Collaborative Sciences Center for ROAD SAFETY

www.roadsafety.unc.edu

R4: Completing the Picture of Traffic Injuries: Understanding Data Needs and Opportunities for Road Safety

THE UNIVERSITY OF TENNESSEE KNOXVILLE

Project Start Date: 2017-03-01

Project End Date: 2018-02-28

Collaborative Sciences Center for ROAD SAFETY

Berkeley UNIVERSITY OF CALIFORNIA

THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

FAU FLORIDA ATLANTIC UNIVERSITY

September 4, 2019

Primary Objectives:

- Develop a complete the picture of crashes and determine which elements of data that exist outside of conventional crash data that can contribute to this picture. These elements likely include EMS, ED, DMV, Health Expenditure, Census, and Land Use, among others.
- Identify innovative statistical, probabilistic, and spatial data visualization tools to link crashes with other records, either by record-matching, or augmenting datasets based on spatial or temporal indicators to perform more-advanced safety analysis
- Perform five applications

Previous Example

- In USA at State level:
  - Crash Outcome Data Evaluation System (CODES)
  - Crash Medical Outcomes Data Project (CMOD)

CODES

- Aim is to Link crash, vehicle, and behavior characteristics to their specific medical and financial outcomes
- Provide a comprehensive understanding of motor vehicle crash outcomes
- CODES data reside in the States where the linkage originated, and NHTSA does not disseminate CODES data.
- Conducted in 15 states (2013)
- Methodology varies by states (Probabilistic and Deterministic) .
- Limitations of CODES or similar program (e.g., CMOD)
  - Only considered health-oriented database and police databases

Literature Review

Linkage Methods:

- Interface
  - Real time interaction between databases
  - Need Compilation of data from multiple agencies
- Direct Method
  - Databases Share a unique single identifier
  - E.g., SSN, License Number
  - Difficulty to access data; Health Insurance Portability and Accountability Act (HIPAA)

Literature Review

- Deterministic Linkage
  - Multiple quasi-unique fields that describe an individual who was involved in an MVC: Time and Geographical data elements, gender, age
  - Need a scoring system (based on researchers' Judgment) to identify the matches
- Probabilistic Linkage
  - Aim: generate the probability that a pair of records describe same person and event.
  - Address the Judgments' concerns
  - It is the current practice in CODES program
- Spatial method

