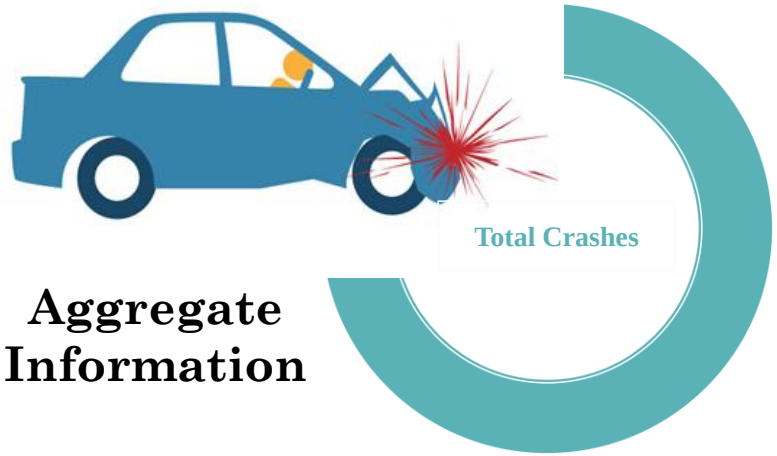
A New Econometric Approach for Modeling Several Count Variables: A Case Study of Crash Frequency Analysis by Crash Type and Severity

Naveen Eluru
Professor, Department of Civil, Environmental and Construction Engineering
University of Central Florida

UNIVERSITY OF CENTRAL FLORIDA

BACKGROUND



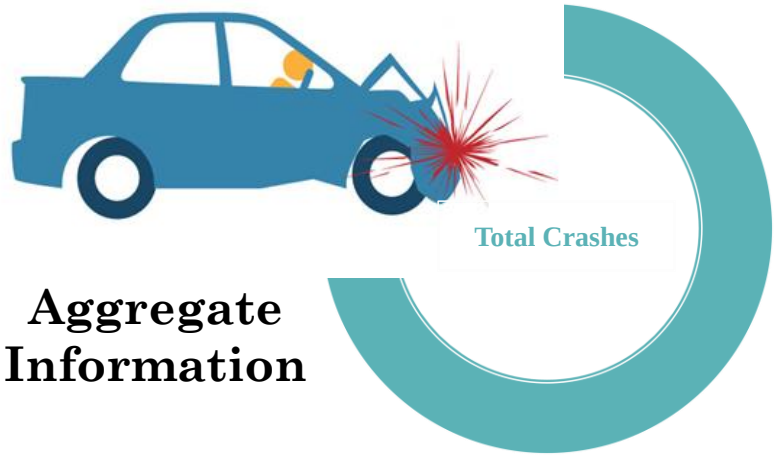
Aggregate Information

Rear-end

Sideswipe

Head-on

BACKGROUND

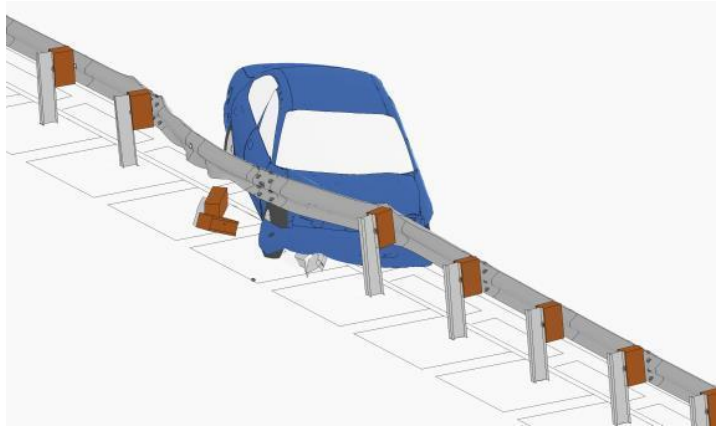


Aggregate Information

Crash profile is different

- Rear-end
- Sideswipe
- Head-on

Presence of Guardrail



- Head-on
- Total Crash
  Rear-end

BACKGROUND

Rear-end	No Injury	Minor Injury	Severe Injury
Angular	No Injury	Minor Injury	Severe Injury
Sideswipe	No Injury	Minor Injury	Severe Injury
Non-motorized	No Injury	Minor Injury	Severe Injury

Severity profile is different

BACKGROUND

- **We usually do not focus on all these dimensions**
- **Maximum: 3 crash types, 3 crash severities**

