” stands for single county statistics with key variables). The third column is dedicated to the percent improvement in the pseudo R^2 measures ($\Delta I\%$), so the value of the additional information, obtained by using the key BE variables. In this setting, the betterment of the fitted models compared to the intercept-only models are compared. The fourth column is dedicated to the goodness-of-fit statistics for the models solved with the full dataset (“SC” stands for single county statistics with the original dataset). The final column is dedicated to the percent improvement in the pseudo R^2 measures ($\Delta II\%$) when the entire dataset was used rather than the Konya case variables. For example, in the injury crash frequency model (S2), when the Konya case's significant variables were used to solve the single county case, the obtained pseudo R^2 measure was 0.70. It was 0.76 if the key variables were used. Thus, the model improved by 8.95%. Moreover, if the entire dataset was used, the pseudo R^2 measure (Corrected R^2) was 0.85, and the model improved by 21.78% compared to the solution with the variable set of the Konya case.

Table 5.12 Measuring Data Contribution in Single County Frequency Models

Frequency (NB)	Single county				
S2	SC-KcV	SC-KV	ΔI (%)	SC	ΔII (%)
R ² - L-ratio based	0.70	0.76	8.95	0.85	21.78
R ² - Corrected	0.70	0.76	8.95	0.85	21.78
AIC	1608.11	1578.77		1513.28	
BIC	1636.33	1622.67		1579.13	
Loglikelihood	-795.06	-775.38		-735.64	
Residual deviance	178.16	178.81		179.19	
ALL	SC-KcV	SC-KV	ΔI (%)	SC	ΔII (%)
R ² - L-ratio based	0.69	0.76	10.64	0.84	22.98
R ² - Corrected	0.69	0.76	10.64	0.84	22.98
AIC	1893.85	1858.66		1798.52	
BIC	1922.07	1902.56		1864.37	
Loglikelihood	-937.93	-915.33		-878.26	
Residual deviance	176.35	176.49		175.83	
MOT	SC-KcV	SC-KV	ΔI (%)	SC	ΔII (%)
R ² - L-ratio based	0.69	0.75	8.86	0.84	21.59
R ² - Corrected	0.69	0.75	8.86	0.84	21.59
AIC	1883.20	1851.67		1793.07	
BIC	1917.69	1895.57		1862.06	
Loglikelihood	-930.60	-911.83		-874.53	
Residual deviance	176.46	176.79		176.18	
VRU	SC-KcV	SC-KV	ΔI (%)	SC	ΔII (%)
R ² - L-ratio based	0.49	0.69	42.48	0.82	68.49
R ² - Corrected	0.49	0.70	42.48	0.82	68.49
AIC	881.66	805.89		729.05	
BIC	906.75	849.79		794.90	
Loglikelihood	-432.83	-388.94		-343.52	
Residual deviance	175.11	168.32		175.35	

The same calculations were done for the tri-county crash frequency models, and those models' results are presented in Table 5.13.

Table 5.13 Measuring Data Contribution in Tri-County Frequency Models

Frequency (NB)	Tri-county				
S2	TC-KcV	TC-KV	ΔI (%)	TC	ΔII (%)
R ² - L-ratio based	0.54	0.61	14.70	0.68	27.48
R ² - Corrected	0.54	0.61	14.70	0.68	27.48
AIC	32661.48	32044.52		31526.06	
BIC	32716.61	32130.28		31777.34	
Loglikelihood	-16321.74	-16008.26		-15722.03	
Residual deviance	3670.24	3662.30		3669.56	
ALL	TC-KcV	TC-KV	ΔI (%)	TC	ΔII (%)
R ² - L-ratio based	0.50	0.59	19.46	0.72	44.91
R ² - Corrected	0.50	0.59	19.46	0.72	44.91
AIC	40262.44	39547.43		38436.07	
BIC	40317.57	39633.19		38662.84	
Loglikelihood	-20122.22	-19759.71		-19181.04	
Residual deviance	3668.43	3631.32		3598.39	
MOT	TC-KcV	TC-KV	ΔI (%)	TC	ΔII (%)
R ² - L-ratio based	0.50	0.59	18.59	0.72	45.15
R ² - Corrected	0.50	0.59	18.59	0.72	45.15
AIC	40017.55	39341.98		38200.96	
BIC	40084.93	39427.75		38433.86	
Loglikelihood	-19997.77	-19656.99		-19062.48	
Residual deviance	3681.09	3645.52		3610.52	
VRU	TC-KcV	TC-KV	ΔI (%)	TC	ΔII (%)
R ² - L-ratio based	0.18	0.47	153.18	0.57	212.42
R ² - Corrected	0.18	0.47	153.18	0.58	212.42
AIC	20452.97	19034.89		18364.52	
BIC	20501.97	19120.65		18597.42	
Loglikelihood	-10218.48	-9503.44		-9144.26	
Residual deviance	3723.27	3682.33		3721.14	

In Table 5.14 and in Table 5.15, the crash density model results of single and tri-county cases are presented, respectively. Similar to the frequency model representation, the first column is dedicated to the goodness-of-fit statistics for the models solved according to the Konya case's variables, the second column is dedicated to the results of the model employing key BE variables, and the fourth column is dedicated to the modeling results with the full datasets. The third and the fifth columns are dedicated to the improvement percentages in the pseudo R^2 measure.

Table 5.14 Measuring Data Contribution in Single County Density Models

Density (GA)	Single county				
S2	SC-KcV	SC-KV	ΔI (%)	SC	ΔII (%)
R^2 - L-ratio based	0.66	0.80	22.29	0.87	32.10
R^2 - Corrected	0.66	0.80	22.29	0.87	32.10
AIC	1382.20	1303.59		1250.17	
BIC	1401.01	1347.49		1316.02	
Loglikelihood	-685.10	-637.80		-604.08	
Residual deviance	1370.20	1275.59		1208.17	
ALL	SC-KcV	SC-KV	ΔI (%)	SC	ΔII (%)
R^2 - L-ratio based	0.64	0.79	24.61	0.86	35.52
R^2 - Corrected	0.64	0.79	24.61	0.86	35.52
AIC	1683.31	1604.16		1549.47	
BIC	1702.12	1648.06		1615.32	
Loglikelihood	-835.65	-788.08		-753.74	
Residual deviance	1671.31	1576.16		1507.47	
MOT	SC-KcV	SC-KV	ΔI (%)	SC	ΔII (%)
R^2 - L-ratio based	0.63	0.79	23.88	0.86	35.04
R^2 - Corrected	0.63	0.79	23.88	0.86	35.04
AIC	1668.83	1596.04		1542.01	
BIC	1684.51	1639.94		1607.86	
Loglikelihood	-829.42	-784.02		-750.00	
Residual deviance	1658.83	1568.04		1500.01	
VRU	SC-KcV	SC-KV	ΔI (%)	SC	ΔII (%)
R^2 - L-ratio based	0.18	0.26	47.71	0.27	53.08
R^2 - Corrected	0.20	0.30	47.71	0.31	53.08
AIC	342.05	341.48		325.25	
BIC	357.73	385.38		347.21	
Loglikelihood	-166.03	-156.74		-155.63	
Residual deviance	332.05	313.48		311.25	

Table 5.15 Measuring Data Contribution in Tri-County Density Models

Density (GA)	Tri-county				
S2	TC-KcV	TC-KV	ΔI (%)	TC	ΔII (%)
R ² - L-ratio based	0.50	0.61	21.48	0.68	36.17
R ² - Corrected	0.50	0.61	21.48	0.68	36.17
AIC	27360.78	26566.51		25990.17	
BIC	27397.54	26652.27		26223.07	
Loglikelihood	-13674.39	-13269.25		-12957.09	
Residual deviance	27348.78	26538.51		25914.17	
ALL	TC-KcV	TC-KV	ΔI (%)	TC	ΔII (%)
R ² - L-ratio based	0.46	0.65	41.60	0.76	66.81
R ² - Corrected	0.46	0.65	41.60	0.76	66.81
AIC	34972.90	33542.47		32363.52	
BIC	35009.65	33628.23		32614.80	
Loglikelihood	-17480.45	-16757.23		-16140.76	
Residual deviance	34960.90	33514.47		32281.52	
MOT	TC-KcV	TC-KV	ΔI (%)	TC	ΔII (%)
R ² - L-ratio based	0.43	0.63	46.43	0.75	73.66
R ² - Corrected	0.43	0.63	46.43	0.75	73.66
AIC	34823.77	33366.39		32197.77	
BIC	34854.40	33452.16		32442.93	
Loglikelihood	-17406.89	-16669.20		-16058.89	
Residual deviance	34813.77	33338.39		32117.77	
VRU	TC-KcV	TC-KV	ΔI (%)	TC	ΔII (%)
R ² - L-ratio based	0.09	0.21	131.01	0.26	183.26
R ² - Corrected	0.09	0.22	131.01	0.27	183.34
AIC	12947.81	12479.60		12281.33	
BIC	12978.44	12565.37		12452.94	
Loglikelihood	-6468.90	-6225.80		-6112.67	
Residual deviance	12937.81	12451.60		12225.33	

The models employing the key built environment variables and the models employing the full dataset were all the better than the base case (the models that employed Konya case's variables) for both the single county and tri-county cases. The minimum improvement was observed when the single county case motor vehicle involved crash frequencies were modeled by employing key BE variables, by 8.86% (see Table 5.12). Whereas, the maximum improvement was observed when the tri-county case vulnerable road user involved crash frequencies were modeled by employing full dataset, by 212.42% (see Table 5.13). The average improvement for the models employing the key BE variables was 39.74%, and the average improvement for the models employing the full datasets was 61.28%.

All the models had improvements in the pseudo R^2 measures and other goodness-of-fit statistics such as AIC and BIC. However, it should be mentioned that though the improvements were salient, the number of variables used in the original cases was much higher than the ones in the Konya case. For example, in the tri-county S2 frequency model, there were 34 significant variables (whereas it was 7 in the Konya case). These improvements depended on many variables so that the emphasis would be given to the key BE factors. Continuing with the number of variables employed in modeling, the improvement percentages were always higher in the tri-county case. The population, the number of block groups, the number of crashes, and the number of BE variables used in the model were higher in the tri-county case than the single county case. More variables were found significant and consistent in the tri-county case models since more BE variables were used to explain the variability in TS variables. This difference also explains the difference between the improvements in the single and tri-county cases. The improvement percentages were consistently higher in the tri-county case (compared to the single county case). This result was because the number of critical variables in the original case was higher than the number of critical variables in the single county case. The explanatory power of the BE variables from the Konya case was smaller in the tri-county case.

Moreover, the improvements were the most with vulnerable road user involved crashes. For the frequency models, the lowest contribution was 42.48 % with the

single county case model, and the highest contribution was 212.42 % with the tri-county case. About the density models, the lowest contribution was 47.71 % with the single county case model, and the highest contribution was 183.34% for the tri-county case model. The most remarkable improvements could be achieved with vulnerable road user crashes since they are rare events compared to the motor vehicle involved crashes for the developed country cases. It is harder to find good fitting models for rare events; thus, an additional significant variable's marginal contribution would be the most among all crash types. Even an additional key variable would improve the models remarkably (similar results would be obtained if the models for the fatal crashes in the Konya case would have converged that significant contributions could be achieved by introducing additional explanatory variables).

The distribution of the crashes into severity and involvement categories is given in Table 5.16. In this table, there are two sections. In the first section, on the left-hand side, the crashes were distributed to the respective categories given in Chapter 3. The severity typology consisted of property damage only crashes (PDO), possible injury or non-incapacitating injury crashes, and incapacitating or fatal crashes. The involvement typology consisted of total crashes, motor vehicle involved crashes and vulnerable road user involved crashes.

On the right-hand side of Table 5.16, the distribution was done according to the crash typology valid for the Konya case. Since there was no record for the property damage only (POD) crashes in the Konya case, total crashes consisted of injury and fatal crashes. Since more than half of the crashes were of PDO type, reduction of the PDO crashes at least doubled the percentages of other types. For example, when PDO crashes were included, 61.4% of total crashes were PDO crashes in the tri-county case, and 34% of the crashes were injury crashes. The remaining 4.5% were fatal and incapacitating injury crashes. When PDO crashes were not accounted for, 88.2% of the crashes were injury crashes, and the remaining 11.8% were fatal crashes.

Table 5.16 Distribution of Crash Types in Varying Cases

	SINGLE COUNTY	TRI- COUNTY	SINGLE COUNTY	TRI- COUNTY	KONYA	
CRASH TYPE	%	%	%	%	%	CRASH TYPE
S1-Property damage only (PDO) crashes	54.3	61.4	-	-	-	-
S2-Possible injury or non-incapacitating injury crashes	41.5	34.0	90.7	88.2	99.2	S2-Injury crashes
S3-Incapacitating injury or fatal crashes	4.2	4.5	9.3	11.8	0.8	S3-Fatal crashes
ALL- Motorized and non-motorized	100	100	100**	100**	100**	ALL- Motorized and non-motorized*
MOT-Motor-vehicle crashes only	96.5	96.0	92.3**	89.5**	67.5**	MOT-Motor-vehicle crashes only*
VRU-Vulnerable road user crashes	3.5	4.0	7.7**	10.5**	32.5**	VRU-Vulnerable road user crashes*

*PDO crashes were not considered **Assuming that all the PDO crashes were of motor-vehicle involved crashes

Moreover, a similar shift would also be expected for the crash distribution according to involvement. The figures represented in Table 5.16 were calculated with the assumption that all the PDO crashes were motor vehicle involved crashes. Therefore, the maximum percentages for the vulnerable road user (VRU) involved crashes were calculated. Accordingly, the highest rates for the VRU crashes for the single and tri-county cases were 7.7% and 10.5%, respectively. In comparison, the percentage of VRU crashes was 32.5% in the Konya case. This solid contrast represents the contextual difference in traffic safety: the rarity of the VRU crashes in the developed country context and the high share of VRU crashes in the developing country context. There are probably many factors contributing to this difference, and the built environment and travel behaviors would be stated among the contributing factors.

Public transportation and walking are still the leading travel modes in the Turkish case (in 2012, the percentage of commuters using public transportation was 59.4% in Konya; Konya Greater Municipality, 2012, as cited in Cirit, 2014). Moreover, the urban form is still on the compactness side of the sprawl vs. compactness continuum; therefore, it is expected to have more VRU crashes in the Turkish case; while, in the US case, the leading transportation mode is private automobile. Another supporting figure depicted in Table 5.16 is that the percentage of fatal crashes was much higher in the US case. Usually, fatal and incapacitating injury crashes are associated with automobile usage. Therefore, this difference in the fatal crash rate between the two contexts was expected. If the results were replicated in another developing country case, the most significant improvements would be expected in the models for fatal crashes (followed by the VRU crashes), depending on the crash distributions.

To conclude, the explanatory power of all BE-TS models improved by employing additional variables drastically. The improvements were the highest with the VRU crashes because the VRU crashes were the least observed crashes in the developed country cases. In the developing country context, VRU crashes constituted a much higher percentage of all crashes (excluding the PDO crashes). Therefore, there are variations from one context to another. If the results of these modeling efforts are used in transportation planning, it would be promising for cutting humanitarian costs; saving lives and public resources.

CHAPTER 6

CONCLUSION

After the Second World War, the increased urbanization rate was provoked by increased car ownership. Increased car ownership and usage were the primary drivers that led to scattered urban development, namely “urban sprawl.” This type of urbanization resulted in some negative externalities, such as increased vehicle miles traveled, traffic accidents, and traffic congestion.

Among those externalities, special attention should be given to traffic accidents since they are directly related to human life. Traffic crashes are amongst the leading causes of death globally, with an increasing rate of crashes occurring in urban areas. This trend points out a rising risk of crashes necessitate assessing the relationship between the built environment (BE) and traffic safety (TS). Therefore, this study aimed at examining the association between the BE and TS.

6.1 Overview of the Research

In order to examine the relationship between the BE and TS, several objectives were defined. These objectives were formulated after a thorough literature review presented in Chapter 2. According to this comprehensive review, it was observed that the variables used to operationalize or define the BE could be classified through D-categories (demographic and socioeconomic properties, density, diversity, design, transit, destination accessibility, and travel). However, the variables under those categories vary from one study to another. Thus, the variable selection was a significant concern, which was one of this study’s objectives. Also, the BE variables’ associations could be mixed, so the context should also be regarded. Moreover, it was observed that only a minor part of the studies considered developing country contexts.

Accordingly, the following objectives were set and introduced in Chapter 3. The first objective was to identify the built environment factors having significant relationships with various traffic crashes. The second objective was to analyze the BE-TS relationship findings within multiple case studies and contexts (developed countries vs. developing countries). The third objective was to propose a generalizable methodology to evaluate those different cases in various contexts. The final objective was to propose an evaluation scheme for the value of information since data availability is critical in the developing country context.

As there are various datasets collected by different agencies, including numerous parameters regarding the BE and TS aspects, it is crucial to determine those that are primary indicators of the complex BE-TS relation in order to take necessary precautions and take countermeasures properly in advance. Moreover, it would help developing countries in their recent ongoing attempts to establish data collection systems in light of the rapid increase in motorization and urban growth, thus, increased fatalities. Therefore, the methodology presented in Chapter 3 had a concern about dataset preparation. The first step of the methodology was the extraction and classification of geospatial data. Then, in the second step, the dataset was reduced to a subset of variables by correlation analysis. A meta-heuristics method (tabu-search) was then applied to determine the best subset of BE variables explaining the variability in TS variables. Finally, a comparison-based evaluation scheme was proposed to decide on the key BE variables for explaining the variability in TS variables.

One difference in this study was evaluating a comprehensive set of variables (there were 100 variables before the correlation analysis). Additionally, to observe BE variables' impacts on different crash types, two crash typologies were used (severity and involvement). Also, crash representation was diversified as frequency and density. For each representation, different statistical distributions were used: negative binomial regression model was used for the crash frequency modeling, and Gamma, Lognormal and Weibull regression models were used for the crash density modeling.

Other than the extended set of BE and TS variables, this study's main contribution was implementing the meta-heuristic method, namely the tabu-search methodology, to improve the subset selection process, in which the best explanatory subset of BE variables was determined through an iterative algorithm.

The proposed methodology was applied to various cases to make a comparative evaluation, as stated in the objectives. Two particular cases were considered from the developed country context and a case from the developing country context. Developed country cases were introduced in Chapter 4. Leon County in the State of Florida was the first case, and Broward, Miami-Dade, and Palm Beach Counties in the state of Florida constituted the second case from the developed country context. Also, in Chapter 5, the case of Konya from Turkey was introduced as a case study from the developing country context. Considering varying case studies, varying crash representations, in total, 58 different BE-TS models were solved. Of those 58 models, 24 models were related to the single county case, 24 models were related to the tri-county case, and the remaining 10 models were related to the Konya case. Moreover, for the developed and developing country cases, there were at least 7500 and 1950 sub-models for each BE-TS model, respectively.

6.2 Summary of the Findings

According to the proposed methodology, all the BE-TS models were solved. For the single county case among 100 built environment variables, 50 variables were used in the analyses as the reduced set, whereas the number of variables for the tri-county case after dimension reduction was 57.

For each TS variable, both frequency and density models were compared, and then all the TS variables were compared, considering the significance levels of the BE variables. The variables that were significant for at least four of the six TS variables were considered as significant variables for the case. Then, all BE variables were evaluated according to their performances during the optimization (tabu-search)

procedure. The ones selected to be in the subset for at least 80% of the iterations and having the same relation with the TS variable (positive or negative) at least 90% of those iterations were highlighted. Moreover, the variables highlighted for at least four of the TS variables were considered consistent. The final selection was made by intersecting those two sets of variables. The significant and consistent variables were determined as the key BE variables for each case study. These evaluations were done for single county and tri-county cases. There were 14 key variables for the single county case and 22 key variables for the tri-county case.

Then the overall key BE variables were determined by comparing the key variables of single and tri-county cases. There were 9 key BE variables (significant and consistent) shared by single and tri-county cases. The following variables were determined as the key BE variables for the developed country cases:

- Total population
- Percentage of housing units built before 1990
- Land area
- Gross employment density
 - Gross industrial employment density
 - Gross entertainment employment density
 - Gross retail employment density
- Employment entropy
- Percentage of primary roads
- Number of bus stops
- Mode choice
 - Percentage of motorized modes for commuting (excluding transit)
 - Percentage of non-motorized modes for commuting

- Annual average daily traffic

The percentage of old housing stock, the number of bus stops in the census block group, and mode choice would be related to TS by the populations' economic status. Deprived communities are captive transit and non-motorized mode users in general. On the other hand, wealth and higher income are usually associated with lower crash rates. Total population, employment density, percent of primary roads, the number of bus stops, and traffic volume are surrogates of exposure. The increases in those variables increase the likelihood of crashes of any type. Total population, employment density, employment entropy, the number of bus stops, and mode choice indicate movement and activity, especially for vulnerable road users. Likewise, those variables also indicate higher motorized vehicle trips and entropy. Thus, these variables would be the key built environment variables to understand traffic safety variability from a macro-level perspective.

Determination of the key BE variables associated with TS was one of the objectives in this study. These key BE variables were both significant and consistent in both cases. However, it should be mentioned that variations among the cases existed, as formulated in the objectives. For example, in the single county case, industrial employment density was a key BE variable having negative associations with all crash types, but in the tri-county case, entertainment and retail employment densities were among the key BE variables having positive associations with crash types. This variance could be explained by the sectoral differentiation of different urban settings. Moreover, more variables were found to be significant and consistent for the tri-county case compared to the single county case, that more variables were required to explain the variability in TS variables in the tri-county case.

In addition to the variability among cases, there were within case variabilities. For example, in the tri-county case, the jobs within 45 minutes of auto travel time had positive associations with PDO crashes, non-incapacitating injury crashes, total, and motor vehicle involved crashes, while had negative associations with fatal and vulnerable road user involved crashes. The negative relation to the fatal crashes was

expected as the jobs were clustered at the city centers. Therefore, fatal crashes usually happen on primary roads at the periphery. However, the negative association between the vulnerable road user crashes, and employment locations were not expected. The reason for this relation is the metro service. Jobs are concentrated at the urban core of Miami-Dade County, and the core is served by the metro system, which is safer for vulnerable road users. The percentage of the primary roads had mixed associations with the crash types as just mentioned, having positive associations with crash types except vulnerable road user involved crashes since VRU involved crashes usually happen where there are more pedestrian activities such as urban centers and shopping centers with slower traffic.

Additionally, in this study, the importance of information was evaluated by comparing developed and developing country cases. Data availability is a critical issue for developing countries. It requires investment and funding, so at least critical data should be collected for safety analyses. Therefore, in the subsequent section, the same procedures were applied to the Konya case, where there were limited data. The prepared dataset had 18 BE variables, and after the correlation analysis, the number of variables was reduced to 13. After model solving, significance, and consistency tests, there were 3 key BE variables:

- Jobs per person
- Road network density
- Total trips (traffic volume)

Those key BE variables in the Konya case had replacements in the developed country context's key BE variables. The total trips variable used in the Konya case was a substitute variable of traffic volume used in the single and tri-county cases. Traffic volume seems to be the most significant explanatory variable of traffic safety, as expected. It was significant and consistent in both the developed and the developing county cases. Traffic volume is a direct measure of exposure, and as exposure increases, the probability of crashes increases. The total trips variable represented

the BE's travel category. Total road network density would replace the percentage of primary roads variable of the single and tri-county cases and represented the BE's design category. The jobs per person variable would substitute the employment density and entropy variables in the single and tri-county cases and represented the BE's diversity category.

In addition to determining the key BE variables in the Konya case, data availability in safety research was discussed in this study. Data availability is a critical issue, especially for developing countries that do not collect or share data properly. Data collection and sharing require investment and funding, so at least the critical data should be collected for safety analysis. That would significantly contribute to the associations' generalizability, restructuring of the confounding ones, and analyzing the BE associations in developing county context; thus, planning safer environments.

Measuring the value of data and so information was the last objective of this study. The additional data's contribution in explaining the variability of TS variables was evaluated. The developed country case BE-TS modes were resolved using the nine key BE variables determined for the single and tri-county cases in order to measure how much variability in the TS variables could be explained only by those key BE variables. The same TS variables were then modeled using the BE variables available to the Konya case to measure how much variability in the TS variables of the single and tri-county cases could be explained by only those BE variables available in the Konya case. Finally, the results were compared based on goodness of fit statistics. Accordingly, the models employing the key built environment variables and the models employing the full dataset were all the better than the base case (the models that used Konya case's variables only) for both the single and tri-county cases. The improvement measures (pseudo- R^2) were used to measure the models' explanatory power and the improvements obtained. The pseudo- R^2 measure of the base model (model with Konya case variables) was compared with the pseudo- R^2 measure of the two other sets of models (models with the original dataset and the models with the key BE variables). The minimum improvement in the pseudo- R^2 measure was 8.86%, and the maximum improvement was 212.42%. The average improvement for

the models employing the key BE variables was 39.74%, and the average improvement for the models employing the complete datasets was 61.28%. All the models had improvements in the pseudo- R^2 measures and other goodness-of-fit statistics (AIC and BIC). However, the most improvement was observed with the models of vulnerable road user involved crashes and then fatal crashes. This result was expected since VRU crashes are usually fewer compared to the motor vehicle involved crashes. Thus, the worst goodness of fit statistics was obtained in the models with VRU crashes as the dependent variable. Also, fewer variables than other crash types (except fatal crashes) were significant in those models. Therefore, the contribution of additional data would be invaluable since explaining those crashes would improve livability, as active transportation is the most sustainable mode of transportation in all dimensions.

6.3 Alternative Perspectives and the Flexibility of the Methodology

This study aimed to examine the relationship between the urban built environment and traffic safety. The built environment was operationalized through an extensive set of variables that were categorized under several built environment dimensions (e.g., density, destination accessibility, and transit service). Traffic safety was operationalized through two major crash typologies severity (property damage only crashes, possible injury or non-incapacitating injury crashes, and incapacitating injury or fatal crashes) and involvement (total crashes, motor-vehicle involved crashes, and vulnerable road user involved crashes).

Following the aim of this study, one objective was to propose a generalizable methodology. The methodology presented in this study was applied to several settings for all crash types (categorized regarding severity and involvement), all crash representations (count and density), and study areas (two from the US and one from Turkey). To comply with the stated objective, the built environment variables identified as the key variables (significant and consistent) were highlighted in most settings. Though all significant variables in each setting were discussed, the focus

was on the overall significance and consistency. One reason was to identify some of the key factors that impact all crash types, so the efforts could be directed to tackle the problems related to those factors to reduce all crashes. One other reason was related to data availability in the developing country cases. Data collection and sharing are expensive, so with limited financial resources, at least the critical data should be collected. Therefore, one of the secondary aims was to contribute to that knowledge.

The methodology presented in this study was flexible. It had several assumptions, and within different settings and perspectives, the same methodology could be applied with minor modifications. For example, the key factors could only be determined by checking significance rather than both significance and consistency. Another modification would be done for the definition of consistency. During the tabu-search procedure, the number of times a variable was in the model and its association with the dependent variable were the indicators of consistency.

There were a few thresholds determined for these indicators. For example, the minimum percentage of the iterations in which a variable is basic was set to 80%. This level was due to the empirical observations of this study. It was observed that 20% of the iterations would be passed for reaching plausible solutions. Also, as accepting a minimum significance level of 90%, the variables should have the same association (positive or negative) at 90% of the iterations, which could be considered consistent. Those thresholds would also be reset for more conservative or tolerant foci.

One other alternative would be related to the crash typology. It is common to define equivalence weights for the severity typology. The aim is to proposing weights to define property damage only equivalence of injury and fatal crashes (Ma et al., 2016; Oh et al., 2010; Washington et al., 2014). Therefore, rather than solving separate models for each severity level, the crashes are aggregated and solved in a single model that consider a property damage only (PDO) equivalence of the severity levels.

Another alternative would be changing the focus from all crashes to a specific crash type; for example, the vulnerable road user involved crashes. There is growing literature examining pedestrian and bicycle accidents, and the topic is a hot one in safety research. Rather than aiming at identifying the overall key built environment factors, defining the factors for VRU involved crashes could be aimed.

As an example, the Konya case is reevaluated. The results of all models for TS variables according to count and density models were presented in the related sections. As mentioned, the significance of the variables was extracted from the optimum iterations, and the consistency levels were extracted by tracing overall tabu-search processes by employing the stated threshold levels. Accordingly, the key BE variables for the Konya case were identified as:

- Jobs per person (D2A_JP_HH) (- association)
- Road network density (D3A) (+ association)
- Total trips (traffic volume) (D6B_AA) (+ association)

Suppose the focus is VRU crashes, and only the count model (negative binomial model) is considered, with the same significance and consistency levels. In that case, the following variables are identified (also given in Table 6.1):

- Urban indicator (D0A) (+ association)
- Land area (D1A_LA) (- association)
- Jobs per person (D2A_JP_HH) (- association)
- Trip production and attraction equilibrium index (D2C_TRIP_EQ) (+ association)
- Road network density (D3A) (+ association)
- Total trips (traffic volume) (D6B_AA) (+ association)

Table 6.1 Factor Significance and Consistency: VRU Crashes in Konya Case

	VRU (count)		Significance	Consistency
	+	-		
IV				
D0A				
D1A_LA				
D1B				
D1C				
D2A_JP_HH				
D2C_TRIP_EQ				
D2C_WRK_MX_LN				
D2D				
D3A				
D6B_AA				

When the focus has shifted from all crashes to VRU crashes, the number of key variables increased in the Konya case. The reason is that only the count model was used (in the former case, count model results were compared to density model results). Moreover, the percentage of VRU crashes was high in this case (32.5%). The reverse was observed in the developed county cases. In the developing country cases, the percentage of VRU crashes was the lowest (at most 10.5%). The number of significant and consistent variables was the least for the VRU crashes, even with the count model. As the number of observations decreases, the fit values worsen, and significant variables decrease.

Three more variables were identified as key variables for the built environment and vulnerable road user involved crash relationship in addition to jobs per person, road network density, and total trips variables. The additional variables were urban indicator, land area, and trip production and attraction index. Thus, the key variables and their associations would vary within the proposed methodology. The methodology presented in this study is flexible and could be applied with varying foci, and the thresholds and assumptions could be determined accordingly.

6.4 A Discussion for the Planning Practice

The “Sustainability” concept emerged with the Brundtland Report in 1987 after the turmoil realized in the 1970s with the oil crises when OPEC set strict quotas on the oil exports. All the countries egregiously realized their dependence on oil with this shock. The oil limit is soon to be reached, so striking changes happened after those crises even the governmental regimes were affected. The understanding of development has changed drastically due to the realization of oil dependence (the scarcity of fossil fuels) and the harm given to the natural environment (emission of greenhouse gases). Since then, the concept of *sustainability* became a vital element of many policy documents at both global and local levels.

As one of the most polluting sectors with energy production, industry, and housing, transportation was also influenced by this change. According to EPA (EPA, n.d.), based on 2010 global emission data, transportation had a significant share with 14%. Increased car ownership (initially was seen as an advantage for congested cities) became a significant problem. Car usage never decreased due to the paucity of reasonable alternatives, as mentioned in the first question, and the symbolic privilege of car ownership. Any improvement was offset (in the worst manner) by the increased volume of traffic.

Moreover, cities were dispersed in an uncontrolled manner with the mobility potential of automobiles. Mobility is an essential element that should not be regressed; however, the benefits must be considered with its negative impacts. Increased mobility by cars would, in turn, yield increased traffic congestion, traffic accidents, pollution, and sprawl. Thus, the problem consisted of reducing carbon emissions by the transport sector and the issues on urban life. Therefore, developed countries tried to change their planning perspectives according to sustainability principles. They replaced their former demand-driven transport policies (predict and provide) with supply-constrained ones (demand management).

This trend is not particular to the developed countries but also the developing ones and should be; however, the story is not the same for the developing countries. There are differences in the development processes of the developing countries. The developing countries could not replicate the high economic and environmental burdens of the industrialization period of the developed countries on underdeveloped countries and earth. The problems were realized after the developed countries' industrialization. From the sustainability perspective, developing countries should also be cautious about the outcomes. So, they should develop within the limitations that did not exist before.

For Turkey's case, as a developing country, the urbanization process could be summarized in three distinct periods (Şengül, 2012). In the first period (urbanization of the nation-state), the process was related to the young nation-state's self-actualization (Şengül, 2012). During this period, the government endeavored to construct the nation-state, took the initiative role, and established its national industry to accumulate capital despite its scarce production resources. Deliberate choices of locations of industrial settlements and the construction of railways to support industrial sites were the main drivers of the nation-state. Those drivers were amalgamated with the modernist planning perspective. That is to say, development in predetermined locations would trigger other places' development, so the given priority to some areas would yield better results for all (Tekeli, 1998). Significant transportation investments aimed at supporting industrial development during that period.

In the second period, namely the urbanization of the labor power, the Marshall Plan (European Recovery Program) triggered the emigration of rural people who lost their jobs due to the mechanization process in agriculture since this period's development strategy depended on agricultural development (Şengül, 2012). Vast numbers of people became idle in the rural areas since a tractor could perform the same processes with less cost than many workers. By the 1960s, the government's policy was *import substitution led to growth* in which by tariffs and quotas, the levels of imported goods were constrained. The main aim was to support Turkey's private entrepreneurs by

effectively protecting them from the external markets and competitors since it was hard for the government to provide all the necessary investments. Protection was assumed to encourage domestic private investment.

Meanwhile, the massive influx of population was not considered to be an important problem since developing industries required cheap labor power as much as possible. Those immigrants settled around industrial areas and on urban thresholds. The state's resources were dedicated to industrial development, so the newcomers – the emerging community – had to supply their housing requirements (*gecekondu*) and transportation modes (*dolmuş*). This type of development was assumed to be surmountable, and the problems were left aside; however, the results were catastrophic. Deregulations and legalization of illegal properties were implemented. Cities had to face uncontrolled growth. The infrastructure investments were not able to confront the needs of the migrants both in housing and transportation, and there were not any substantial urban transport planning efforts in this period. The government's import substitution strategy was transformed into an export-led development following the worldwide oil-led economic crises in the 1970s brings up the third period of urbanization, urbanization of capital.