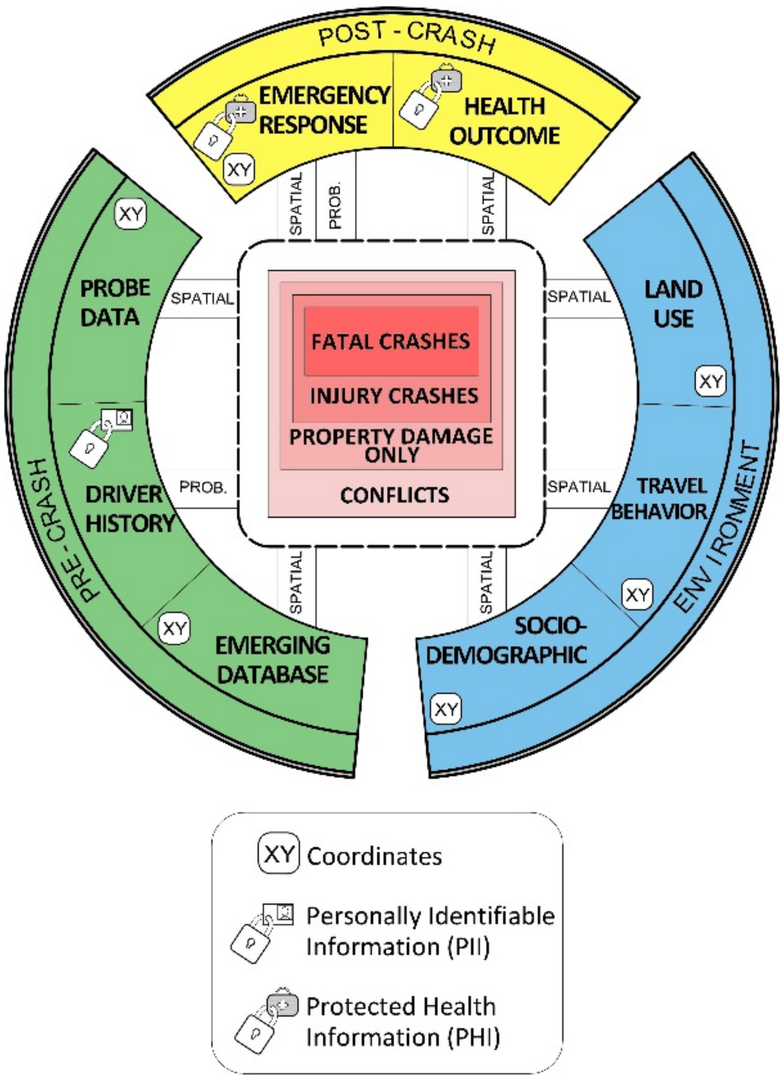
Studies Classification

Linking Police Crash databases and Health Oriented Data Applications

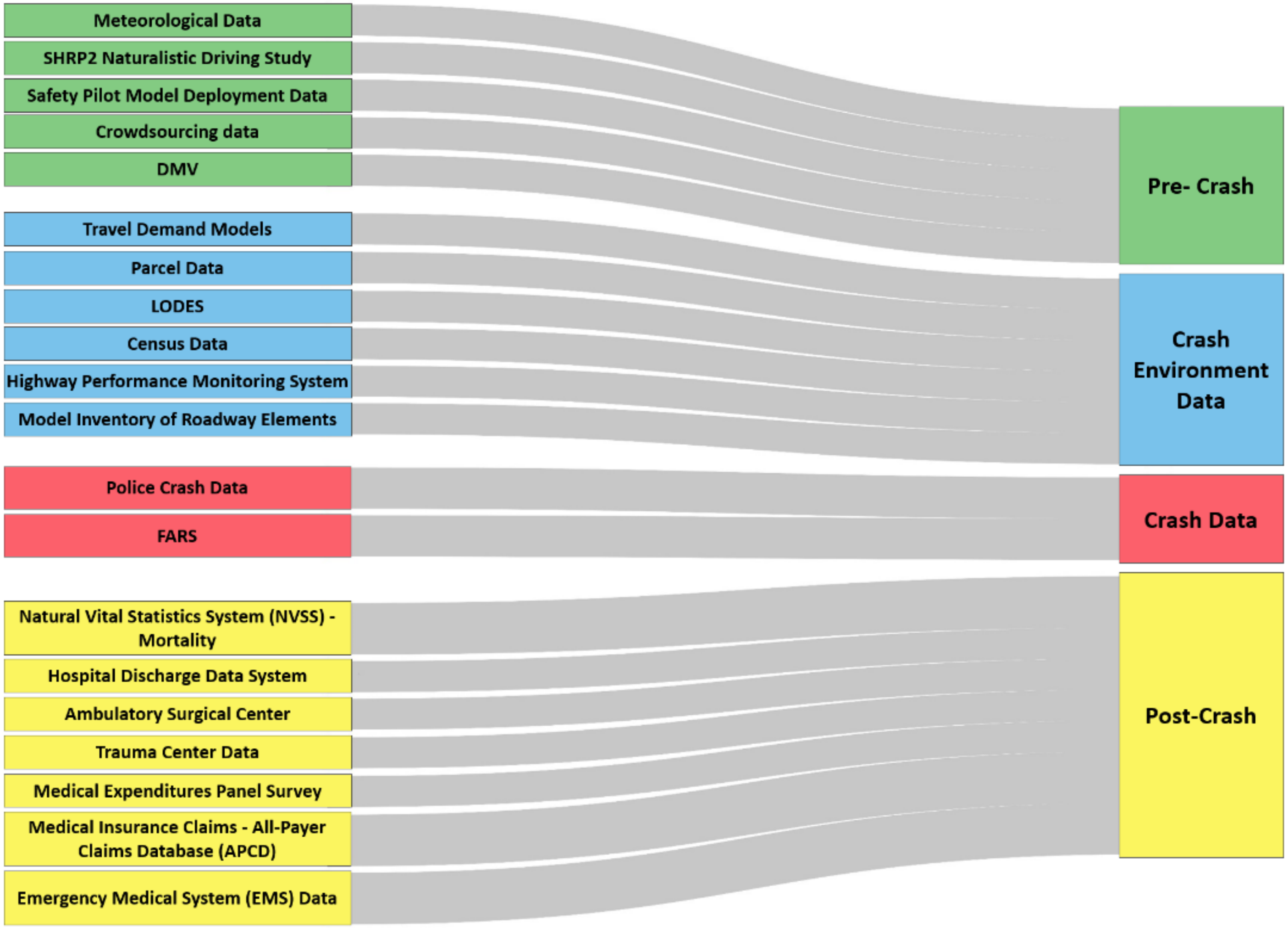
- The studies could be categorized into:
  1. Comparison Of The KABCO Scale and AIS Injury Severity Scale.
  2. Factors Influencing Injury Severity
  3. Underreporting of Traffic Crashes
  4. Substance Abuse and Motor-Vehicle Crashes
  5. Evaluation of Safety Equipment
  6. Analyzing Specific Road User traffic crashes

Introducing Databases

- Description of the database
- How to access the database w/wo PHI
- Consistency between state/local
- Variables within the dataset