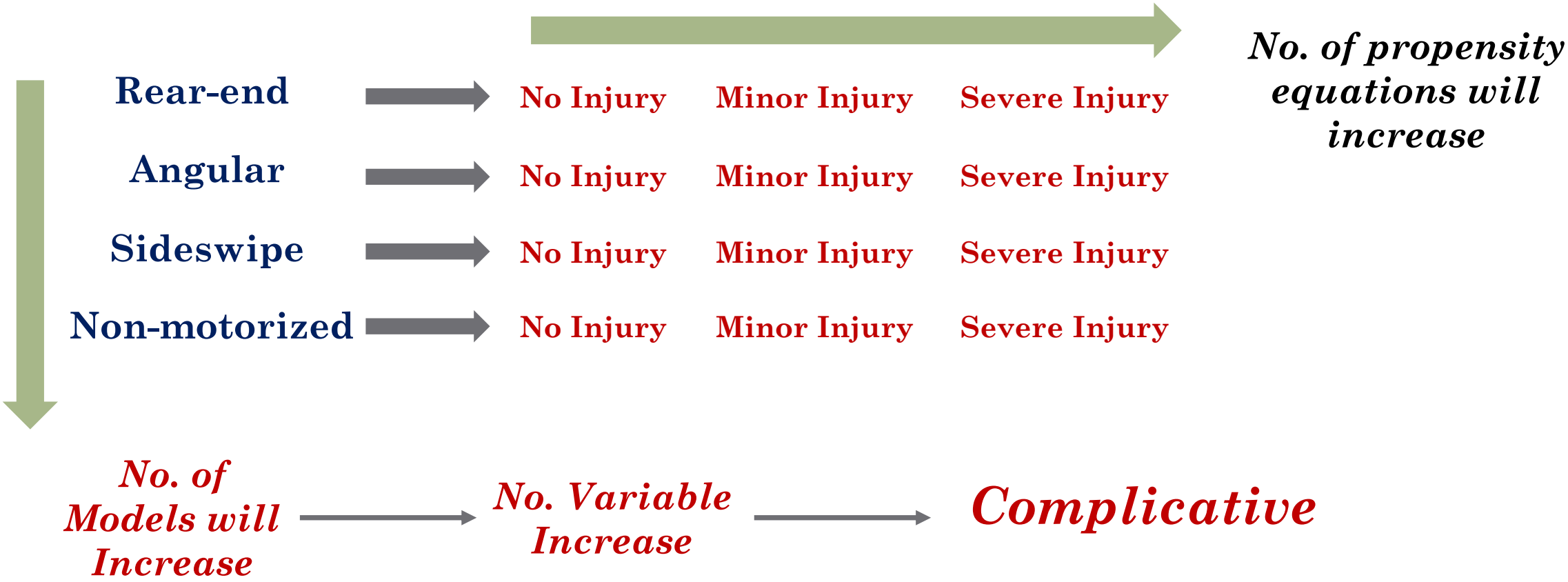
MOTIVATION


Rear-end
No Injury
Minor Injury
Severe Injury
No. of propensity equations will increase

MOTIVATION

No. of propensity equations will increase

MOTIVATION


No. of propensity equations will increase
Rear-end
Angular
Sideswipe
Non-motorized
No Injury
Minor Injury
Severe Injury
No. of Models will Increase
No. Variable Increase
Complicative

MOTIVATION

Simulation Based Multivariate NB Model (Challenges)

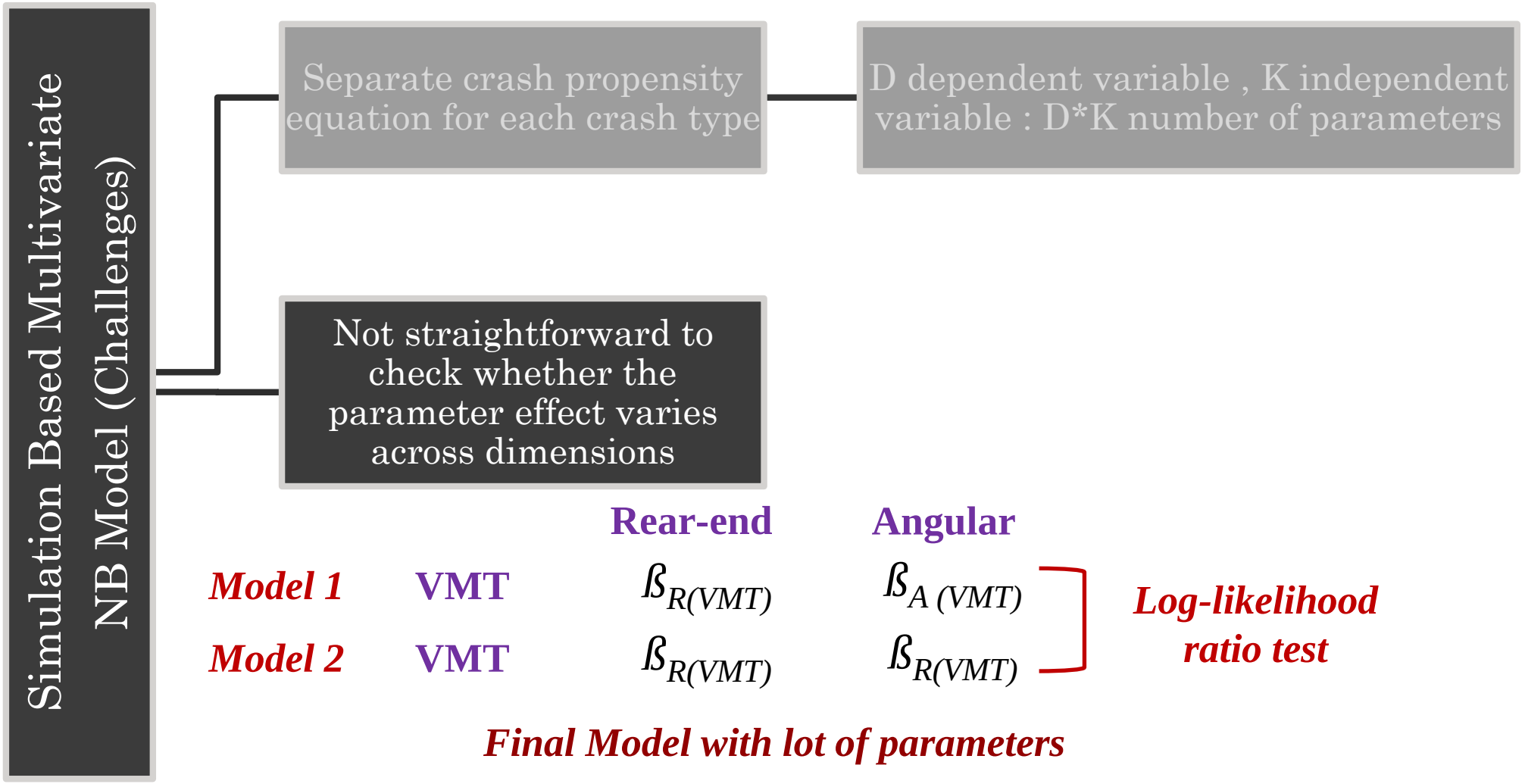
Separate crash propensity equation for each crash type

	Rear-end	Angular	Sideswipe	Non-motorized
VMT	$ß_{R(VMT)}$	$ß_{A\ (VMT)}$	$ß_{S\ (VMT)}$	$ß_{N\ (VMT)}$
Urban Road	$ß_{R(URB)}$	$ß_{A\ (URB)}$	$ß_{S\ (URB)}$	$ß_{N\ (URB)}$