In the third period of urbanization (urbanization of capital), neoliberal policies were important in line with the global applications. What is not seen in the first two periods other than the legal framework is the rapid accumulation of capital and switching it into the built environment. This aim could easily be attached to the legislation such as Development Law and Mass Housing Acts through which the government strongly supports the urbanization process. Local governments were given the authority to develop the land, and cheaper credits were available to the households. With its endeavors, Turkey tried to take its part in the global cyberspace through the development of necessary institutions such as financial markets and free trade zones in addition to the legal regulations (Tekeli, 1998). Competition in the global environment and capital investment into the built environment peaked in recent years.

In terms of urban form and development, the first two periods could be associated with compact settlements served by public transportation (paratransit mode where bus services were insufficient). However, both the political frame (neoliberal policies) and physical conditions (especially increased population and pollution) triggered the cities' expansion. Capital accumulation in the third urbanization period was followed by the investments in the construction sector and transportation infrastructure (prioritizing motor vehicles) motivated car usage and new settlements at the periphery. The economic development is reflected in major road projects. Moreover, the public sector prefers to invest in rail systems that are costly and neglects non-motorized modes and bus services (also the paratransit mode) that were the backbone of the former compact settlements (Babalık-Sutcliffe, 2017).

The urbanization experiences of developed (e.g., the US case) and developing countries differed remarkably. Till the 1970s, developing country cases would be defined as public transportation-oriented compact settlements, and developed countries would be defined as car-oriented sprawling settlements. There is still a gap between the development processes. Currently, the supply-side investments in road projects resemble the road investments that developed countries used to favor but try to curb now. The growth of cities combined with those investments urges private car usage.

Nevertheless, public transportation (considering bus, rail, paratransit, institutions' services, and taxis) still has a more significant share in commuter trips than private cars. Since the car ownership rates increase in developing countries, planning institutions should consider supporting its sustainable modes. The gap between the developed and developing countries in urban development and transportation may be narrowed in the future. However, it is crucial achieving convergence at a point favoring sustainability.

Policies and strategies should be implemented to reduce the travel demand of unsustainable modes and lessen transportation's incurable impacts on the environment. So, both macro and micro scale sets of principles should be applied for

urban transportation. By macro scale, the form, development, and functioning of cities should be decided. By micro scales (in accordance with micro-scale policies), the design and urban scale decisions such as the development of pedestrian ways should be decided. Moreover, the policies should consider the environmental, economic, and social sustainability aspects formerly mentioned (Khan, 1995).

For the sake of environmental friendliness, the emissions should be reduced. So, clean fuels and technological advances should be used and encouraged. Government incentives should be given for developing those technologies, and subsidies could be provided for the users. Moreover, those systems should be used in public modes with the corroboration of reducing automobile usage. Rather than automobiles and fossil-fueled vehicles (air transport), rail and sea transit should be encouraged. For the economic aspect, the dependence on fossil fuels should be reduced. Costs are not all monetary; the accidents result in deaths. So, public transit and non-motorized transportation should be supported. The same improvements are required for the equity aspect, and accessibility could be increased by improvements in public transport frequency and convenience. That is to say, for all aspects of sustainability, the important thing is to reduce the traffic volume and automobile usage while using energy-efficient and environmentally friendly modes and fuels. From this discussion, the following principles would lead sustainable transportation could be deduced:

Equitable and affordable transport: Transportation should be provided to eliminate the adverse effects of increased car dependence and set the frame of equity. The system should be planned and designed to satisfy the mobility needs of disadvantaged groups such as children, the elderly, and disabled people. In addition to the availability, public transportation should be affordable for all segments of the population. Management policies should be developed (e.g., fare integration, congestion charging, and parking policies) to achieve affordable transportation.

Increased mobility and accessibility: Transportation is one of the fundamental rights of people. In the current situation, the most convenient mode of mobility is automobiles. Thus, transportation services should be provided so that mobility and

accessibility levels are not reduced. The same mobility and accessibility level should be provided by the green modes (public transportation, cycling, and walking).

Efficient, clean, and safe transportation: To increase transportation efficiency, both operational and physical actions should be taken. So, not only the share of public transportation (especially the mass transit alternatives) should be increased, but also energy-efficient and cleaner modes (walking, cycling) should be favored and integrated into the transit system.

Safe transportation: Road safety should be considered. Since the most significant share of traffic accidents belongs to car accidents, there are factors to consider, such as speed limitations and modal shifts.

Pedestrian oriented transportation: The dominance of automobile-oriented transportation and planning should be replaced with pedestrian-oriented transportation. Public and private transportation alternatives should be integrated into a pedestrian-oriented planning perspective.

In this study, the built environment and traffic safety were formulated as representatives of urban development and urban transportation, respectively. The aim was to examine the relationship between the built environment and traffic safety. Safety is one of the sustainable-transportation principles, and all the presented principles indicate the importance of green modes (public transit, walking, and cycling).

Determining the key built environment factors and testing their significance and consistency would be a part of transportation and urban planning. This study's findings could be used in the planning of a green mode dominant transportation framework. The findings did not directly propose the design principles but put forward the relations between the built environment variables and traffic safety variables.

The results and variables should be considered in planning as safety considerations. For example, the number of bus stops was positively associated with all types of

crashes. Moreover, the number of bus stops was a consistent variable in the tabu-search process; besides, former studies came up with the same associations. This finding does not mean that to plan a safe environment, the number of bus stops should be reduced. Public transportation is essential for both environmental and social sustainability and must be sustained. Therefore, other aspects that would increase the likelihood of crashes in locations with a higher number of bus stops should be critically inspected rather than reducing bus stops. Bus stops are conflict points for both pedestrians and motor vehicle users. The number of bus stops is expected to be higher around trip attraction and generation sites or uses. Hence, the traffic volume (pedestrian and motor vehicle) is also higher around those areas increases the likelihood of crashes. For example, visible traffic signs should be used, and stop areas should be designed for safer transfers for the areas with a higher likelihood of crashes to increase traffic safety. Traffic should be controlled by some regulations such as enforcing speed limits around those areas. That is to say, the number of bus stops could be positively associated with increased crashes, but this information should be evaluated from a holistic point of view.

Similarly, the share of public transit was found to be positively associated with vulnerable road user crashes in the tri-county case. This association would also be undesirable since the share of transit should be increased according to the sustainability principles. This fact has relations with socioeconomic status. In some cities of the US (e.g., New York), public transportation has a significant share, and it is appealing to all segments of the population, but transit is usually associated with income deprivation and higher crash rates (Amoh-Gyimah et al., 2017; Bao et al., 2017; Jermprapai & Srinivasan, 2014). Therefore, providing reliable, affordable, and quality transit service is essential in terms of equity and safety. Another aspect of the increased crashes within high transit preference areas would be the concentration of activities and higher pedestrian volumes. This aspect also requires safety precautions just explained in the relation of the number of bus stops variable.

Furthermore, the non-motorized modes were found to be positively associated with all crash types in the tri-county case. It is expected to have higher crash rates in urban

areas with higher pedestrian and cyclist activities (Sener et al., 2019). However, as discussed in the background chapters, non-motorized modes (as the most sustainable travel mode) must be prioritized. Sidewalks and bike lanes should be planned as indispensable elements of transportation.

The outputs of this study could be used to signify the problematic aspects. A positive relation between traffic crashes and the green modes (non-motorized modes and transit) should be evaluated to signal the need for safety improvements.

To sum up, the resulting associations could be used in planning, especially in proactive planning. The potential conflict locations and issues should be considered while regarding area-specific properties. All the cities have their own economies, demographic structures, and functionality. It is essential to approach each case with both globally accepted facts and local features.

6.5 Limitations and Future Research Directions

There are several limitations to this study. First of all, parts of the dataset belonged to different periods. For example, the crash data belonged to 2009-2011; the census data was from 2010, whereas AADT values were calculated using the 2013-2015 data. The same was true for the smart location database (SLD). In SLD, the 2010 US census was used as the prime data source, but community surveys (ACS), employment indices, and transportation data from other sources with various periods were used. Using varying sources for data was sometimes incompatible due to changes. However, it could be assumed that the built environment does not change drastically in the short term. That is why similar properties could be valid for more extended periods. Also, AADT in the region did not change drastically year by year, like other metropolitan areas. Minor changes would not affect the general picture; nevertheless, it is desirable to employ matching data.

Related to the data availability issue, one of the objectives of this study was to highlight the importance of data in terms of the explanatory contributions of traffic

accidents. Thus, data availability is a primary concern for developing countries. Data limitation prohibits detailed analyses for the developing country cases.

The other limitation would be the generalizability of the results. The model would be repeated to cities of various sizes and from different states and countries. Therefore, all the variables could be evaluated within a more extensive set of results, which would also contribute to the variables' consistency.

One other limitation would be about the assumptions. Throughout this study's methodology, several assumptions were made, such as the buffer distances for crash frequency assignments, bandwidth and cell size assumptions for density calculations, and the selection of the objective function value. All those assumptions were based on either previously used values or calibrated values in the model. However, it should be mentioned that a change in any assumption may yield different results. It is impossible to determine optimal values for those values, so the high number of trials would offset the results' potential variance, at least in the optimization part.

The final limitation is concerning the modeling approach. The models used in this study did not consider spatial variability. That is to say, for each case and each BE-TS model, only a single coefficient set was optimized. For example, the single county case's fatal crash frequency model results were a set of BE variables' coefficients. Each population block group has its observations for those variables, so the resulting estimate for that block group could be calculated using the coefficients and the observations for that block group. For another block group, the same coefficients were also valid. The spatial relation of those block groups (if any relation exists) could only be represented in this modeling approach by the crash frequency and density calculations. For example, a crash was assigned to these two block groups if they were within the buffer zone of that specific crash occurrence. Other than that, no spatial relation was considered. This limitation is a direction to the subsequent studies of this thesis. One alternative would be geographically weighted regression models that consider spatial correlation, in which each block group would have its coefficients. The other choice would be multilevel modeling, in which a lower-level

unit is assumed to be nested under an upper-level unit. When applied to this study's crash modeling, population block groups could be nested under counties; however, that type of modeling requires much more data for the upper level. For example, for reliable research, around twenty counties should be considered. While in this study, just four counties were considered.

Another future research direction would be employing models jointly or sequentially estimating crash severity and involvement. In this study, these dimensions were analyzed separately. Bayesian approaches that use conditional probabilities are popular ones, as shown in the literature review part. A better evaluation could be realized by joint modeling, and the results would be compared.

Finally, a conceptual and modeling aspect could be highlighted for future research. In this study, BE variables were modeled in direct relation to TS variables. Moreover, travel variables were considered under the BE variables. Thus, in this study, travel and other BE variables were directly associated with TS variables. However, several studies in the literature examined the impacts of BE on travel, as presented in the dedicated literature review section. In a subsequent formulation, the (in)direct effects of BE variables on travel variables could also be considered. Therefore, in this alternative formulation, BE variables other than the travel dimension would be associated with TS variables both directly and indirectly through travel variables.

REFERENCES

- Abdel-Aty, M. A., & Radwan, A. E. (2000). Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention*, 32(5), 633–642. [https://doi.org/10.1016/S0001-4575\(99\)00094-9](https://doi.org/10.1016/S0001-4575(99)00094-9)
- Abdel-Aty, M., Lee, J., Siddiqui, C., & Choi, K. (2013). Geographical unit based analysis in the context of transportation safety planning. *Transportation Research Part A: Policy and Practice*, 49, 62–75. <https://doi.org/10.1016/j.tra.2013.01.030>
- Aguero-Valverde, J., & Jovanis, P. P. (2010). Spatial correlation in multilevel crash frequency models: Effects of different neighboring structures. *Transportation Research Record*, 2165, 21–32. <https://doi.org/10.3141/2165-03>
- Alarifi, S. A., Abdel-Aty, M. A., Lee, J., & Park, J. (2017). Crash modeling for intersections and segments along corridors: A Bayesian multilevel joint model with random parameters. *Analytic Methods in Accident Research*, 16, 48–59. <https://doi.org/10.1016/j.amar.2017.08.002>
- Alarifi, S. A., Abdel-Aty, M. A., Lee, J., & Wang, X. (2018). Exploring the Effect of Different Neighboring Structures on Spatial Hierarchical Joint Crash Frequency Models. *Transportation Research Record*, 2672(38), 210–222. <https://doi.org/10.1177/0361198118776759>
- Amoh-Gyimah, R., Saberi, M., & Sarvi, M. (2016). Macroscopic modeling of pedestrian and bicycle crashes: A cross-comparison of estimation methods. *Accident Analysis & Prevention*, 93, 147–159. <https://doi.org/10.1016/j.aap.2016.05.001>
- Amoh-Gyimah, R., Saberi, M., & Sarvi, M. (2017). The effect of variations in spatial units on unobserved heterogeneity in macroscopic crash models. *Analytic Methods in Accident Research*, 13, 28–51. <https://doi.org/10.1016/j.amar.2016.11.001>
- Angel, S., Parent, J., & Civco, D. (2007). Urban sprawl metrics: an analysis of global urban expansion using GIS. In *ASPRS 2007 Annual Conference Tampa, Florida* (pp. 1–12).

AR5 Climate Change 2014: Mitigation of Climate Change — IPCC. (n.d.). Retrieved October 9, 2020, from <https://www.ipcc.ch/report/ar5/wg3/>

Babalik-Sutcliffe, E. (2017). Turkey. In D. Pojani & D. Stead (Eds.), *The Urban Transport Crisis in Emerging Economies* (pp. 247–266). Springer International Publishing. https://doi.org/10.1007/978-3-319-43851-1_12

Bae, J., Kim, J., & Kim, J. (2016). Subset selection in multiple linear regression : An improved Tabu search. *Journal of the Korean Society of Marine Engineering*, 40(2), 138–145. <https://doi.org/10.5916/jkosme.2016.40.2.138>

Bagley, M. N., & Mokhtarian, P. L. (2002). The impact of residential neighborhood type on travel behavior: A structural equations modeling approach. *Annals of Regional Science*, 36(2002), 279–297. <https://doi.org/DOI10.1007/s001680200083>

Bao, J., Liu, P., Qin, X., & Zhou, H. (2018). Understanding the effects of trip patterns on spatially aggregated crashes with large-scale taxi GPS data. *Accident Analysis and Prevention*, 120(August 2017), 281–294. <https://doi.org/10.1016/j.aap.2018.08.014>

Bao, J., Liu, P., Yu, H., & Xu, C. (2017). Incorporating twitter-based human activity information in spatial analysis of crashes in urban areas. *Accident Analysis and Prevention*, 106(December 2016), 358–369. <https://doi.org/10.1016/j.aap.2017.06.012>

Bartholomew, K., & Ewing, R. (2008). Land Use–Transportation Scenarios and Future Vehicle Travel and Land Consumption: A Meta-Analysis. *Journal of the American Planning Association*, 75(1), 13–27. <https://doi.org/10.1080/01944360802508726>

Barua, S., El-Basyouny, K., & Islam, M. T. (2014). A Full Bayesian multivariate count data model of collision severity with spatial correlation. *Analytic Methods in Accident Research*, 3–4, 28–43. <https://doi.org/10.1016/j.amar.2014.09.001>

Barua, S., El-Basyouny, K., & Islam, M. T. (2016). Multivariate random parameters collision count data models with spatial heterogeneity. *Analytic Methods in Accident Research*, 9, 1–15. <https://doi.org/10.1016/j.amar.2015.11.002>

- Basiago, A. D. (1998). Economic, social, and environmental sustainability in development theory and urban planning practice. *Environmentalist*, 19(2), 145–161. <https://doi.org/10.1023/A:1006697118620>
- BBBike. (n.d.). *BBBike extracts OpenStreetMap*. Retrieved January 9, 2021, from <https://extract.bbbike.org/>
- Bento, A. M., Cropper, M. L., Mobarak, A. M., & Vinha, K. (2005). The Effects of Urban Spatial Structure on Travel Demand in the United States. *Review of Economics and Statistics*, 87(3), 466–478. <https://doi.org/10.1162/0034653054638292>
- Berrigan, D., Tatalovich, Z., Pickle, L. W., Ewing, R., & Ballard-Barbash, R. (2014). Urban sprawl, obesity, and cancer mortality in the United States: cross-sectional analysis and methodological challenges. *International Journal of Health Geographics*, 13(1), 3. <https://doi.org/10.1186/1476-072X-13-3>
- Bishai, D., Quresh, A., James, P., & Ghaffar, A. (2006). National road casualties and economic development. *Health Economics*, 15(1), 65–81. <https://doi.org/10.1002/hec.1020>
- Blazquez, C. A., & Celis, M. S. (2013). A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile. *Accident Analysis and Prevention*, 50, 304–311. <https://doi.org/10.1016/j.aap.2012.05.001>
- Blincoe, L., Miller, T. R., Zaloshnja, E., & Lawrence, B. A. (2015). *The economic and societal impact of motor vehicle crashes, 2010 (Revised)*. <https://trid.trb.org/view/1311862>
- Boarnet, M. G., Greenwald, M., & McMillan, T. E. (2008). Walking, urban design, and health: Toward a cost-benefit analysis framework. *Journal of Planning Education and Research*, 27(3), 341–358. <https://doi.org/10.1177/0739456X07311073>
- Boarnet, M. G., Nesamani, K. S., & Smith, C. S. (2003). *Comparing the influence of land use on network trip generation and vehicle distance traveled: An analysis using travel diary data*.
- Boarnet, M. G., & Wang, X. (2019). Urban spatial structure and the potential for

- vehicle miles traveled reduction: the effects of accessibility to jobs within and beyond employment sub-centers. *The Annals of Regional Science*, 381–404. <https://doi.org/10.1007/s00168-019-00900-7>
- Boer, R., Zheng, Y., Overton, A., Ridgeway, G. K., & Cohen, D. A. (2007). Neighborhood Design and Walking Trips in Ten U.S. Metropolitan Areas. *American Journal of Preventive Medicine*, 32(4), 298–304. <https://doi.org/10.1016/j.amepre.2006.12.012>
- Boğaziçi Proje. (2020). *Konya Büyükşehir Alanı Kentiçi ve Yakın Çevre Ulaşım Master Planı Çalışması*.
- Broadhurst, D., Goodacre, R., Jones, A., Rowland, J. J., & Kell, D. B. (1997). Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry. *Analytica Chimica Acta*, 348(1–3), 71–86. [https://doi.org/10.1016/S0003-2670\(97\)00065-2](https://doi.org/10.1016/S0003-2670(97)00065-2)
- Brueckner, J. K. (2000). Urban Sprawl: Diagnosis and Remedies. *International Regional Science Review*, 23(2), 160–171. <https://doi.org/10.1177/016001700761012710>
- Brunsdon, C. (1995). Estimating probability surfaces for geographical point data: An adaptive kernel algorithm. *Computers & Geosciences*, 21(7), 877–894. [https://doi.org/10.1016/0098-3004\(95\)00020-9](https://doi.org/10.1016/0098-3004(95)00020-9)
- Burchell, R. W., Shad, N. A., Listokin, D., & Phillips, H. (1998). *Transit Cooperative Research Program Report 39: The Costs of Sprawl—Revisited*.
- Cadar, R. D., Boitor, M. R., & Dumitrescu, M. (2017). Effects of traffic volumes on accidents: The case of Romania’s national roads. *Geographia Technica*, 12(2), 20–29. https://doi.org/10.21163/GT_2017.122.03
- Cai, Q., Abdel-Aty, M., & Lee, J. (2017). Macro-level vulnerable road users crash analysis: A Bayesian joint modeling approach of frequency and proportion. *Accident Analysis and Prevention*, 107(July), 11–19. <https://doi.org/10.1016/j.aap.2017.07.020>
- Cai, Q., Abdel-Aty, M., Lee, J., & Eluru, N. (2017). Comparative analysis of zonal

systems for macro-level crash modeling. *Journal of Safety Research*, 61, 157–166. <https://doi.org/10.1016/j.jsr.2017.02.018>

Cai, Q., Abdel-Aty, M., Lee, J., & Huang, H. (2019). Integrating macro- and micro-level safety analyses: a Bayesian approach incorporating spatial interaction. *Transportmetrica A: Transport Science*, 15(2), 285–306. <https://doi.org/10.1080/23249935.2018.1471752>

Cai, Q., Abdel-Aty, M., Lee, J., Wang, L., & Wang, X. (2018). Developing a grouped random parameters multivariate spatial model to explore zonal effects for segment and intersection crash modeling. *Analytic Methods in Accident Research*, 19, 1–15. <https://doi.org/10.1016/j.amar.2018.05.001>

Cai, Q., Abdel-Aty, M., Sun, Y., Lee, J., & Yuan, J. (2019). Applying a deep learning approach for transportation safety planning by using high-resolution transportation and land use data. *Transportation Research Part A: Policy and Practice*, 127(August 2018), 71–85. <https://doi.org/10.1016/j.tra.2019.07.010>

Cai, Q., Lee, J., Eluru, N., & Abdel-Aty, M. (2016). Macro-level pedestrian and bicycle crash analysis: Incorporating spatial spillover effects in dual state count models. *Accident Analysis and Prevention*, 93, 14–22. <https://doi.org/10.1016/j.aap.2016.04.018>

Çalışkan, O. (2004). Sürdürülebilir Kent Formu: Derişik Kent. *Planlama Dergisi*, 3, 33–54.

Cao, X. (Jason), Mokhtarian, P. L., & Handy, S. L. (2009). The relationship between the built environment and nonwork travel: A case study of Northern California. *Transportation Research Part A: Policy and Practice*, 43(5), 548–559. <https://doi.org/10.1016/j.tra.2009.02.001>

Cao, X., Mokhtarian, P. L., & Handy, S. L. (2007). Do changes in neighborhood characteristics lead to changes in travel behavior? A structural equations modeling approach. *Transportation*, 34(5), 535–556. <https://doi.org/10.1007/s11116-007-9132-x>

Center for Neighborhood Technology. (n.d.-a). *AllTransit Rankings*. Retrieved January 24, 2021, from <https://alltransit.cnt.org/rankings/>

- Center for Neighborhood Technology. (n.d.-b). *Housing + Transportation Affordability Index*. Retrieved November 11, 2020, from <https://www.cnt.org/tools/housing-and-transportation-affordability-index>
- Cervero, R. (2002). Built environments and mode choice: Toward a normative framework. *Transportation Research Part D: Transport and Environment*, 7(4), 265–284. [https://doi.org/10.1016/S1361-9209\(01\)00024-4](https://doi.org/10.1016/S1361-9209(01)00024-4)
- Cervero, R., & Duncan, M. (2006). Which Reduces Vehicle Travel More: Jobs-Housing Balance or Retail-Housing Mixing? *Journal of the American Planning Association*, 72(4), 475–490. <https://doi.org/10.1080/01944360608976767>
- Cervero, R., & Kockelman, K. (1997). Travel demand and the 3Ds: Density, diversity, and design. *Transportation Research Part D: Transport and Environment*, 2(3), 199–219. [https://doi.org/10.1016/S1361-9209\(97\)00009-6](https://doi.org/10.1016/S1361-9209(97)00009-6)
- Cervero, R., & Murakami, J. (2010). Effects of Built Environments on Vehicle Miles Traveled: Evidence from 370 US Urbanized Areas. *Environment and Planning A: Economy and Space*, 42(2), 400–418. <https://doi.org/10.1068/a4236>
- Chatman, D. G. (2003). How Density and Mixed Uses at the Workplace Affect Personal Commercial Travel and Commute Mode Choice. *Transportation Research Record: Journal of the Transportation Research Board*, 1831(1), 193–201. <https://doi.org/10.3141/1831-22>
- Chatman, D. G. (2008). Deconstructing development density: Quality, quantity and price effects on household non-work travel. *Transportation Research Part A: Policy and Practice*, 42(7), 1008–1030. <https://doi.org/10.1016/j.tra.2008.02.003>
- Chatman, D. G. (2009). Residential Choice, the Built Environment, and Nonwork Travel: Evidence Using New Data and Methods. *Environment and Planning A: Economy and Space*, 41(5), 1072–1089. <https://doi.org/10.1068/a4114>
- Chen, C., & Xie, Y. (2016). Modeling the effects of AADT on predicting multiple-vehicle crashes at urban and suburban signalized intersections. *Accident Analysis and Prevention*, 91, 72–83. <https://doi.org/10.1016/j.aap.2016.02.016>
- Chen, P. (2015). Built environment factors in explaining the automobile-involved

- bicycle crash frequencies: A spatial statistic approach. *Safety Science*, 79, 336–343. <https://doi.org/10.1016/j.ssci.2015.06.016>
- Chen, P., & Shen, Q. (2016). Built environment effects on cyclist injury severity in automobile-involved bicycle crashes. *Accident Analysis & Prevention*, 86, 239–246. <https://doi.org/10.1016/j.aap.2015.11.002>
- Chen, P., & Shen, Q. (2019). Identifying high-risk built environments for severe bicycling injuries. *Journal of Safety Research*, 68, 1–7. <https://doi.org/10.1016/j.jsr.2018.11.002>
- Chen, P., & Zhou, J. (2016). Effects of the built environment on automobile-involved pedestrian crash frequency and risk. *Journal of Transport & Health*, 3(4), 448–456. <https://doi.org/10.1016/j.jth.2016.06.008>
- Chi Nguyen, T. M., Kato, H., & Phan, L. B. (2020). Is built environment associated with travel mode choice in developing cities? Evidence from Hanoi. *Applied Sciences (Switzerland)*, 12(14), 1–16. <https://doi.org/10.3390/su12145773>
- Chiou, Y. C., & Fu, C. (2013). Modeling crash frequency and severity using multinomial-generalized Poisson model with error components. *Accident Analysis and Prevention*, 50, 73–82. <https://doi.org/10.1016/j.aap.2012.03.030>
- Cho, G., Rodríguez, D. A., & Khattak, A. J. (2009). The role of the built environment in explaining relationships between perceived and actual pedestrian and bicyclist safety. *Accident Analysis & Prevention*, 41(4), 692–702. <https://doi.org/10.1016/j.aap.2009.03.008>
- Choi, K. (2018). The influence of the built environment on household vehicle travel by the urban typology in Calgary, Canada. *Cities*, 75(January), 101–110. <https://doi.org/10.1016/j.cities.2018.01.006>
- Chowdhury, T., & Scott, D. M. (2020). An analysis of the built environment and auto travel in Halifax, Canada. *Transport Policy*, 94(May), 23–33. <https://doi.org/10.1016/j.tranpol.2020.05.003>
- Chung, S., Park, Y. W., & Cheong, T. (2020). A mathematical programming approach for integrated multiple linear regression subset selection and validation. *Pattern Recognition*, 108, 107565.

<https://doi.org/10.1016/j.patcog.2020.107565>

- Cirit, F. (2014). *Sürdürülebilir Kentiçi Ulaşım Politikaları ve Toplu Taşıma Sistemlerinin Karşılaştırılması* [T.C. Kalkınma Bakanlığı]. <https://www.sbb.gov.tr/wp-content/uploads/2018/11/FarukCirit.pdf>
- Clark, D. E., & Cushing, B. M. (2004). Rural and urban traffic fatalities, vehicle miles, and population density. *Accident Analysis & Prevention*, 36(6), 967–972. <https://doi.org/10.1016/j.aap.2003.10.006>
- Clifton, K. J., Burnier, C. V., & Akar, G. (2009). Severity of injury resulting from pedestrian-vehicle crashes: What can we learn from examining the built environment? *Transportation Research Part D: Transport and Environment*, 14(6), 425–436. <https://doi.org/10.1016/j.trd.2009.01.001>
- Clifton, K. J., & Kreamer-Fults, K. (2007). An examination of the environmental attributes associated with pedestrian-vehicular crashes near public schools. *Accident Analysis & Prevention*, 39(4), 708–715. <https://doi.org/10.1016/j.aap.2006.11.003>
- Cottrill, C. D., & Thakuriah, P. (2010). Evaluating pedestrian crashes in areas with high low-income or minority populations. *Accident Analysis and Prevention*, 42(6), 1718–1728. <https://doi.org/10.1016/j.aap.2010.04.012>
- Cox, D. R., & Snell, E. J. (1998). Analysis of Binary Data, Second Edition. In *Analysis of Binary Data, Second Edition*. Chapman and Hall.
- D’Agostino, R. B., & Stephens, M. A. (Eds.). (1986). *Goodness-of-Fit Techniques*. Marcel Dekker.
- Dai, D., & Jaworski, D. (2016). Influence of built environment on pedestrian crashes: A network-based GIS analysis. *Applied Geography*, 73, 53–61. <https://doi.org/10.1016/j.apgeog.2016.06.005>
- Dai, D., Taquechel, E., Steward, J., & Strasser, S. (2010). The impact of built environment on pedestrian crashes and the identification of crash clusters on an urban university campus. *The Western Journal of Emergency Medicine*, 11(3), 294–301.

- de-Graft Acquah, H. (2010). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *Journal of Development and Agricultural Economics*, 2(1), 001–006.
- Deepika, S., & Saradha, R. (2014). Clustering Crash Hotspots to Organize Police Dispatch Routes Using GIS. *International Journal of Science and Research (IJSR)*, 3(2), 44–50.
- Department of Transport. (2012). *Statistical Release-Reported Road Casualties in Great Britain: 2012 Annual Report*. <https://www.gov.uk/transport-statistics-notes-and-guidance-road-accident-and-safety>
- Diao, M. i, & Ferreira, J. (2014). Vehicle Miles Traveled and the Built Environment: Evidence from Vehicle Safety Inspection Data. *Environment and Planning A: Economy and Space*, 46(12), 2991–3009. <https://doi.org/10.1068/a140039p>
- Ding, C., Chen, P., & Jiao, J. (2018). Non-linear effects of the built environment on automobile-involved pedestrian crash frequency: A machine learning approach. *Accident Analysis and Prevention*, 112(December 2017), 116–126. <https://doi.org/10.1016/j.aap.2017.12.026>
- Ding, C., Wang, Y., Tang, T., Mishra, S., & Liu, C. (2018). Joint analysis of the spatial impacts of built environment on car ownership and travel mode choice. *Transportation Research Part D: Transport and Environment*, 60, 28–40. <https://doi.org/10.1016/j.trd.2016.08.004>
- Dohoo, I. R., Ducrot, C., Fourichon, C., Donald, A., & Hurnik, D. (1997). An overview of techniques for dealing with large numbers of independent variables in epidemiologic studies. *Preventive Veterinary Medicine*, 29(3), 221–239. [https://doi.org/10.1016/S0167-5877\(96\)01074-4](https://doi.org/10.1016/S0167-5877(96)01074-4)
- Domijan, K. (2018). *tabuSearch: Tabu-search Algorithm for Binary Configurations (1.1.1)*. R. <https://doi.org/10.1111/j.1751-5823.2002.tb00174.x>
- Dong, N., Huang, H., Lee, J., Gao, M., & Abdel-Aty, M. (2016). Macroscopic hotspots identification: A Bayesian spatio-temporal interaction approach. *Accident Analysis and Prevention*, 92, 256–264. <https://doi.org/10.1016/j.aap.2016.04.001>

- Dong, N., Huang, H., Xu, P., Ding, Z., & Wang, D. (2014). Evaluating Spatial-Proximity Structures in Crash Prediction Models at the Level of Traffic Analysis Zones. *Transportation Research Record: Journal of the Transportation Research Board*, 2432(1), 46–52. <https://doi.org/10.3141/2432-06>
- Dong, N., Huang, H., & Zheng, L. (2015). Support vector machine in crash prediction at the level of traffic analysis zones: Assessing the spatial proximity effects. *Accident Analysis and Prevention*, 82, 192–198. <https://doi.org/10.1016/j.aap.2015.05.018>
- Downs, A. (1999). Some Realities about Sprawl and Urban Decline. *Housing Policy Debate*, 10(4), 955–974. <https://doi.org/10.1080/10511482.1999.9521356>
- Drezner, Z., Marcoulides, G. A., & Salhi, S. (1999). Tabu search model selection in multiple regression analysis. *Communications in Statistics Part B: Simulation and Computation*, 28(2), 349–367. <https://doi.org/10.1080/03610919908813553>
- Dumbaugh, E., & Li, W. (2011). Designing for the safety of pedestrians, cyclists, and motorists in urban environments. *Journal of the American Planning Association*, 77(1), 69–88. <https://doi.org/10.1080/01944363.2011.536101>
- Dumbaugh, E., & Rae, R. (2009). Safe Urban Form: Revisiting the Relationship Between Community Design and Traffic Safety. *Journal of the American Planning Association*, 75(3), 309–329. <https://doi.org/10.1080/01944360902950349>
- Dumbaugh, E., Rae, R., & Wunneberger, D. (2011). Using GIS to develop a performance-based framework for evaluating urban design and crash incidence. *Urban Design International*, 16(1), 63–71. <https://doi.org/10.1057/udi.2010.16>
- El-Basyouny, K., & Sayed, T. (2009). Urban Arterial Accident Prediction Models with Spatial Effects. *Transportation Research Record: Journal of the Transportation Research Board*, 2102(1), 27–33. <https://doi.org/10.3141/2102-04>
- EPA. (n.d.). *Global Greenhouse Gas Emissions Data*. Retrieved October 9, 2020, from <https://www.epa.gov/ghgemissions/global-greenhouse-gas-emissions->

data

Erdogan, S. (2009). Explorative spatial analysis of traffic accident statistics and road mortality among the provinces of Turkey. *Journal of Safety Research*, 40(5), 341–351. <https://doi.org/10.1016/j.jsr.2009.07.006>

ESRI. (2011). *ArcGIS Desktop: Release 10*. Environmental Systems Research Institute.