Exhaustive List of Databases



Data Descriptions

- Each dataset was outlined in report
  - Access (PII)
  - National Standards and State Consistency
  - Variables types in database
  - Linking Methods

Case Studies

- Case Study 1: Underreporting Bike/Ped
- Case Study 2: EMS Response Time
- Case Study 4: Aggregate Crash Prediction Model
- Case Study 3: Accessibility Measures and Safety
- Case Study 5: Seat Belt Use

Case Study 1
Evaluating Research On Data Linkage to Assess Underreporting Pedestrian and Bicyclist Injury In Police Crash Data

Sarah Doggett[a]*, David Ragland[a], Grace Felschundneff

[a] Safe Transportation Research and Education Center (SafeTREC). University of California, Berkeley

*Corresponding Author: doggett_sarah@berkeley.edu

Complete Picture – UC Berkeley Component

- Builds on existing effort to perform road safety research that explores core safety issues. This project addresses post-crash issues by considering EMS, ED, and hospital data.
- The project will support the development of data sets, i.e., linked crash and medical data, which are designed to (i) clarify the true burden of pedestrian and bicyclist injury (Case 1) and (ii) improve post-crash management of injury (Case 2).

CASE STUDY 1

Linking Crash and Post-Crash Data to Get a "Complete Picture" of Pedestrian/Bicyclist Injury

- Rationale:

- Crash reports submitted by police are primary sources of data to assess pedestrian and bicyclist injury and to develop countermeasures.

- A number of studies have identified pedestrian and bicyclist injuries that are not recorded in police reports.

- Linking police reported and medical data can provide a more "complete picture" of pedestrian and bicyclist injury

CASE STUDY 1

Linking Crash and Post-Crash Data to Get a “Complete Picture” of Pedestrian/Bicyclist Injury

- Year 1 Activity

- Literature review and bibliographic summary of previous articles/reports linking police reported pedestrian/bicyclist injury and medical data describing pedestrian/bicyclist injury

- Critical review of this literature focusing on findings and methodological issues and solutions related to matching procedures.

Case Study 2

Pre-Hospital Response Time and Traumatic Injury—A Review

Sarah Doggett[a]*, David R. Ragland[a], Grace Felschundneff[a]

[a]Safe Transportation Research and Education Center (SafeTREC), University of California, Berkeley

*corresponding author
Email: doggett_sarah@berkeley.edu

CASE STUDY 2

Develop measures of EMS response times (time from crash to dispatch, time from dispatch to arrival of EMS crew, time on site, time to ED, etc.) as a function of rural versus urban, cell phone coverage, trauma center location, etc.

- EMS response time has been identified in some studies as a factor influencing degree of injury and probably of fatality.
- A review of distances between crashes in California and the nearest trauma center/ER indicates potential times of up to three hours.
- There is a least anecdotal evidence of even longer times bases on communication and other issues.
- There is a need to document actual response times as a function of distance and other factors.

CASE STUDY 2

Develop measures of EMS response times (time from crash to dispatch, time from dispatch to arrival of EMS crew, time on site, time to ED, etc.) as a function of rural versus urban, cell phone coverage, trauma center location, etc.

- Year 1 Activities

- Literature review and bibliographic summary of articles/reports that address impact of extended response time and factors influencing response time and other quality of on-site care.

- Critical review of this literature focusing on findings and methodological issues in studies of response times and in studies of implications of response time and other factors on outcomes.

Case Study 4
Home-Based Approach: A Complementary Definition of Road Safety

Amin Mohamadi Hezaveh[a], Christopher R. Cherry[a*]

[a] Civil and Environmental Engineering, University of Tennessee, Knoxville, TN, United States,

* Corresponding Author: Cherry@utk.edu

BIG ORANGE BIG IDEAS

Aims of the study

1. Identify neighborhoods that have higher risk of involvement in traffic crashes (hotspots)
2. Investigate the relationship between socio-demographic variables and risk of involvement in traffic crashes.
3. Compare new definition with traditional definition of the road safety

   Current definition of road safety (i.e., **Location-Based Approach**)

   "*the number of accidents (crashes) by kind and severity, expected to occur on the entity during a specified period.”*(Hauer 1997)

   Instead we used (**Home-Based Approach**):

   The expected number of crashes that road users who lives in a certain geographic area have during a specified period.

