"Introducing Latent Psychological Constructs in Injury Severity Modeling:
A Multi-Vehicle and Multi-Occupant Approach"

By Patrícia S. Lavieri, Chandra R. Bhat, Ram M. Pendyala and Venu M. Garikapati

Clarification on the latent variable labels

In the literature, problematic driving behaviors are usually classified as distracted, careless, risky, aggressive and reckless driving. The definitions and use of these five terms overlap, and they can be broadly categorized as: (i) distracted and careless driving; (ii) risky driving, and (iii) aggressive and reckless[1] driving. The scopes of risky and aggressive driving behaviors are used interchangeably in some studies because commonly aggressive behaviors are considered risky as well. Examples of risky driving behavior would include not wearing a seatbelt, speeding and driving under drug or alcohol impairment, while examples of aggressive driving behavior would include cursing and shouting, constantly trying to pass other vehicles, and tailgating. Linking to the psychology literature, risky driving behavior derives from the sensation seeking trait, aggressive driving behavior derives from the aggression trait, and careless driving from the contentiousness trait[2]. Again, it is important to mention that classifications of certain actions as aggressive or risky may vary across studies, but we opt for the above definition based on studies that associate driving behaviors with personality traits (Jonah (*1*)*;* Taubman-Ben-Ari and Yehiel (*2*)).

In the model framework, we focus on two of the three types of behavior identified above, *distracted/careless driving* behavior and *risky driving* behavior as shown in Figure 1 (the choice to focus only on two types of behavior was based solely on the availability of variables in the dataset that could be associated with each type of behavior). Each behavior is represented by a latent variable and is assumed to be a consequence of *distal* factors; however some *proximal* factors such as the presence of children in the vehicle may also affect driving behavior constructs. The figure also shows the personality traits that are assumed to impact driving behaviors (distracted/careless driving should be a consequence of low levels of contentiousness trait, while risky driving should be a consequence of high levels of sensation seeking trait). But since these are not measured in crash databases, their presence in the figure is merely a theoretical representation. The social-psychological literature suggests that these personality traits influence crash risk and outcomes not directly, but indirectly through their impact on driving behavior characteristics (*3, 4*).

For the empirical application conducted in the paper, the indicators associated with careless/distracted behavior are soft violations and inattention, while the indicators associated risky behavior are no seat belt use and alcohol impairment (the endogeneity of these last two

---

[1] There is no universal definition of these terms and ideally their scope should be adequately defined in each study to avoid misinterpretations and allow for comparisons. The term reckless, for example, is usually a synonym of careless but in the GES analytical user's guide it is associated with aggressive offenses such as manslaughter or homicide, willful reckless driving, driving to endanger, negligent driving fleeing or eluding police. Therefore, we will avoid the term reckless and just use the term aggressive in this paper.

[2] For more information on personality trait definitions, please refer to the Big Five Trait Model (*5*) and The Alternative Big Five Trait Model (*6*).