European Commission. (n.d.). *Sustainable transport | Mobility and Transport*. Retrieved October 9, 2020, from https://ec.europa.eu/transport/themes/sustainable_en

Ewing, R. (1997). Is Los Angeles-Style Sprawl Desirable? *Journal of the American Planning Association*, 63(1), 107–126. <https://doi.org/10.1080/01944369708975728>

Ewing, R., & Cervero, R. (2001). Travel and the Built Environment: A Synthesis. *Transportation Research Record: Journal of the Transportation Research Board*, 1780(1), 87–114. <https://doi.org/10.3141/1780-10>

Ewing, R., & Cervero, R. (2010). Travel and the built environment. *Journal of the American Planning Association*, 76(3), 265–294. <https://doi.org/10.1080/01944361003766766>

Ewing, R., & Dumbaugh, E. (2009). The Built Environment and Traffic Safety. *Journal of Planning Literature*, 23(4), 347–367. <https://doi.org/10.1177/0885412209335553>

Ewing, R., & Hamidi, S. (2014). Measuring Sprawl 2014. *Smart Growth America*, 45, 51. <https://doi.org/10.1111/j.1728-4465.2014.00390.x>

Ewing, R., Hamidi, S., & Grace, J. B. (2016). Urban sprawl as a risk factor in motor vehicle crashes. *Urban Studies*, 53(2), 247–266. <https://doi.org/10.1177/0042098014562331>

Ewing, R., Hamidi, S., Grace, J. B., & Dennis, Y. (2016). Does urban sprawl hold down upward mobility? *Landscape and Urban Planning*, 148, 80–88.

<https://doi.org/10.1016/j.landurbplan.2015.11.012>

Ewing, R., Pendall, R., & Chen, D. (2002). *Measuring Sprawl and its Impact. I.*

Ewing, R., Pendall, R., & Chen, D. (2003). Measuring Sprawl and Its Transportation Impacts. *Transportation Research Record: Journal of the Transportation Research Board*, 1831(1), 175–183. <https://doi.org/10.3141/1831-20>

Ewing, R., Schieber, R. A., & Zegeer, C. V. (2003). Urban Sprawl as a Risk Factor in Motor Vehicle Occupant and Pedestrian Fatalities. *American Journal of Public Health*, 93(9), 1541–1545. <https://doi.org/10.2105/AJPH.93.9.1541>

Ewing, R., Tian, G., & Lyons, T. (2018). Does compact development increase or reduce traffic congestion? *Cities*, 72(August 2016), 94–101. <https://doi.org/10.1016/j.cities.2017.08.010>

FDOT. (2015). *Traffic Data*. Florida Department of Transportation.

Federal Highway Administration. (n.d.-a). *Definition, Interpretation, and Calculation of Traffic Analysis Tools Measures of Effectiveness*. Retrieved October 10, 2020, from <https://ops.fhwa.dot.gov/publications/fhwahop08054/sect2.htm>

Federal Highway Administration. (n.d.-b). *Highway Safety Improvement Program Manual*. Retrieved October 11, 2020, from <https://safety.fhwa.dot.gov/hsip/resources/fhwasa09029/sec2.cfm>

Flahaut, B. (2004). Impact of infrastructure and local environment on road unsafety. *Accident Analysis & Prevention*, 36(6), 1055–1066. <https://doi.org/10.1016/j.aap.2003.12.003>

Flask, T., & Schneider, W. (2013). A Bayesian analysis of multi-level spatial correlation in single vehicle motorcycle crashes in Ohio. *Safety Science*, 53, 1–10. <https://doi.org/10.1016/j.ssci.2012.08.008>

Florida Transit Data Exchange - FTIS: Florida Transit Information System. (n.d.). Retrieved December 13, 2020, from <https://ftis.org/Posts.aspx>

- Fouskakis, D., & Draper, D. (2002). Stochastic optimization: A review. *International Statistical Review*, 70(3), 315–349. <https://doi.org/10.1111/j.1751-5823.2002.tb00174.x>
- Frenkel, A., & Ashkenazi, M. (2008). Measuring urban sprawl: How can we deal with it? *Environment and Planning B: Planning and Design*, 35(1), 56–79. <https://doi.org/10.1068/b32155>
- Galster, G., Hanson, R., Ratcliffe, M. R., Wolman, H., Coleman, S., & Freihage, J. (2001). Wrestling Sprawl to the Ground: Defining and measuring an elusive concept. *Housing Policy Debate*, 12(4), 681–717. <https://doi.org/10.1080/10511482.2001.9521426>
- Gao, J., Kamphuis, C. B. M., Helbich, M., & Ettema, D. (2020). What is ‘neighborhood walkability’? How the built environment differently correlates with walking for different purposes and with walking on weekdays and weekends. *Journal of Transport Geography*, 88(September), 102860. <https://doi.org/10.1016/j.jtrangeo.2020.102860>
- García-Palomares, J. C. (2010). Urban sprawl and travel to work: the case of the metropolitan area of Madrid. *Journal of Transport Geography*, 18(2), 197–213. <https://doi.org/10.1016/j.jtrangeo.2009.05.012>
- Gehrke, S. R., & Wang, L. (2020). Operationalizing the neighborhood effects of the built environment on travel behavior. *Journal of Transport Geography*, 82(October 2018), 102561. <https://doi.org/10.1016/j.jtrangeo.2019.102561>
- Gerike, R., & Koszowski, C. (2017). Sustainable Urban Transportation. In *Encyclopedia of Sustainable Technologies* (Vol. 2). Elsevier. <https://doi.org/10.1016/B978-0-12-409548-9.10176-9>
- Ghnatios, C., Hage, R.-M., & Hage, I. (2019). An efficient Tabu-search optimized regression for data-driven modeling. *Comptes Rendus Mécanique*, 347(11), 806–816. <https://doi.org/10.1016/j.crme.2019.11.006>
- Gillham, O., & MacLean, A. S. (2002). *The limitless city: A primer on the urban sprawl debate*. Island Press.
- Glover, F. (1986). Future Paths for Integer Programming and Links to Artificial

- Intelligence. *Computers and Operations Research*, 13(5), 533–549. [https://doi.org/http://dx.doi.org/10.1016/0305-0548\(86\)90048-1](https://doi.org/http://dx.doi.org/10.1016/0305-0548(86)90048-1)
- Gomes, M. J. T. L., Cunto, F., & Silva, A. R. (2017). Geographically weighted negative binomial regression applied to zonal level safety performance models. *Accident Analysis and Prevention*, 106(February), 254–261. <https://doi.org/10.1016/j.aap.2017.06.011>
- Gordon, P., & Richardson, H. W. (1997). Are Compact Cities a Desirable Planning Goal? *Journal of the American Planning Association*, 63(1), 95–107. <https://doi.org/10.1080/01944369708975727>
- Greenwald, M. J., & Boarnet, M. G. (2001). Built Environment as Determinant of Walking Behavior: Analyzing Nonwork Pedestrian Travel in Portland, Oregon. *Transportation Research Record: Journal of the Transportation Research Board*, 1780(1), 33–41. <https://doi.org/10.3141/1780-05>
- Guan, X., & Wang, D. (2019). Influences of the built environment on travel: A household-based perspective. *Transportation Research Part A: Policy and Practice*, 130(October), 710–724. <https://doi.org/10.1016/j.tra.2019.10.003>
- Guerra, E., Dong, X., & Kondo, M. (2019). Do Denser Neighborhoods Have Safer Streets? Population Density and Traffic Safety in the Philadelphia Region. *Journal of Planning Education and Research*, 0739456X1984504. <https://doi.org/10.1177/0739456X19845043>
- Guo, F., Wang, X., & Abdel-Aty, M. A. (2010). Modeling signalized intersection safety with corridor-level spatial correlations. *Accident Analysis and Prevention*, 42(1), 84–92. <https://doi.org/10.1016/j.aap.2009.07.005>
- Guo, Q., Xu, P., Pei, X., Wong, S. C., & Yao, D. (2017). The effect of road network patterns on pedestrian safety: A zone-based Bayesian spatial modeling approach. *Accident Analysis and Prevention*, 99, 114–124. <https://doi.org/10.1016/j.aap.2016.11.002>
- Guo, Z. (2013). Does residential parking supply affect household car ownership? The case of New York City. *Journal of Transport Geography*, 26, 18–28. <https://doi.org/10.1016/j.jtrangeo.2012.08.006>

- Guzman, L. A., Peña, J., & Carrasco, J. A. (2020). Assessing the role of the built environment and sociodemographic characteristics on walking travel distances in Bogotá. *Journal of Transport Geography*, 88(January), 102844. <https://doi.org/10.1016/j.jtrangeo.2020.102844>
- Habibi, S., & Asadi, N. (2011). Causes, Results and Methods of Controlling Urban Sprawl. *Procedia Engineering*, 21, 133–141. <https://doi.org/10.1016/j.proeng.2011.11.1996>
- Hadayeghi, A., Shalaby, A. S., & Persaud, B. N. (2003). Macrolevel Accident Prediction Models for Evaluating Safety of Urban Transportation Systems. *Transportation Research Record: Journal of the Transportation Research Board*, 1840(1), 87–95. <https://doi.org/10.3141/1840-10>
- Hadayeghi, A., Shalaby, A. S., & Persaud, B. N. (2007). Safety Prediction Models. *Transportation Research Record: Journal of the Transportation Research Board*, 2019(1), 225–236. <https://doi.org/10.3141/2019-27>
- Hadayeghi, A., Shalaby, A. S., & Persaud, B. N. (2010). Development of planning level transportation safety tools using Geographically Weighted Poisson Regression. *Accident Analysis and Prevention*, 42(2), 676–688. <https://doi.org/10.1016/j.aap.2009.10.016>
- Hage, R.-M., Hage, I., Ghnatios, C., & Hamade, R. (2018). Statistically validated and optimized tabu search estimation of cutting tool life in turning. *ASME International Mechanical Engineering Congress and Exposition, Proceedings (IMECE)*, 2, 1–6. <https://doi.org/10.1115/IMECE2018-86232>
- Hamidi, S., & Ewing, R. (2014). A longitudinal study of changes in urban sprawl between 2000 and 2010 in the United States. *Landscape and Urban Planning*, 128, 72–82. <https://doi.org/10.1016/j.landurbplan.2014.04.021>
- Hamidi, S., & Ewing, R. (2015). Is Sprawl Affordable for Americans? *Transportation Research Record: Journal of the Transportation Research Board*, 2500(2500), 75–79. <https://doi.org/10.3141/2500-09>
- Hamidi, S., Ewing, R., Preuss, I., & Dodds, A. (2015). Measuring Sprawl and Its Impacts. *Journal of Planning Education and Research*, 35(1), 35–50. <https://doi.org/10.1177/0739456X14565247>

- Hamidi, S., Sabouri, S., & Ewing, R. (2020). Does Density Aggravate the COVID-19 Pandemic?: Early Findings and Lessons for Planners. *Journal of the American Planning Association*, 86(4), 495–509. <https://doi.org/10.1080/01944363.2020.1777891>
- Hamidi, S., & Zandiatashbar, A. (2021). Compact development and adherence to stay-at-home order during the COVID-19 pandemic: A longitudinal investigation in the United States. *Landscape and Urban Planning*, 205(May 2020), 103952. <https://doi.org/10.1016/j.landurbplan.2020.103952>
- Han, C., Huang, H., Lee, J., & Wang, J. (2018). Investigating varying effect of road-level factors on crash frequency across regions: A Bayesian hierarchical random parameter modeling approach. *Analytic Methods in Accident Research*, 20(June), 81–91. <https://doi.org/10.1016/j.amar.2018.10.002>
- Hart, T., & Zandbergen, P. (2014). Kernel density estimation and hotspot mapping. *Policing: An International Journal of Police Strategies & Management*, 37(2), 305–323. <https://doi.org/10.1108/PIJPSM-04-2013-0039>
- Highway Traffic Safety Administration. (2019). *2017 Data: Rural/Urban Comparison*. www.census.gov/programs-surveys/geography/guidance/
- Highway Traffic Safety Administration, N., & Department of Transportation, U. (2019). *2017 Data: Rural/Urban Comparison*. www.census.gov/programs-surveys/geography/guidance/
- Holcombe, R. G., & Williams, D. W. (2010). Urban Sprawl and Transportation Externalities. *Review of Regional Studies*, 40(3), 257–272.
- Hong, W. C., Pai, P. F., Yang, S. L., & Theng, R. (2006). Highway traffic forecasting by support vector regression model with tabu search algorithms. *IEEE International Conference on Neural Networks - Conference Proceedings*, 1617–1621. <https://doi.org/10.1109/ijcnn.2006.246627>
- Huang, H., Abdel-Aty, M. A., & Darwiche, A. L. (2010). County-Level Crash Risk Analysis in Florida: Bayesian Spatial Modeling. *Transportation Research Record: Journal of the Transportation Research Board*, 2148(1), 27–37. <https://doi.org/10.3141/2148-04>

- Huang, H., Song, B., Xu, P., Zeng, Q., Lee, J., & Abdel-Aty, M. (2016). Macro and micro models for zonal crash prediction with application in hot zones identification. *Journal of Transport Geography*, 54, 248–256. <https://doi.org/10.1016/j.jtrangeo.2016.06.012>
- Huang, H., Zhou, H., Wang, J., Chang, F., & Ma, M. (2017). A multivariate spatial model of crash frequency by transportation modes for urban intersections. *Analytic Methods in Accident Research*, 14, 10–21. <https://doi.org/10.1016/j.amar.2017.01.001>
- Huang, Y., Wang, X., & Patton, D. (2018). Examining spatial relationships between crashes and the built environment: A geographically weighted regression approach. *Journal of Transport Geography*, 69(October 2017), 221–233. <https://doi.org/10.1016/j.jtrangeo.2018.04.027>
- IBM Corp. (2019). *IBM SPSS* (26.0). IBM Corp.
- Indicators - Eurostat. (n.d.). Retrieved October 9, 2020, from <https://ec.europa.eu/eurostat/web/sdi/indicators>
- ITF. (2018). *Road Safety Annual Report 2018*. <https://www.itf-oecd.org/road-safety-annual-report-2018>
- Jang, K., Kim, K., Kim, H. Y., & Kang, S. (2017). A multi-level analysis of the relationship between urban built environment and severe injury traffic crashes. *KSCCE Journal of Civil Engineering*, 22, 1427–1435. <https://doi.org/10.1007/s12205-017-1251-0>
- Jenks, M., Burton, E., & Williams, K. (2005). Compact Cities and Sustainability: An Introduction. In E. Burton, M. Jenks, & K. Williams (Eds.), *The Compact City: A Sustainable Urban Form?* (pp. 2–6). Routledge. <https://doi.org/10.4324/9780203362372>
- Jermprapai, K., & Srinivasan, S. (2014). Planning-Level Model for Assessing Pedestrian Safety. *Transportation Research Record: Journal of the Transportation Research Board*, 2464(1), 109–117. <https://doi.org/10.3141/2464-14>
- Jiang, X., Abdel-Aty, M., Hu, J., & Lee, J. (2016). Investigating macro-level hotzone

- identification and variable importance using big data: A random forest models approach. *Neurocomputing*, 181, 53–63. <https://doi.org/10.1016/j.neucom.2015.08.097>
- Jovanis, P. P., & Chang, H. (1986). Modeling the Relationships of Accidents to Miles Traveled. *Transportation Research Record*, 1068, 42–51.
- Kahn, M. E. (2001). Does sprawl reduce the black/white housing consumption gap? *Housing Policy Debate*, 12(4), 681–717. <https://doi.org/10.1080/10511482.2001.9521399>
- Kahn, M. E. (2008). Green Cities: Urban Growth and the Environment. *Global Environmental Politics*, 8(1), 147–148. <https://doi.org/10.1162/glep.2008.8.1.147>
- Kaygisiz, Ö., Senbil, M., & Yildiz, A. (2017). Influence of urban built environment on traffic accidents: The case of Eskisehir (Turkey). *Case Studies on Transport Policy*, 5(2), 306–313. <https://doi.org/10.1016/j.cstp.2017.02.002>
- Khan, M. A. (1995). Sustainable development: The key concepts, issues and implications. Keynote paper given at the international sustainable development research conference, 27–29 march 1995, Manchester, UK. *Sustainable Development*, 3(2), 63–69. <https://doi.org/10.1002/sd.3460030203>
- Khattak, A. J., & Rodriguez, D. (2005). Travel behavior in neo-traditional neighborhood developments: A case study in USA. *Transportation Research Part A: Policy and Practice*, 39(6), 481–500. <https://doi.org/10.1016/j.tra.2005.02.009>
- Khattak, A. J., Wang, X., & Zhang, H. (2010). Spatial Analysis and Modeling of Traffic Incidents for Proactive Incident Management and Strategic Planning. *Transportation Research Record: Journal of the Transportation Research Board*, 2178(1), 128–137. <https://doi.org/10.3141/2178-14>
- Khondakar, B., Sayed, T., & Lovegrove, G. (2010). Transferability of Community-Based Collision Prediction Models for Use in Road Safety Planning Applications. *Journal of Transportation Engineering*, 136(10), 871–880. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000153](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000153)

- Kim, H., Yang, J., & Lee, K.-D. (2009). Vehicle routing in reverse logistics for recycling end-of-life consumer electronic goods in South Korea. *Transportation Research Part D: Transport and Environment*, 14(5), 291–299. <https://doi.org/10.1016/j.trd.2009.03.001>
- Kim, J., & Brownstone, D. (2013). The impact of residential density on vehicle usage and fuel consumption: Evidence from national samples. *Energy Economics*, 40, 196–206. <https://doi.org/10.1016/j.eneco.2013.06.012>
- Kim, J. K., Ulfarsson, G. F., Shankar, V. N., & Mannering, F. L. (2010). A note on modeling pedestrian-injury severity in motor-vehicle crashes with the mixed logit model. *Accident Analysis and Prevention*, 42(6), 1751–1758. <https://doi.org/10.1016/j.aap.2010.04.016>
- Kim, K., Made Brunner, I., Yamashita, E. Y., Brunner, I., & Yamashita, E. Y. (2006). Influence of Land Use, Population, Employment, and Economic Activity on Accidents. *Transportation Research Record: Journal of the Transportation Research Board*, 1953(1), 56–64. <https://doi.org/10.3141/1953-07>
- Kim, K., Pant, P., & Yamashita, E. (2010). Accidents and Accessibility: Measuring Influences of Demographic and Land Use Variables in Honolulu, Hawaii. *Transportation Research Record: Journal of the Transportation Research Board*, 2147(1), 9–17. <https://doi.org/10.3141/2147-02>
- Kim, K., & Yamashita, E. (2002). Motor Vehicle Crashes and Land Use: Empirical Analysis from Hawaii. *Transportation Research Record: Journal of the Transportation Research Board*, 1784(1), 73–79. <https://doi.org/10.3141/1784-10>
- Kitamura, R., Mokhtarian, P. L., & Laidet, L. (1997). A micro-analysis of land use and travel in five neighborhoods in the San Francisco Bay Area. *Transportation*, 24, 125–158. <https://doi.org/10.1023/A:1017959825565>
- Kloog, I., Haim, A., & Portnov, B. A. (2009). Using kernel density function as an urban analysis tool: Investigating the association between nightlight exposure and the incidence of breast cancer in Haifa, Israel. *Computers, Environment and Urban Systems*, 33(1), 55–63. <https://doi.org/10.1016/j.compenvurbsys.2008.09.006>

- Ladrón de Guevara, F., Washington, S. P., & Oh, J. (2004). Forecasting Crashes at the Planning Level: Simultaneous Negative Binomial Crash Model Applied in Tucson, Arizona. *Transportation Research Record: Journal of the Transportation Research Board*, 1897(1), 191–199. <https://doi.org/10.3141/1897-25>
- LaScala, E. A., Gerber, D., & Gruenewald, P. J. (2000). Demographic and environmental correlates of pedestrian injury collisions: a spatial analysis. *Accident Analysis & Prevention*, 32(5), 651–658. [https://doi.org/10.1016/S0001-4575\(99\)00100-1](https://doi.org/10.1016/S0001-4575(99)00100-1)
- LaScala, E. A., Gruenewald, P. J., & Johnson, F. W. (2004). An ecological study of the locations of schools and child pedestrian injury collisions. *Accident Analysis and Prevention*, 36(4), 569–576. [https://doi.org/10.1016/S0001-4575\(03\)00063-0](https://doi.org/10.1016/S0001-4575(03)00063-0)
- LaScala, E. A., Johnson, F. W., & Gruenewald, P. J. (2001). Neighborhood characteristics of alcohol-related pedestrian injury collisions; a geostatistical analysis. *Prevention Science*, 2(2), 123–134. <https://doi.org/10.1023/A:1011547831475>
- Lee, J., & Abdel-Aty, M. (2018). Macro-level analysis of bicycle safety: Focusing on the characteristics of both crash location and residence. *International Journal of Sustainable Transportation*, 12(8), 553–560. <https://doi.org/10.1080/15568318.2017.1407973>
- Lee, J., Abdel-Aty, M., & Cai, Q. (2017). Intersection crash prediction modeling with macro-level data from various geographic units. *Accident Analysis and Prevention*, 102, 213–226. <https://doi.org/10.1016/j.aap.2017.03.009>
- Lee, J., Abdel-Aty, M., Cai, Q., Wang, L., & Huang, H. (2018). Integrated Modeling Approach for Non-Motorized Mode Trips and Fatal Crashes in the Framework of Transportation Safety Planning. *Transportation Research Record*, 2672(32), 49–60. <https://doi.org/10.1177/0361198118772704>
- Lee, J., Abdel-Aty, M., Choi, K., & Huang, H. (2015). Multi-level hot zone identification for pedestrian safety. *Accident Analysis & Prevention*, 76, 64–73. <https://doi.org/10.1016/j.aap.2015.01.006>

- Lee, J., Abdel-Aty, M., de Blasiis, M. R., Wang, X., & Mattei, I. (2019). International transferability of macro-level safety performance functions: a case study of the United States and Italy. *Transportation Safety and Environment*, 1(1), 68–78. <https://doi.org/10.1093/transp/tdz001>
- Lee, J., Abdel-Aty, M., Huang, H., & Cai, Q. (2019). Transportation Safety Planning Approach for Pedestrians: An Integrated Framework of Modeling Walking Duration and Pedestrian Fatalities. *Transportation Research Record*, 2673(4), 898–906. <https://doi.org/10.1177/0361198119837962>
- Lee, J., Abdel-Aty, M., & Jiang, X. (2014). Development of zone system for macro-level traffic safety analysis. *Journal of Transport Geography*, 38, 13–21. <https://doi.org/10.1016/j.jtrangeo.2014.04.018>
- Lee, J., Abdel-Aty, M., & Jiang, X. (2015). Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level. *Accident Analysis & Prevention*, 78, 146–154. <https://doi.org/10.1016/j.aap.2015.03.003>
- Li, Zhenning, Chen, X., Ci, Y., Chen, C., & Zhang, G. (2019). A hierarchical Bayesian spatiotemporal random parameters approach for alcohol/drug impaired-driving crash frequency analysis. *Analytic Methods in Accident Research*, 21, 44–61. <https://doi.org/10.1016/j.amar.2019.01.002>
- Li, Zhibin, Wang, W., Liu, P., Bigham, J. M., & Ragland, D. R. (2013). Using Geographically Weighted Poisson Regression for county-level crash modeling in California. *Safety Science*, 58, 89–97. <https://doi.org/10.1016/j.ssci.2013.04.005>
- Liu, C., & Sharma, A. (2017). Exploring spatio-temporal effects in traffic crash trend analysis. *Analytic Methods in Accident Research*, 16, 104–116. <https://doi.org/10.1016/j.amar.2017.09.002>
- Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5), 291–305. <https://doi.org/10.1016/j.tra.2010.02.001>
- Loukaitou-Sideris, A., Liggett, R., & Sung, H. G. (2007). Death on the crosswalk: A study of pedestrian-automobile collisions in Los Angeles. *Journal of Planning*

Education and Research, 26(3), 338–351.
<https://doi.org/10.1177/0739456X06297008>

Lovegrove, G., Lim, C., & Sayed, T. (2010). Community-Based, Macrolevel Collision Prediction Model Use with a Regional Transportation Plan. *Journal of Transportation Engineering*, 136(2), 120–128.
[https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000081](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000081)

Lovegrove, G. R., & Sayed, T. (2006). Macro-level collision prediction models for evaluating neighbourhood traffic safety. *Canadian Journal of Civil Engineering*, 33(5), 609–621. <https://doi.org/10.1139/106-013>

Lovegrove, G. R., & Sayed, T. (2007). Macrolevel collision prediction models to enhance traditional reactive road safety improvement programs. *Transportation Research Record*, 2019, 65–73. <https://doi.org/10.3141/2019-09>

Lovegrove, G., & Sayed, T. (2006). Using Macrolevel Collision Prediction Models in Road Safety Planning Applications. *Transportation Research Record: Journal of the Transportation Research Board*, 1950(1950), 73–82.
<https://doi.org/10.3141/1950-09>

Ludlow, D. (2006). *Urban sprawl in Europe: The ignored challenge*. <https://uwe-repository.worktribe.com/output/1042555>

Ma, L., Yan, X., Wei, C., & Wang, J. (2016). Modeling the equivalent property damage only crash rate for road segments using the hurdle regression framework. *Analytic Methods in Accident Research*, 11, 48–61.
<https://doi.org/10.1016/j.amar.2016.07.001>

Marshall, W. E., & Garrick, N. W. (2011). Does street network design affect traffic safety? *Accident Analysis & Prevention*, 43(3), 769–781.
<https://doi.org/10.1016/j.aap.2010.10.024>

Martin, J. L. (2002). Relationship between crash rate and hourly traffic flow on interurban motorways. *Accident Analysis and Prevention*, 34(5), 619–629.
[https://doi.org/10.1016/S0001-4575\(01\)00061-6](https://doi.org/10.1016/S0001-4575(01)00061-6)

Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50(4), 370–396. <https://doi.org/10.1037/h0054346>

- Merlin, L. A., Guerra, E., & Dumbaugh, E. (2020). Crash risk, crash exposure, and the built environment: A conceptual review. *Accident Analysis & Prevention*, 134(May 2019), 105244. <https://doi.org/10.1016/j.aap.2019.07.020>
- Miaou, S.-P. (1994). The relationship between truck accidents and geometric design of road sections: Poisson versus negative binomial regressions. *Accident Analysis & Prevention*, 26(4), 471–482. [https://doi.org/10.1016/0001-4575\(94\)90038-8](https://doi.org/10.1016/0001-4575(94)90038-8)
- Milestones: 1969–1976 - Office of the Historian*. (n.d.). Retrieved October 5, 2020, from <https://history.state.gov/milestones/1969-1976/oil-embargo>
- Miller, A. J. (1984). Selection of Subsets of Regression Variables. *Journal of the Royal Statistical Society. Series A (General)*, 147(3), 389. <https://doi.org/10.2307/2981576>
- Milton, J., & Mannering, F. (1998). The relationship among highway geometrics, traffic-related elements and motor-vehicle accident frequencies. *Transportation*, 25(4), 395–413. <https://doi.org/10.1023/a:1005095725001>
- Miranda-Moreno, L. F., Morency, P., & El-Geneidy, A. M. (2011). The link between built environment, pedestrian activity and pedestrian-vehicle collision occurrence at signalized intersections. *Accident Analysis and Prevention*, 43(5), 1624–1634. <https://doi.org/10.1016/j.aap.2011.02.005>
- Moeinaddini, M., Asadi-Shekari, Z., & Zaly Shah, M. (2014). The relationship between urban street networks and the number of transport fatalities at the city level. *Safety Science*, 62, 114–120. <https://doi.org/10.1016/j.ssci.2013.08.015>
- Morency, P., Gauvin, L., Plante, C., Fournier, M., & Morency, C. (2012). Neighborhood Social Inequalities in Road Traffic Injuries: The Influence of Traffic Volume and Road Design. *American Journal of Public Health*, 102(6), 1112–1119. <https://doi.org/10.2105/AJPH.2011.300528>
- Moudon, A. V., Lin, L., Hurvitz, P., & Reeves, P. (2008). Risk of Pedestrian Collision Occurrence. *Transportation Research Record: Journal of the Transportation Research Board*, 2073(1), 25–38. <https://doi.org/10.3141/2073-04>

- Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3), 691–692.
<https://doi.org/10.1093/biomet/78.3.691>
- Najaf, P., Thill, J.-C., Zhang, W., & Fields, M. G. (2018). City-level urban form and traffic safety: A structural equation modeling analysis of direct and indirect effects. *Journal of Transport Geography*, 69(May), 257–270.
<https://doi.org/10.1016/j.jtrangeo.2018.05.003>
- Narayanamoorthy, S., Paleti, R., & Bhat, C. R. (2013). On accommodating spatial dependence in bicycle and pedestrian injury counts by severity level. *Transportation Research Part B: Methodological*, 55, 245–264.
<https://doi.org/10.1016/j.trb.2013.07.004>
- Nashad, T., Yasmin, S., Eluru, N., Lee, J., & Abdel-Aty, M. A. (2016). Joint Modeling of Pedestrian and Bicycle Crashes: Copula-Based Approach. *Transportation Research Record*, 2601(1), 119–127.
<https://doi.org/10.3141/2601-14>
- Nkeki, F. N., & Asikhia, M. O. (2019). Geographically weighted logistic regression approach to explore the spatial variability in travel behaviour and built environment interactions: Accounting simultaneously for demographic and socioeconomic characteristics. *Applied Geography*, 108(May), 47–63.
<https://doi.org/10.1016/j.apgeog.2019.05.008>
- Noland, R. B., & Oh, L. (2004). The effect of infrastructure and demographic change on traffic-related fatalities and crashes: A case study of Illinois county-level data. *Accident Analysis and Prevention*, 36(4), 525–532.
[https://doi.org/10.1016/S0001-4575\(03\)00058-7](https://doi.org/10.1016/S0001-4575(03)00058-7)
- Noland, R. B., & Quddus, M. A. (2004a). A spatially disaggregate analysis of road casualties in England. *Accident Analysis and Prevention*, 36(6), 973–984.
<https://doi.org/10.1016/j.aap.2003.11.001>
- Noland, R. B., & Quddus, M. A. (2004b). Analysis of Pedestrian and Bicycle Casualties with Regional Panel Data. *Transportation Research Record: Journal of the Transportation Research Board*, 1897(1), 28–33.
<https://doi.org/10.3141/1897-04>

- Noland, R. B., & Quddus, M. A. (2005). Congestion and safety: A spatial analysis of London. *Transportation Research Part A: Policy and Practice*, 39(7–9), 737–754. <https://doi.org/10.1016/j.tra.2005.02.022>
- Nowrouzian, R., & Srinivasan, S. (2013). Modeling the Effect of Land Use on Person Miles Traveled by Using Geographically Weighted Regression. *Transportation Research Record: Journal of the Transportation Research Board*, 2397(1), 108–116. <https://doi.org/10.3141/2397-13>
- Oh, J., Washington, S., & Lee, D. (2010). Property Damage Crash Equivalency Factors to Solve Crash Frequency–Severity Dilemma: Case Study on South Korean Rural Roads. *Transportation Research Record: Journal of the Transportation Research Board*, 2148(1), 83–92. <https://doi.org/10.3141/2148-10>
- Örkcü, H. H. (2013). Subset selection in multiple linear regression models: A hybrid of genetic and simulated annealing algorithms. In *Applied Mathematics and Computation* (Vol. 219, Issue 23, pp. 11018–11028). Elsevier Inc. <https://doi.org/10.1016/j.amc.2013.05.016>
- Osama, A., & Sayed, T. (2017). Macro-spatial approach for evaluating the impact of socio-economics, land use, built environment, and road facility on pedestrian safety. *Canadian Journal of Civil Engineering*, 44(12), 1036–1044. <https://doi.org/10.1139/cjce-2017-0145>
- Ouyang, Y., & Bejleri, I. (2014). Geographic Information System–Based Community-Level Method to Evaluate the Influence of Built Environment on Traffic Crashes. *Transportation Research Record: Journal of the Transportation Research Board*, 2432(1), 124–132. <https://doi.org/10.3141/2432-15>
- Pacheco, J., Casado, S., & Núñez, L. (2009). A variable selection method based on Tabu search for logistic regression models. *European Journal of Operational Research*, 199(2), 506–511. <https://doi.org/10.1016/j.ejor.2008.10.007>
- Park, K., Ewing, R., Scheer, B. C., & Tian, G. (2018). The impacts of built environment characteristics of rail station areas on household travel behavior. *Cities*, 74(November 2017), 277–283. <https://doi.org/10.1016/j.cities.2017.12.015>

- Pendall, R. (1999). Do land-use controls cause sprawl? *Environment and Planning B: Planning and Design*, 26(4), 555–571. <https://doi.org/10.1068/b260555>
- Pirdavani, A., Bellemans, T., Brijs, T., & Wets, G. (2014). Application of geographically weighted regression technique in spatial analysis of fatal and injury crashes. *Journal of Transportation Engineering*, 140(8), 1–10. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000680](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000680)
- Pirdavani, A., Brijs, T., Bellemans, T., Kochan, B., & Wets, G. (2013). Evaluating the road safety effects of a fuel cost increase measure by means of zonal crash prediction modeling. *Accident Analysis and Prevention*, 50, 186–195. <https://doi.org/10.1016/j.aap.2012.04.008>
- Pseudo R-Squared Measures*. (n.d.). Retrieved January 10, 2021, from https://www.ibm.com/support/knowledgecenter/SSLVMB_24.0.0/spss/tutorial/s/plum_germcr_rsquare.html
- Pulugurtha, S. S., Duddu, V. R., & Kotagiri, Y. (2013). Traffic analysis zone level crash estimation models based on land use characteristics. *Accident Analysis & Prevention*, 50, 678–687. <https://doi.org/10.1016/j.aap.2012.06.016>
- Quistberg, D. A., Howard, E. J., Ebel, B. E., Moudon, A. V., Saelens, B. E., Hurvitz, P. M., Curtin, J. E., & Rivara, F. P. (2015). Multilevel models for evaluating the risk of pedestrian–motor vehicle collisions at intersections and mid-blocks. *Accident Analysis & Prevention*, 84, 99–111. <https://doi.org/10.1016/j.aap.2015.08.013>
- Quistberg, D. A., Howard, E. J., Hurvitz, P. M., Moudon, A. V., Ebel, B. E., Rivara, F. P., & Saelens, B. E. (2017). The Relationship Between Objectively Measured Walking and Risk of Pedestrian–Motor Vehicle Collision. *American Journal of Epidemiology*, 185(9), 810–821. <https://doi.org/10.1093/aje/kwx020>
- R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. <https://www.r-project.org/>
- Renne, J. L., & Tolford, T. (2018). A planning tool for evaluating vehicles miles travelled and traffic safety forecasts of growth management scenarios: A case study of Baton Rouge and New Orleans. *Transportation Research Part D: Transport and Environment*, 59(February), 237–245.