Database

- Databases
    1. Census Tract Data of TN
    2. Highway Performance Monitoring System
    3. Police Crash Report in TN
- We used Spatial Join to merge databases
- Model: Geographically Weighted Poisson Regression
    - to investigate the relationship between sociodemographic variables and risk of involvement in traffic crashes at zonal level

New Definition

Current definition of road safety (i.e., **Location-Based Approach**)

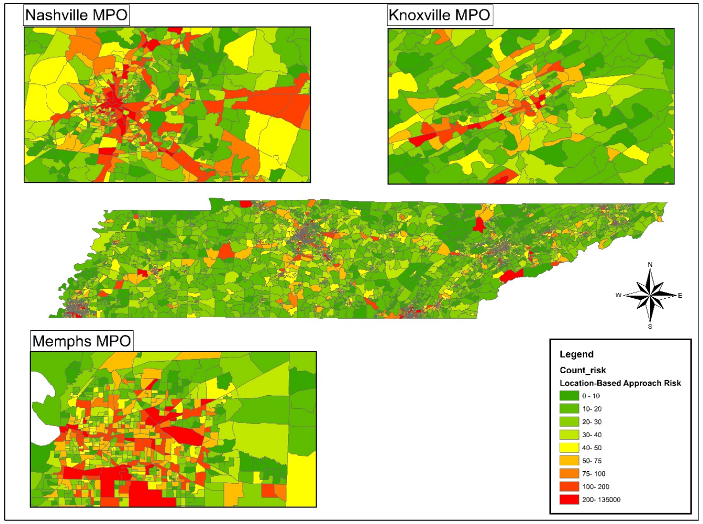
"*the number of accidents (crashes) by kind and severity, expected to occur on the entity during a specified period.”*(Hauer 1997)

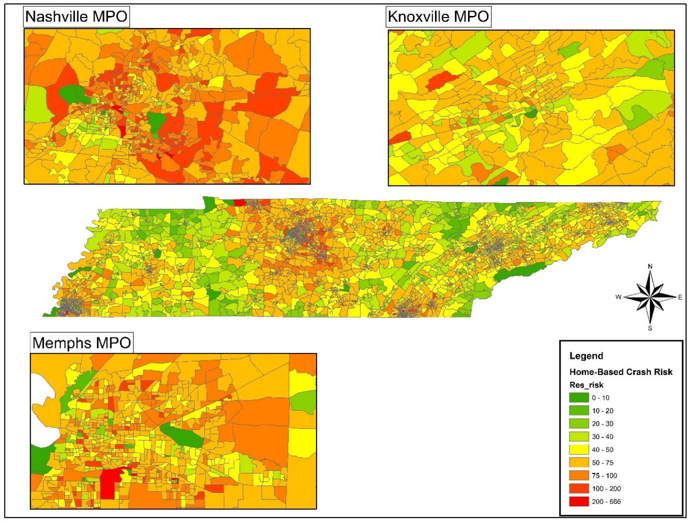
Instead we used (**Home-Based Approach**):

The expected number of crashes that road users who lives in a certain geographic area have during a specified period.

BIG ORANGE BIG IDEAS

Comparing Crash Risk*





Location-Based Approach vs Home-Based Approach

*Crash Risk: Crash Frequency divided by 1000 population

Correlation between HBA and LBA crash frequency: 0.19 (p value = 0.000)

Results of Poisson and GWPR model for LBA

DV: LBA: Number of Crashes that Occurred in a Census Tract

	Estimate	Standard Error	z(Est/SE)	Mean	STD	Min	Lower Quartile	Median	Upper Quartile	Max
Intercept	3.956	0.025	158.781	3.697	4.038	-18.294	1.083	3.760	6.397	17.028
Population	0.177	0.001	126.602	0.226	0.259	-0.958	0.083	0.213	0.397	1.452
Age cohorts proportion										
Under 16 years	-0.935	0.021	-45.545	-1.225	2.327	-10.305	-2.642	-1.152	0.224	6.550
between 16-42	0.188	0.017	11.119	-0.207	1.943	-8.032	-1.634	-0.192	1.180	6.176
between 43-59	-0.339	0.025	-13.612	-0.862	2.434	-13.003	-2.372	-0.604	0.826	7.871
White Race Proportion	0.200	0.006	32.099	-0.013	1.893	-10.070	-0.826	-0.055	0.716	10.462
Average Travel Time to Work	-0.025	0.000	-99.451	-0.021	0.034	-0.187	-0.043	-0.017	0.000	0.103
Household Income	-2.15E-03	8.60E-05	-24.944	-4.08E-03	1.25E-02	-5.78E-02	-1.21E-02	-2.54E-03	3.97E-03	4.78E-02
Vehicle Ownership										
Household With No-Vehicle	1.709	0.018	96.098	1.278	3.236	-16.932	-0.501	1.519	3.286	12.571
Household With 1 or 2 vehicles	0.858	0.012	69.457	0.690	1.642	-7.321	-0.267	0.592	1.632	6.787
Daily-VMT (10,000 Miles)	0.005	0.000	482.648	0.007	0.004	-0.004	0.005	0.007	0.009	0.025
Travel Model To work										
Personal Vehicle	0.047	0.019	2.539	0.872	2.906	-8.966	-0.812	0.789	2.425	16.626
Active More	1.100	0.028	38.701	1.349	5.889	-29.868	-1.899	1.085	4.488	38.338
Education										
College Degree	0.862	0.018	48.644	0.420	1.733	-6.645	-0.681	0.398	1.461	7.921
Bachelor Degree	0.441	0.016	28.430	0.250	1.889	-5.514	-1.108	0.262	1.609	6.258
Classic AIC:	281805.3			78007.81						
AICc:	281805.5			80603.01						
Percent deviance explained	0.46			0.86						
Deviance:	281775.3			74508.61						
MAD	73.6			35.8						
R^2 Poisson	0.59			0.92						
Lagrange Multiplier	0.28			0.04						
Moran's I of residuals	0.08			-0.01						
Bandwidth	Not applicable			70.00						