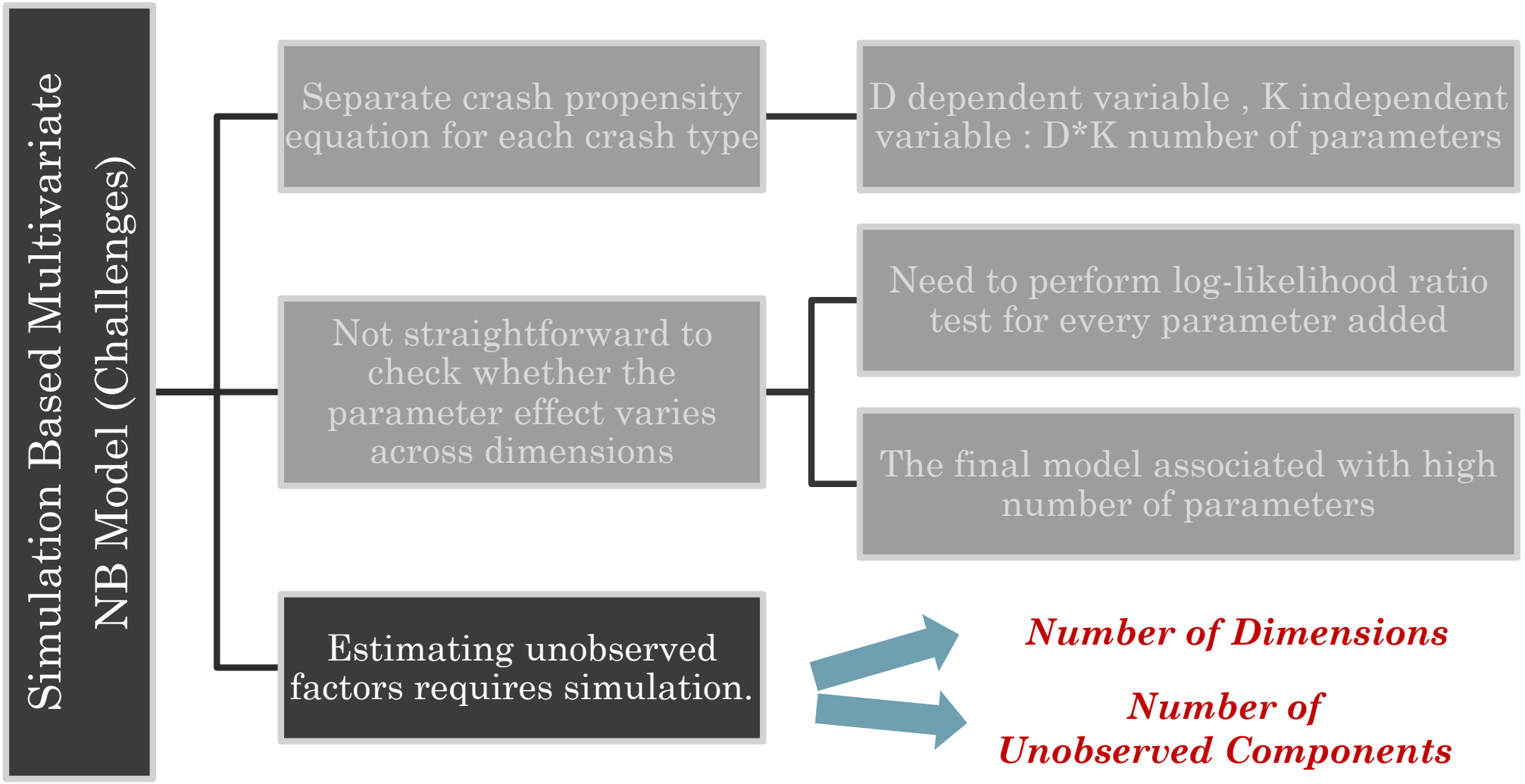
Number of estimated parameter will increase

D dependent variable , K independent variable :
D*K number of parameters

MOTIVATION


Simulation Based Multivariate NB Model (Challenges)
Separate crash propensity equation for each crash type
D dependent variable , K independent variable : D*K number of parameters
Not straightforward to check whether the parameter effect varies across dimensions
Rear-end
Angular
Model 1
VMT
$ß_{R(VMT)}$
$ß_{A\ (VMT)}$
Model 2
VMT
$ß_{R(VMT)}$
$ß_{R(VMT)}$
Log-likelihood ratio test
Final Model with lot of parameters

MOTIVATION


Simulation Based Multivariate NB Model (Challenges)
Separate crash propensity equation for each crash type
D dependent variable , K independent variable : D*K number of parameters
Not straightforward to check whether the parameter effect varies across dimensions
Need to perform log-likelihood ratio test for every parameter added
The final model associated with high number of parameters
Estimating unobserved factors requires simulation.
Number of Dimensions
Number of Unobserved Components

CURRENT STUDY

Traditional Count Approach

ID	TAZ	Rear-end	Angular	Sideswipe	Non-motorized	VMT
1	1	10	12	8	2	100

Proposed Count Approach

ID	TAZ	Crash Type	Crash	VMT
1	1	Rear-end	10	100
2	1	Angular	12	100
3	1	Sideswipe	8	100
4	1	Non-motorized	2	100

CURRENT STUDY

Traditional Approach

ID	TAZ	Rear-end	Angular	Sideswipe	Non-motorized	VMT
1	1	10	12	8	2	100

Model $_{SW}$ → $ß_{S0} + ß_{sV}*VMT$

Proposed Count Approach

					Dummy			***Interaction***		
ID	**TAZ**	**Crash Type**	**Crash**	**VMT**	**A**	**S**	**NM**	**VMT*A**	**VMT*S**	**VMT*NM**
1	1	Rear-end	10	100	0	0	0	0	0	0
2	1	Angular	12	100	1	0	0	100	0	0
3	1	Sideswipe	8	100	0	1	0	0	100	0
4	1	Non-motorized	2	100	0	0	1	0	0	100

Model → $ß_0 + ß_1*VMT + ß_2*A + ß_3*S + ß_4*NM + ß_6*VMT*S + ß_5*VMT*A + ß_7*VMT*NM$

$ß_{S0} = ß_0 + ß_3;\ ß_{SV} = ß_1 + ß_6$

CURRENT STUDY

Proposed Severity Approach

				Counts			Fractions		
ID	TAZ	Crash Type	Crash	PDO	Injury	Fatal	PDO	Injury	Fatal
1	1	Rear-end	10	6	4	0	0.6	0.4	0
2	1	Angular	12	6	5	1	0.5	0.42	0.08
3	1	Sideswipe	8	5	2	1	0.63	0.25	0.12
4	1	Non-motorized	4	0	3	1	0	0.75	0.25

1

CURRENT STUDY

Proposed Approach

ID	TAZ	Crash Type	Crash	PDO	Injury	Fatal	PDO	Injury	Fatal
1	1	Rear-end	10	6	4	0	0.6	0.4	0
2	1	Angular	12	6	5	1	0.5	0.42	0.08
3	1	Sideswipe	8	5	2	1	0.63	0.25	0.12
4	1	Non-motorized	4	0	3	1	0	0.75	0.25