<https://doi.org/10.1016/j.trd.2018.01.001>

- Rhee, K. A., Kim, J. K., Lee, Y. I., & Ulfarsson, G. F. (2016). Spatial regression analysis of traffic crashes in Seoul. *Accident Analysis and Prevention*, 91, 190–199. <https://doi.org/10.1016/j.aap.2016.02.023>
- Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3), 507–554. <https://doi.org/10.1111/j.1467-9876.2005.00510.x>
- Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., & Firth, D. (2020). *MASS: Support Functions and Datasets for Venables and Ripley's MASS* (7.3-53). R. <http://www.stats.ox.ac.uk/pub/MASS4/>
- Rodríguez, D. A., & Joo, J. (2004). The relationship between non-motorized mode choice and the local physical environment. *Transportation Research Part D: Transport and Environment*, 9(2), 151–173. <https://doi.org/10.1016/j.trd.2003.11.001>
- Rothman, L., Macarthur, C., To, T., Buliung, R., & Howard, A. (2014). Motor Vehicle-Pedestrian Collisions and Walking to School: The Role of the Built Environment. *PEDIATRICS*, 133(5), 776–784. <https://doi.org/10.1542/peds.2013-2317>
- Saha, D., Dumbaugh, E., & Merlin, L. A. (2020). A conceptual framework to understand the role of built environment on traffic safety. *Journal of Safety Research*, 75(xxxx), 41–50. <https://doi.org/10.1016/j.jsr.2020.07.004>
- Salon, D. (2015). Heterogeneity in the relationship between the built environment and driving: Focus on neighborhood type and travel purpose. *Research in Transportation Economics*, 52, 34–45. <https://doi.org/10.1016/j.retrec.2015.10.008>
- Santamariña-Rubio, E., Pérez, K., Olabarria, M., & Novoa, A. M. (2014). Gender differences in road traffic injury rate using time travelled as a measure of exposure. *Accident Analysis and Prevention*, 65, 1–7. <https://doi.org/10.1016/j.aap.2013.11.015>

- Sargut, F. Z., Altuntaş, C., & Tulazoğlu, D. C. (2017). Multi-objective integrated acyclic crew rostering and vehicle assignment problem in public bus transportation. *OR Spectrum*, 39(4), 1071–1096. <https://doi.org/10.1007/s00291-017-0485-z>
- Sener, I. N., Lee, K., Hudson, J. G., Martin, M., & Dai, B. (2019). The challenge of safe and active transportation: Macrolevel examination of pedestrian and bicycle crashes in the Austin District. *Journal of Transportation Safety & Security*, 1–27. <https://doi.org/10.1080/19439962.2019.1645778>
- Şengül, T. (2012). Türkiye'nin Kentleşme Deneyiminin Dönemlenmesi. In F. Alpkaya & B. Duru (Eds.), *1920'den Günümüze Türkiye'de Toplumsal Yapı ve Değişim* (pp. 405–451). Phoenix Yayınevi.
- Šenk, P., & Ambros, J. (2011). Estimation of Accident Frequency at Newly-built Roundabouts in the Czech Republic. *Transactions on Transport Sciences*, 4(4), 199–206. <https://doi.org/10.2478/v10158-011-0018-4>
- Shay, E., Fan, Y., Rodríguez, D. A., & Khattak, A. J. (2006). Drive or Walk?: Utilitarian Trips Within a Neotraditional Neighborhood. *Transportation Research Record: Journal of the Transportation Research Board*, 1985(1985), 154–161. <https://doi.org/10.3141/1985-17>
- Shay, E., & Khattak, A. J. (2005). Automobile Ownership and Use in Neotraditional and Conventional Neighborhoods. *Transportation Research Record: Journal of the Transportation Research Board*, 1902(1902), 18–25. <https://doi.org/10.3141/1902-03>
- Siddiqui, C., & Abdel-Aty, M. (2012). Nature of modeling boundary pedestrian crashes at zones. *Transportation Research Record*, 2299, 31–40. <https://doi.org/10.3141/2299-04>
- Siddiqui, C., Abdel-Aty, M., & Choi, K. (2012). Macroscopic spatial analysis of pedestrian and bicycle crashes. *Accident Analysis & Prevention*, 45, 382–391. <https://doi.org/10.1016/j.aap.2011.08.003>
- Siswadi, & Quesenberry, C. P. (1982). Selecting among weibull, lognormal and gamma distributions using complete and censored samples. *Naval Research Logistics Quarterly*, 29(4), 557–569. <https://doi.org/10.1002/nav.3800290404>

- Šliupas, T. (2009). The impact of road parameters and the surrounding area on traffic accidents. *TRANSPORT*, 24(1), 42–47. <https://doi.org/10.3846/1648-4142.2009.24.42-47>
- Smart Growth EPA. (n.d.). *Smart Location Mapping*. Retrieved November 11, 2020, from <https://www.epa.gov/smartgrowth/smart-location-mapping#SLD>
- Sohn, J., Choi, S., Lewis, R., & Knaap, G. (2012). Characterising urban sprawl on a local scale with accessibility measures. *Geographical Journal*, 178(3), 230–241. <https://doi.org/10.1111/j.1475-4959.2012.00468.x>
- Song, Y., & Zenou, Y. (2006). Property tax and urban sprawl: Theory and implications for US cities. *Journal of Urban Economics*, 60, 519–534. <https://doi.org/10.1016/j.jue.2006.05.001>
- Statistical Distributions*. (n.d.). Retrieved November 1, 2020, from http://people.stern.nyu.edu/adamodar/New_Home_Page/StatFile/statdistns.htm
- Steenberghen, T., Aerts, K., & Thomas, I. (2010). Spatial clustering of events on a network. *Journal of Transport Geography*, 18(3), 411–418. <https://doi.org/10.1016/j.jtrangeo.2009.08.005>
- Stevens, M. R. (2017). Response to Commentaries on Does Compact Development Make People Drive Less? *Journal of the American Planning Association*, 83(2), 151–158. <https://doi.org/10.1080/01944363.2016.1240044>
- Stimpson, J. P., Wilson, F. A., Araz, O. M., & Pagan, J. A. (2014). Share of Mass Transit Miles Traveled and Reduced Motor Vehicle Fatalities in Major Cities of the United States. *Journal of Urban Health*, 91(6), 1136–1143. <https://doi.org/10.1007/s11524-014-9880-9>
- Sustainable Mobility for All. (2017). Sustainable Mobility for All. In *Global Mobility Report 2017*. <https://sum4all.org/publications/global-mobility-report-2017>
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using Multivariate Statistics*. Pearson. <https://lccn.loc.gov/2017040173>

- Tasic, I., Elvik, R., & Brewer, S. (2017). Exploring the safety in numbers effect for vulnerable road users on a macroscopic scale. *Accident Analysis and Prevention*, 109(July), 36–46. <https://doi.org/10.1016/j.aap.2017.07.029>
- Tekeli, İ. (1998). Türkiye’de Cumhuriyet Döneminde Kentsel Gelişme ve Kent Planlaması. In Y. Sey (Ed.), *75. Yılda Değişen Kent ve Mimarlık* (pp. 1–25). Tarih Vakfı Yayınları.
- Thao, V. T., & Ohnmacht, T. (2019). The impact of the built environment on travel behavior: The Swiss experience based on two National Travel Surveys. *Research in Transportation Business and Management*, July, 100386. <https://doi.org/10.1016/j.rtbm.2019.100386>
- Traffic Department. (n.d.). *Traffic Statistics*. Retrieved January 9, 2021, from <http://trafik.gov.tr/>
- Trowbridge, M. J., & McDonald, N. C. (2008). Urban Sprawl and Miles Driven Daily by Teenagers in the United States. *American Journal of Preventive Medicine*, 34(3), 202–206. <https://doi.org/10.1016/j.amepre.2007.11.013>
- Turkish Statistics Institute. (n.d.). *Adrese Dayalı Nüfus Kayıt Sistemi*. Retrieved January 9, 2021, from <https://biruni.tuik.gov.tr/medas/?kn=95&locale=tr>
- U.S. Census Bureau. (2011). *2010 US Census Block Groups in Florida (with Selected Fields from 2010 Redistricting Summary File)*. https://www.fgdl.org/metadata/metadata_archive/fgdc_html/cenblkgrp2010_mar11.fgdc.htm
- Ukkusuri, S., Hasan, S., & Aziz, H. M. A. A. (2011). Random Parameter Model Used to Explain Effects of Built-Environment Characteristics on Pedestrian Crash Frequency. *Transportation Research Record: Journal of the Transportation Research Board*, 2237(1), 98–106. <https://doi.org/10.3141/2237-11>
- Ukkusuri, S., Miranda-Moreno, L. F., Ramadurai, G., & Isa-Tavarez, J. (2012). The role of built environment on pedestrian crash frequency. *Safety Science*, 50(4), 1141–1151. <https://doi.org/10.1016/j.ssci.2011.09.012>
- Ulak, M. B. (2018). *Geographical Information Systems-based Spatial and Statistical Investigation to Enhance Transportation Safety for the Aging Population*.

Florida State University.

Ulak, M. B., Ozguven, E. E., Spainhour, L., & Vanli, O. A. (2017). Spatial investigation of aging-involved crashes: A GIS-based case study in Northwest Florida. *Journal of Transport Geography*, 58, 71–91. <https://doi.org/10.1016/j.jtrangeo.2016.11.011>

United Nations-Population Division. (2019). *World Urbanization Prospects*. <https://population.un.org/wup/DataQuery/>

USCB. (1994). *Geographic Areas Reference Manual*. <https://www.census.gov/programs-surveys/geography/guidance/geographic-areas-reference-manual.html>

USHUD. (1999). *The State of the Cities: 1999*. <https://www.huduser.gov/Publications/PDF/soc99.pdf>

Vance, C., & Hedel, R. (2007). The impact of urban form on automobile travel: Disentangling causation from correlation. *Transportation*, 34(5), 575–588. <https://doi.org/10.1007/s11116-007-9128-6>

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth edi). Springer. <https://www.stats.ox.ac.uk/pub/MASS4/>

Wang, C. (2010). The Relationship Between Traffic Congestion and Road Accidents : An Econometric Approach Using GIS. *Loughborough University's Institutional Repository*, January, 1–253. <https://doi.org/10.13140/RG.2.1.3180.0087>

Wang, J., & Huang, H. (2016). Road network safety evaluation using Bayesian hierarchical joint model. *Accident Analysis and Prevention*, 90, 152–158. <https://doi.org/10.1016/j.aap.2016.02.018>

Wang, L., Abdel-Aty, M., Lee, J., & Shi, Q. (2019). Analysis of real-time crash risk for expressway ramps using traffic, geometric, trip generation, and socio-demographic predictors. *Accident Analysis and Prevention*, 122(December 2016), 378–384. <https://doi.org/10.1016/j.aap.2017.06.003>

- Wang, X., Song, Y., Yu, R., & Schultz, G. G. (2014). Safety modeling of suburban arterials in Shanghai, China. *Accident Analysis and Prevention*, 70, 215–224. <https://doi.org/10.1016/j.aap.2014.04.005>
- Wang, Y., & Kockelman, K. M. (2013). A Poisson-lognormal conditional-autoregressive model for multivariate spatial analysis of pedestrian crash counts across neighborhoods. *Accident Analysis & Prevention*, 60, 71–84. <https://doi.org/10.1016/j.aap.2013.07.030>
- Washington, S., Haque, M. M., Oh, J., & Lee, D. (2014). Applying quantile regression for modeling equivalent property damage only crashes to identify accident blackspots. *Accident Analysis & Prevention*, 66, 136–146. <https://doi.org/10.1016/j.aap.2014.01.007>
- WCED. (1987). *Our Common Future*. <http://www.un-documents.net/our-common-future.pdf>
- Wei, F., & Lovegrove, G. (2013). An empirical tool to evaluate the safety of cyclists: Community based, macro-level collision prediction models using negative binomial regression. *Accident Analysis & Prevention*, 61, 129–137. <https://doi.org/10.1016/j.aap.2012.05.018>
- Wen, H., Sun, J., Zeng, Q., Zhang, X., & Yuan, Q. (2018). The Effects of Traffic Composition on Freeway Crash Frequency by Injury Severity: A Bayesian Multivariate Spatial Modeling Approach. *Journal of Advanced Transportation*, 2018, 1–7. <https://doi.org/10.1155/2018/6964828>
- WHO. (2018). *Global Status Report on Road Safety 2018* (Issue 10 April 2019). https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/
- Wier, M., Weintraub, J., Humphreys, E. H., Seto, E., & Bhatia, R. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning. *Accident Analysis & Prevention*, 41(1), 137–145. <https://doi.org/10.1016/j.aap.2008.10.001>
- Xie, K., Ozbay, K., & Yang, H. (2019). A multivariate spatial approach to model crash counts by injury severity. *Accident Analysis & Prevention*, 122(September 2018), 189–198. <https://doi.org/10.1016/j.aap.2018.10.009>

- Xie, K., Wang, X., Huang, H., & Chen, X. (2013). Corridor-level signalized intersection safety analysis in Shanghai, China using Bayesian hierarchical models. *Accident Analysis and Prevention*, 50, 25–33. <https://doi.org/10.1016/j.aap.2012.10.003>
- Xie, K., Wang, X., Ozbay, K., & Yang, H. (2014). Crash frequency modeling for signalized intersections in a high-density urban road network. *Analytic Methods in Accident Research*, 2, 39–51. <https://doi.org/10.1016/j.amar.2014.06.001>
- Xie, S. Q., Dong, N., Wong, S. C., Huang, H., & Xu, P. (2018). Bayesian approach to model pedestrian crashes at signalized intersections with measurement errors in exposure. *Accident Analysis and Prevention*, 121(August), 285–294. <https://doi.org/10.1016/j.aap.2018.09.030>
- Xie, Z., & Yan, J. (2013). Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. *Journal of Transport Geography*, 31, 64–71. <https://doi.org/10.1016/j.jtrangeo.2013.05.009>
- Xu, C., Li, H., Zhao, J., Chen, J., & Wang, W. (2017). Investigating the relationship between jobs-housing balance and traffic safety. *Accident Analysis and Prevention*, 107(December 2016), 126–136. <https://doi.org/10.1016/j.aap.2017.08.013>
- Xu, P., & Huang, H. (2015). Modeling crash spatial heterogeneity: Random parameter versus geographically weighting. *Accident Analysis and Prevention*, 75, 16–25. <https://doi.org/10.1016/j.aap.2014.10.020>
- Xu, P., Huang, H., & Dong, N. (2018). The modifiable areal unit problem in traffic safety: Basic issue, potential solutions and future research. *Journal of Traffic and Transportation Engineering (English Edition)*, 5(1), 73–82. <https://doi.org/10.1016/j.jtte.2015.09.010>
- Xu, P., Huang, H., Dong, N., & Abdel-Aty, M. (2014). Sensitivity analysis in the context of regional safety modeling: Identifying and assessing the modifiable areal unit problem. *Accident Analysis and Prevention*, 70, 110–120. <https://doi.org/10.1016/j.aap.2014.02.012>
- Xu, P., Huang, H., Dong, N., & Wong, S. C. (2017). Revisiting crash spatial

- heterogeneity: A Bayesian spatially varying coefficients approach. *Accident Analysis and Prevention*, 98, 330–337. <https://doi.org/10.1016/j.aap.2016.10.015>
- Yasmin, S., & Eluru, N. (2016). Latent segmentation based count models: Analysis of bicycle safety in Montreal and Toronto. *Accident Analysis & Prevention*, 95, 157–171. <https://doi.org/10.1016/j.aap.2016.07.015>
- Yeo, J., Park, S., & Jang, K. (2015). Effects of Urban Sprawl and Vehicle Miles Traveled on Traffic Fatalities. *Traffic Injury Prevention*, 16(4), 397–403. <https://doi.org/10.1080/15389588.2014.948616>
- Yu, C.-Y. (2014). Environmental supports for walking/biking and traffic safety: Income and ethnicity disparities. *Preventive Medicine*, 67, 12–16. <https://doi.org/10.1016/j.ypmed.2014.06.028>
- Yu, C.-Y., & Xu, M. (2018). Local Variations in the Impacts of Built Environments on Traffic Safety. *Journal of Planning Education and Research*, 38(3), 0739456X1769603. <https://doi.org/10.1177/0739456X17696035>
- Yu, C.-Y. Y. (2015). Built environmental designs in promoting pedestrian safety. *Sustainability (Switzerland)*, 7(7), 9444–9460. <https://doi.org/10.3390/su7079444>
- Yu, C.-Y., & Zhu, X. (2016). Planning for Safe Schools. *Journal of Planning Education and Research*, 36(4), 476–486. <https://doi.org/10.1177/0739456X15616460>
- Yu, L., Xie, B., & Chan, E. H. W. (2019). Exploring impacts of the built environment on transit travel: Distance, time and mode choice, for urban villages in Shenzhen, China. *Transportation Research Part E: Logistics and Transportation Review*, 132(October), 57–71. <https://doi.org/10.1016/j.tre.2019.11.004>
- Zegras, C. (2010). The built environment and motor vehicle ownership and use: Evidence from Santiago de Chile. *Urban Studies*, 47(8), 1793–1817. <https://doi.org/10.1177/0042098009356125>
- Zeng, Q., & Huang, H. (2014). Bayesian spatial joint modeling of traffic crashes on

an urban road network. *Accident Analysis and Prevention*, 67, 105–112.
<https://doi.org/10.1016/j.aap.2014.02.018>

Zhai, X., Huang, H., Xu, P., & Sze, N. N. (2019). The influence of zonal configurations on macro-level crash modeling. *Transportmetrica A: Transport Science*, 15(2), 417–434. <https://doi.org/10.1080/23249935.2018.1493550>

Zhao, P. (2010). Sustainable urban expansion and transportation in a growing megacity: Consequences of urban sprawl for mobility on the urban fringe of Beijing. *Habitat International*, 34(2), 236–243.
<https://doi.org/10.1016/j.habitatint.2009.09.008>

Zhao, P., Li, S., & Liu, D. (2020). Unequable spatial accessibility to hospitals in developing megacities: New evidence from Beijing. *Health and Place*, 65(August), 102406. <https://doi.org/10.1016/j.healthplace.2020.102406>

Zheng, Z. (2012). Empirical Analysis on Relationship between Traffic Conditions and Crash Occurrences. *Procedia - Social and Behavioral Sciences*, 43, 302–312. <https://doi.org/10.1016/j.sbspro.2012.04.103>

Ziakopoulos, A., & Yannis, G. (2020). A review of spatial approaches in road safety. *Accident Analysis & Prevention*, 135(September 2019), 105323.
<https://doi.org/10.1016/j.aap.2019.105323>

Zolnik, E. J. (2011). The effect of sprawl on private-vehicle commuting outcomes. *Environment and Planning A*, 43(8), 1875–1893.
<https://doi.org/10.1068/a42466>