Results of Poisson and GWPR model for HBA

DV: HBA: Number of Crashes per population among residents of a Census Tract

	Estimate	Standard Error	z(Est/SE)	Mean	STD	Min	Lower Quartile	Median	Upper Quartile	Max
Intercept	3.044	0.024	127.830	3.871	1.942	-10.303	2.864	4.020	5.027	9.430
Population	0.337	0.001	390.872	0.503	0.125	0.180	0.422	0.501	0.582	0.909
Age cohorts proportion										
Under 16 years	0.218	0.016	13.678	-0.457	0.816	-3.816	-0.999	-0.425	0.091	2.625
between 16-42	0.583	0.014	41.992	-0.179	0.902	-5.505	-0.704	-0.128	0.420	3.776
between 43-59	0.736	0.020	36.364	0.244	0.938	-3.313	-0.361	0.282	0.784	4.155
White Race Proportion	-0.203	0.005	-44.791	-0.010	1.037	-3.758	-0.429	-0.112	0.290	12.610
Average Travel Time to Work	0.010	0.000	56.306	0.005	0.014	-0.052	-0.003	0.005	0.013	0.075
Household Income	0.001	0.000	14.479	0.001	0.005	-0.016	-0.002	0.000	0.003	0.021
Vehicle Ownership										
Household With No-Vehicle	1.56E-04	1.73E-04	0.903	0.001	0.011	-0.059	-0.005	0.001	0.008	0.049
Household With 1 or 2 vehicles	0.278	0.010	27.266	0.287	0.647	-2.177	-0.091	0.285	0.680	3.775
Daily-VMT (10,000 Miles)	5.69E-03	1.30E-04	43.770	0.005	0.011	-0.059	-0.001	0.004	0.010	0.064
Travel Model To work										
Personal Vehicle	0.857	0.020	43.346	0.422	1.245	-4.234	-0.276	0.330	1.072	7.583
Active More	0.018	0.029	0.609	-0.520	2.230	-8.374	-1.933	-0.540	0.801	8.372
Education										
College Degree	0.865	0.014	63.390	0.281	0.802	-2.160	-0.218	0.206	0.725	3.768
Bachelor Degree	0.499	0.012	42.024	0.117	0.713	-3.130	-0.293	0.117	0.537	2.749
Classic AIC:	118441.6			29716.08						
AICc:	118441.7			32681.39						
Percent deviance explained	0.66			0.92						
Deviance:	118411.6			26044.5						
MAD	64.53			29.06						
R^2 **Poisson**	0.74			0.94						
Lagrange Multiplier	0.07			0.01						
Moran's I of residuals	0.16			-0.005						
Bandwidth	Not applicable			72.00						

Case Study 3
An Approach to Assess Residential Neighborhood Accessibility and Safety: A Case Study of Charlotte, North Carolina.

Louis Merlin[a*], Eric Dumbaugh[a], Amin Mohamadi Hezaveh[b], Christopher R. Cherry[b]

[a] School of Urban and Regional Planning, Florida Atlantic University, Boca Raton FL

[b] Civil and Environmental Engineering, University of Tennessee, Knoxville, TN

* Corresponding Author: lmerlin@fau.edu

Why does accessibility matter for safety?

- Research on sprawl suggests that those who living in more sprawling counties are more likely to be in fatal accidents
- The primary expected mechanism for this is higher VMT
- Greater sprawl -> higher VMT -> greater exposure to fatal crashes
- Accessibility is the built environment variable with the strongest relationship with VMT
  - Higher accessibility environments are associated with reduced VMT
- Therefore high accessibility at the residential location may be associated with reduced vehicular crash risk
- High pedestrian and bike accessibility at one's residential location may be associated with greater pedestrian/bike crash risk