variables is also considered). The choice of which variable adequately captures each type of behavior was based on empirical results from the driving behavior and psychology literature (*1, 7-10*).
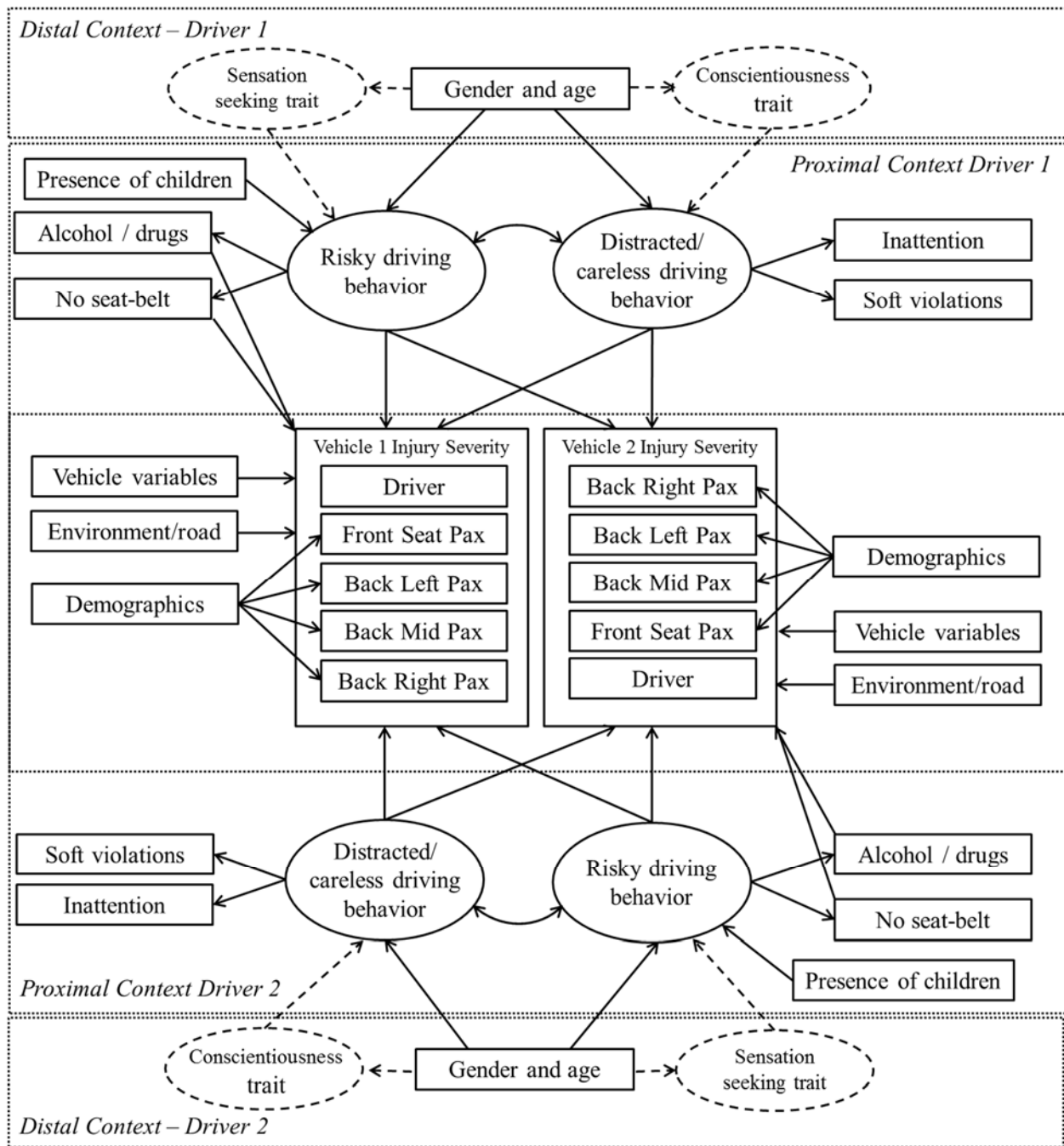

FIGURE 1 Conceptual framework of injury severity model system.

Model Goodness-of-Fit and Elasticity Effects Assessment

To assess the performance of the GHDM specification used in this study, the model used in the paper is compared to one that does not consider latent constructs, maintaining the same specification of the final model. The proof model is an independent model in which the error term correlations across the dimensions are ignored, but the best specification of the explanatory variables (including those used in the GHDM model in the structural equation system to explain the latent constructs) is considered to explain the injury severity of the vehicle occupants. The model that has no latent constructs takes the form of a multivariate ordered probit model. This may be referred to as an independent heterogeneous data model (IHDM). The GHDM and the IHDM specifications are not nested, but they may be compared using the composite likelihood information criterion (CLIC) introduced by Varin and Vidoni (*11*). The CLIC takes the following form:

$$\log L_{CML}^{*}(\hat{\boldsymbol{\theta}}) = \log L_{CML}(\hat{\boldsymbol{\theta}}) - tr\left[\hat{\boldsymbol{J}}(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{H}}(\hat{\boldsymbol{\theta}})^{-1}\right] \tag{4}$$

The model that provides a higher value of CLIC is preferred. The performance of the two models may also be compared through the likelihood values $\mathscr{L}(\hat{\boldsymbol{\theta}})$. The corresponding IHDM predictive log-likelihood value may also be computed. The measures of goodness of fit are presented at the bottom of Table 1. The average probability of correct prediction presented in the table provides evidence of the importance of jointly modeling the unobserved parts common to people sharing the same vehicle and involved in the same crash.