Count Part Panel NB Model

Crash Type

X

Proportion Part Ordered Fractional Split Model

Crash Severity

Crash Counts of Each Severities by Crash Types

CURRENT STUDY

Proposed Approach

ID	TAZ	Crash Type	Crash	PDO	Injury	Fatal	PDO	Injury	Fatal
1	1	Rear-end	10	6	4	0	0.6	0.4	0
2	1	Angular	12	6	5	1	0.5	0.42	0.08
3	1	Sideswipe	8	5	2	1	0.63	0.25	0.12
4	1	Non-motorized	4	0	3	1	0	0.75	0.25

Count Part
Panel NB Model

Crash Type

Proportion Part
Ordered Fractional Split Model

Crash Severity

Advantage

- A single model system
- 2 components only
- Retain ordering nature of severities
- Parsimonious Model Structure

MODELING FRAMEWORK

- Methodology **(Joint Panel NB-GOPFS Model)**

$$v_{ir} = E(y_{ir}|\boldsymbol{x}_{ir}) = exp\left((\boldsymbol{\beta} + \boldsymbol{\theta}_i + \boldsymbol{\Phi}_{ir} \pm \eta_{irk})\boldsymbol{x}_{ir} + \varepsilon_{ir}\right)$$

$$P(y_{ir}|v_{ir}, \lambda') = \frac{\Gamma\left(y_{ir} + \frac{1}{\lambda'}\right)}{\Gamma(y_{ir}+1)\Gamma\left(\frac{1}{\lambda'}\right)}\left(\frac{1}{1+\lambda' v_{ir}}\right)^{\frac{1}{\lambda'}}\left(1 - \frac{1}{1+\lambda' v_{ir}}\right)^{y_{ir}}$$

NB Model

$$y^*_{irk} = (\alpha_r + \gamma_{irk} + \delta_{ir} \pm \eta_{irk})z_{ir} + \xi_{irk}$$

$$\psi_{rk} = \psi_{r,k-1} + exp((\beta_{rk} + \theta_{irk} + \varsigma_{ir} \pm \eta_{irk})s_{irk})$$

$$P_{irk} = G\left[(\psi_{rk} - \{(\alpha + \gamma_i + \delta_{irk} \pm \eta_{irk})z_{ir}\}\right] - G\left[(\psi_{r,k-1} - \{(\alpha + \gamma_i + \delta_{irk} \pm \eta_{irk})z_{ir}\}\right]$$

GOPFS Model

MODELING FRAMEWORK

- Methodology **(Joint Panel NB-GOPFS Model)**

$$L_i = \int_{\Omega} \prod_{r=1}^{R} \left[\left(P(c_{ir}) \right) \times \prod_{k=1}^{K} (P_{irk})^{\varpi_{ir} d_{irk}} \right] d\Omega$$

Joint Model Estimation

Correlation between Crash Types

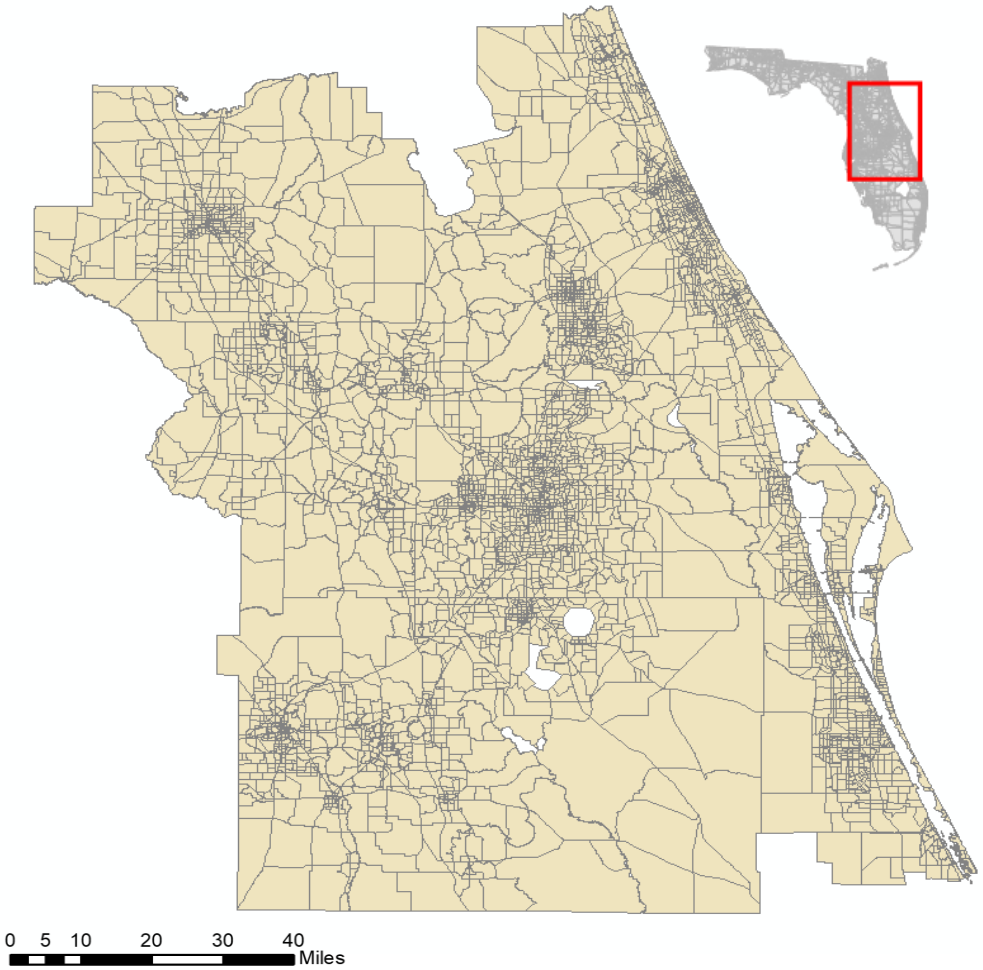
Correlation between Crash Severity Proportions