Accessibility vs. Density

- Density is a highly localized measure of the built environment
- Accessibility is a regional measure that indicates overall regional proximity to destinations
- Therefore density may be a more relevant built environment measure for crash locations
- Accessibility may be a more relevant measure for residential locations because most people's activity space spans significantly beyond their home location
- As a regional measure, accessibility may also correlate with a person's generalized exposure to regional traffic
    - Persons who live in a high accessibility environment are surrounded by many destinations, and therefore travel in high-traffic environments
    - Persons who live in a low accessibility environment are surrounded with few destinations, and therefore travel in low-traffic environments

Aims of the study

- Investigate the relation between accessibility (job and population) and Safety
- Investigate the relation between density (job and population) and Safety

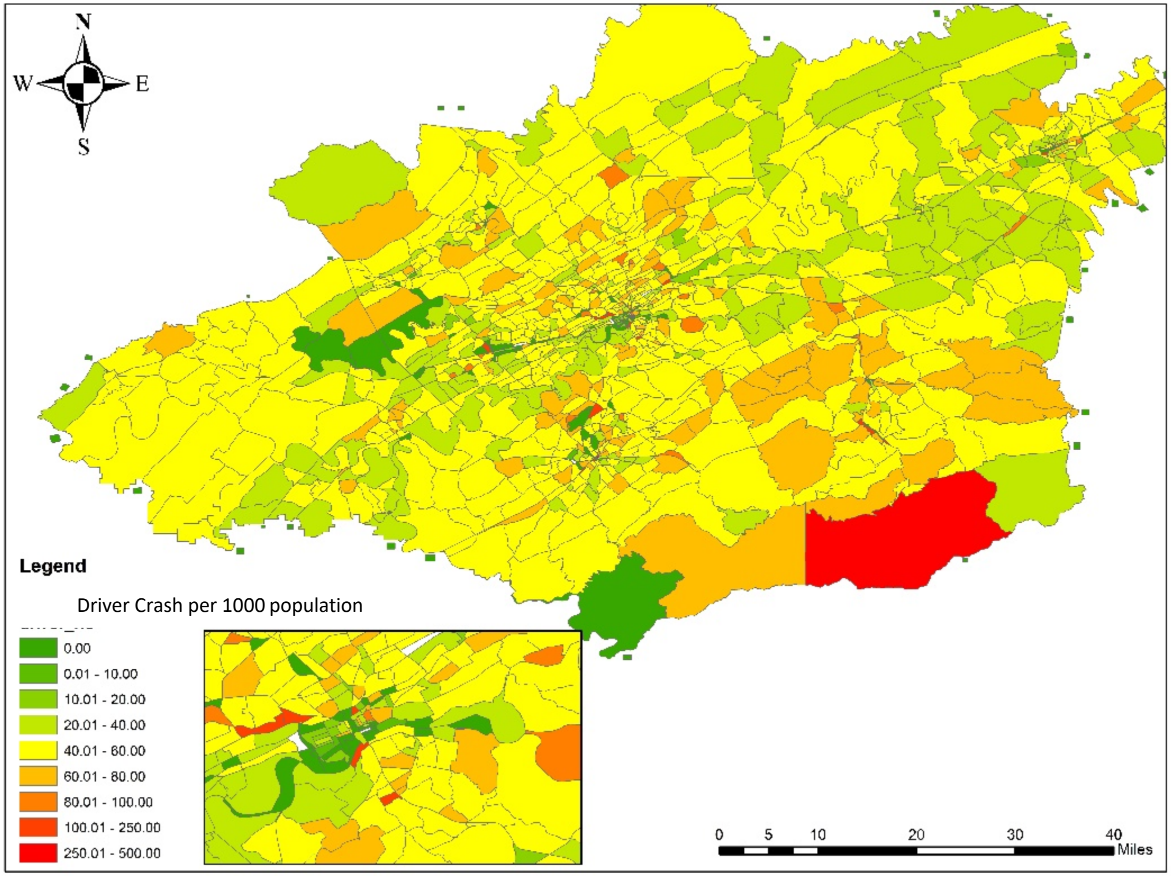
- Model: Spatial Error Model
  - to investigate the relationship between built environment and driver crash frequency at zonal level

Databases

- Knoxville Regional Travel Demand Model
- Police Crash Report in TN 2016
- Highway Performance Monitoring System

- Spatial Join
- Geocoding process; similar to the previous case study

Driver likelihood of involvement in traffic crash



Spatial Error Model

Dependent Variable: Driver Crash Freq. at TAZ Level

Variable	Coef.	Std. Err.	z	P>|z|
Vehicle per Household	3.308	1.006	3.290	0.001
Total Population	0.050	0.001	98.870	0.000
Average Median Household Income	0.000	0.000	-6.360	0.000
university Student Population	-0.008	0.002	-3.190	0.001
Tourist attractuib	6.594	2.327	2.830	0.005
Percent Pay Parking	-28.638	11.154	-2.570	0.010
Population Density	-0.002	0.000	-9.610	0.000
Employment Denisty	0.000	0.000	2.680	0.007
Job acc. Wihtin 10 minutes	-0.001	0.001	-1.710	0.088
Population acc. Wihtin 10 minutes	0.000	0.000	3.330	0.001
Cosntant	-4.798	5.377	-0.890	0.372
lambda	0.90554	0.08530	10.62	0