APPENDICES

A. Single County Case Results

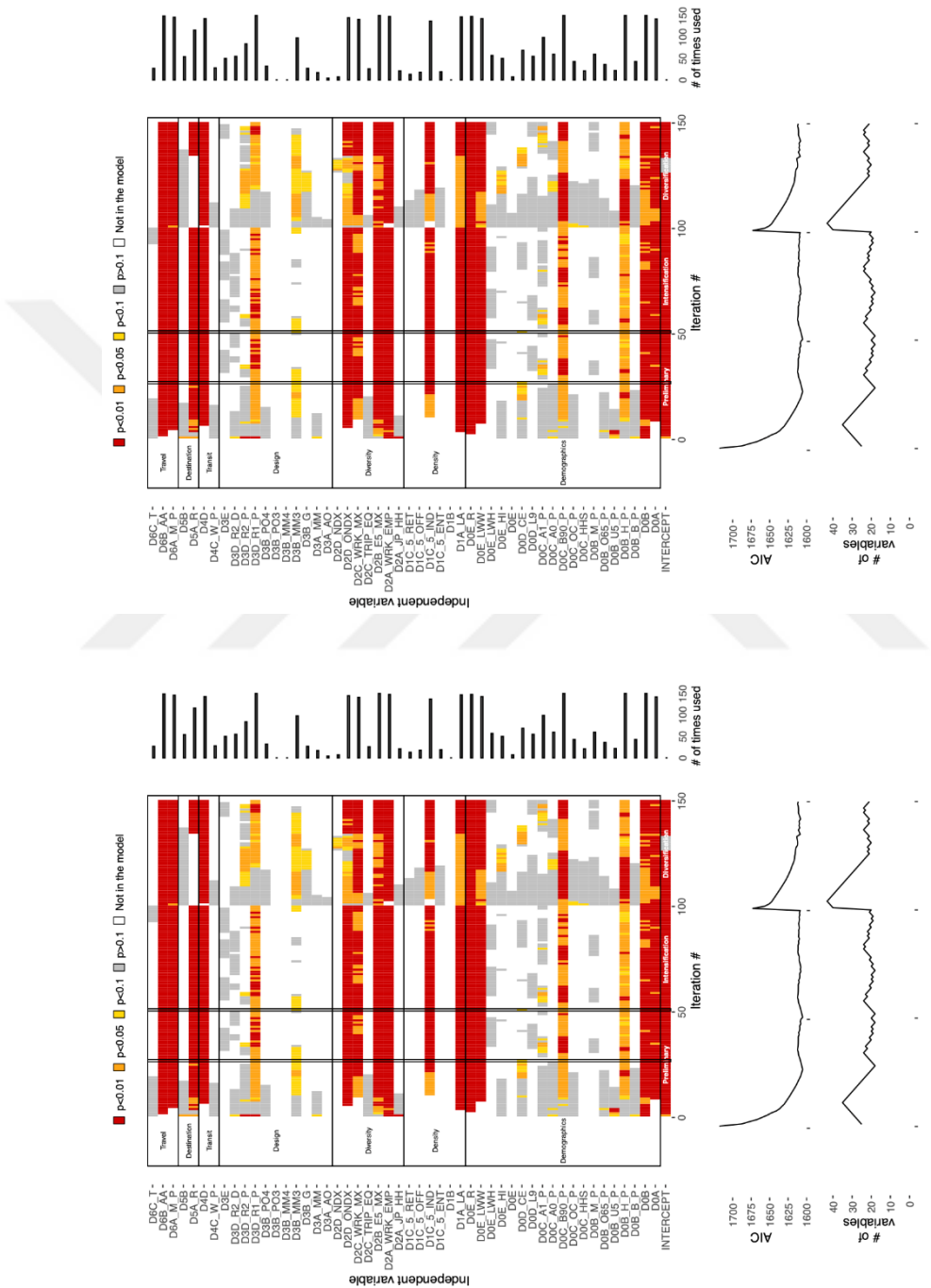


Figure A.1. FQ_S1 Negative Binomial

Figure A.2. FQ_S2 Negative Binomial

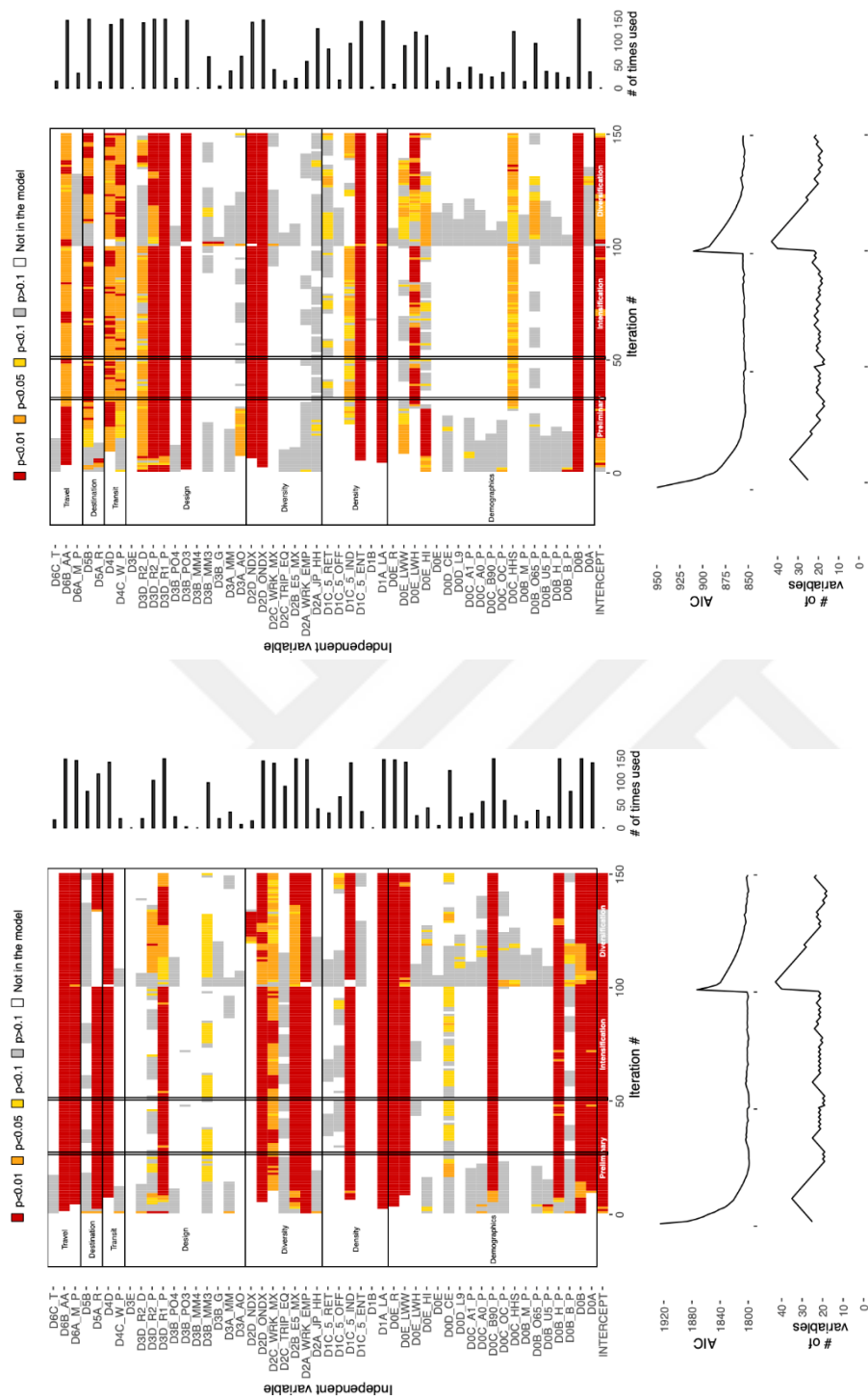


Figure A.3 FQ_S3 Negative Binomial

Figure A.4. FQ_ALL Negative Binomial

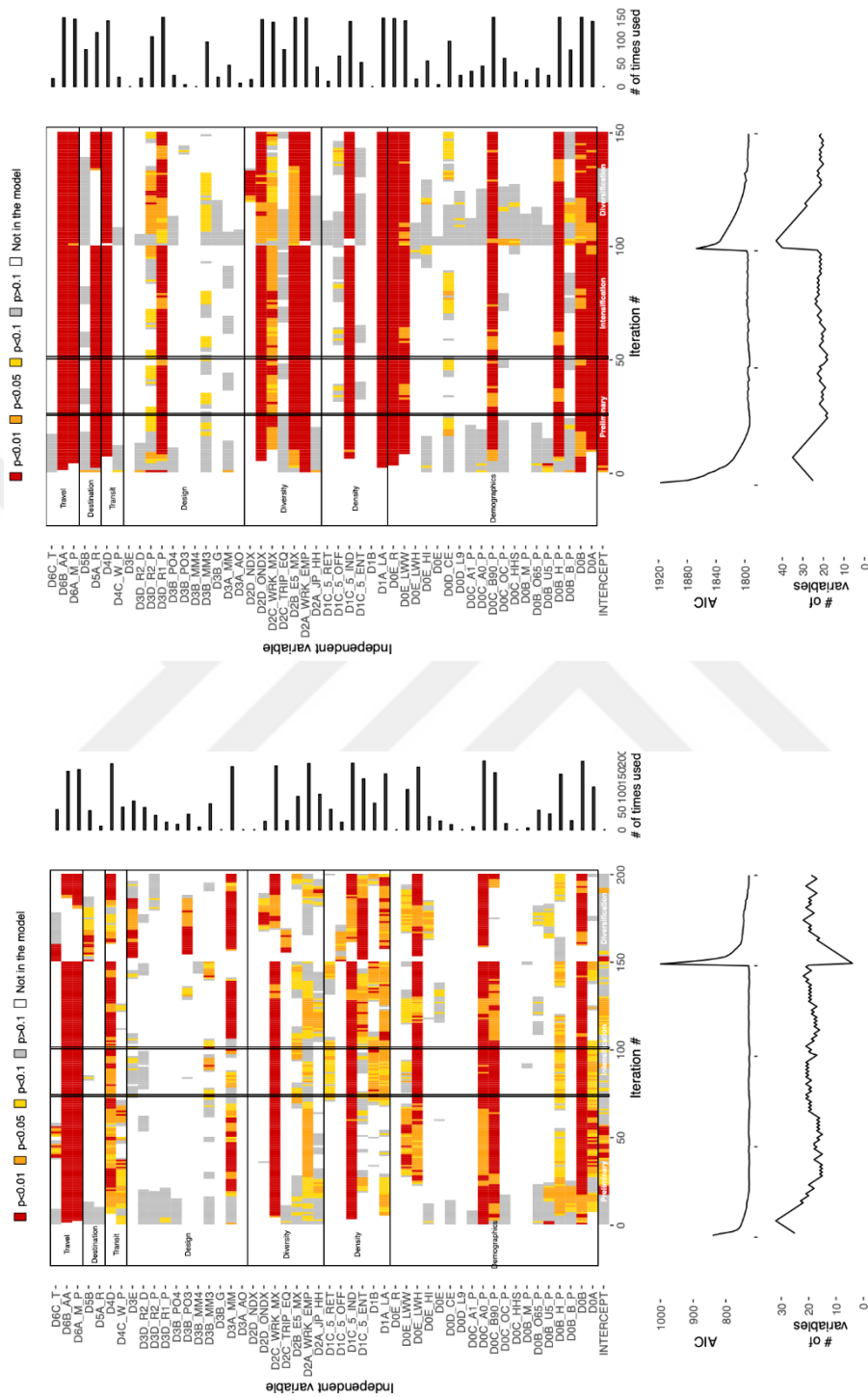


Figure A.5. FQ_MOT Negative Binomial

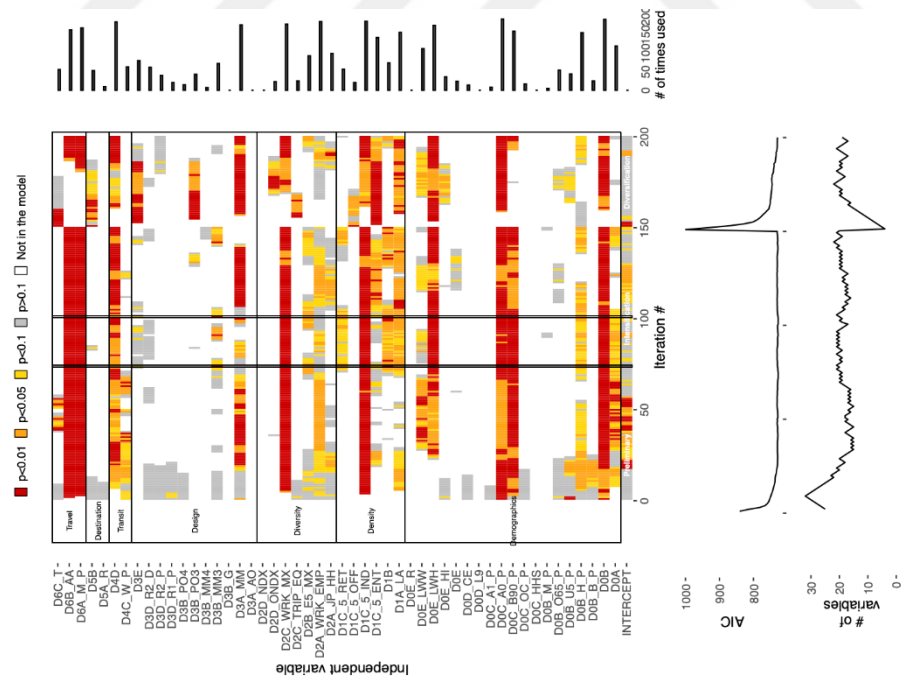


Figure A.6. FQ_VRU Negative Binomial

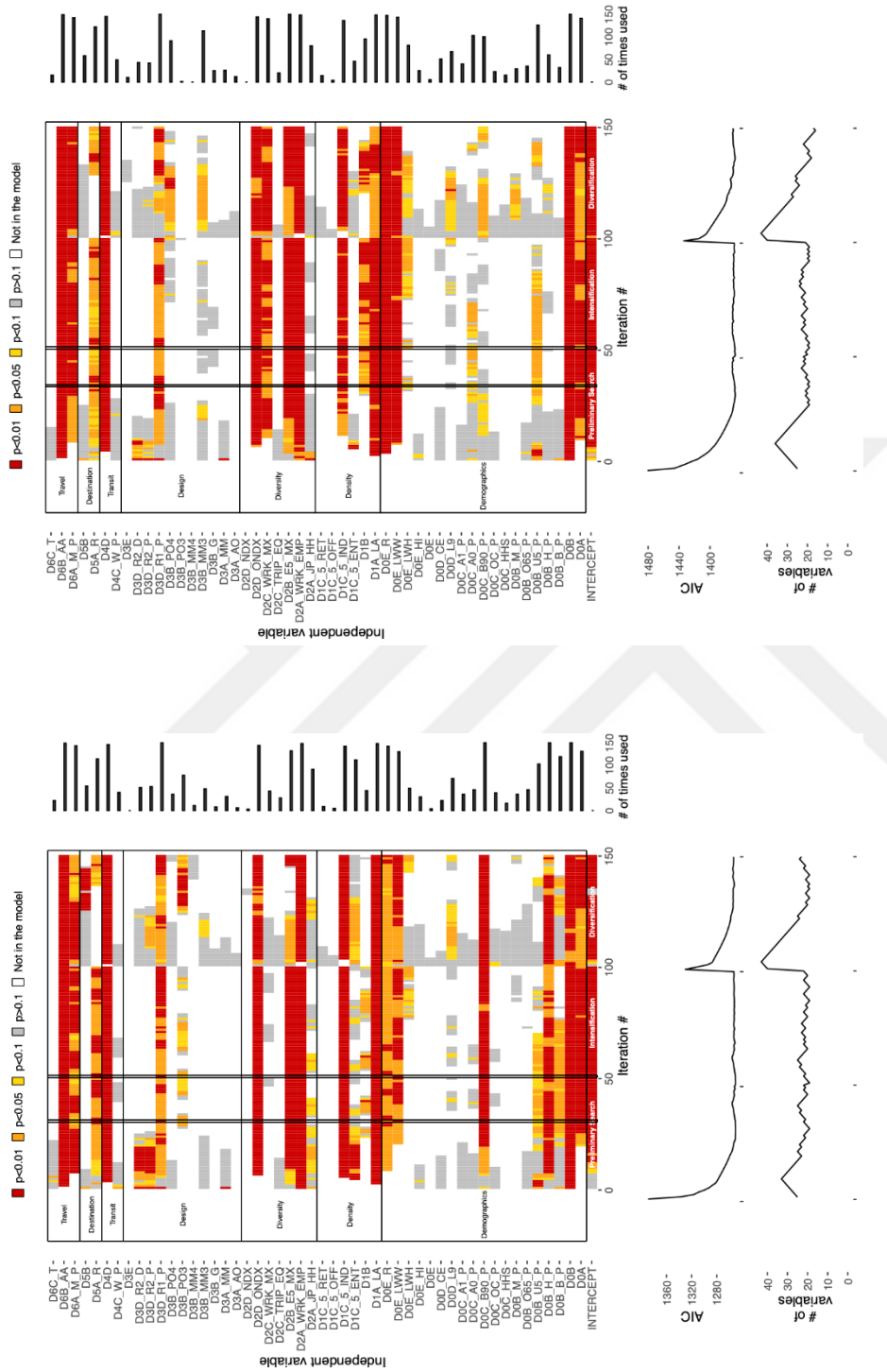


Figure A.7. KD_S1 Gamma

Figure A.8. KD_S2 Gamma

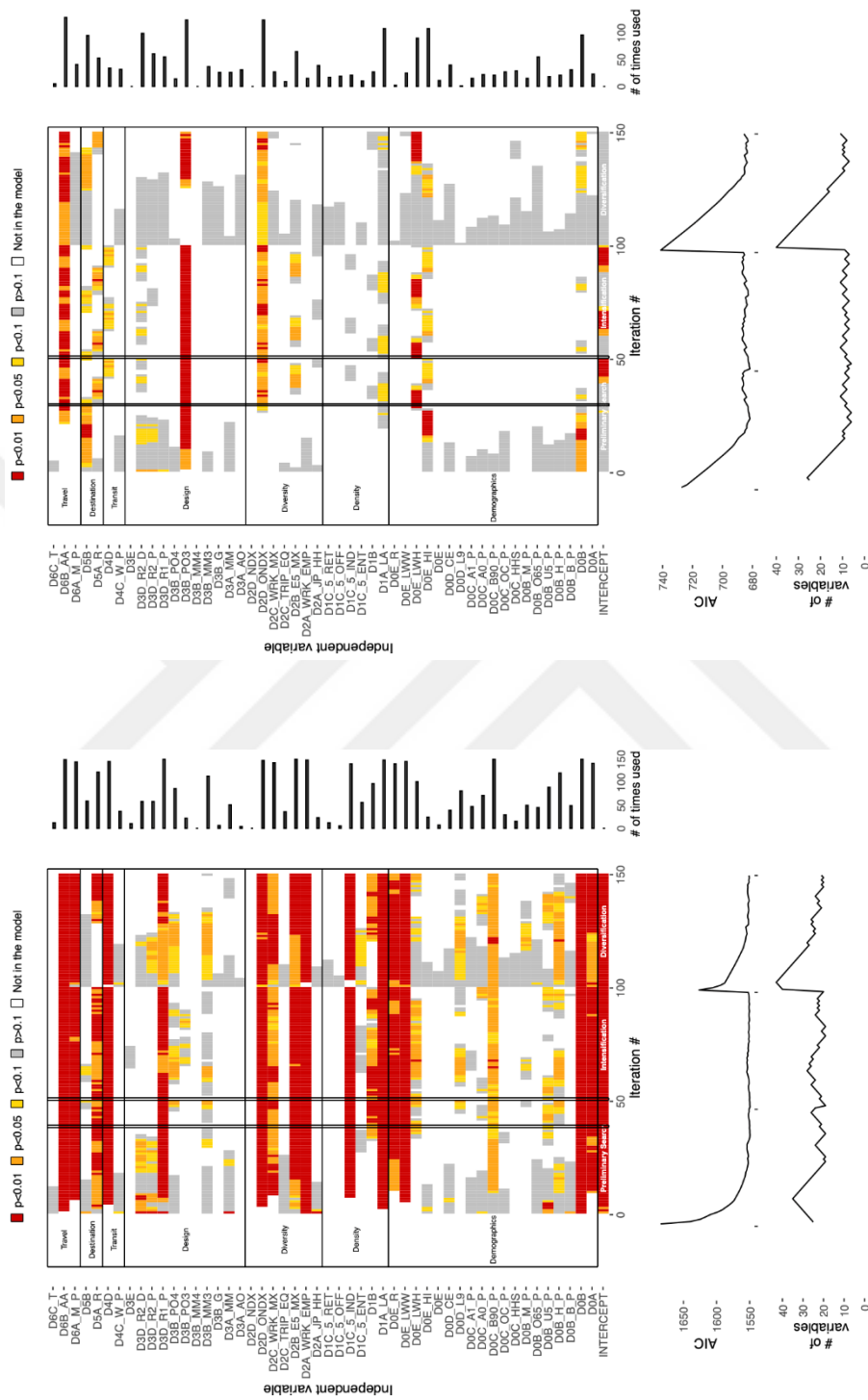


Figure A.9. KD_S3 Gamma

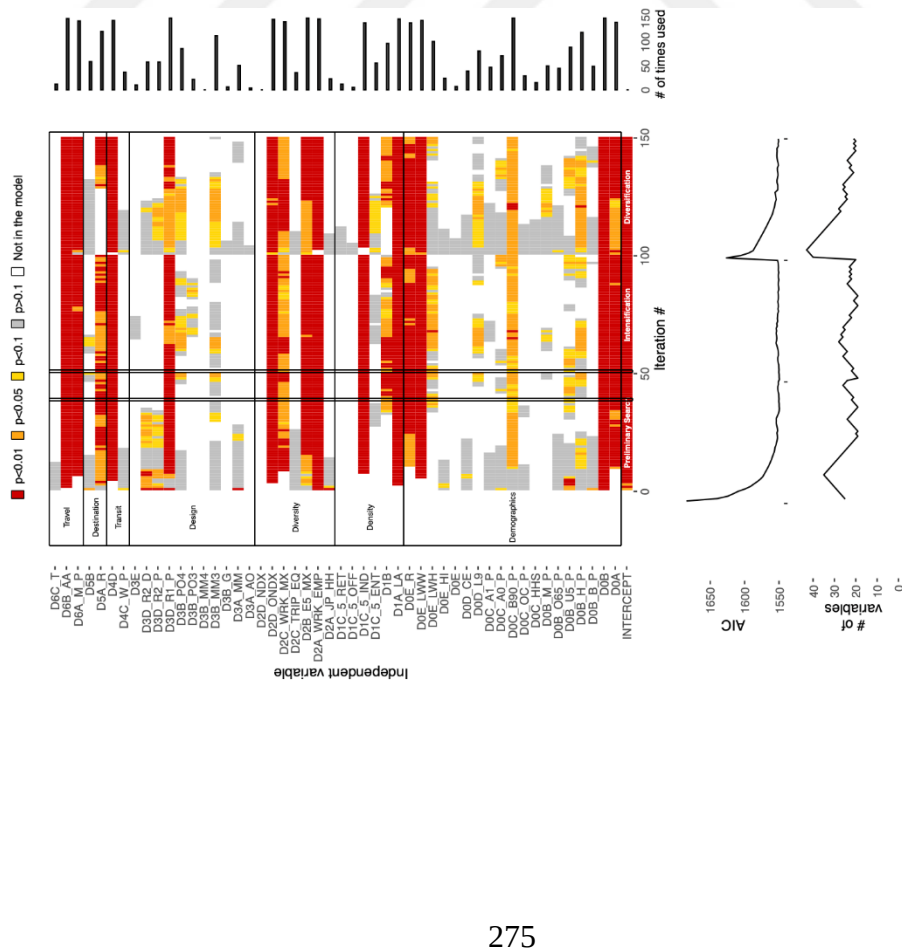


Figure A.10. KD_ALL Gamma

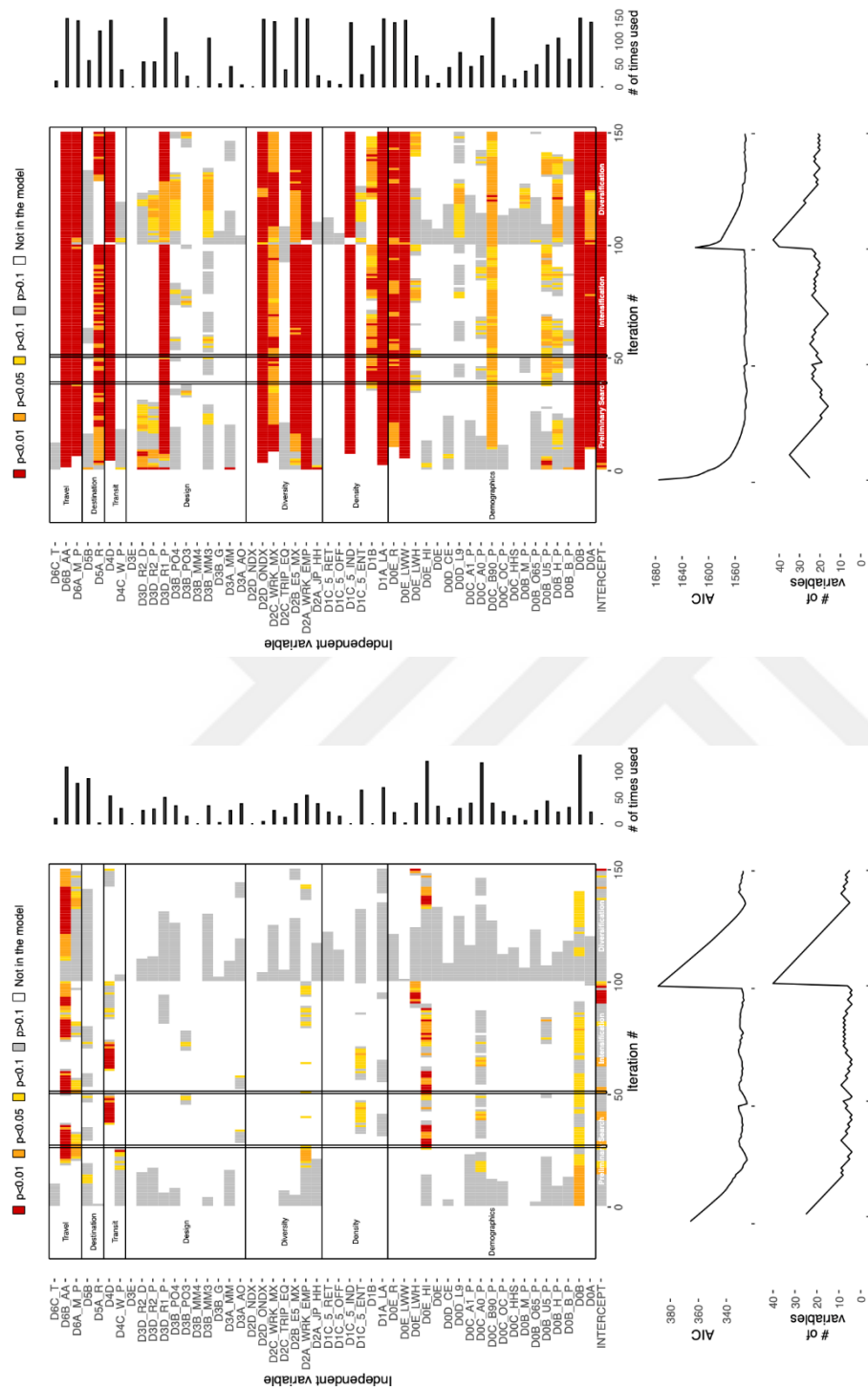


Figure A.11 KD_MOT Gamma

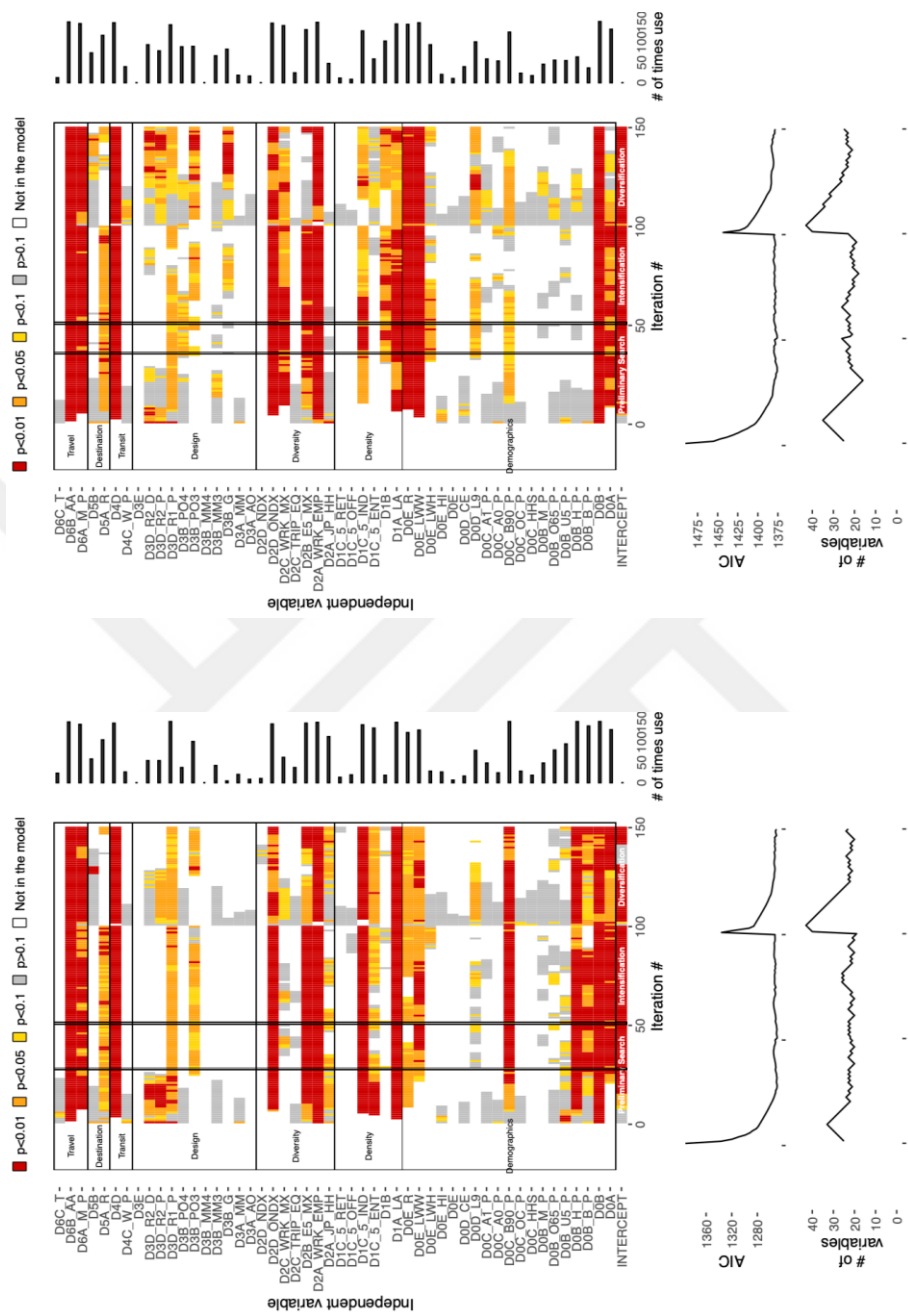


Figure A.14. KD_S2 Lognormal

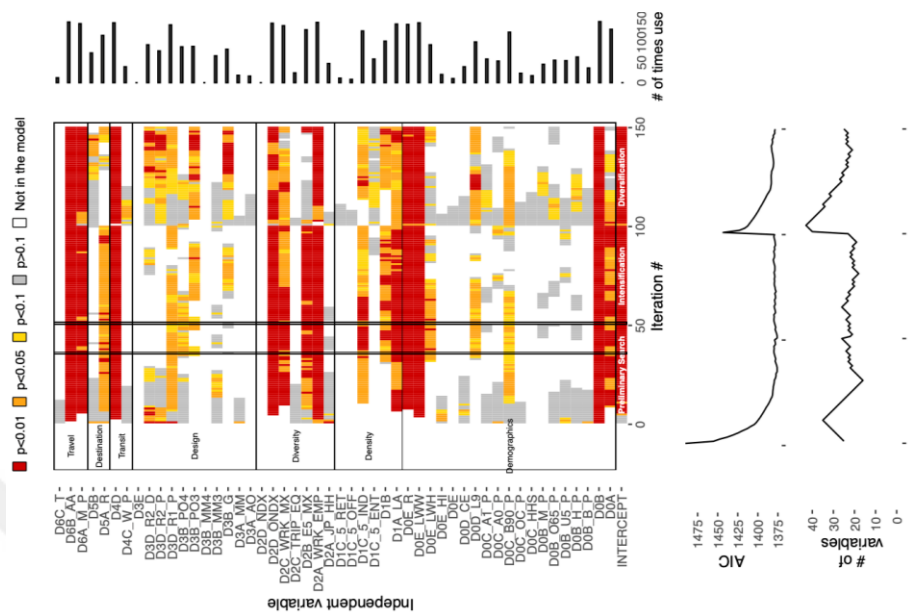


Figure A.13. KD_S1 Lognormal



Figure A.15. KD_S3 Lognormal

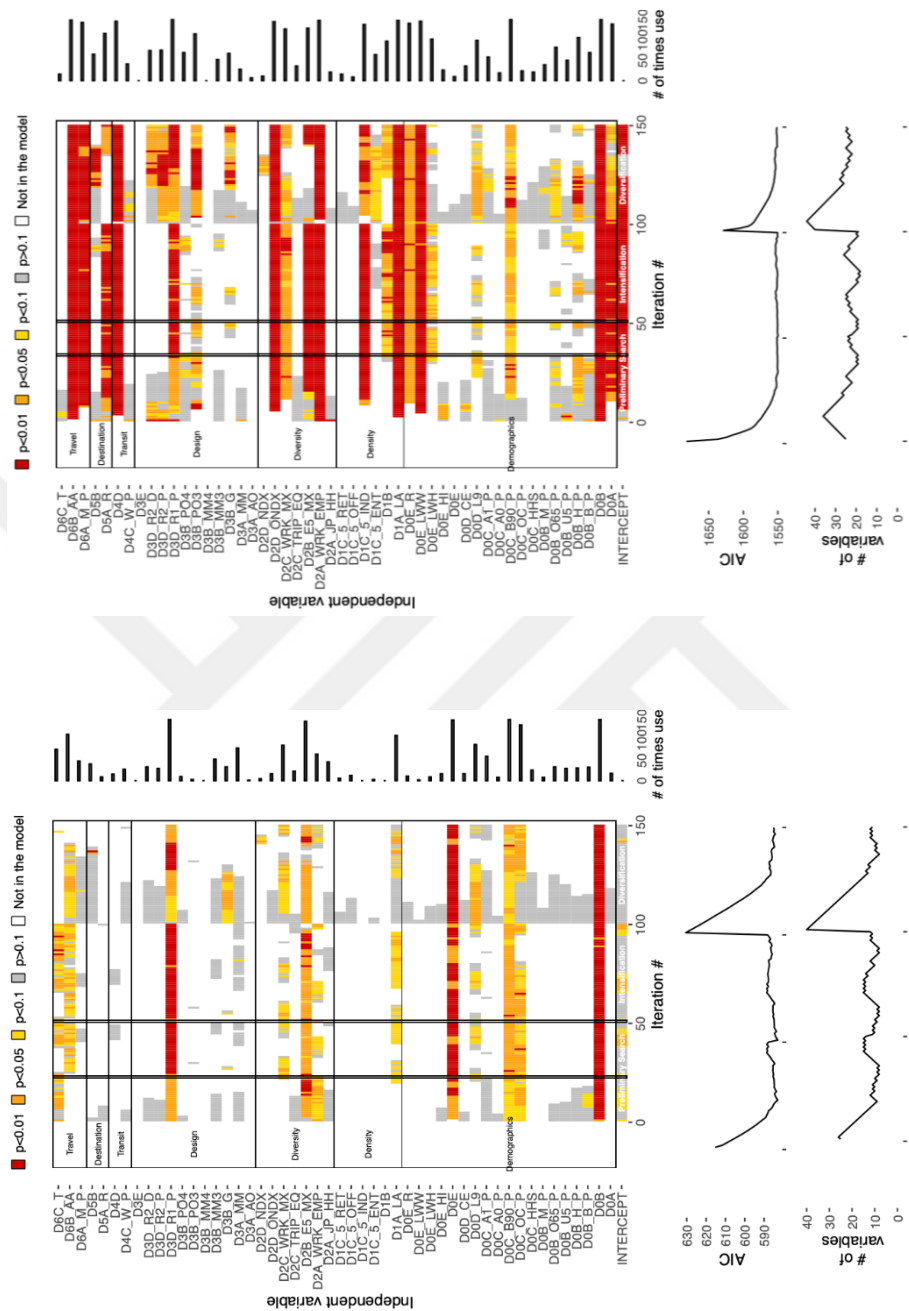


Figure A.17. KD_MOT Lognormal

Figure A.18. KD_VRU Lognormal

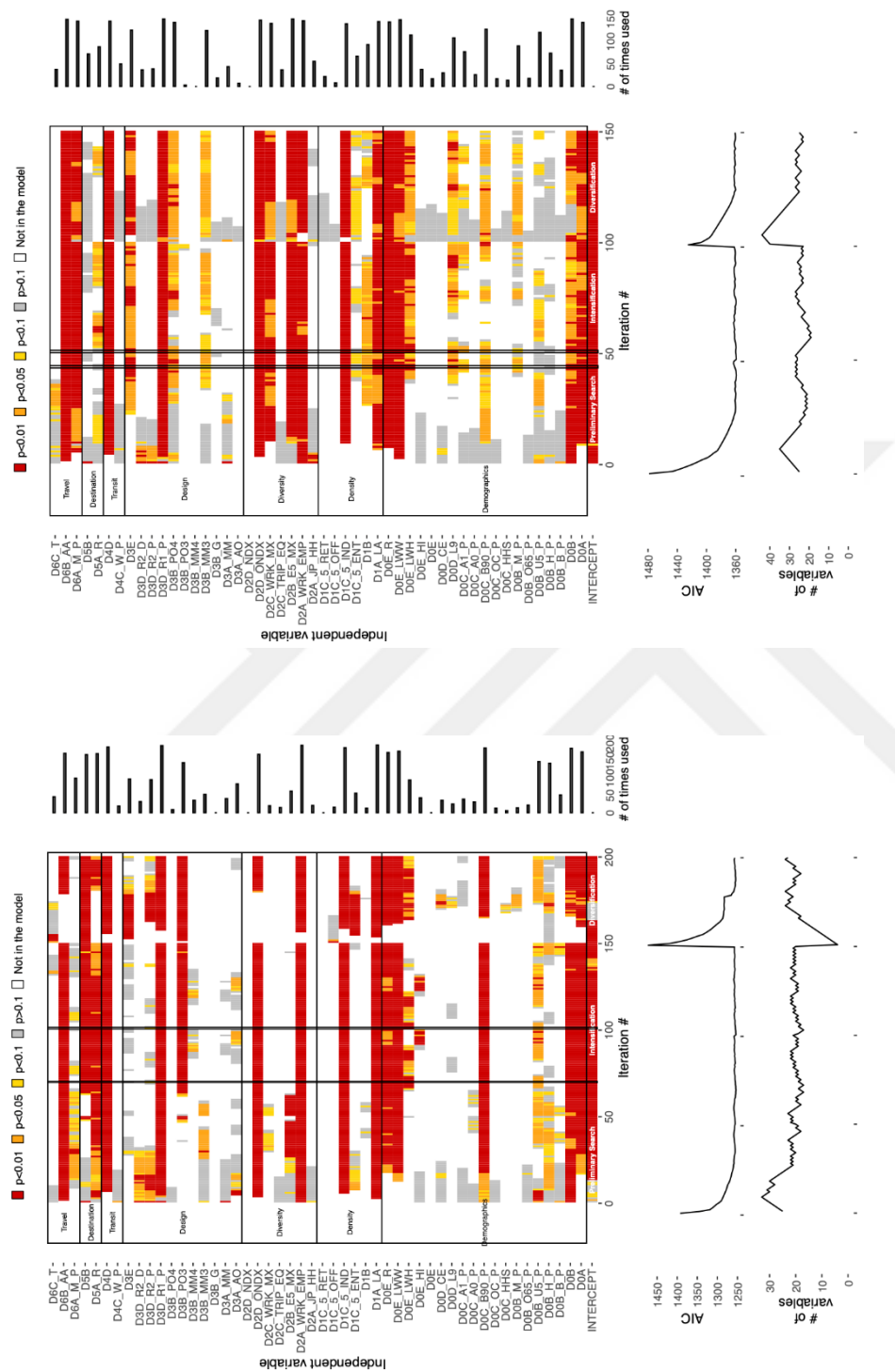


Figure A.19. KD_S1 Weibull

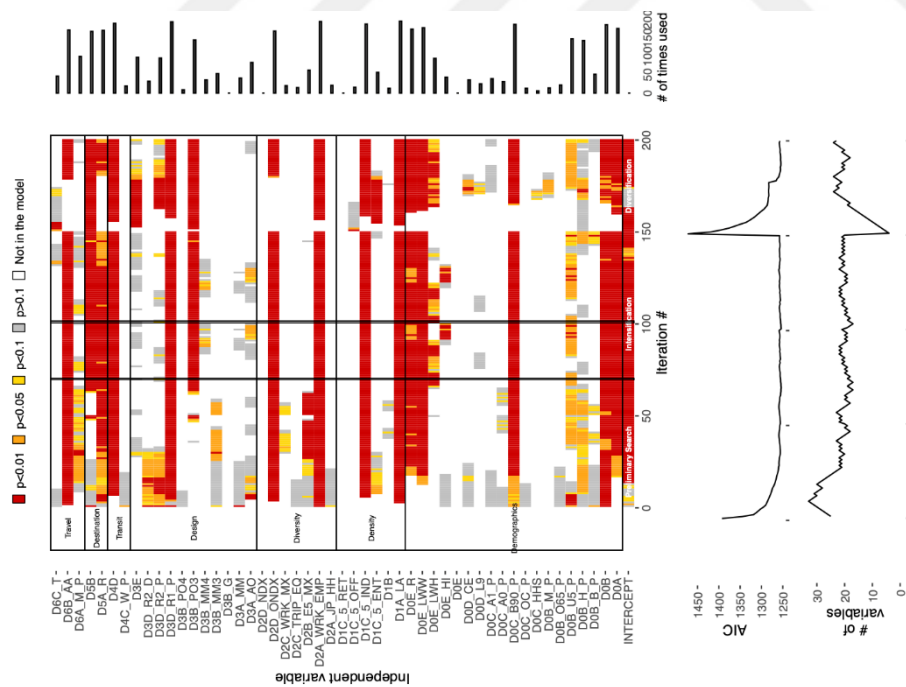


Figure A.20. KD_S2 Weibull

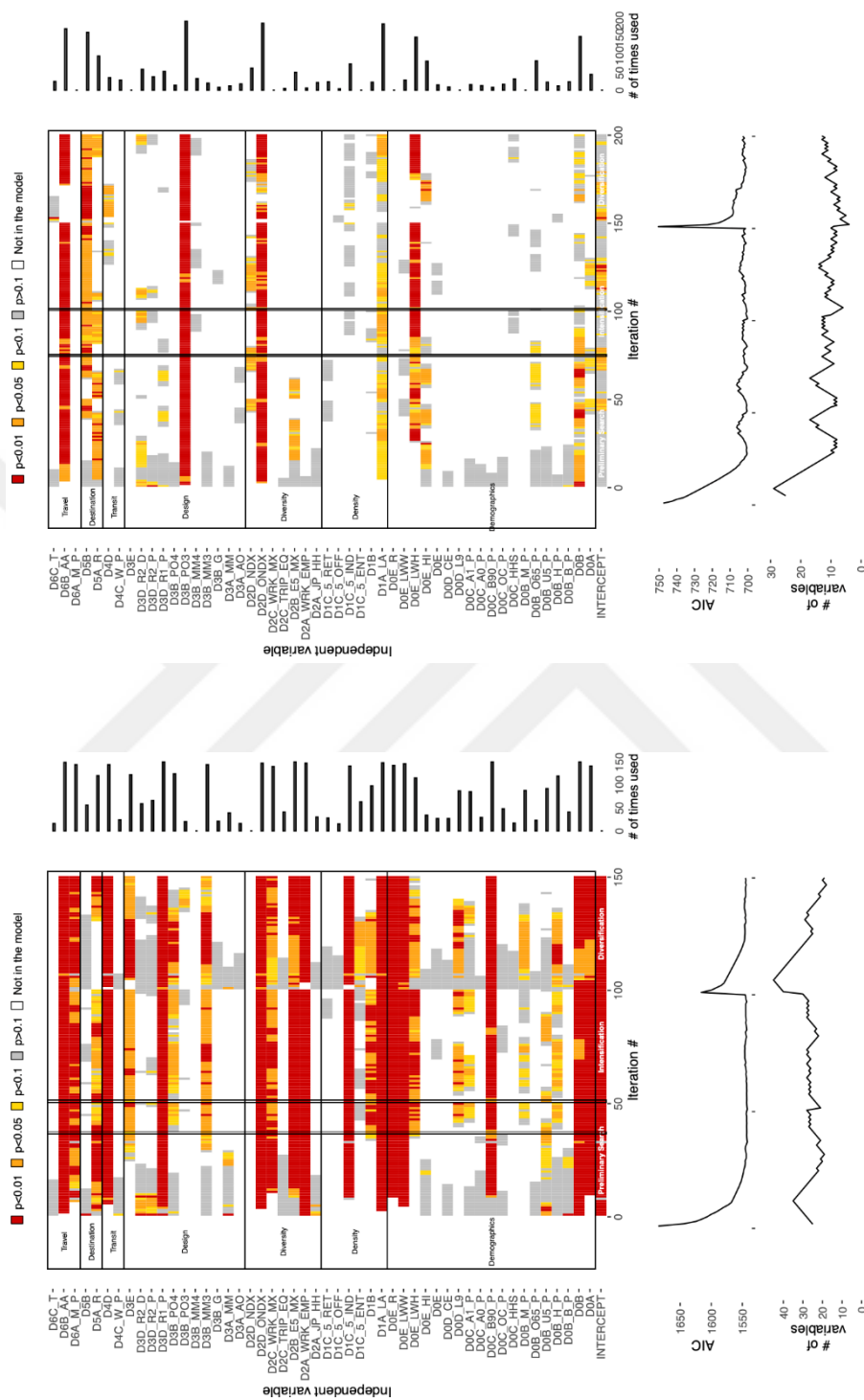


Figure A.21. KD_S3 Weibull

Figure A.22. KD_ALL Weibull

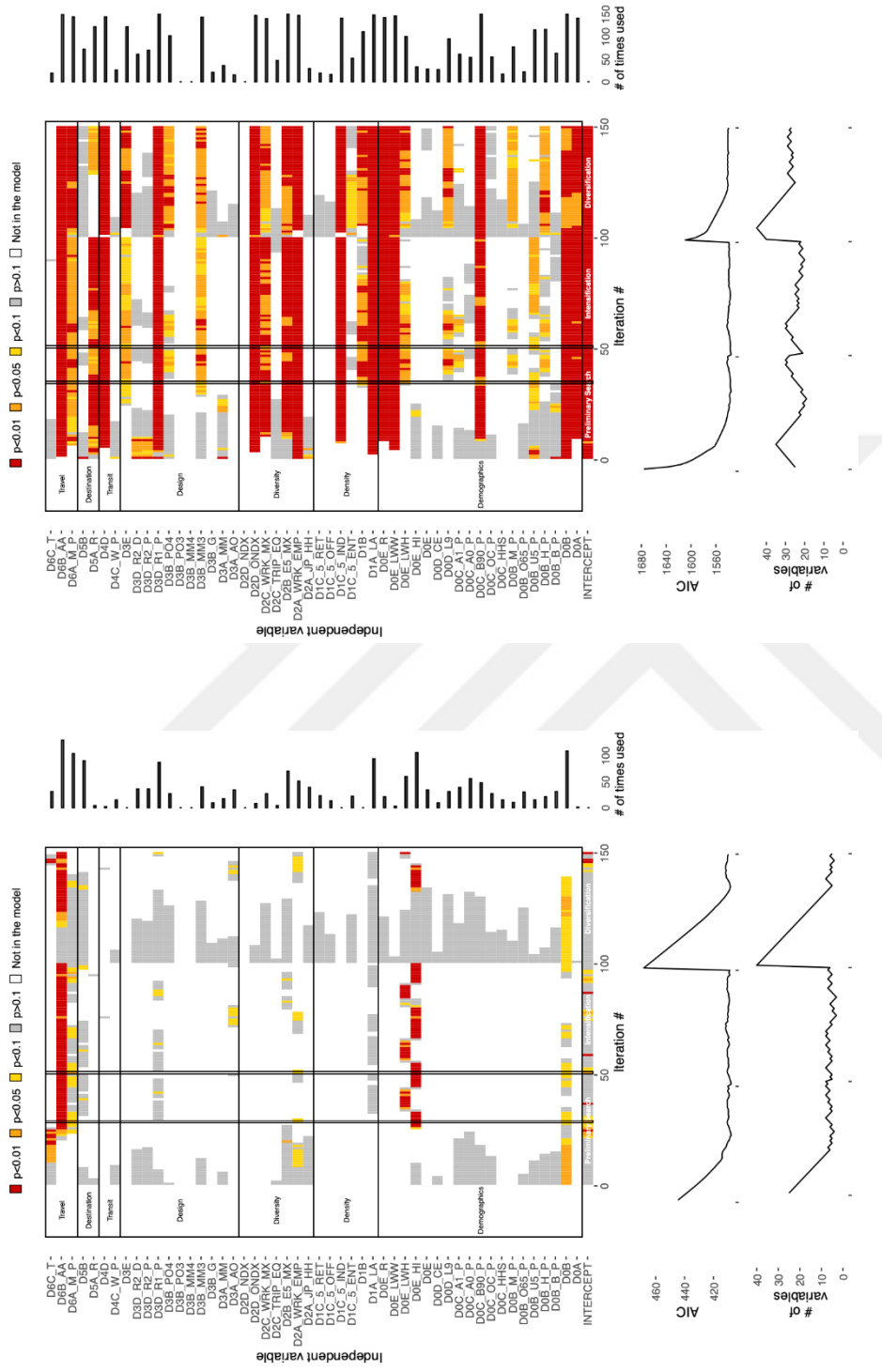


Figure A.23. KD_MOT Weibull

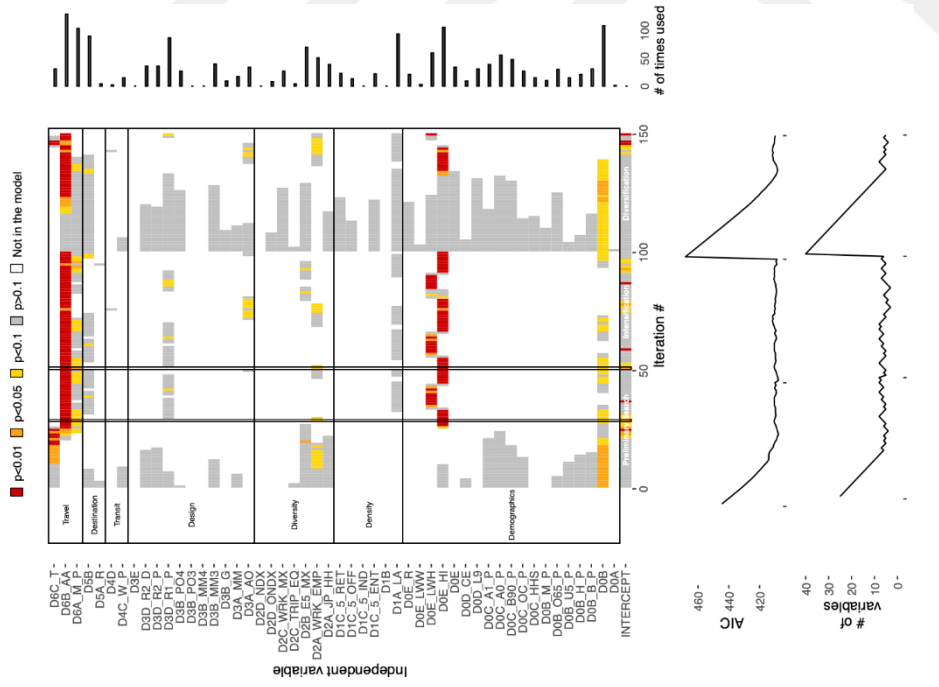


Figure A.24. KD_VRU Weibull

B. Tri-County Case Results

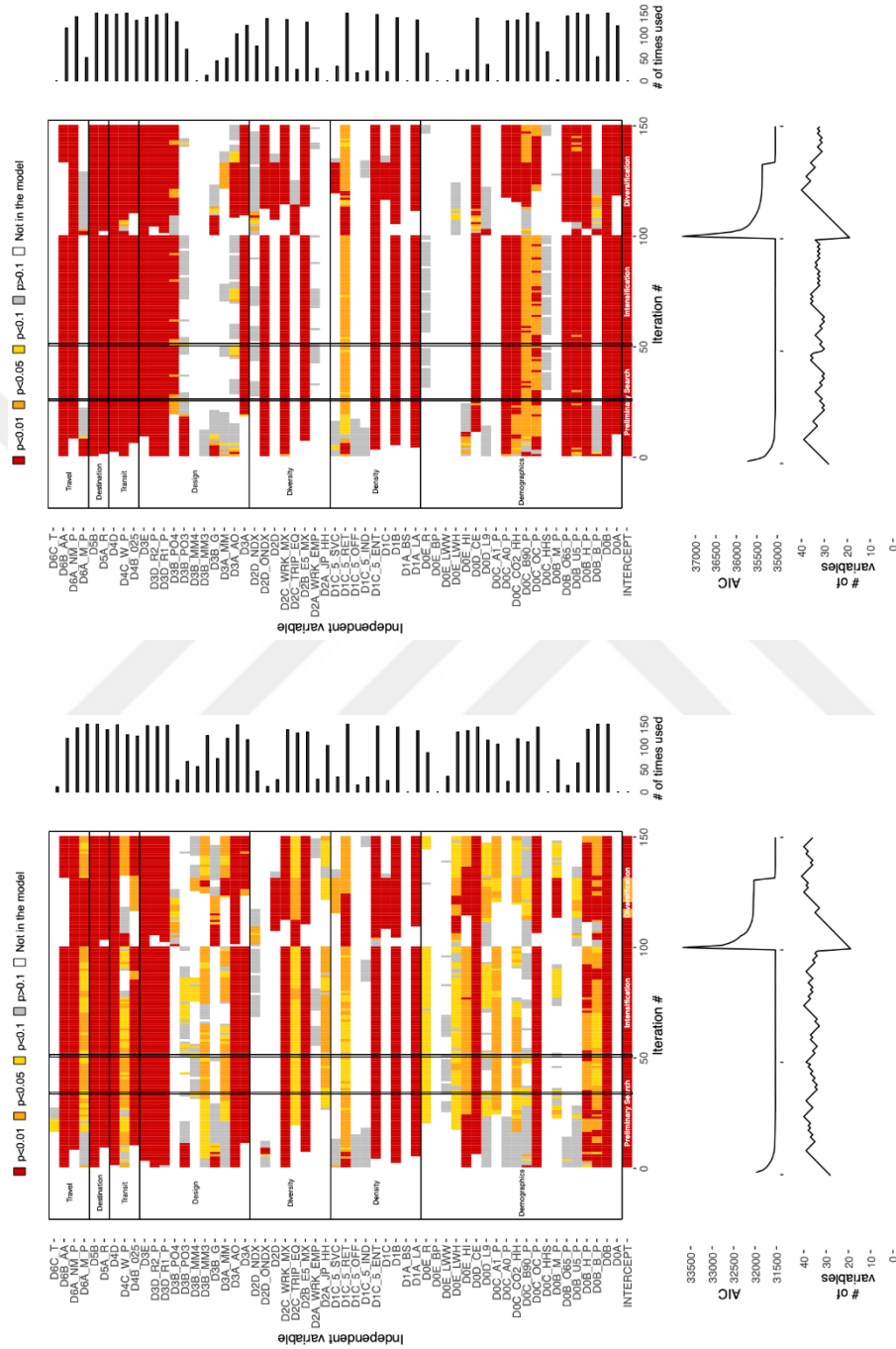


Figure B.1. FQ_S1 Negative Binomial

Figure B.2. FQ_S2 Negative Binomial

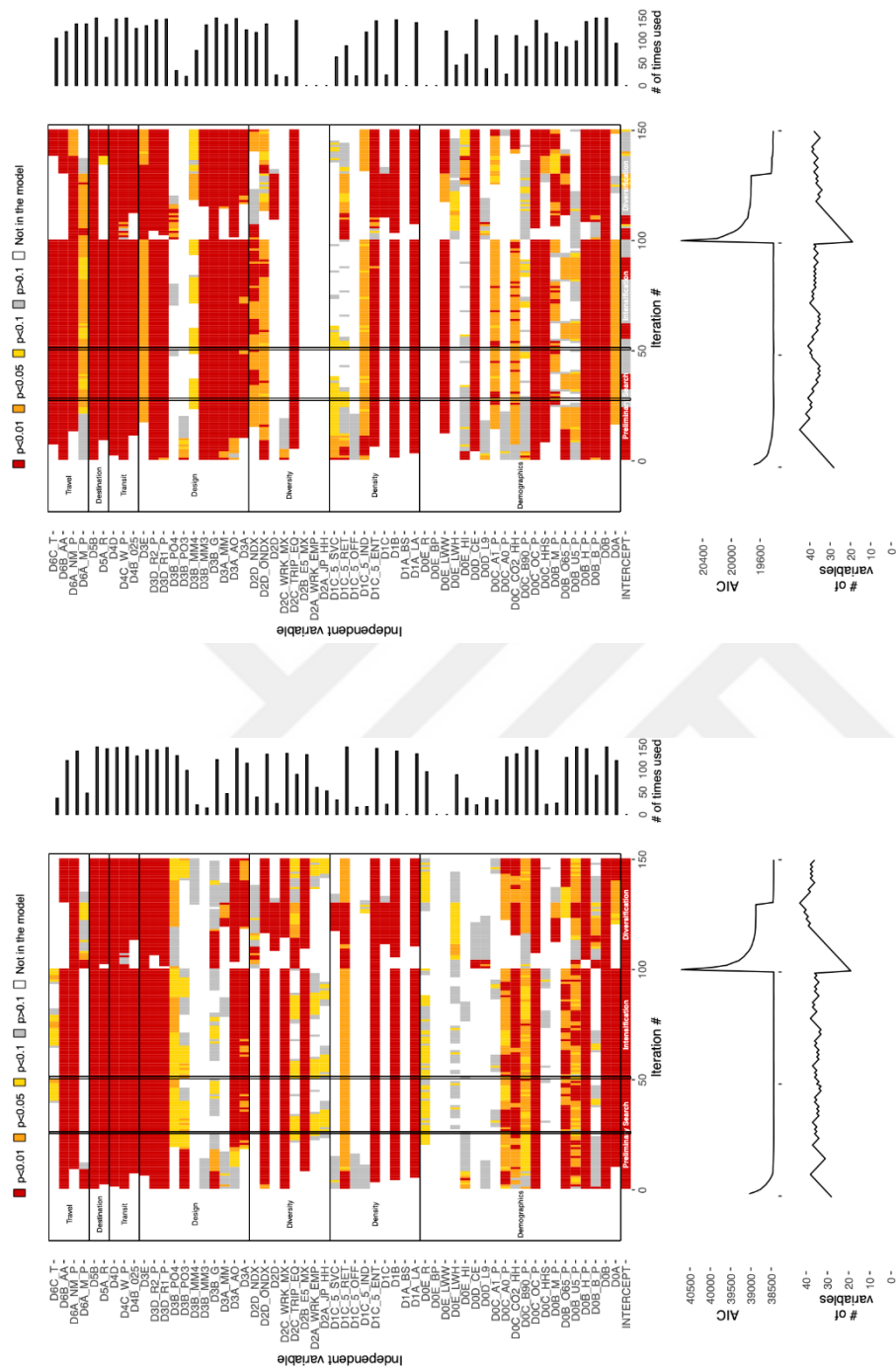


Figure B.4. FQ_ALL Negative Binomial

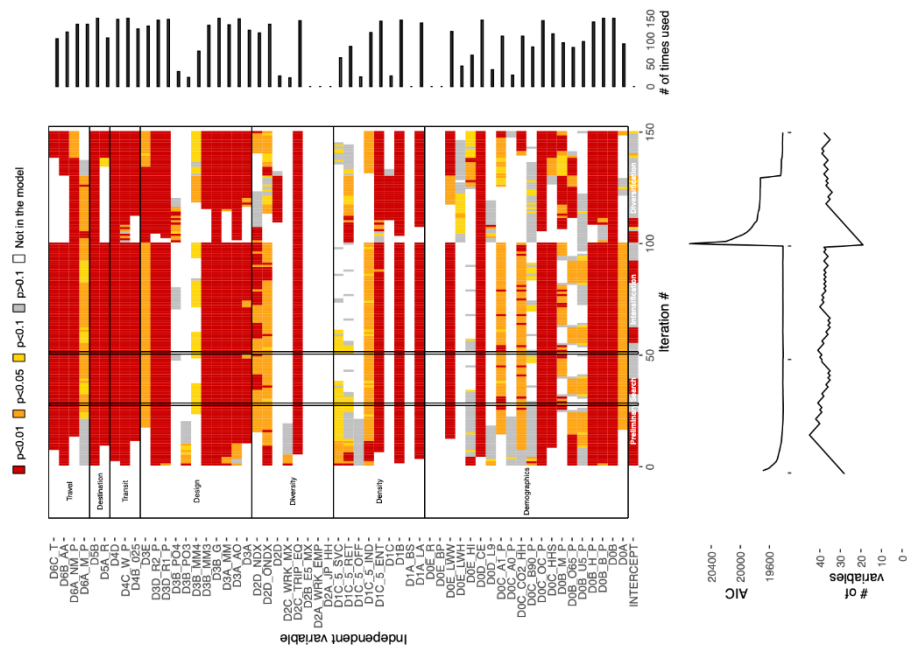


Figure B.3. FQ_S3 Negative Binomial

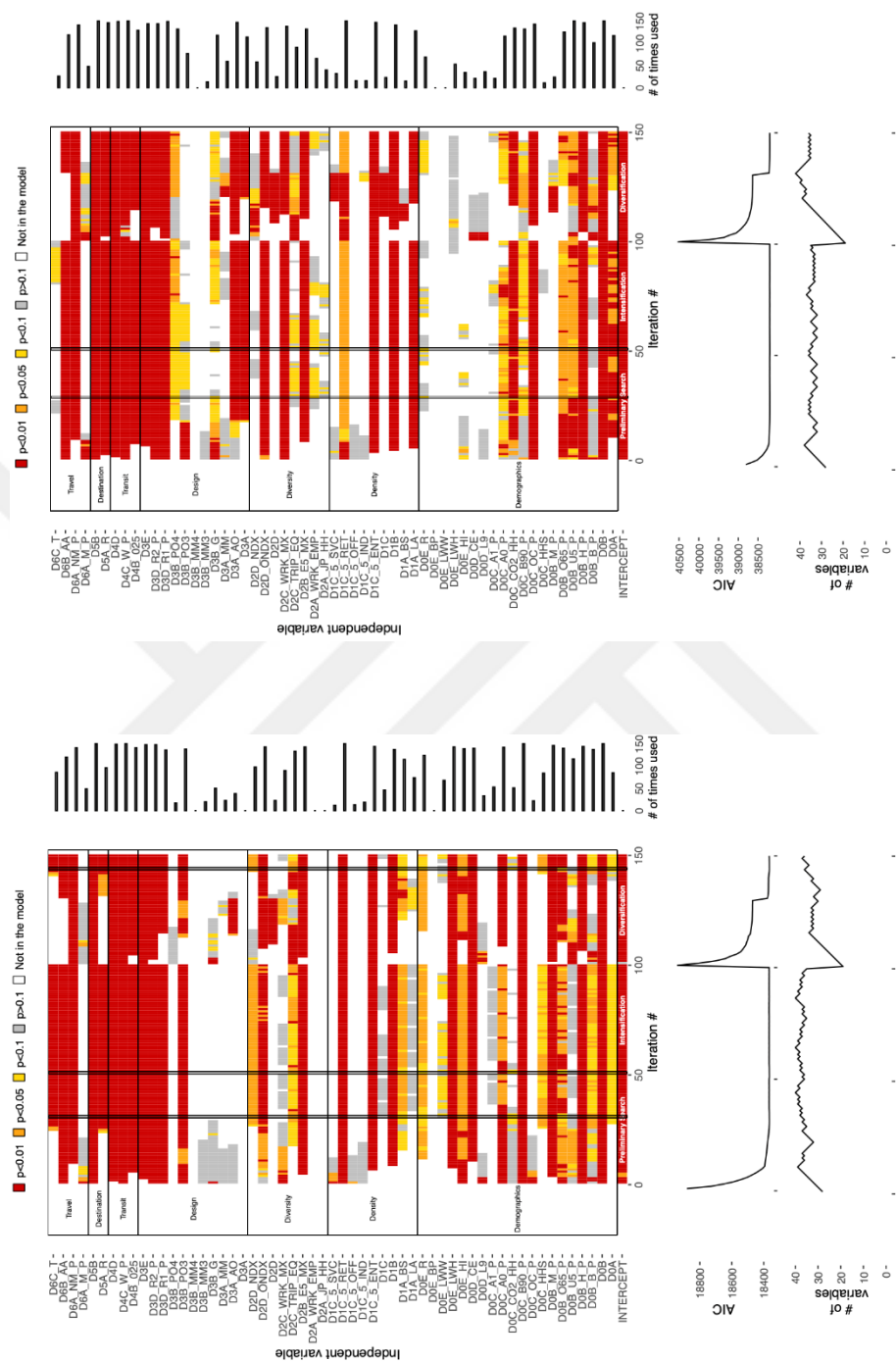


Figure B.6. FQ_VRU Negative Binomial

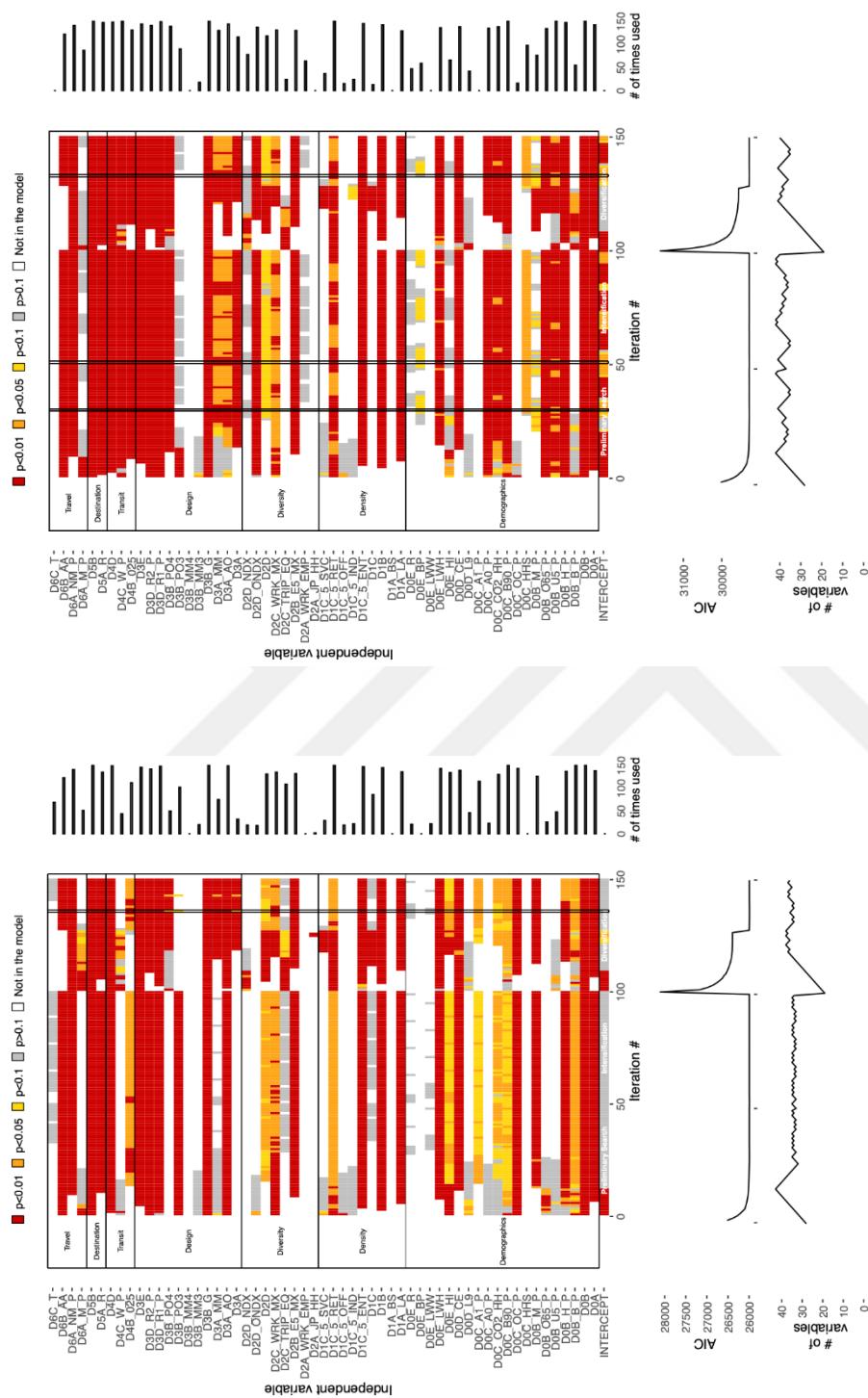


Figure B.8 KD_S2 Gamma

Figure B.7 KD_S1 Gamma