Correlation between Crash Severity Proportions

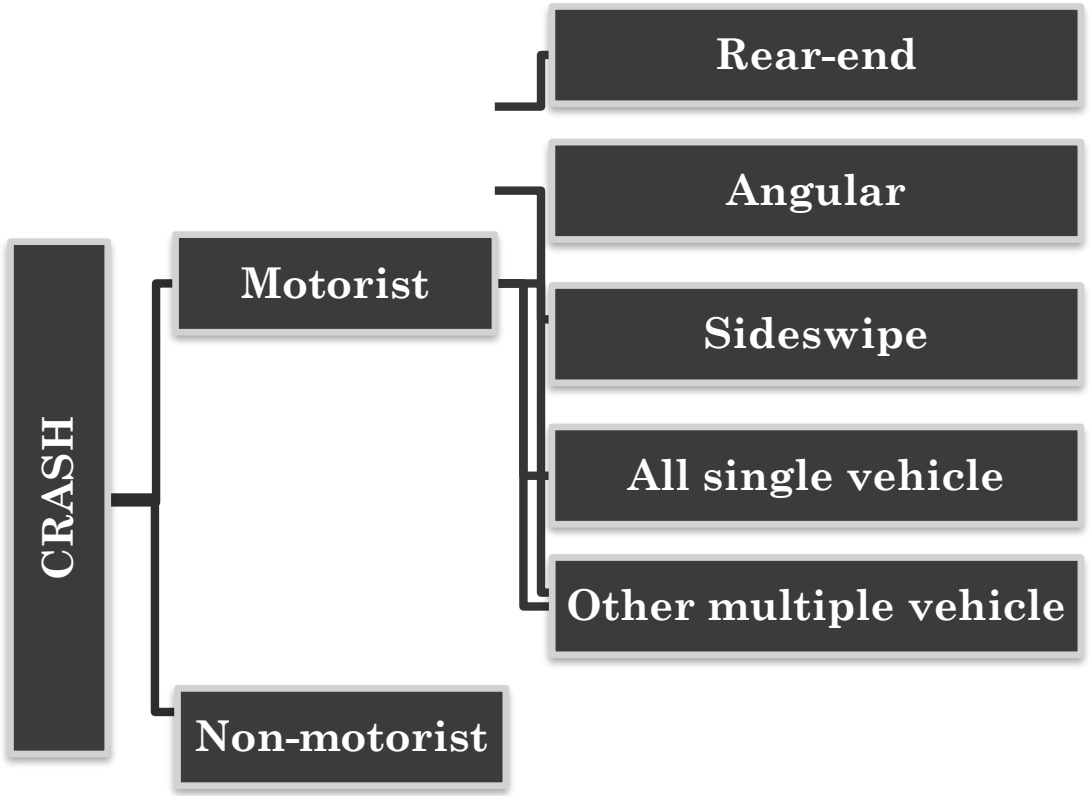
$$\mathbb{A} = \begin{array}{c|cc} & \text{Crash Types } (1,2,\ldots J,\ J=6) & \text{Crash Severity } (1..K,\ K=4) \\ \hline \text{Crash Types } (1,2,\ldots J,\ J=6) & \mathbb{A}_1 \ \varrho_{ir} & \mathbb{A}_3 \ \eta_{irk} \\ \text{Crash Severity } (1..K,\ K=4) & \mathbb{A}_3 \ \eta_{irk} & \mathbb{A}_2 \ \delta_{ir} \end{array}$$

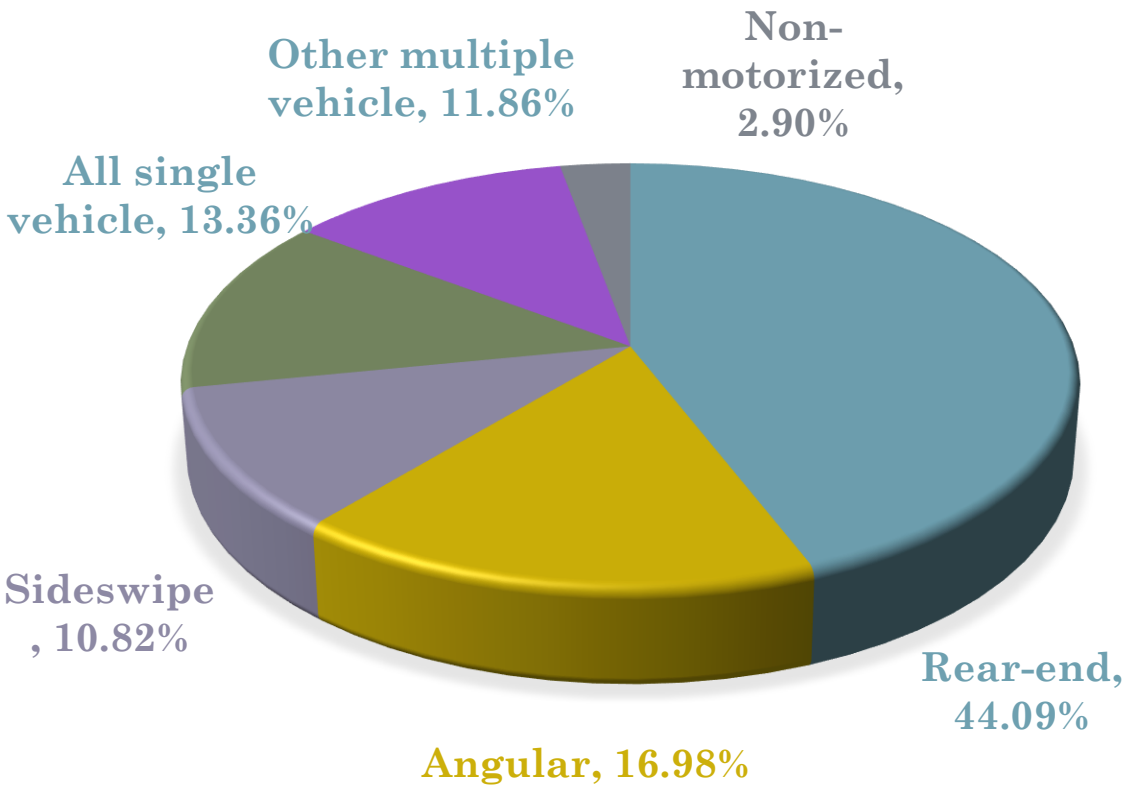
STUDY AREA

- **Central Florida Region**
  - Total 11 counties
- Crash data- Year 2016
- **117K crashes**
- 4,747 TAZs
  - 3,815 TAZs - Estimation Sample
  - 932 TAZs – Validation Sa
- Data source: FDOT, CARS, S