Findings are discussed in details in the case studies

BIG ORANGE BIG IDEAS

Case Study 5
Neighborhood-Level Factors Affecting Seat Belt Use

Amin Mohamadi Hezaveh[a], Christopher R. Cherry[a*]

[a] Civil and Environmental Engineering, University of Tennessee, Knoxville, TN, United States,

* Corresponding Author: Cherry@utk.edu

Published in the journal of the Accident Analysis & Prevention

https://doi.org/10.1016/j.aap.2018.10.005

Aims of the study

1. Identify seat belt use hotspots in TN at zonal level
2. Investigating the relationship between sociodemographic variables and seat belt use rate at zonal level based on the home address of the individual

- Study group:
  - Road users over 16 years old who were involved in traffic crash in TN in 2016 (i.e., driver or passenger)

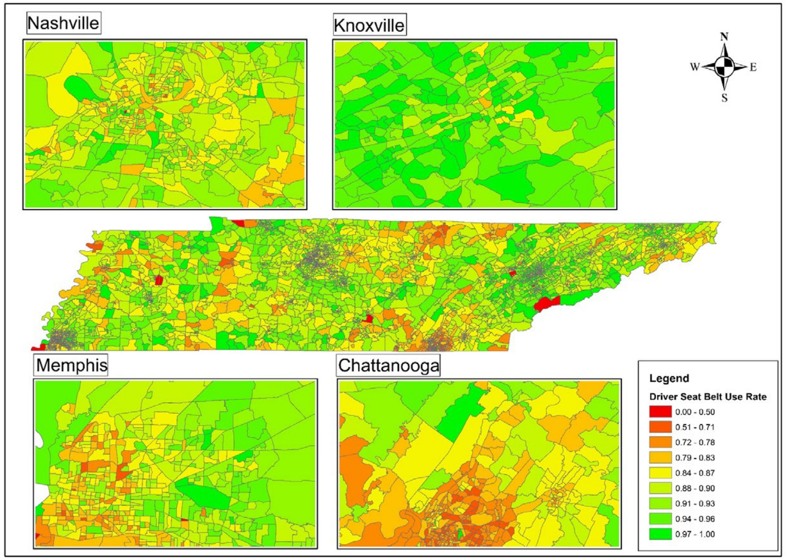
- Model: Tobit Model
- to investigate the relationship between sociodemographic variables and driver/passenger seat belt use rate at zonal level

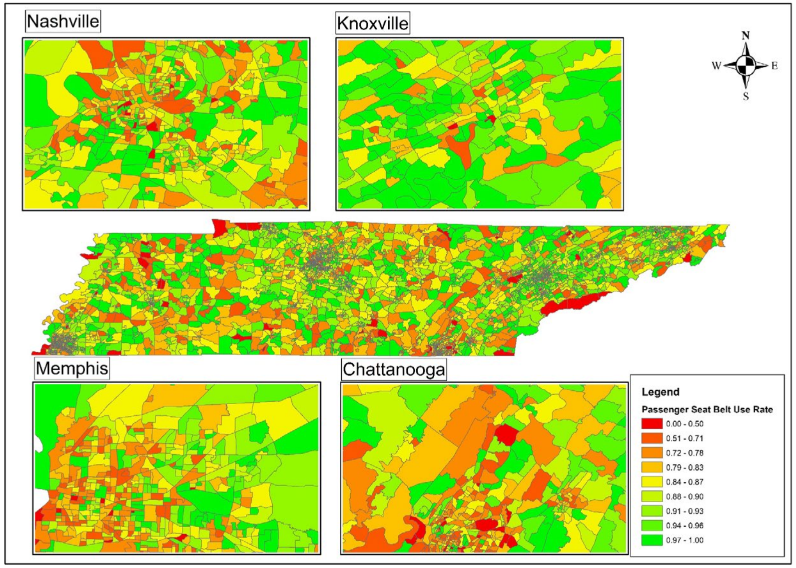
Seat Belt Use Distribution

- Databases:
  1. Police Crash Report
  2. US Census

- Spatial Join
- Geocoding process; similar to the previous case study

Seat Belt Spatial Distribution





Driver vs Passenger

DV: Seat Belt Rate Rate for

Findings are discussed in details in the case studies
DV: Driver seat belt use rate at zonal level
Passenger seat belt use rate at zonal level