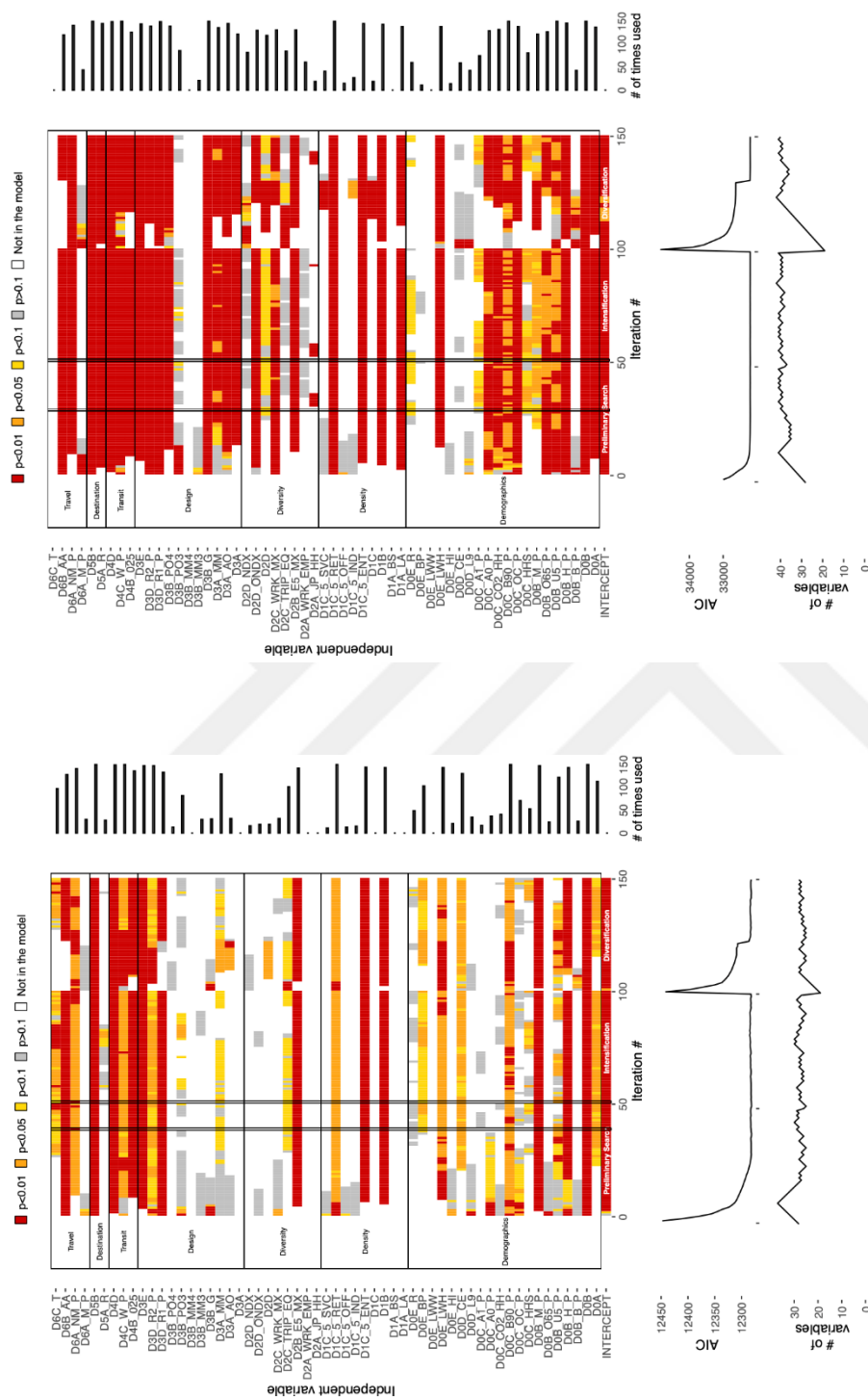


Figure B.12. KD_VRU Gamma

Figure B.11. KD_MOT Gamma

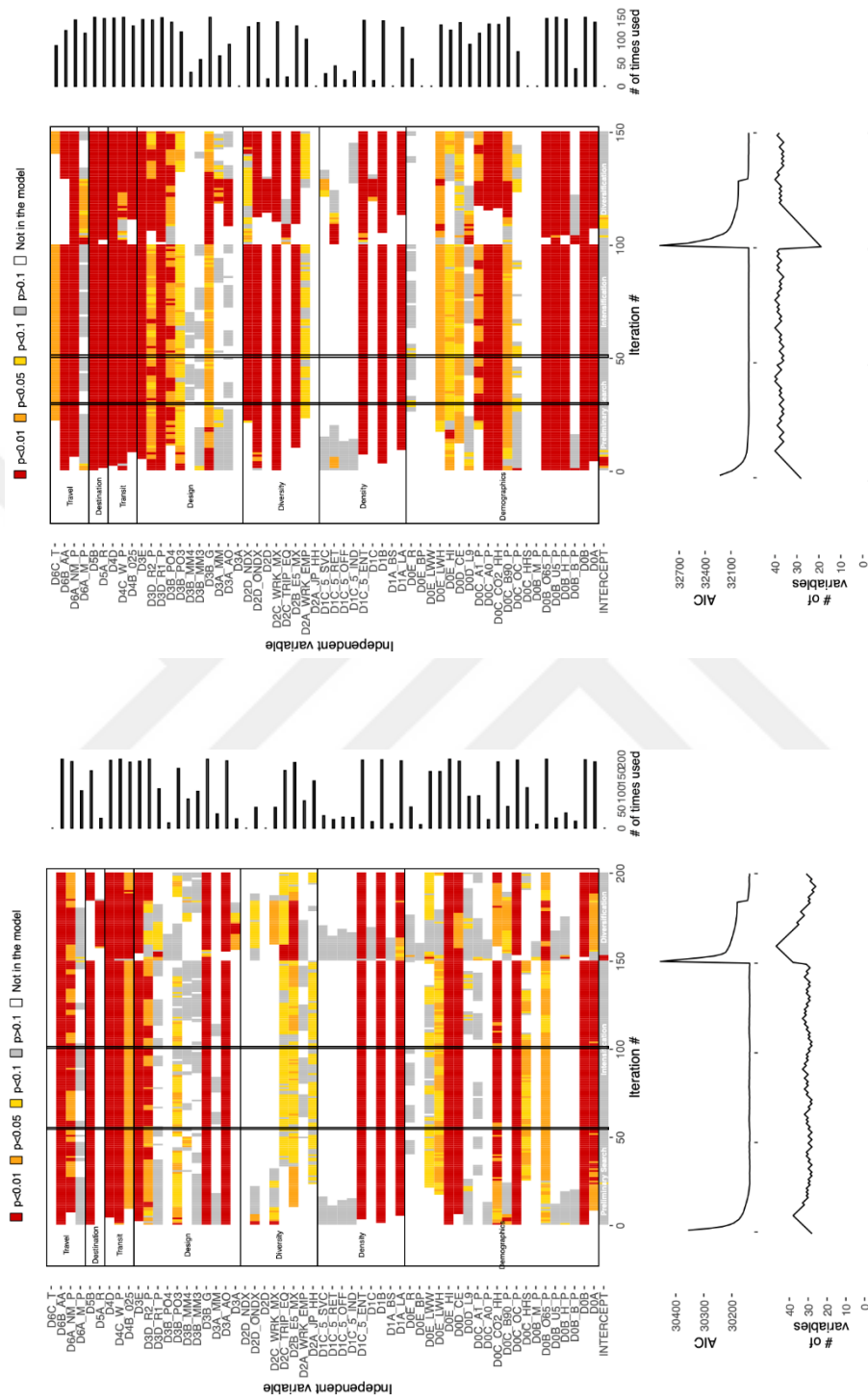


Figure B.13. KD_S1 Lognormal

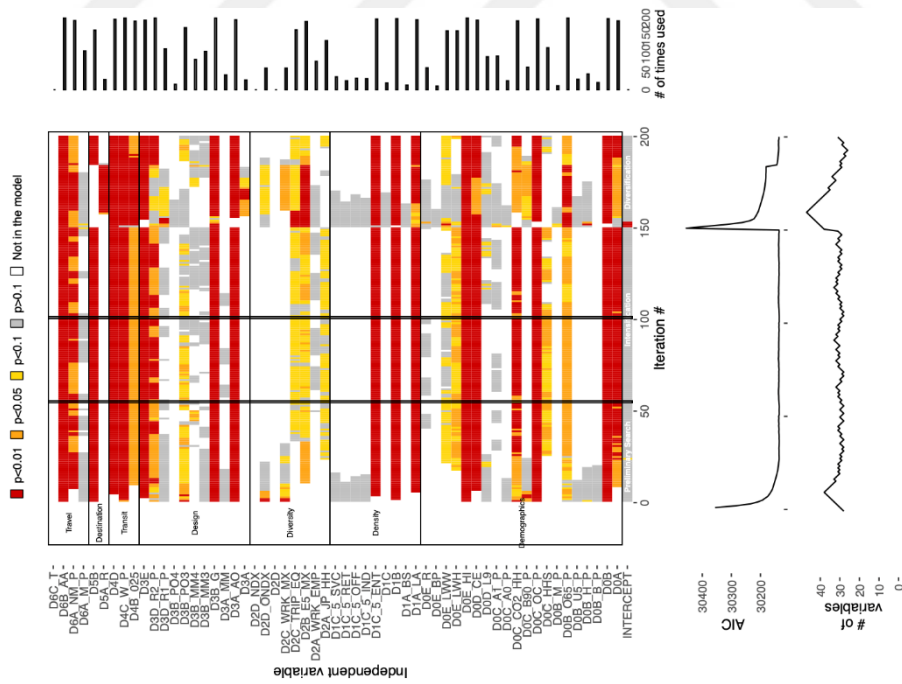


Figure B.14. KD_S2 Lognormal

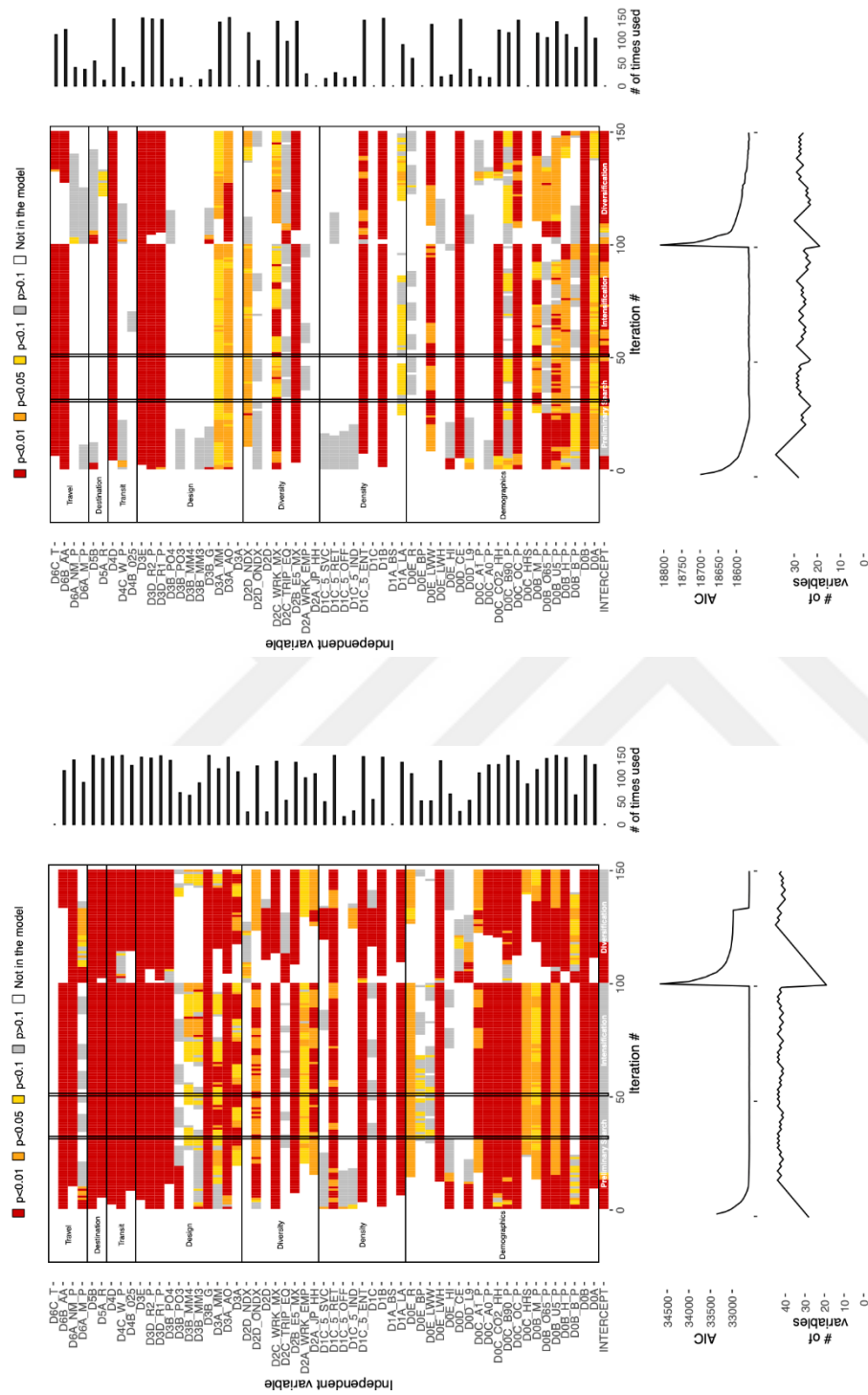


Figure B.15. KD_S3 Lognormal

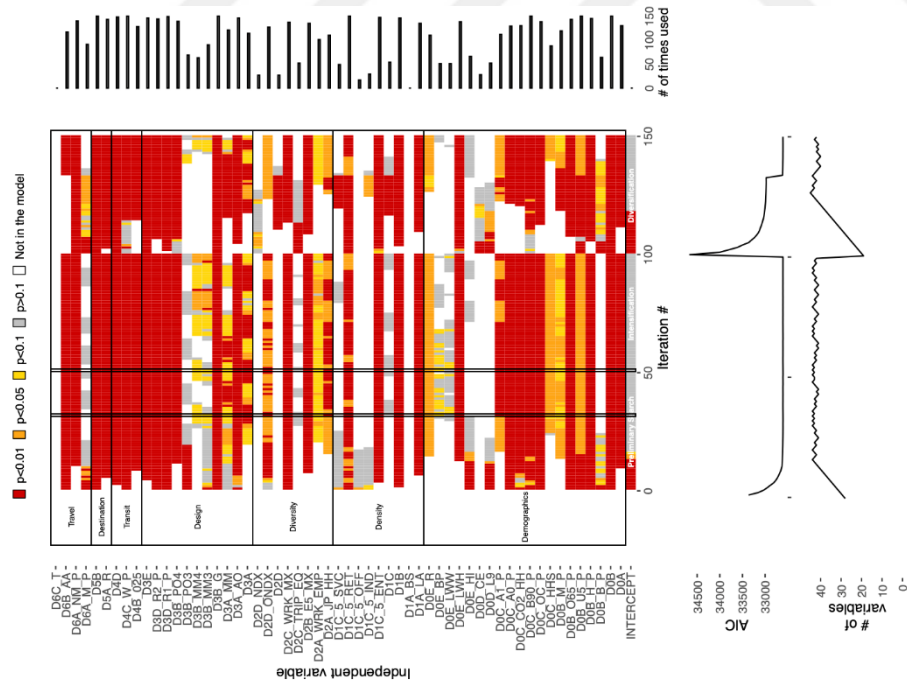


Figure B.16. KD_ALL Lognormal

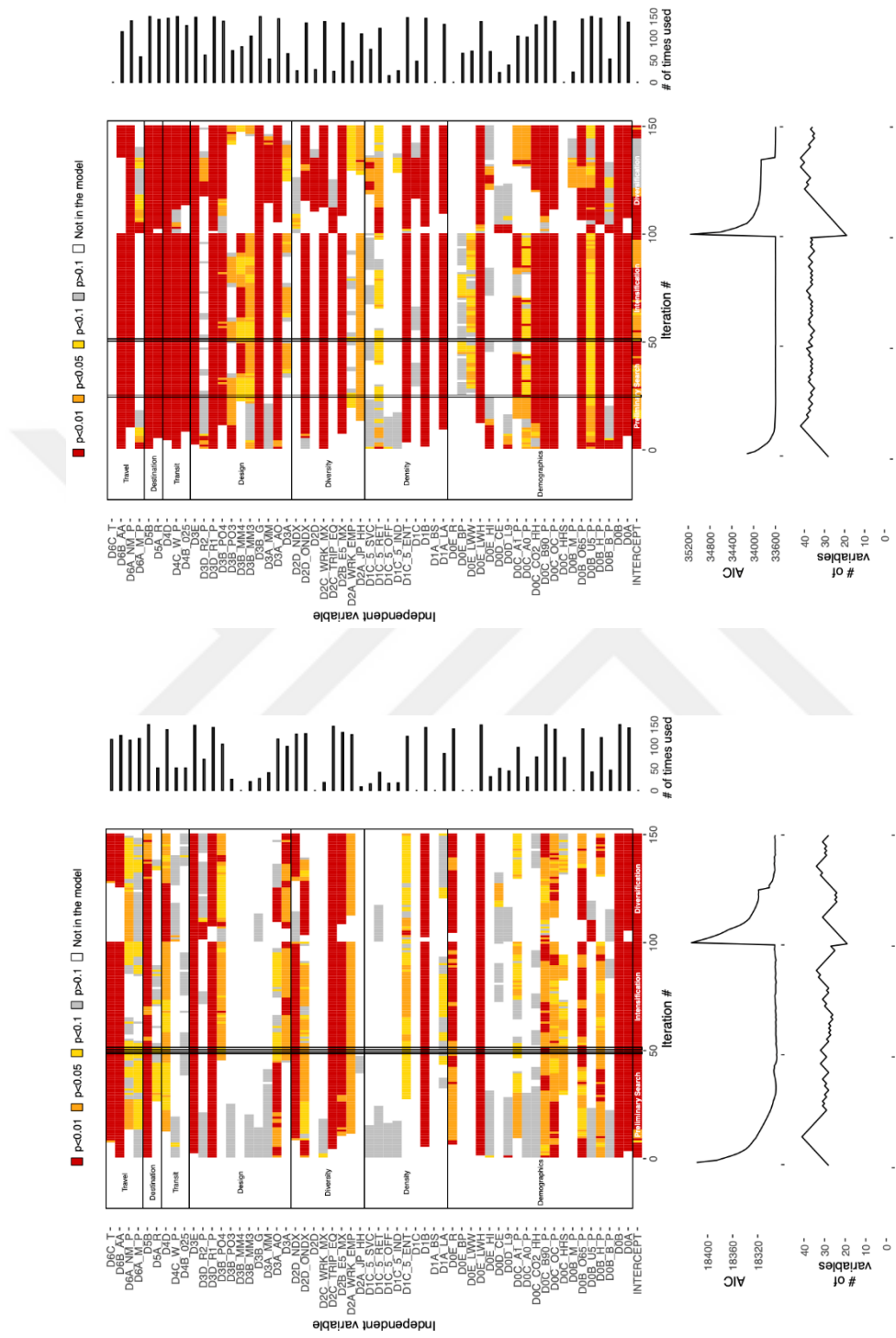


Figure B.18. KD_VRU Lognormal

Figure B.17. KD_MOT Lognormal

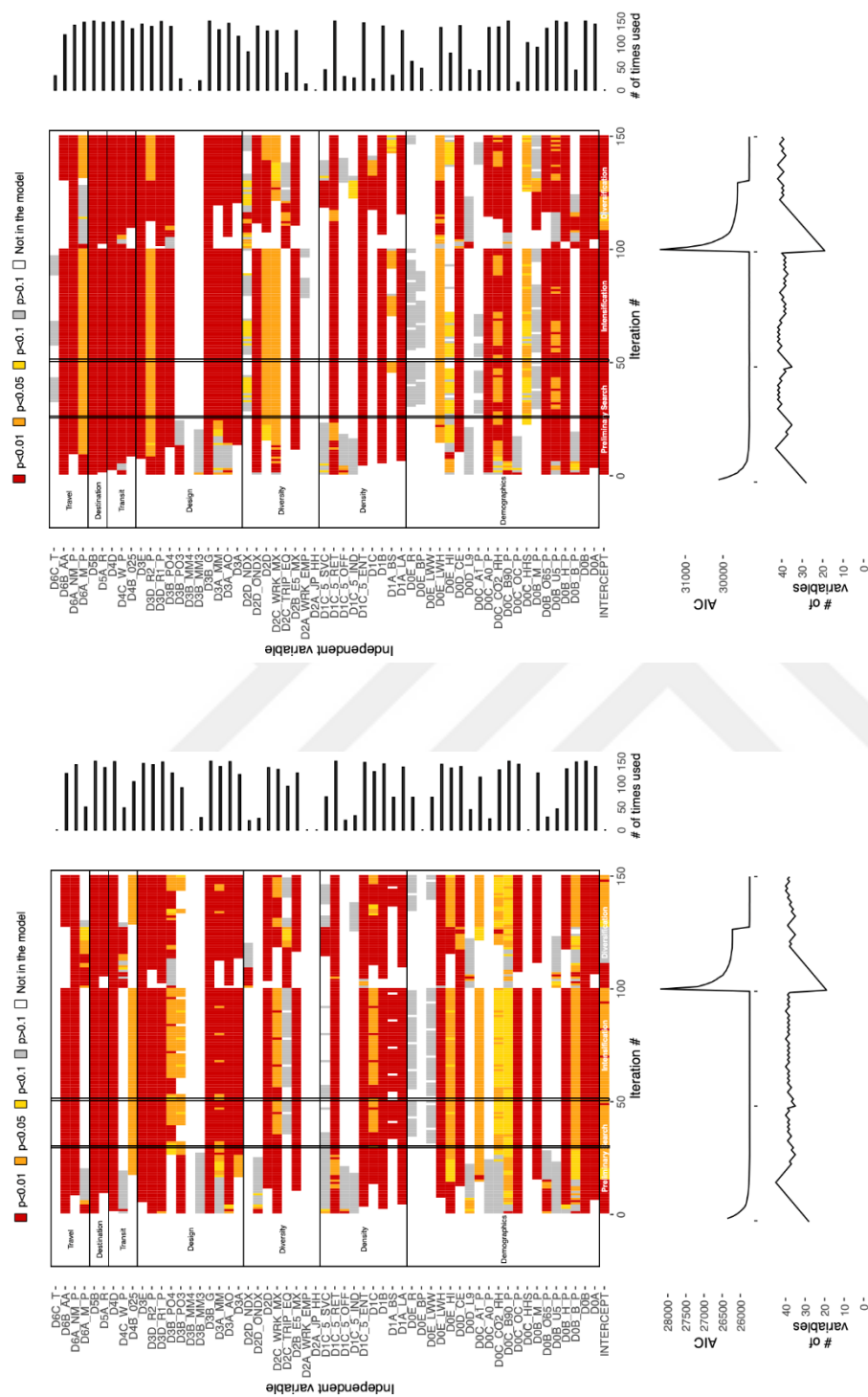


Figure B.20. KD_S2 Weibull

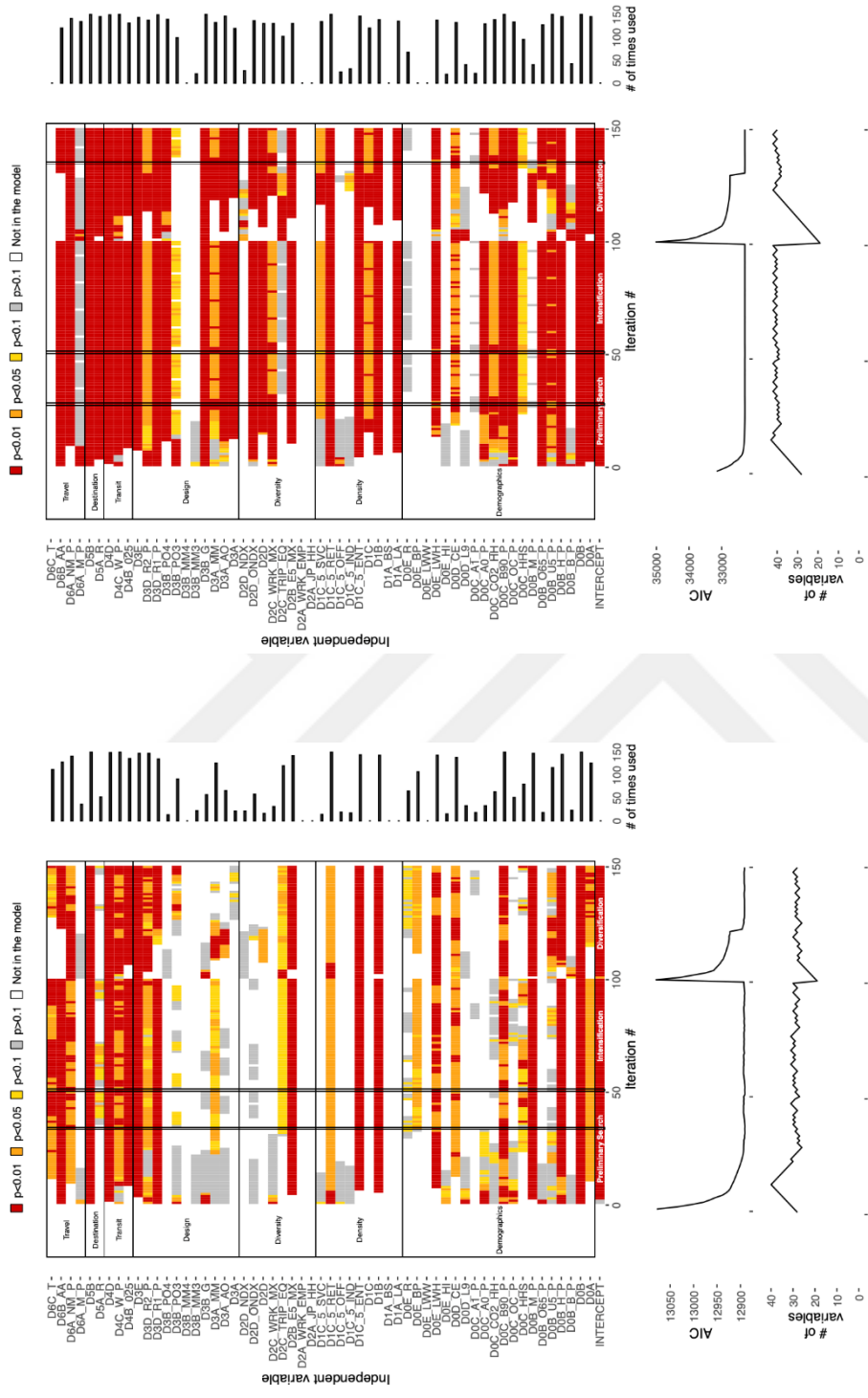


Figure B.24. KD_VRU Weibull

Figure B.23. KD_MOT Weibull

C. Konya Case Results

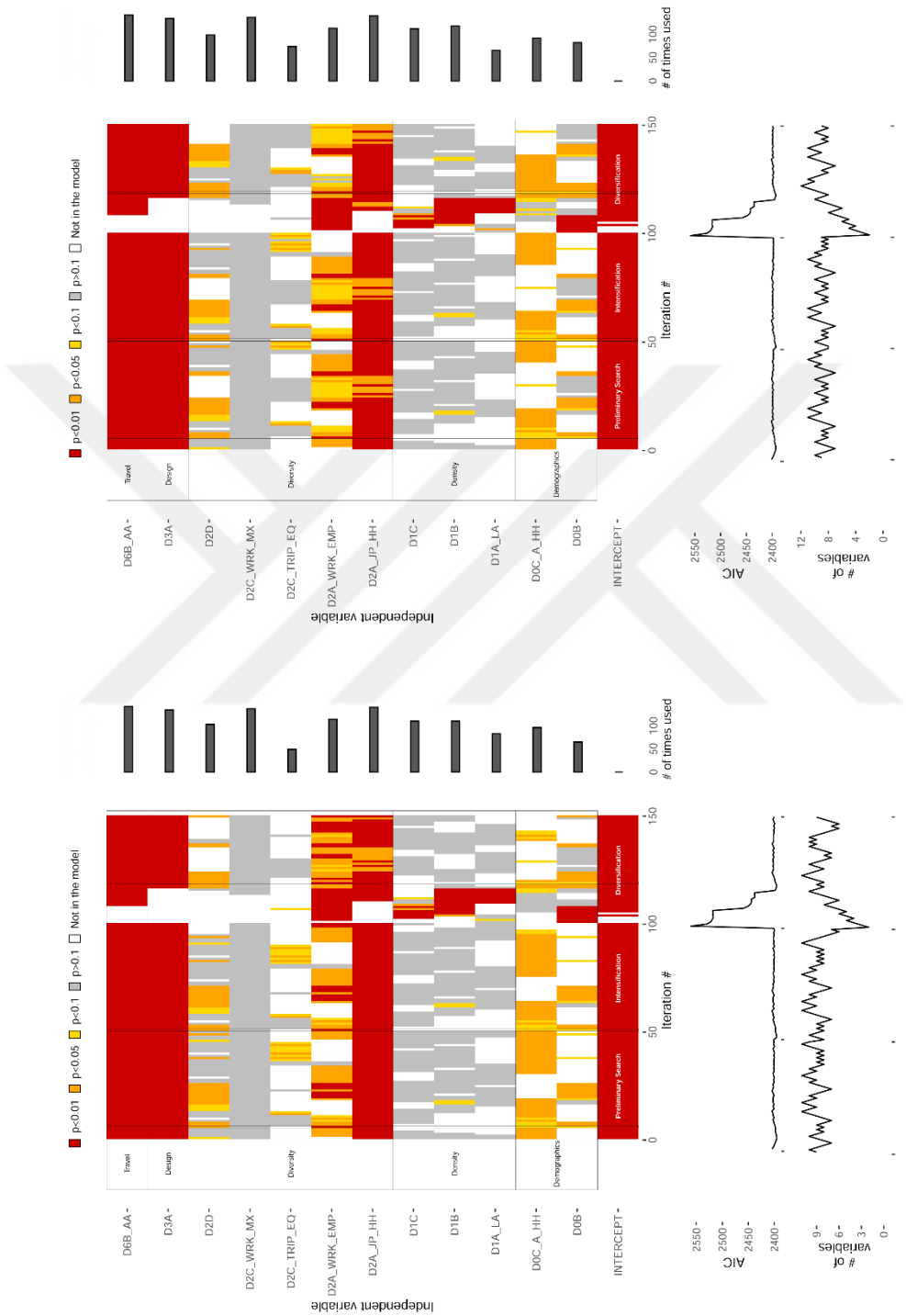


Figure C.2. FQ_ALL Negative Binomial

Figure C.1. FQ_S2 Negative Binomial

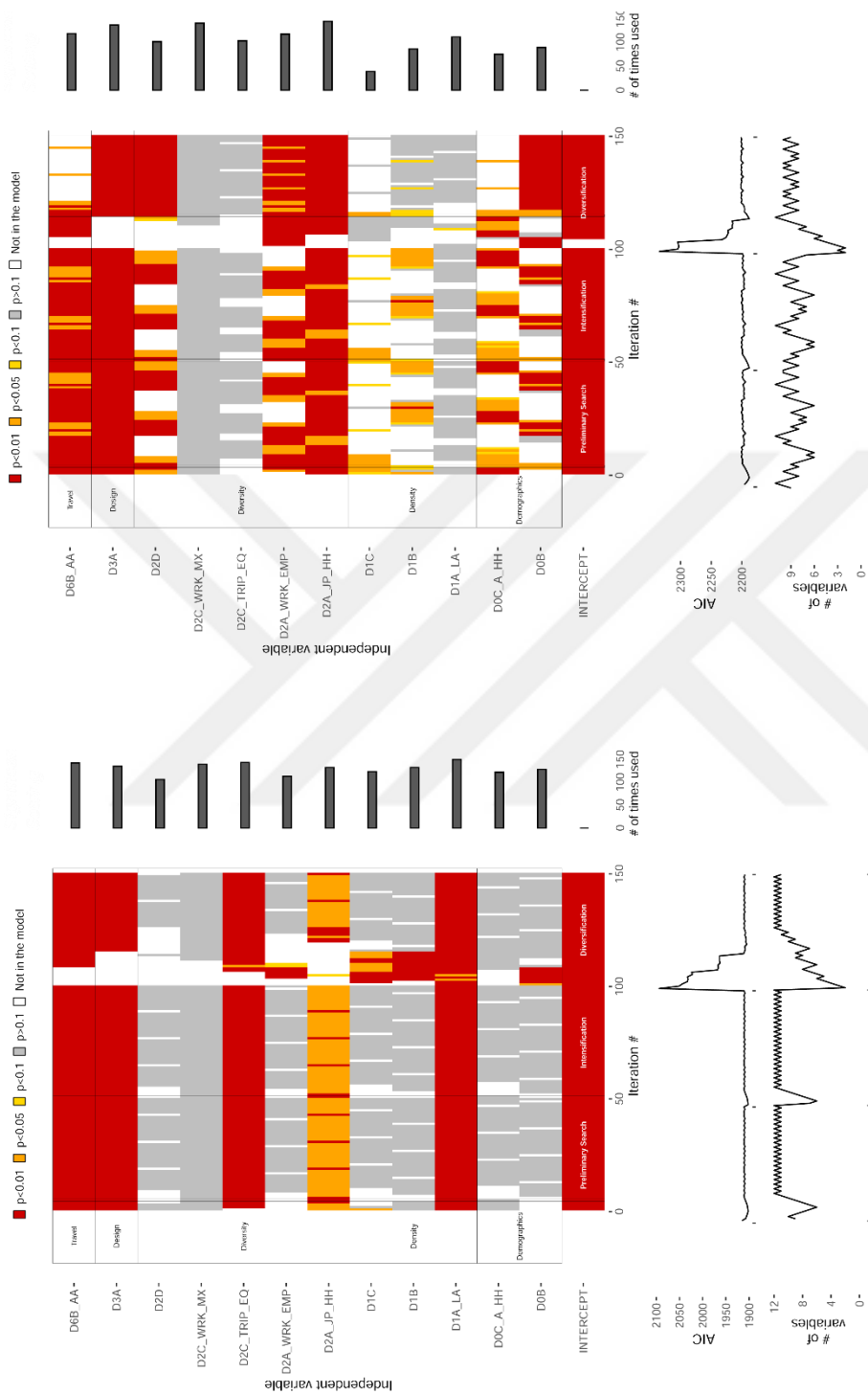


Figure C.4. FQ_VRU Negative Binomial

Figure C.3. FQ_MOT Negative Binomial

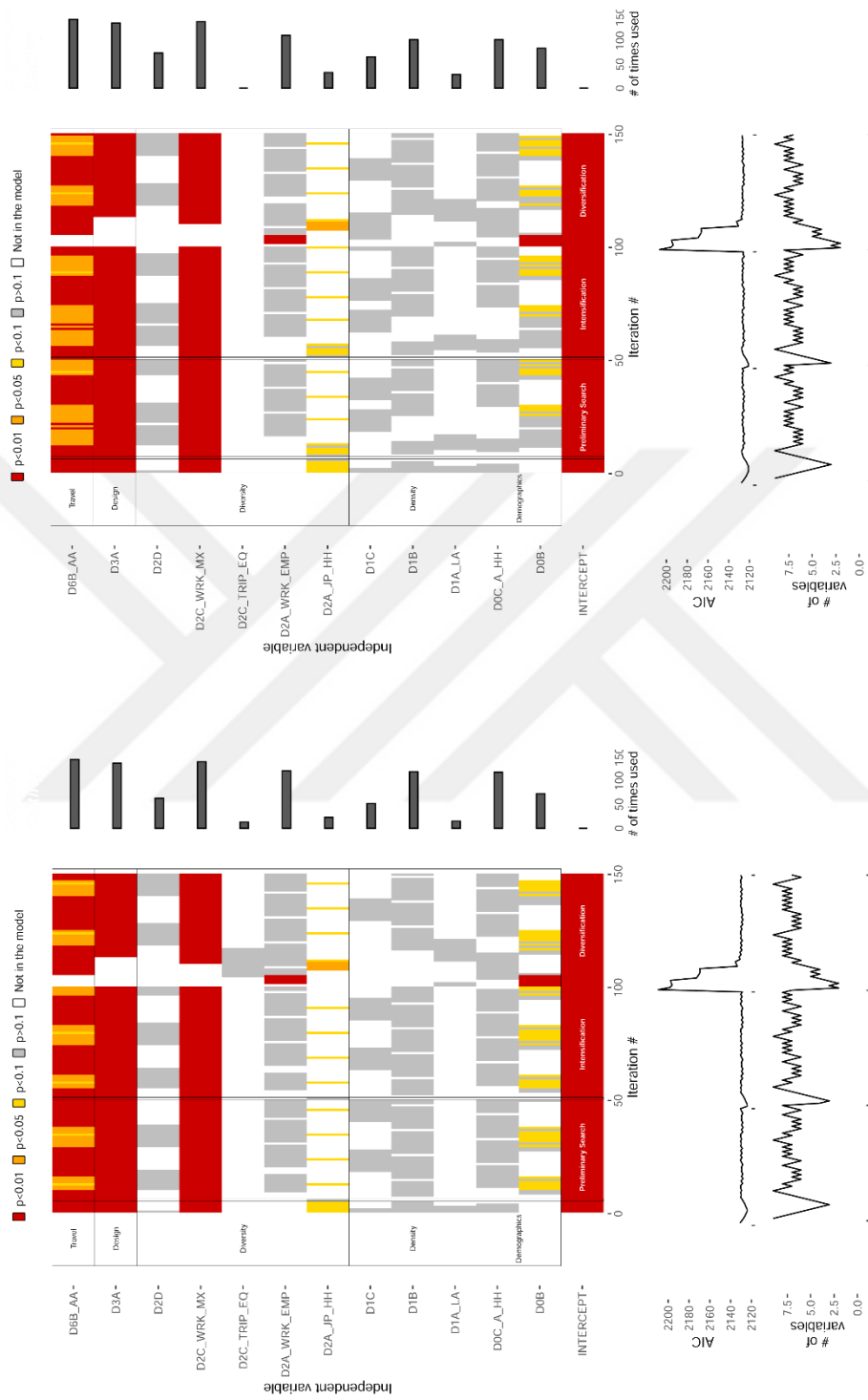


Figure C.6. KD_ALL Gamma

Figure C.5. KD_S2 Gamma

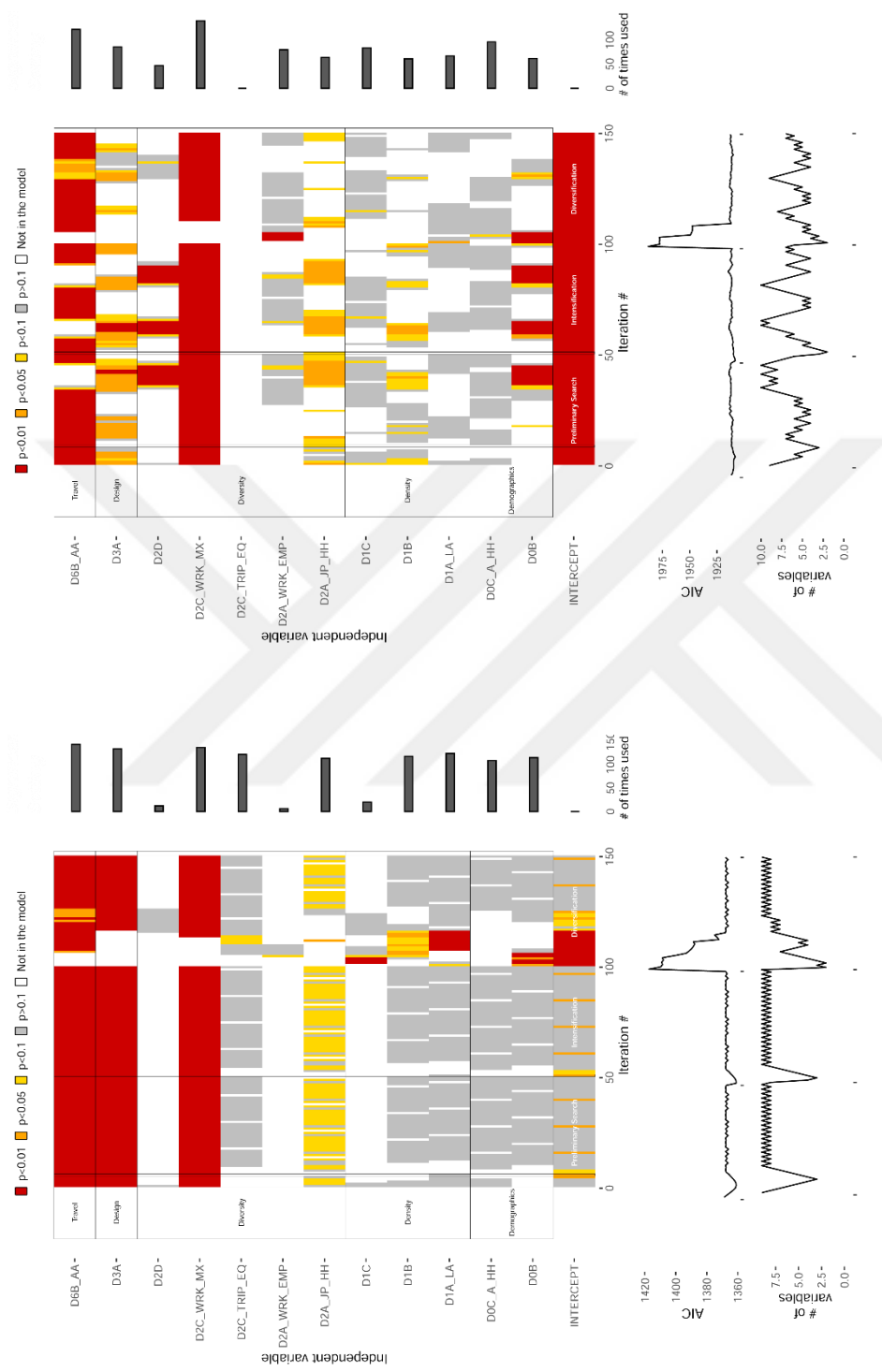


Figure C.8. KD_VRU Gamma

Figure C.7. KD_MOT Gamma

D. Results of Single and Tri-County Case Models with Key BE Variables

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.0907831	0.5378140	5.747	9.09e-09 ***
D0B	0.1247326	0.0531615	2.346	0.01896 *
D0C_B90_P	0.0027620	0.0017547	1.574	0.11547
D1A_LA	0.0105011	0.0039669	2.647	0.00812 **
D1C_5_IND	-0.2275009	0.1005721	-2.262	0.02369 *
D1C_5_ENT	0.1924782	0.0695829	2.766	0.00567 **
D1C_5_RET	0.0664507	0.0390274	1.703	0.08863 .
D2B_E5_MX	0.4407331	0.1125881	3.915	9.06e-05 ***
D3D_R1_P	0.0028347	0.0038840	0.730	0.46549
D4D	0.0127683	0.0029684	4.301	1.70e-05 ***
D6A_M_P	-0.0040949	0.0052391	-0.782	0.43445
D6A_NM_P	0.0044779	0.0069812	0.641	0.52125
D6B_AA	0.0039919	0.0004871	8.194	2.52e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial (7.5763) family taken to be 1)