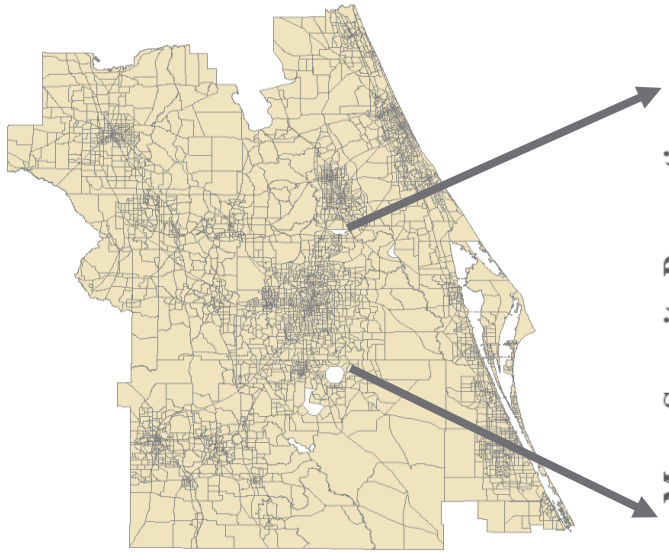


DATA DESCRIPTION



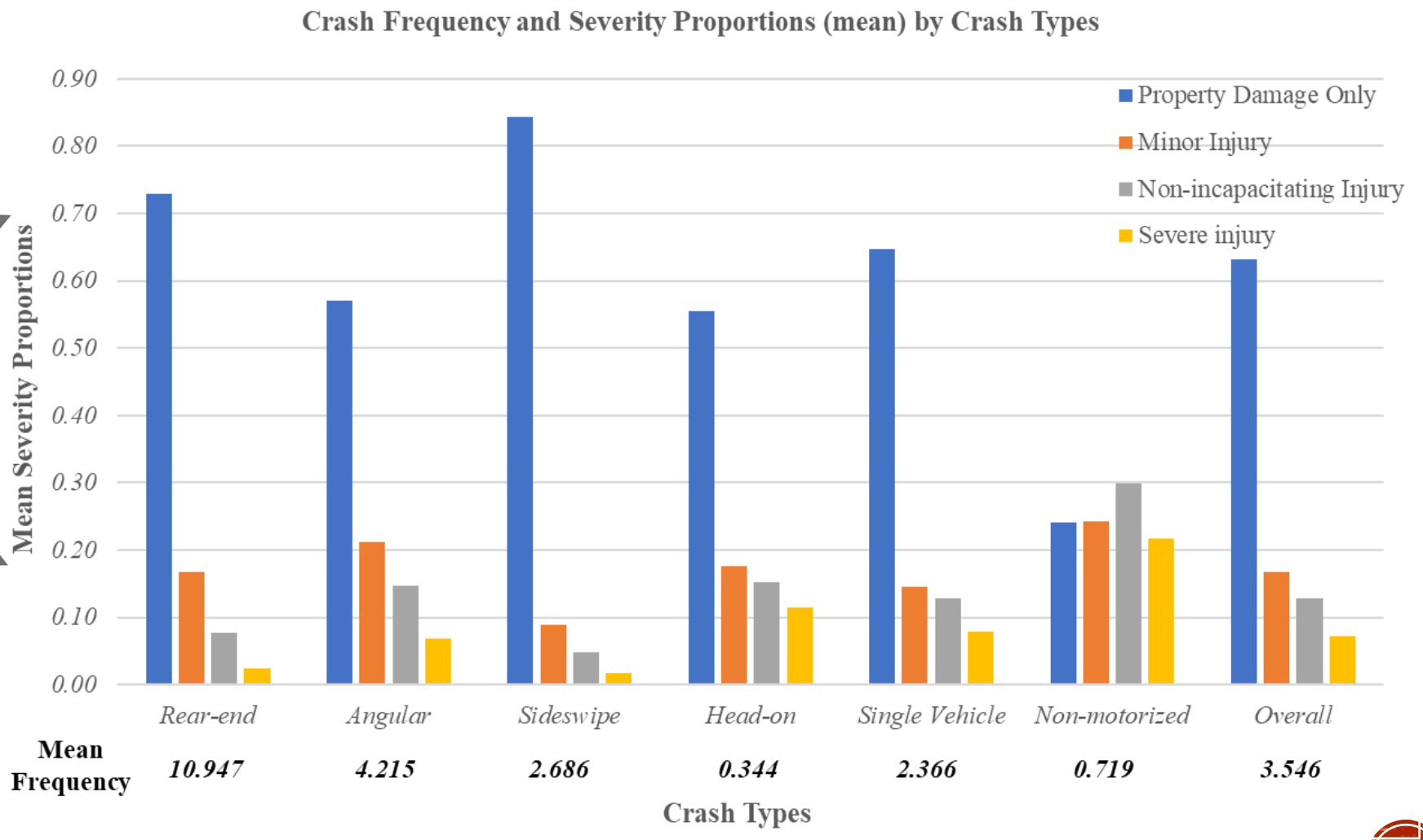


DATA DESCRIPTION



Central Florida Region

Crash data- Year 2016



MODEL FIT

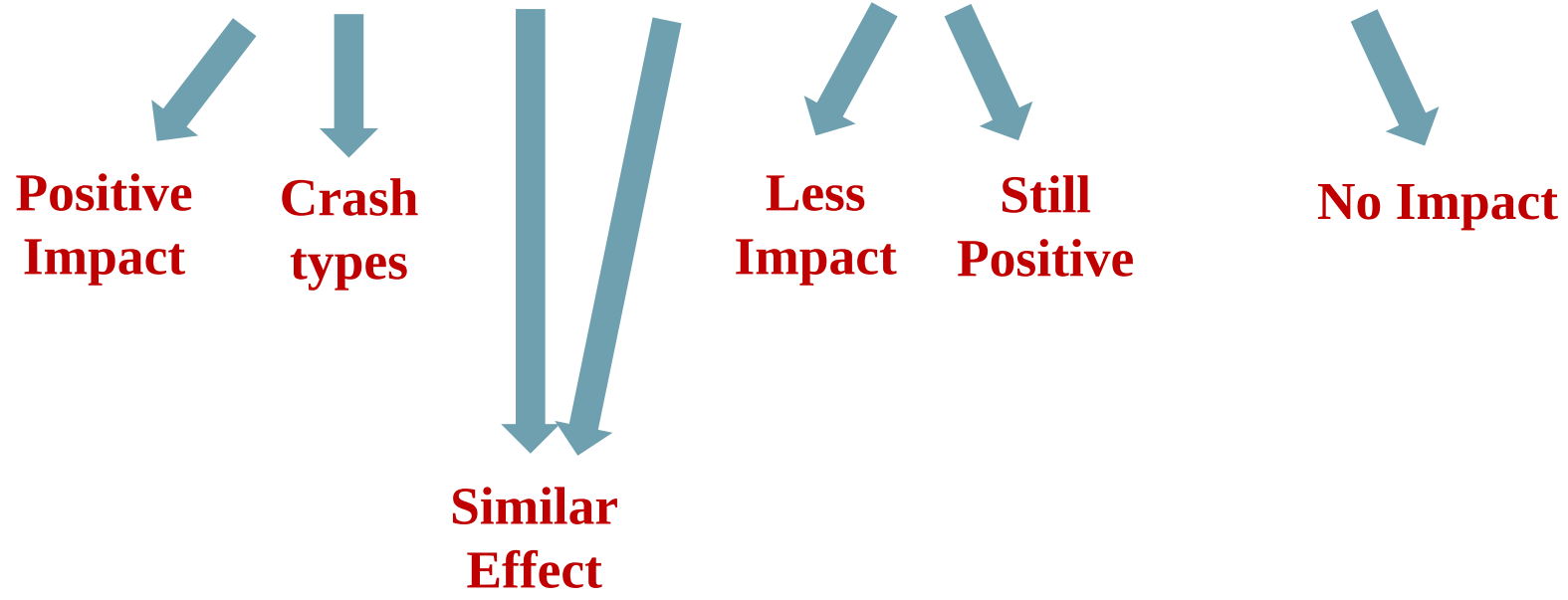
	Parameters	AIC	BIC
Traditional NB Model	251	104,070.90	105,123.93
Proposed Panel NB-OPFS Model	105	102,101.63	102,757.53

MODEL RESULTS (NB PART)

Variable	Overall Crash Risk	Deviations					
		Rear-end (1)	Angular (2)	Sideswipe (3)	Head-on (4)	Single Vehicle (5)	Non-motorized (6)
		(Deviation, Overall) – N/I = No Impact					
Roadway Characteristic							
Proportion of arterial roads	▲ (1,2,5,6)	--	--	N/I	N/I	▼ , ▼	--
Number of intersections	▲ (2,4,6)	N/I	--	N/I	--	N/I	--
Signal Intensity	▲ (1,3,5,6)	--	N/I	▼ , ▼	N/I	▼ , ▼	--
Road length over 55mph	▲ (1-6)	--	▼ , ▼	--	▼ , ▼	▲ , ▲	▼ , ▼
Variance of speed limit	▲ (1-3)	--	--	▲ , ▲	N/I	N/I	N/I
Road with median	▲ (1-4)	--	--	--	▼ , ▼	N/I	N/I
Width of outside shoulder	▼ (1-5)	--	▼ , ▼	▼ , ▼	--	▲ , ▼	N/I
Average sidewalk width	▼ (6)	N/I	N/I	N/I	N/I	N/I	▼
Land Use Characteristic							
Urban area	▲ (1-4,6)	--	--	▼ , ▲	▼ , ▲	N/I	▼ , ▲
Office area	▲ (1,3,6)	--	N/I	--	N/I	N/I	▼ , ▲
Residential area	▼ (3,4)	N/I	N/I	--	--	N/I	N/I

MODEL RESULTS (NB PART)

Variable	Overall Crash Risk	Deviations					
		Rear-end (1)	Angular (2)	Sideswipe (3)	Head-on (4)	Single Vehicle (5)	Non-motorized (6)
		(Deviation, Overall) – N/I = No Impact					
Land Use Characteristic							
Urban Area	▲ (1-4,6)	--	--	▼ , ▲	▼ , ▲	N/I	▼ , ▲



MODEL RESULTS (NB PART)

Variable	Overall Crash Risk	Deviations					