Variable	DSBUR			PSBUR		
	Coef.	Standard Error	Elasticity	Coef.	Standard Error	Elasticity
Population (1,000)	0.006^{***}	0.001	0.010	0.005^{*}	0.003	0.008
% Children	0.023^{*}	0.012	0.005	0.085^{***}	0.024	0.037
% Race White	0.036^{***}	0.004	0.031	0.042^{***}	0.008	0.019
Vehicle Ownership						
% Household with no Vehicle	-0.078^{***}	0.013	-0.006	-0.041^{*}	0.025	-0.003
% Household with One or Two Vehicles	-0.025^{***}	0.008	-0.020	0.036^{**}	0.016	0.029
Education						
% College degree	-0.032^{**}	0.013	-0.007			
% Bachelor Degree	0.016^{*}	0.009	0.004	0.058^{***}	0.018	0.013
Metropolitan Indicator	0.007^{***}	0.002	0.005	0.015^{***}	0.005	0.012
Household Size				-0.001^{***}	0.000	-0.004
Density (1,000 population per square km)				$-1.46E-06^{***}$	2.24E-07	-0.011
Constant	0.863^{***}	0.008		0.773^{***}	0.016	
Scale parameter	0.004^{***}	7.97E-05		0.014^{***}	3.03E-04	
χ^2	328.37			233.50		
LL_0	5,563.87			2,841.95		
LL_M	5,728.06			2,958.70		
Maddala Pseudo-R^2	0.077			0.056		
N	4,114			4,103		
AIC	-11,436.12			-5,897.41		
*** p<.10; ** p<.05; *** p<.01**						
Source: Authors' analysis of TITAN data and the US Census						

Reporting and Next Steps:

- Finish report (March-April)
- Publish, Publish, Publish!
- Present work at technical meetings (e.g. ITE)
- Disseminate results to stakeholders through webinars and CSCRS/SafeTREC/CTR educational and professional development outlets.

Follow-up Work

CASE STUDY 1

Linking Crash and Post-Crash Data to Get a "Complete Picture" of Pedestrian/Bicyclist Injury

- Proposed Year 2 Activities

- Obtain data files linking crash and medical data (CMOD, or Crash Medical Outcome Data) developed by the California State DPH to evaluate the degree to which crash data (i.e., police collision reports) under-report crash injuries.

- Focus on pedestrian/bicyclist injury, identifying factors (e.g., age, ethnicity, geographic area) associated with level of reporting.

- Focus on evaluating level of crash reporting of pedestrian/bicyclist injury on tribal areas in California

CASE STUDY 2

Develop measures of EMS response times (time from crash to dispatch, time from dispatch to arrival of EMS crew, time on site, time to ED, etc.) as a function of rural versus urban, cell phone coverage, trauma center location, etc.

- Year 2 Proposed Activities

- Begin analysis of data already obtained from CEMIS (California EMS Information System) to evaluate time elements in EMS response from the time of the crash to the time of arrival at an emergency department or trauma center.

- Obtain addition data listed in the NEMSIS Uniform EMS Dataset as needed from CEMSIS to explore how factors such as location of EMS unit, type of treatment provided at the scene, etc. impact time elements.

- Prepare a detailed report showing EMS response times as a function of crash location, ED/trauma center location, and other factors. Highlight the factors that might be modified (e.g., cell phone coverage, placement of EMS response unites, etc.) to improve EMS response. This could take the form of a statistical model of EMS response in California that can identify the factors most likely to have a beneficial impact on improved injury outcomes.

- As a subpart of the above goals, look specifically at EMS response times in tribal areas in California (note: in a study of traffic safety in tribal areas in California, EMS response has been noted as a particular issue).

CASE STUDY 3 & 4

- Year 2: **Integrating Spatial Safety Data into Planning Processes**

- Extend the Home-based Safety safety approach integrate into planning process
- Expand attribution of crash causal behaviors to neighborhood profiles.
- Integrate ”crash generation” concepts into transportation planning processes (akin to “trip generation” concepts) and test on one metro planning model.

CASE STUDY X

- Year 2: **Opioids at the health and transportation safety nexus.**

- Explore integration of crash, health, and prescription drug monitoring datasets across critical states

- Health system map: opioid → traffic safety → opioid → health outcome

- Identify capabilities of datasets and institutions to answer questions related to opioid health system map.

Contacts

- Co-Author's
  - University of Tennessee
    - Christopher Cherry, cherry@utk.edu
    - Amin Mohamadi Hezaveh
    - Melanie Nolteneus
    - Asad Khattak
  - Florida Atlantic University
    - Louis Merlin, lmerlin@fau.edu
    - Eric Dumbaugh
  - University of California, Berkeley
    - David Ragland, davidr@berkeley.edu
  - University of North Carolina, Chapel Hill.
    - Laura Sandt, sandt@hsrc.unc.edu

Questions