Null deviance: 708.98 on 169 degrees of freedom

Residual deviance: 178.81 on 157 degrees of freedom

AIC: 1578.8

Number of Fisher Scoring iterations: 1

Theta: 7.576

Std. Err.: 0.955

2 x log-likelihood: -1550.767

Figure D.1. Single County Frequency S2

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	4.009908	0.508488	7.886	3.12e-15	***
D0B	0.139242	0.050158	2.776	0.00550	**
D0C_B90_P	0.002080	0.001651	1.260	0.20776	
D1A_LA	0.011012	0.003759	2.930	0.00339	**
D1C_5_IND	-0.210010	0.095656	-2.195	0.02813	*
D1C_5_ENT	0.156586	0.067095	2.334	0.01961	*
D1C_5_RET	0.057437	0.037776	1.520	0.12840	
D2B_E5_MX	0.481644	0.105734	4.555	5.23e-06	***
D3D_R1_P	0.001768	0.003678	0.481	0.63065	
D4D	0.012014	0.002825	4.253	2.11e-05	***
D6A_M_P	-0.003985	0.004957	-0.804	0.42148	
D6A_NM_P	0.007937	0.006639	1.196	0.23188	
D6B_AA	0.003711	0.000466	7.965	1.65e-15	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial (7.857) family taken to be 1)

Null deviance: 695.74 on 169 degrees of freedom

Residual deviance: 176.49 on 157 degrees of freedom

AIC: 1858.7

Number of Fisher Scoring iterations: 1

Theta: 7.857

Std. Err.: 0.908

2 x log-likelihood: -1830.656

Figure D.2. Single County Frequency ALL

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.9608790	0.5131923	7.718	1.18e-14 ***
D0B	0.1350774	0.0506061	2.669	0.00760 **
D0C_B90_P	0.0019207	0.0016662	1.153	0.24902
D1A_LA	0.0110554	0.0037919	2.916	0.00355 **
D1C_5_IND	-0.2078827	0.0965104	-2.154	0.03124 *
D1C_5_ENT	0.1501176	0.0677020	2.217	0.02660 *
D1C_5_RET	0.0565869	0.0381154	1.485	0.13764
D2B_E5_MX	0.4863014	0.1066858	4.558	5.16e-06 ***
D3D_R1_P	0.0020904	0.0037100	0.563	0.57314
D4D	0.0118050	0.0028503	4.142	3.45e-05 ***
D6A_M_P	-0.0035611	0.0050028	-0.712	0.47658
D6A_NM_P	0.0074950	0.0067010	1.119	0.26335
D6B_AA	0.0037200	0.0004701	7.913	2.52e-15 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial (7.7215) family taken to be 1)