		Rear-end (1)	Angular (2)	Sideswipe (3)	Head-on (4)	Single Vehicle (5)	Non-motorized (6)
		(Deviation, Overall) – N/I = No Impact					
Built Environment Characteristic							
No. of restaurants	▲ (13,6)	--	N/I	▼ , ▲	N/I	N/I	▼ , ▲
No. of shopping centers	▲ (1,3)	--	N/I	--	N/I	N/I	N/I
Traffic Characteristic							
VMT	▲ (2-4,6)	N/I	--	▲ , ▲	▲ , ▲	N/I	▼ , ▲
Truck VMT	▲ (1,5)	--	N/I	N/I	N/I	▲ , ▲	N/I
Socio-demographic Characteristic							
Non-motorist commuters	▲ (1-3,6)	--	▲ , ▲	▲ , ▲	N/I	N/I	--
Transit Users	▲ (1,6)	--	N/I	N/I	N/I	N/I	--
Random Parameters							
Road with speed over 55mph – Angular Crash							

MODEL RESULTS (GOPFS PART)

Variable	Overall Severity Risk	Deviations					
		Rear-end (1)	Angular (2)	Sideswipe (3)	Head-on (4)	Single Vehicle (5)	Non-motorized (6)
		(Deviation, Overall) – N/I = No Impact					
Roadway Characteristic							
Proportion of arterial roads	▲ (1,2,5)	--	▲ , ▲	N/I	N/I	--	N/I
Possible and NIC injury	--	▼	N/I	N/I	N/I	N/I	N/I
Local Roads	▼ (4,6)	N/I	N/I	N/I	--	N/I	--
Number of intersections	▼ (4,5)	N/I	N/I	N/I	--	--	N/I
Signal Intensity	▼ (2-4)	N/I	--	--	--	N/I	N/I
Width of inside shoulder	▼ (3)	N/I	N/I	--	N/I	N/I	N/I
Width of outside shoulder	▼ (1)	--	N/I	N/I	N/I	N/I	N/I
Road length over 55mph	▲ (1,2,4-6)	--	--	N/I	▲ , ▲	--	--
NIC and Severe injury	--	▼	N/I	N/I	N/I	N/I	N/I
Poor Pavement Condition	▲ (3)	N/I	N/I	--	N/I	N/I	N/I
Land Use Characteristic							
Urban area	▲ (1-4,6)	--	--	▼ , ▲	▼ , ▲	N/I	▼ , ▲
Office area	▲ (1,3,6)	--	N/I	--	N/I	N/I	▼ , ▲
Residential area	▼ (3,4)	N/I	N/I	--	--	N/I	N/I