TABLE 1 Goodness of Fit Assessment

Total number of passengers including drivers in two cars	Average probability of correct prediction of injury severity		Number of cases (crashes)	Share
	GHDM	IHDM		
2	0.462	0.349	1968	57.39%
3	0.197	0.086	890	25.96%
4	0.104	0.026	369	10.76%
5	0.063	0.011	138	4.02%
6	0.052	0.004	48	1.40%
7	0.119	0.008	11	0.32%
8	0.052	0.0001	5	0.15%
Overall average probability of correct prediction	0.331	0.226		
Measure of Fit	**GHDM**		**IHDM**	
Composite marginal log-likelihood value at convergence	-140977.8		-164343	
Number of parameters	63		45	
Composite Likelihood Information Criterion (CLIC)	-142202.4		-165818	

Elasticity effects

Another way to show the difference between the GHDM and the IHDM is to compute the aggregate level elasticity effects corresponding to the driver's sociodemographic characteristics. We undertook such an investigation for the variables that were common and statistically significant in both specifications. As shown in Table 2, these variables are gender and age.

Gender is a dummy variable (male is the base category) for which the elasticity effect can be calculated in the following manner:

1) Set the value of the gender variable to 0 for all the drivers in the sample and calculate the probability of belonging to each injury severity category.
2) Set the value of the gender variable to 1 for all the drivers in the sample and calculate the probability of belonging to each injury severity category.
3) Subtract the result of step 2 by the result of step 1 and divide by the result of step 1 and calculate the average across all observations.

On the other hand, age appears as a multinomial variable (base is from 16 to 25 years old; other categories are 26 to 35, 36 to 65, 66 or more) in the GHDM specification for which the average elasticity effects can be calculated as follows:

1) Set the value of the categorical age variable to zero for all the categories and calculate the probability of belonging to each injury severity category.
2) Set the value of the age category of interest to one keeping all the other categories fixed to zero and calculate the probability of belonging to each injury severity category. Repeat this step for each of the age categories.
3) Subtract the result of step 2 by the result of step 1 and divide by the result of step 1 and calculate the average of all observations for each age category.

The results show that in the GHDM, being a female driver reduces the likelihood of severe injuries which is a consequence of females being less risky and less distracted/careless compared to men (please refer to the results in the original paper). On the other hand, the IHDM provides completely contrary results with females being very likely to suffer high levels of injury. These differences occur because, while the GHDM is able to capture behavioral effects, the IHDM only captures the mechanical effect of gender (possibly body strength based on the direction of the elasticity effects). Further, the elasticity values obtained from the IHDM model for the three upper levels of injury seem pretty unreasonable. A similar observation can be made of the age variable. Overall, given the better data fit and the behavioral foundation of the framework, results of the GHDM are found to be more realistic.

TABLE 2 Aggregate Elasticity Effects for Driver's Sociodemographic Characteristics

	Elasticity effects (%)			
Injury severity	Female		66 years or older	
	GHDM	IHDM	GHDM	IHDM
No apparent injury	33.48	-28.52	20.28	-45.88
Possible injury	-54.38	210.42	-41.82	348.39
Minor injury	-69.98	288.43	-55.96	941.72
Serious or fatal injury	-83.51	422.37	-70.41	1095.30

Finally, to illustrate the need for joint modeling of all vehicle occupants involved in a crash, we build a hypothetical crash scenario and calculate the probability of each injury severity level associated with each seat position (Table 3). For this purpose, we consider a two-vehicle crash involving ten people. One of the drivers is a young male (16 to 25 years old) under the influence of alcohol and the other is a middle age female (36 to 65 years old). Each vehicle carries 4 passengers with the same characteristics: the front seat passenger is a young (16 to 25 years old)

4

female, and, in the back seats, there is a female child and two elderly women. The vehicle, crash and environment related variables are also fixed for both vehicles (vehicle type is sedan and age is 5 to 10 years, frontal collision and frontal impact, daylight condition, clear weather, 35 mph speed limit, not in a junction, all vehicle occupants are wearing seat belt). The idea of having the same configuration of passengers and vehicles is to show that, if the crash is modeled jointly as done in the GHDM, the injury severity of the passengers can be different depending on the characteristics of the drivers and the behaviors associated with these characteristics (as seen in Table 3). It is evident from the table that the GHDM approach is able to identify and predict the differences in injury severity levels by seat position of the passenger (between vehicle 1 and vehicle 2), while the IHDM approach predicts the same probability of injury severity for a given seat position in either of the vehicles involved in the crash.