Null deviance: 675.50 on 169 degrees of freedom

Residual deviance: 176.79 on 157 degrees of freedom

AIC: 1851.7

Number of Fisher Scoring iterations: 1

Theta: 7.722

Std. Err.: 0.893

2 x log-likelihood: -1823.667

Figure D.3. Single County Frequency MOT

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	0.7911794	0.8535183	0.927	0.353946
D0B	0.3045307	0.0905238	3.364	0.000768 ***
D0C_B90_P	0.0088620	0.0030729	2.884	0.003927 **
D1A_LA	0.0079058	0.0069971	1.130	0.258533
D1C_5_IND	-0.3814927	0.1679136	-2.272	0.023089 *
D1C_5_ENT	0.3357832	0.0968043	3.469	0.000523 ***
D1C_5_RET	0.0756106	0.0523360	1.445	0.148537
D2B_E5_MX	0.3240713	0.1996315	1.623	0.104515
D3D_R1_P	-0.0070646	0.0069826	-1.012	0.311666
D4D	0.0197349	0.0045984	4.292	1.77e-05 ***
D6A_M_P	-0.0151912	0.0082324	-1.845	0.064995 .
D6A_NM_P	0.0139042	0.0103086	1.349	0.177402
D6B_AA	0.0027704	0.0007439	3.724	0.000196 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial (6.0326) family taken to be 1)

Null deviance: 570.08 on 169 degrees of freedom

Residual deviance: 168.32 on 157 degrees of freedom

AIC: 805.89

Number of Fisher Scoring iterations: 1

Theta: 6.03

Std. Err.: 1.34

2 x log-likelihood: -777.889

Figure D.4. Single County Frequency VRU

Family: c("GA", "Gamma")

Mu link function: log

Mu Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.2203589	0.5078724	4.372	2.24e-05 ***
D0B	0.1597779	0.0552023	2.894	0.00434 **
D0C_B90_P	0.0024758	0.0018370	1.348	0.17969
D1A_LA	0.0208211	0.0044349	4.695	5.80e-06 ***
D1C_5_IND	-0.2370663	0.1004948	-2.359	0.01956 *
D1C_5_ENT	0.1514315	0.0709675	2.134	0.03442 *
D1C_5_RET	0.0433168	0.0403150	1.074	0.28428
D2B_E5_MX	0.5519699	0.1132428	4.874	2.66e-06 ***
D3D_R1_P	0.0050491	0.0039375	1.282	0.20163
D4D	0.0137851	0.0030808	4.475	1.47e-05 ***
D6A_M_P	-0.0051209	0.0048570	-1.054	0.29337
D6A_NM_P	0.0045256	0.0070146	0.645	0.51976
D6B_AA	0.0043391	0.0004963	8.742	3.42e-15 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sigma link function: log

Sigma Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.93982	0.05314	-17.68	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Global Deviance: 1275.593

AIC: 1303.593

SBC: 1347.494

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.220358875	0.5078723538	4.3718837	2.240651e-05
D0B	0.159777936	0.0552022726	2.8944087	4.343332e-03
D0C_B90_P	0.002475824	0.0018369974	1.3477557	1.796910e-01
D1A_LA	0.020821120	0.0044348559	4.6948809	5.802790e-06
D1C_5_IND	-0.237066341	0.1004947769	-2.3589917	1.956340e-02
D1C_5_ENT	0.151431516	0.0709675427	2.1338137	3.442133e-02
D1C_5_RET	0.043316824	0.0403149662	1.0744601	2.842756e-01
D2B_E5_MX	0.551969930	0.1132427509	4.8742187	2.664801e-06
D3D_R1_P	0.005049117	0.0039374550	1.2823300	2.016292e-01
D4D	0.013785135	0.0030807788	4.4745617	1.468876e-05
D6A_M_P	-0.005120867	0.0048570475	-1.0543169	2.933680e-01
D6A_NM_P	0.004525602	0.0070145838	0.6451704	5.197645e-01
D6B_AA	0.004339104	0.0004963367	8.7422602	3.420649e-15
(Intercept)	-0.939820237	0.0531447731	-17.6841518	4.197571e-39

Figure D.5. Single County Kernel Density S2

Family: c("GA", "Gamma")

Mu link function: log

Mu Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.2592128	0.4931823	6.609	5.81e-10 ***
D0B	0.1804069	0.0525136	3.435	0.000758 ***
D0C_B90_P	0.0012199	0.0017377	0.702	0.483734
D1A_LA	0.0194418	0.0042721	4.551	1.07e-05 ***
D1C_5_IND	-0.2186916	0.0962023	-2.273	0.024376 *
D1C_5_ENT	0.1241152	0.0681444	1.821	0.070468 .
D1C_5_RET	0.0262241	0.0381614	0.687	0.492984
D2B_E5_MX	0.5150317	0.1074928	4.791	3.83e-06 ***
D3D_R1_P	0.0032607	0.0037788	0.863	0.389522
D4D	0.0124436	0.0029134	4.271	3.37e-05 ***
D6A_M_P	-0.0047458	0.0047382	-1.002	0.318082
D6A_NM_P	0.0083588	0.0066888	1.250	0.213293
D6B_AA	0.0039515	0.0004734	8.348	3.50e-14 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sigma link function: log

Sigma Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.9866	0.0533	-18.51	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Global Deviance: 1576.162

AIC: 1604.162

SBC: 1648.063

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.259212762	0.4931822867	6.6085357	5.811200e-10
D0B	0.180406933	0.0525135769	3.4354341	7.581914e-04
D0C_B90_P	0.001219850	0.0017377144	0.7019855	4.837342e-01
D1A_LA	0.019441847	0.0042720626	4.5509273	1.068281e-05
D1C_5_IND	-0.218691587	0.0962023193	-2.2732465	2.437601e-02
D1C_5_ENT	0.124115219	0.0681443794	1.8213567	7.046844e-02
D1C_5_RET	0.026224065	0.0381613972	0.6871883	4.929844e-01
D2B_E5_MX	0.515031653	0.1074928161	4.7913123	3.827932e-06
D3D_R1_P	0.003260686	0.0037787887	0.8628919	3.895215e-01
D4D	0.012443631	0.0029134337	4.2711222	3.368545e-05
D6A_M_P	-0.004745791	0.0047381504	-1.0016126	3.180818e-01
D6A_NM_P	0.008358781	0.0066888029	1.2496678	2.132925e-01
D6B_AA	0.003951455	0.0004733503	8.3478439	3.500999e-14
(Intercept)	-0.986634543	0.0532996355	-18.5110936	3.310052e-41

Figure D.6. Single County Kernel Density ALL

Family: c("GA", "Gamma")

Mu link function: log

Mu Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.2017446	0.4962660	6.452	1.32e-09 ***
D0B	0.1767971	0.0529467	3.339	0.00105 **
D0C_B90_P	0.0010837	0.0017509	0.619	0.53683
D1A_LA	0.0195530	0.0043148	4.532	1.16e-05 ***
D1C_5_IND	-0.2174483	0.0969272	-2.243	0.02628 *
D1C_5_ENT	0.1173570	0.0687198	1.708	0.08967 .
D1C_5_RET	0.0254276	0.0383737	0.663	0.50855
D2B_E5_MX	0.5223429	0.1082796	4.824	3.32e-06 ***
D3D_R1_P	0.0036411	0.0038089	0.956	0.34058
D4D	0.0122530	0.0029325	4.178	4.87e-05 ***
D6A_M_P	-0.0042862	0.0047669	-0.899	0.36996
D6A_NM_P	0.0077939	0.0067126	1.161	0.24738
D6B_AA	0.0039578	0.0004768	8.300	4.63e-14 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sigma link function: log

Sigma Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.97898	0.05328	-18.38	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Global Deviance: 1568.037

AIC: 1596.037

SBC: 1639.938

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.201744585	0.4962660229	6.4516699	1.322470e-09
D0B	0.176797052	0.0529467055	3.3391511	1.051152e-03
D0C_B90_P	0.001083744	0.0017508618	0.6189776	5.368340e-01
D1A_LA	0.019553028	0.0043148080	4.5316102	1.158301e-05
D1C_5_IND	-0.217448318	0.0969272469	-2.2434179	2.627800e-02
D1C_5_ENT	0.117357031	0.0687198311	1.7077608	8.966998e-02
D1C_5_RET	0.025427590	0.0383737244	0.6626302	5.085452e-01
D2B_E5_MX	0.522342881	0.1082796221	4.8240183	3.319909e-06
D3D_R1_P	0.003641135	0.0038089086	0.9559523	3.405754e-01
D4D	0.012252992	0.0029325231	4.1783106	4.874858e-05
D6A_M_P	-0.004286169	0.0047669439	-0.8991439	3.699618e-01
D6A_NM_P	0.007793856	0.0067125875	1.1610808	2.473826e-01
D6B_AA	0.003957769	0.0004768367	8.3000513	4.629004e-14
(Intercept)	-0.978976266	0.0532765720	-18.3753614	7.288763e-41

Figure D.7. Single County Kernel Density MOT

```

Family: c("GA", "Gamma")
Mu link function: log
Mu Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.223994  2.377568  0.094  0.9251
D0B          0.279235  0.238221  1.172  0.2429
D0C_B90_P    0.005970  0.008219  0.726  0.4687
D1A_LA       0.015917  0.017824  0.893  0.3732
D1C_5_IND    -0.273657  0.461048 -0.594  0.5537
D1C_5_ENT    0.305586  0.310143  0.985  0.3260
D1C_5_RET    0.074919  0.189217  0.396  0.6927
D2B_E5_MX    0.336260  0.514922  0.653  0.5147
D3D_R1_P     -0.014833  0.017463 -0.849  0.3969
D4D          0.021038  0.014224  1.479  0.1411
D6A_M_P      -0.017402  0.023083 -0.754  0.4521
D6A_NM_P     0.018720  0.034270  0.546  0.5857
D6B_AA       0.003839  0.002209  1.738  0.0842 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
-----
Sigma link function: log
Sigma Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.54591  0.04322  12.63 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
-----
Global Deviance:   313.4796
      AIC:    341.4796
      SBC:    385.3808
*****
      Estimate      Std. Error      t value      Pr(>|t|)
(Intercept)  0.223994019  2.377567936  0.09421141  9.250621e-01
D0B          0.279234698  0.238221419  1.17216453  2.429180e-01
D0C_B90_P    0.005970396  0.008219475  0.72637200  4.686992e-01
D1A_LA       0.015917375  0.017823818  0.89303956  3.732115e-01
D1C_5_IND    -0.273657481  0.461048138 -0.59355512  5.536693e-01
D1C_5_ENT    0.305585867  0.310143358  0.98530521  3.259999e-01
D1C_5_RET    0.074919170  0.189217382  0.39594233  6.926886e-01
D2B_E5_MX    0.336260349  0.514921623  0.65303210  5.146968e-01
D3D_R1_P     -0.014833322  0.017462610 -0.84943329  3.969419e-01
D4D          0.021038366  0.014223576  1.47911937  1.411249e-01
D6A_M_P      -0.017401977  0.023083104 -0.75388376  4.520554e-01
D6A_NM_P     0.018719847  0.034270078  0.54624464  5.856782e-01
D6B_AA       0.003838915  0.002208850  1.73796996  8.418939e-02
(Intercept)  0.545905469  0.043220956 12.63057374  1.226111e-25

```

Figure D.8. Single County Kernel Density VRU

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.083e+00	1.059e-01	19.667	< 2e-16	***
D0B	1.564e-01	1.242e-02	12.600	< 2e-16	***
D0C_B90_P	2.530e-03	3.881e-04	6.521	7.00e-11	***
D1A_LA	2.813e-03	5.919e-04	4.753	2.01e-06	***
D1C_5_IND	-6.463e-03	3.265e-03	-1.979	0.047784	*
D1C_5_ENT	9.850e-04	3.094e-03	0.318	0.750258	
D1C_5_RET	4.652e-02	8.519e-03	5.461	4.74e-08	***
D2B_E5_MX	3.979e-01	3.233e-02	12.306	< 2e-16	***
D3D_R1_P	4.124e-03	1.313e-03	3.140	0.001691	**
D4D	1.724e-02	7.375e-04	23.380	< 2e-16	***
D6A_M_P	4.192e-03	1.070e-03	3.918	8.94e-05	***
D6A_NM_P	7.743e-03	2.048e-03	3.781	0.000156	***
D6B_AA	3.182e-03	8.219e-05	38.711	< 2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial (3.0976) family taken to be 1)

Null deviance: 8863.4 on 3380 degrees of freedom

Residual deviance: 3662.3 on 3368 degrees of freedom

AIC: 32045

Number of Fisher Scoring iterations: 1

Theta: 3.0976

Std. Err.: 0.0803

2 x log-likelihood: -32016.5180

Figure D.9. Tri-County Frequency S2

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.321e+00	1.162e-01	28.583	< 2e-16 ***
D0B	1.584e-01	1.379e-02	11.487	< 2e-16 ***
D0C_B90_P	2.944e-03	4.274e-04	6.887	5.70e-12 ***
D1A_LA	2.478e-03	6.653e-04	3.724	0.000196 ***
D1C_5_IND	-6.980e-04	3.390e-03	-0.206	0.836874
D1C_5_ENT	2.243e-02	3.383e-03	6.631	3.34e-11 ***
D1C_5_RET	7.592e-02	9.501e-03	7.991	1.34e-15 ***
D2B_E5_MX	4.095e-01	3.562e-02	11.497	< 2e-16 ***
D3D_R1_P	4.739e-03	1.462e-03	3.240	0.001194 **
D4D	2.480e-02	8.190e-04	30.288	< 2e-16 ***
D6A_M_P	3.552e-05	1.175e-03	0.030	0.975894
D6A_NM_P	9.810e-03	2.261e-03	4.339	1.43e-05 ***
D6B_AA	2.746e-03	9.211e-05	29.816	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial (2.418) family taken to be 1)

Null deviance: 8283.2 on 3380 degrees of freedom

Residual deviance: 3631.3 on 3368 degrees of freedom

AIC: 39547

Theta: 2.4180

Std. Err.: 0.0574

2 x log-likelihood: -39519.4260

Figure D.10. Tri-County Frequency ALL

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	3.238e+00	1.192e-01	27.168	< 2e-16 ***
D0B	1.618e-01	1.414e-02	11.438	< 2e-16 ***
D0C_B90_P	2.870e-03	4.383e-04	6.548	5.84e-11 ***
D1A_LA	2.567e-03	6.822e-04	3.762	0.000168 ***
D1C_5_IND	-5.710e-04	3.476e-03	-0.164	0.869529
D1C_5_ENT	2.245e-02	3.470e-03	6.470	9.80e-11 ***
D1C_5_RET	7.691e-02	9.744e-03	7.893	2.94e-15 ***
D2B_E5_MX	4.003e-01	3.653e-02	10.958	< 2e-16 ***
D3D_R1_P	5.509e-03	1.500e-03	3.674	0.000239 ***
D4D	2.493e-02	8.399e-04	29.685	< 2e-16 ***
D6A_M_P	3.729e-04	1.206e-03	0.309	0.757083
D6A_NM_P	9.647e-03	2.319e-03	4.160	3.18e-05 ***
D6B_AA	2.788e-03	9.446e-05	29.520	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial (2.2989) family taken to be 1)

Null deviance: 8160.1 on 3380 degrees of freedom

Residual deviance: 3645.5 on 3368 degrees of freedom

AIC: 39342

Theta: 2.2989

Std. Err.: 0.0544

2 x log-likelihood: -39313.9840

Figure D.11. Tri-County Frequency MOT

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	9.339e-01	1.234e-01	7.566	3.86e-14 ***
D0B	8.987e-02	1.467e-02	6.127	8.93e-10 ***
D0C_B90_P	4.554e-03	4.782e-04	9.524	< 2e-16 ***
D1A_LA	-7.138e-04	8.543e-04	-0.836	0.403
D1C_5_IND	-3.394e-03	3.592e-03	-0.945	0.345
D1C_5_ENT	2.090e-02	3.220e-03	6.492	8.45e-11 ***
D1C_5_RET	5.444e-02	9.089e-03	5.989	2.11e-09 ***
D2B_E5_MX	6.020e-01	3.919e-02	15.363	< 2e-16 ***
D3D_R1_P	-1.629e-02	1.664e-03	-9.793	< 2e-16 ***
D4D	2.266e-02	8.324e-04	27.223	< 2e-16 ***
D6A_M_P	-5.975e-03	1.231e-03	-4.853	1.21e-06 ***
D6A_NM_P	1.181e-02	2.269e-03	5.205	1.94e-07 ***
D6B_AA	1.438e-03	9.267e-05	15.519	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for Negative Binomial (3.1095) family taken to be 1)

Null deviance: 6645.2 on 3380 degrees of freedom

Residual deviance: 3682.3 on 3368 degrees of freedom

AIC: 19035

Theta: 3.109

Std. Err.: 0.114

2 x log-likelihood: -19006.889

Figure D.12. Tri-County Frequency VRU

Family: c("GA", "Gamma")

Mu link function: log

Mu Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.2125058	0.1254862	9.662	< 2e-16 ***
D0B	0.1950967	0.0157982	12.349	< 2e-16 ***
D0C_B90_P	0.0022843	0.0004583	4.985	6.53e-07 ***
D1A_LA	0.0042947	0.0008139	5.277	1.40e-07 ***
D1C_5_IND	-0.0043454	0.0035590	-1.221	0.22218
D1C_5_ENT	0.0021346	0.0038489	0.555	0.57921
D1C_5_RET	0.0439006	0.0113988	3.851	0.00012 ***
D2B_E5_MX	0.4959970	0.0383194	12.944	< 2e-16 ***
D3D_R1_P	0.0050135	0.0015778	3.177	0.00150 **
D4D	0.0178635	0.0009261	19.288	< 2e-16 ***
D6A_M_P	0.0018603	0.0012737	1.461	0.14423
D6A_NM_P	0.0064298	0.0024430	2.632	0.00853 **
D6B_AA	0.0035984	0.0001131	31.821	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sigma link function: log

Sigma Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.36518	0.01133	-32.23	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

No. of observations in the fit: 3381

Degrees of Freedom for the fit: 14

Residual Deg. of Freedom: 3367

at cycle: 2

Global Deviance: 26538.51

AIC: 26566.51

SBC: 26652.27

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.212505763	0.1254861664	9.6624656	8.330094e-22
D0B	0.195096722	0.0157981836	12.3493135	2.685380e-34
D0C_B90_P	0.002284268	0.0004582736	4.9845065	6.525482e-07
D1A_LA	0.004294718	0.0008139202	5.2765830	1.399358e-07
D1C_5_IND	-0.004345434	0.0035589509	-1.2209872	2.221763e-01
D1C_5_ENT	0.002134603	0.0038489498	0.5545937	5.792095e-01
D1C_5_RET	0.043900626	0.0113988036	3.8513363	1.196612e-04
D2B_E5_MX	0.495997014	0.0383193639	12.9437695	1.963111e-37
D3D_R1_P	0.005013484	0.0015778266	3.1774620	1.499149e-03
D4D	0.017863487	0.0009261364	19.2881806	1.026941e-78
D6A_M_P	0.001860335	0.0012737145	1.4605589	1.442298e-01
D6A_NM_P	0.006429798	0.0024429997	2.6319276	8.528944e-03
D6B_AA	0.003598439	0.0001130851	31.8206384	1.669555e-194
(Intercept)	-0.365184490	0.0113323666	-32.2249095	8.103868e-199

Figure D.13. Tri-County Kernel Density S2

Family: c("GA", "Gamma")

Mu link function: log

Mu Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.4019387	0.1224797	19.611	< 2e-16 ***
D0B	0.2066446	0.0146852	14.072	< 2e-16 ***
D0C_B90_P	0.0029585	0.0004338	6.820	1.08e-11 ***
D1A_LA	0.0039824	0.0007340	5.425	6.19e-08 ***
D1C_5_IND	0.0021163	0.0039470	0.536	0.59187
D1C_5_ENT	0.0238605	0.0038892	6.135	9.50e-10 ***
D1C_5_RET	0.0726904	0.0111678	6.509	8.69e-11 ***
D2B_E5_MX	0.4897791	0.0363638	13.469	< 2e-16 ***
D3D_R1_P	0.0062295	0.0014951	4.166	3.17e-05 ***
D4D	0.0251527	0.0008923	28.189	< 2e-16 ***
D6A_M_P	-0.0015612	0.0012404	-1.259	0.20826
D6A_NM_P	0.0073815	0.0023286	3.170	0.00154 **
D6B_AA	0.0030257	0.0001007	30.053	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sigma link function: log

Sigma Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.42149	0.01141	-36.94	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

No. of observations in the fit: 3381

Degrees of Freedom for the fit: 14

Residual Deg. of Freedom: 3367

at cycle: 2

Global Deviance: 33514.47

AIC: 33542.47

SBC: 33628.23

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.401938677	0.1224797162	19.6109099	3.584412e-81
D0B	0.206644650	0.0146852262	14.0716014	9.623313e-44
D0C_B90_P	0.002958458	0.0004338163	6.8196096	1.077647e-11
D1A_LA	0.003982351	0.0007340135	5.4254463	6.189350e-08
D1C_5_IND	0.002116328	0.0039470437	0.5361806	5.918692e-01
D1C_5_ENT	0.023860508	0.0038891868	6.1350892	9.499958e-10
D1C_5_RET	0.072690353	0.0111677955	6.5089259	8.692100e-11
D2B_E5_MX	0.489779069	0.0363638121	13.4688593	2.591016e-40
D3D_R1_P	0.006229485	0.0014951455	4.1664740	3.171211e-05
D4D	0.025152680	0.0008923012	28.1885538	3.866044e-157
D6A_M_P	-0.001561216	0.0012404191	-1.2586195	2.082551e-01
D6A_NM_P	0.007381479	0.0023286386	3.1698686	1.538761e-03
D6B_AA	0.003025681	0.0001006780	30.0530559	5.391682e-176
(Intercept)	-0.421493629	0.0114114540	-36.9360144	4.976350e-251

Figure D.14. Tri-County Kernel Density ALL

Family: c("GA", "Gamma")

Mu link function: log

Mu Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.3130995	0.1272697	18.175	< 2e-16 ***
D0B	0.2100353	0.0152722	13.753	< 2e-16 ***
D0C_B90_P	0.0029148	0.0004508	6.466	1.15e-10 ***
D1A_LA	0.0040896	0.0007672	5.330	1.05e-07 ***
D1C_5_IND	0.0023561	0.0041524	0.567	0.57047
D1C_5_ENT	0.0237657	0.0040191	5.913	3.69e-09 ***
D1C_5_RET	0.0738844	0.0116003	6.369	2.16e-10 ***
D2B_E5_MX	0.4800127	0.0378028	12.698	< 2e-16 ***
D3D_R1_P	0.0070203	0.0015537	4.518	6.44e-06 ***
D4D	0.0253268	0.0009281	27.287	< 2e-16 ***
D6A_M_P	-0.0012438	0.0012889	-0.965	0.33460
D6A_NM_P	0.0072208	0.0024123	2.993	0.00278 **
D6B_AA	0.0030735	0.0001047	29.359	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sigma link function: log

Sigma Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.38339	0.01136	-33.75	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

No. of observations in the fit: 3381

Degrees of Freedom for the fit: 14

Residual Deg. of Freedom: 3367

at cycle: 2

Global Deviance: 33338.39

AIC: 33366.39

SBC: 33452.16

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.313099531	0.1272696986	18.1747860	1.725236e-70
D0B	0.210035334	0.0152722490	13.7527442	6.522733e-42
D0C_B90_P	0.002914790	0.0004507723	6.4662144	1.149956e-10
D1A_LA	0.004089626	0.0007672441	5.3302805	1.045158e-07
D1C_5_IND	0.002356126	0.0041524195	0.5674104	5.704732e-01
D1C_5_ENT	0.023765731	0.0040190571	5.9132604	3.689247e-09
D1C_5_RET	0.073884399	0.0116003030	6.3691783	2.158093e-10
D2B_E5_MX	0.480012708	0.0378027760	12.6978164	4.041181e-36
D3D_R1_P	0.007020321	0.0015537136	4.5184141	6.444602e-06
D4D	0.025326820	0.0009281479	27.2874815	2.716899e-148
D6A_M_P	-0.001243778	0.0012888512	-0.9650286	3.345998e-01
D6A_NM_P	0.007220798	0.0024123350	2.9932816	2.780008e-03
D6B_AA	0.003073466	0.0001046866	29.3587389	6.869864e-169
(Intercept)	-0.383385234	0.0113591463	-33.7512367	2.339674e-215

Figure D.15. Tri-County Kernel Density MOT

Family: c("GA", "Gamma")

Mu link function: log

Mu Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.0601679	0.2505475	-0.240	0.81023
D0B	0.1525628	0.0296970	5.137	2.95e-07 ***
D0C_B90_P	0.0040016	0.0008846	4.524	6.29e-06 ***
D1A_LA	0.0005403	0.0013136	0.411	0.68087
D1C_5_IND	-0.0023157	0.0064628	-0.358	0.72013
D1C_5_ENT	0.0273045	0.0088992	3.068	0.00217 **
D1C_5_RET	0.0544697	0.0230619	2.362	0.01824 *
D2B_E5_MX	0.6850925	0.0737472	9.290	< 2e-16 ***
D3D_R1_P	-0.0148704	0.0030441	-4.885	1.08e-06 ***
D4D	0.0245411	0.0018155	13.517	< 2e-16 ***
D6A_M_P	-0.0071003	0.0025411	-2.794	0.00523 **
D6A_NM_P	0.0108891	0.0050085	2.174	0.02976 *
D6B_AA	0.0017840	0.0002096	8.513	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Sigma link function: log

Sigma Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.29505	0.01014	29.1	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

No. of observations in the fit: 3381

Degrees of Freedom for the fit: 14

Residual Deg. of Freedom: 3367

at cycle: 2

Global Deviance: 12451.6

AIC: 12479.6

SBC: 12565.37

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.0601678915	0.2505474611	-0.2401457	8.102319e-01
D0B	0.1525628354	0.0296970366	5.1373084	2.945466e-07
D0C_B90_P	0.0040015509	0.0008845980	4.5235812	6.290111e-06
D1A_LA	0.0005402835	0.0013135781	0.4113067	6.808738e-01
D1C_5_IND	-0.0023156980	0.0064628063	-0.3583115	7.201327e-01
D1C_5_ENT	0.0273044642	0.0088992034	3.0681920	2.170718e-03
D1C_5_RET	0.0544697086	0.0230618702	2.3618947	1.823842e-02
D2B_E5_MX	0.6850925004	0.0737471512	9.2897487	2.697468e-20
D3D_R1_P	-0.0148704011	0.0030440942	-4.8850003	1.082498e-06
D4D	0.0245411087	0.0018155445	13.5172169	1.390507e-40
D6A_M_P	-0.0071003422	0.0025410970	-2.7942036	5.232212e-03
D6A_NM_P	0.0108891246	0.0050084513	2.1741500	2.976330e-02
D6B_AA	0.0017840110	0.0002095576	8.5132248	2.511213e-17
(Intercept)	0.2950510284	0.0101389490	29.1007507	2.832295e-166

Figure D.16. Tri-County Kernel Density VRU

E. Additional Goodness-of-Fit Statistics for Single and Tri-County Case Models

Table E.2 Goodness-of-Fit Statistics of Single County Frequency Models

	FQ_S1	FQ_S2	FQ_S3	FQ_ALL	FQ_MOT	FQ_VRU
RMSE	29.83	23.26	3.14	56.54	56.67	2.21
MAE	20.72	16.58	2.22	38.33	37.32	1.65
MAPE	0.30	0.30	0.47	0.28	0.28	0.47
NRMSE_sd	0.40	0.37	0.57	0.40	0.43	0.28

Table E.3 Goodness-of-Fit Statistics of Single County Density Models

	KD_S1			KD_ALL		
	Gamma	Lognormal	Weibull	Gamma	Lognormal	Weibull
RMSE	17.18	15.14	16.00	28.83	27.39	31.98
MAE	10.96	10.63	10.99	19.25	18.50	20.33
MAPE	0.31	0.29	0.37	0.28	0.26	0.33
NRMSE_sd	0.33	0.29	0.31	0.30	0.28	0.33
	KD_S2			KD_MOT		
	Gamma	Lognormal	Weibull	Gamma	Lognormal	Weibull
RMSE	13.83	11.51	15.73	28.01	26.64	31.29
MAE	8.75	8.02	9.62	18.80	18.12	19.86
MAPE	0.30	0.27	0.38	0.28	0.26	0.34
NRMSE_sd	0.32	0.27	0.37	0.30	0.29	0.34
	KD_S3			KD_VRU		
	Gamma	Lognormal	Weibull	Gamma	Lognormal	Weibull
RMSE	2.04	61.02	2.23	2.61	83.77	2.53
MAE	1.51	11.96	1.54	1.41	20.65	1.39
MAPE	5.50	3.03	4.29	3.43	5.36	2.52
NRMSE_sd	0.64	19.07	0.70	0.53	17.13	0.52

Table E.4 Goodness-of-Fit Statistics of Tri-County Frequency Models

	FQ_S1	FQ_S2	FQ_S3	FQ_ALL	FQ_MOT	FQ_VRU
RMSE	3327.18	1706.79	56.22	4956.93	5196.28	13.09
MAE	123.91	60.12	5.62	179.76	183.05	3.82
MAPE	0.90	0.76	0.73	0.76	0.79	0.71
NRMSE_sd	21.10	23.78	5.49	22.06	23.59	1.69

Table E.5 Goodness-of-Fit Statistics of Tri-County Density Models

	KD_S1			KD_ALL		
	Gamma	Lognormal	Weibull	Gamma	Lognormal	Weibull
RMSE	3498.95	2547.68	4601.44	6100.74	2753.93	11359.13
MAE	91.55	71.06	116.03	150.29	85.15	254.67
MAPE	1.27	0.84	1.55	0.77	0.60	0.98
NRMSE_sd	42.94	31.27	56.47	51.45	23.22	95.79
	KD_S2			KD_MOT		
	Gamma	Lognormal	Weibull	Gamma	Lognormal	Weibull
RMSE	2740.39	1873.84	4592.42	6354.62	3083.29	12048.60
MAE	62.90	47.63	98.90	154.39	91.40	266.65
MAPE	0.96	0.72	1.20	0.80	0.62	1.02
NRMSE_sd	71.48	48.88	119.79	54.60	26.49	103.52
	KD_S3			KD_VRU		
	Gamma	Lognormal	Weibull	Gamma	Lognormal	Weibull
RMSE	121.75	3198.92	131.81	14.48	1120.65	16.60
MAE	4.48	60.76	4.61	1.96	24.23	1.99
MAPE	5.53	3.87	5.09	5.10	3.76	4.65
NRMSE_sd	23.24	610.54	25.16	3.90	302.00	4.47

CURRICULUM VITAE

PERSONAL INFORMATION

Surname, Name: Alişan, Onur

EDUCATION

Degree	Institution	Year of Graduation
PhD	METU, City and Regional Planning	2021
BCRP	METU, City and Regional Planning	2014
MS	METU, Industrial Engineering	2008
BS	METU, Business Administration	2004
High School	Gazi Anatolian High School, Ankara	2000

WORK EXPERIENCE

Year	Place	Enrollment
2014-2021	METU Dept. of City and Regional Planning	Research Assistant
2018-2019	Fulbright-VSR	Florida State University
2006-2009	Ministry of Culture and Tourism	Civil Servant

FOREIGN LANGUAGES

Advanced English

PUBLICATIONS

Alisan, Onur, Mahyar Ghorbanzadeh, Mehmet Baran Ulak, Ayberk Kocatepe, Eren Erman Ozguven, Mark Horner, and Wenrui Huang. 2020. "Extending Interdiction and Median Models to Identify Critical Hurricane Shelters." *International Journal of Disaster Risk Reduction* 43(March 2019):101380. doi: 10.1016/j.ijdr.2019.101380.

Alisan, Onur, Hediye Tuydes-Yaman, Ayberk Kocatepe, and Eren Erman Ozguven. 2018. "Developing Performance Measures to Evaluate Impact of Cross-County Collaboration for Populations with Access and Functional Needs." *International Journal of Disaster Risk Reduction* 31(August):873–85. doi: 10.1016/j.ijdr.2018.08.007.

Alisan, Onur, Ayberk Kocatepe, Hediye Tuydes-Yaman, Eren Erman Ozguven, and Hidayet Ozel. 2017. "Benefits of Managing the Capacity of Special Needs Shelters with Cross-County Collaboration." *Transportation Research Record: Journal of the Transportation Research Board* 2604(1):131–43. doi: 10.3141/2604-16.

PROCEEDINGS

Alisan, Onur, Mahyar Ghorbanzadeh, Mehmet Baran Ulak, Ayberk Kocatepe, Eren Erman Ozguven, Mark Horner, and Wenrui Huang. 2020. "Identifying Critical Hurricane Shelters and Alternatives in Case of Interdiction: A Case Study in Southeast Florida".. Transportation Research Board's 99th Annual Meeting, Washington D.C.

Alisan, Onur, Hediye Tuydes-Yaman, Ayberk Kocatepe, and Eren Erman Ozguven. 2018. "Performance and Cost of Cross-County Collaboration in Special Needs Shelter (SPNS) Planning". Transportation Research Board's 97th Annual Meeting, Washington D.C.

Alisan, Onur, Ayberk Kocatepe, Hediye Tuydes-Yaman, Eren Erman Ozguven, and Hidayet Ozel. 2017. "Benefits of Managing Special Needs Shelters (SPNS) Capacities with Utilizing Cross-County Collaboration: A Case Study Application in Florida". Transportation Research Board's 96th Annual Meeting, Washington D.C.