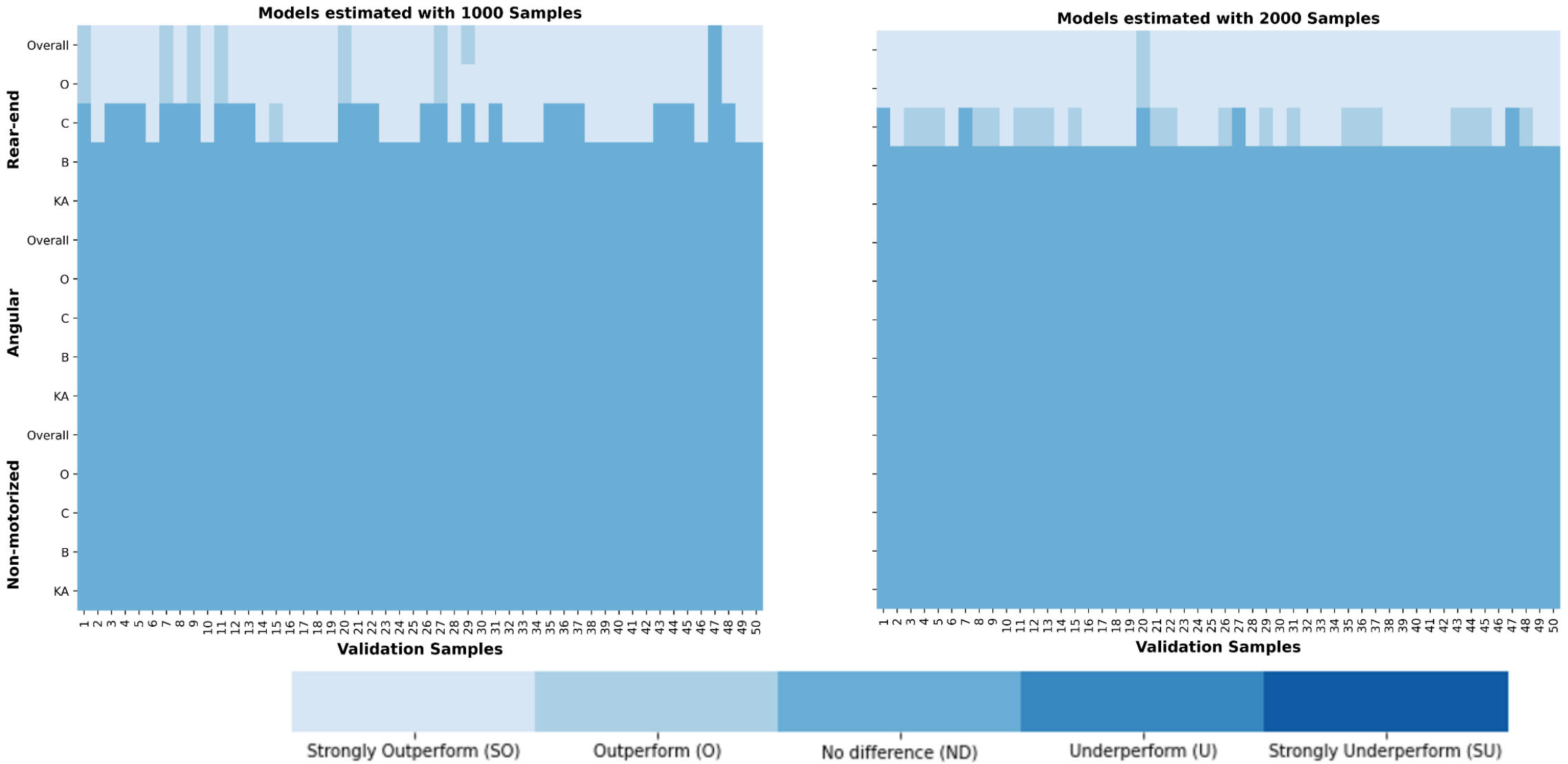
MODEL RESULTS (NB PART)

Variable	Overall Severity Risk	Deviations					
		Rear-end (1)	Angular (2)	Sideswipe (3)	Head-on (4)	Single Vehicle (5)	Non-motorized (6)
		(Deviation, Overall) – N/I = No Impact					
Built Environment Characteristic							
No. of commercial centers	▼ (6)	N/I	N/I	N/I	N/I	N/I	--
No. of recreational centers	▼ (1)	--	N/I	N/I	N/I	N/I	N/I
No. of restaurants	▼ (5)	N/I	N/I	N/I	N/I	--	N/I
NIC and Severe injury	--	N/I	N/I	N/I	N/I	▲	N/I
No. of shopping centers	▼ (2-4)	N/I	--	--	--	N/I	N/I
Possible and NIC injury	--	N/I	N/I	▲	N/I	N/I	N/I
Traffic Characteristic							
Congested condition	▼ (1-2)	--	--	N/I	N/I	N/I	N/I
NIC and Severe injury	--	N/I	▲	N/I	N/I	N/I	N/I
Truck VMT	▲ (3,4)	N/I	N/I	--	--	N/I	N/I

MODEL RESULTS (NB PART)

Variable	Overall Severity Risk	Deviations					
		Rear-end (1)	Angular (2)	Sideswipe (3)	Head-on (4)	Single Vehicle (5)	Non-motorized (6)
		(Deviation, Overall) – N/I = No Impact					
Socio-demographic Characteristic							
Employee	▼ (6)	N/I	N/I	N/I	N/I	N/I	--
Motorcycle users	▲ (2)	N/I	--	N/I	N/I	N/I	N/I
Senior people (>65)	▼ (6)	N/I	N/I	N/I	N/I	N/I	--
HH with no cars	▲ (6)	N/I	N/I	N/I	N/I	N/I	--

MODEL VALIDATION



SUMMARY AND CONCLUSION

Current Research

- A joint model for crash types and severities
- **Panel NB-GOPFS model**

Advantage

- Only need 2 propensity equations
- Less computational time
- Parsimonious specification
- **Can predict several dimensions**

Findings

- **Good performance for both sample**

PAPER

- **Bhowmik T.,** S. Yasmin and N. Eluru (2021). "A New Econometric Approach for Modeling Several Count Variables: A Case Study of Crash Frequency Analysis by Crash Type and Severity", **Transportation Research Part B** Volume 153, November 2021, Pages 172-203