TABLE 3 Injury Severity Probabilities for Illustrative Synthetic Crash Scenario

		GHDM				IHDM			
		No apparent injury	Possible injury	Minor injury	Serious or fatal injury	No apparent injury	Possible injury	Minor injury	Serious or fatal injury
Vehicle 1 (young male)	Driver-1	0.270	0.149	0.236	0.345	0.089	0.197	0.436	0.277
	Passenger-1	0.301	0.182	0.227	0.290	0.055	0.194	0.429	0.322
	Passenger-2	0.110	0.194	0.331	0.365	0.047	0.264	0.512	0.177
	Passenger-3	0.023	0.064	0.198	0.715	0.002	0.028	0.250	0.721
	Passenger-4	0.026	0.057	0.278	0.639	0.002	0.025	0.382	0.591
Vehicle 2 (middle age female)	Driver-2	0.549	0.153	0.169	0.129	0.129	0.233	0.427	0.211
	Passenger-1	0.453	0.188	0.190	0.169	0.055	0.194	0.429	0.322
	Passenger-2	0.213	0.254	0.314	0.219	0.047	0.264	0.512	0.177
	Passenger-3	0.059	0.118	0.269	0.554	0.002	0.028	0.250	0.721
	Passenger-4	0.065	0.105	0.360	0.470	0.002	0.025	0.382	0.591

REFERENCES

1. Jonah, B. A. Sensation seeking and risky driving: a review and synthesis of the literature. *Accident Analysis & Prevention*, Vol. 29, 1997, pp. 651-665.
2. Taubman-Ben-Ari, O., and D. Yehiel. Driving styles and their associations with personality and motivation. *Accident Analysis & Prevention*, Vol. 45, 2012, pp. 416-422.
3. Elander, J., R. West, and D. French. Behavioral correlates of individual differences in road-traffic crash risk: an examination of methods and findings. *Psychological Bulletin*, Vol. 113, 1993, pp. 279–294
4. Sümer, N. Personality and behavioral predictors of traffic accidents: testing a contextual mediated model. *Accident Analysis & Prevention*, Vol. 35, 2003, pp. 949-964.
5. Costa, P. T., Jr., and R. R. McCrae. NEO-PI-R: Revised NEO Personality Inventory (NEO-PI-R). *Psychological Assessment Resources*, Odessa, FL, 1992.
6. Zuckerman, M. An alternative five-factor model for personality. *The developing structure of temperament and personality from infancy to adulthood*, 1994, pp. 53-68.
7. Jonah, B. A., R. Thiessen, and E. Au-Yeung. Sensation seeking, risky driving and behavioral adaptation. *Accident Analysis & Prevention*, Vol. 33, 2001, pp. 679-684.
8. Steinberg, L., D. Albert, E. Cauffman, M. Banich, S. Graham, and J. Woolard. Age differences in sensation seeking and impulsivity as indexed by behavior and self-report: evidence for a dual systems model. *Developmental Psychology,* Vol. 44, No. 6, 2008, p. 1764.
9. Hanley, P. F., and N. Sikka. Bias caused by self-reporting distraction and its impact on crash estimates. *Accident Analysis & Prevention*, Vol 49, 2012, pp. 360-365.
10. Donmez, B., and Z. Liu. Associations of distraction involvement and age with driver injury severities. *Journal of Safety Research*, Vol. 52, 2015, pp. 23-28.
11. Varin, C., and P. Vidoni. A note on composite likelihood inference and model selection. *Biometrika*, Vol. 92, No.3, 2005, pp. 519-528.