Injury Severity Prediction using Functional Data Analysis based on Select Vehicle Category from NASS-CDS

Susumu Ejima, Tsukasa Goto, Zhang Peng, Kristen Cunningham, Stewart Wang

Abstract An Injury Severity Prediction (ISP) algorithm was developed using a logistic regression model to predict the probability of sustaining an Injury Severity Score (ISS) 15+ injury. NASS-CDS (1999-2015) and model year 2000 or later were filters for new case selection criteria, which were based on vehicle body type, to match Subaru vehicle categories. In order to model the effect of PDOF (Principal Direction of Force) as a continuous curve, an analytical method called functional data analysis was employed to create a model of the PDOF curve with two different knot positions (ISP-f1R, ISP-f2R) defined for the periodic basis splines. The most significant variable identified in the current analysis was "Delta V". "Left impact" was high risk compared to other impact directions in the models. "Belt use" shows a significant risk in the model and belt use was safer than unbelted for any impact direction. "Age" was a significant occupant variable in both models and the presence of a female was notable but not selected in the models. When a right-front passenger was present, the effect on side impact (far-side) injury risk is larger than without a right-front passenger. To evaluate model performance, five-fold cross-validation was performed within the training data (NASS-CDS 1999-2015). The area under the receiver operator characteristic curve (AUCs) was used as the metric to evaluate model performances, AUC was 0.847 with the ISP-f1R model, 0.856 with the ISP-f2R model for cross-validation. This study utilizes the field triage of crash subjects and also evaluates the baseline vehicle safety performance of a representative Subaru vehicle categories in real-world crashes.

Keywords Advanced Automatic Collision Notification, Functional data analysis, NASS-CDS

I. INTRODUCTION

With the evolution of automotive telematics technology, Advanced Automatic Collision Notification (AACN), which predicts the severity of a traffic accident and determines the occupants who may have been injured, was expected to not only shorten the time to start treatment for patients, but also have a positive effect on healing after treatment [1]. According to a report by the Centers for Disease Control and Prevention (CDC) Field Triage Task Force [2], the triage process of consists of four steps (Physiologic, Anatomic, Mechanistic, Special Consideration), and of the third Mechanistic, predicting severity based on data transmitted by telematics was the effective method for appropriate triage during emergency medical care. It is recommended that an injury be considered severe when the likelihood of receiving an ISS (Injury Severity Score) [3] ≥ 15 is 20% or greater. Although there were many ways (e.g., Maximum Abbreviated Injury Scale (MAIS)[4], ISS) to dichotomize injury severity, the National Expert Panel in Field Triage chose ISS 15 as its partition when its on-scene triage decision scheme [2].

Prediction of injury severity for automobile occupants has been conducted since around 2000 in the U.S., and various studies [4-7] have been conducted. Kononen et al. [8] used information from the National Automotive Sampling System Crashworthiness Data System (NASS-CDS; 1999-2008) [9] to develop a vehicle-level Injury Severity Prediction (hereinafter referred to as ISP) model. Here, by identifying predictors that affect occupant severity, ISPs using each factor were calculated and applied to practical use as GM OnStar service. Later, ISP was improved by Wang et al. [10], who proposed a method to replace the direction of impact with a continuous function using the Principle Direction of Force (hereafter referred to as PDOF). These models were for all vehicle body types (Car, Utility, Pickup, Van) registered in NASS-CDS and were not limited to Subaru vehicles.

In this study, we attempted to construct an injury severity prediction algorithm for Subaru vehicles in the United

S. Ejima (tel: +1-734-763-1807; e-mail: ejima.susumu@subaru.co.jp) and T. Goto are from Subaru Corporation, P. Zhang is a Research Assistant Professor, K. Cunningham is a Program Director, and S. C. Wang is Endowed Professor, all at the International Center for Automotive Medicine, University of Michigan, USA.

States. Specifically, we focused on the vehicle category to which Subaru vehicles belong as the target vehicle category and conducted a logistics regression analysis based on NASS-CDS data to develop a new ISP model. The algorithm uses the latest accident data, NASS-CDS (1999-2015), which we believe provides a baseline for the safety performance of vehicles in real-world accidents as well as post-crash safety (including emergency medicine) by AACN. By continuously improving this injury severity prediction algorithm, we believe it will play a role toward "Zero" fatal traffic accidents as one of the pillars of a safe traffic system.

II. METHODS

Sample Data (NASS-CDS)

From a practical perspective, the variables used in the algorithm were selected from NASS-CDS based on information transmitted by vehicle telematics, i.e., information obtained from the EDR (Event Data Recorder). The reason for setting constraints on the variables in this way was to reflect the current state of information obtained by the current vehicle telematics systems.

NASS-CDS (1999-2015) and model year 2000 or later were filtered with new case selection criteria, which were based on vehicle body type, to match Subaru vehicle categories. Counting 1999-2015 data, since 1999 data has some model year 2000 cars. That was the reason we have included year 1999. In order to deal with limited sample sizes when focusing on Subaru-make vehicles only, we expanded our scope to investigate the vehicle belong to Subaru vehicle categories. The vehicle category was defined by limiting the study cohort to only the vehicle types recently produced by Subaru. We used the NHTSA standard definitions of passenger cars and sport utility vehicles (SUVs) as defined in the NASS-CDS coding [11]. The cohort was selected using the following specific vehicle codes: 02, 04, 05, and 06 are passenger cars and 14, and 15 are SUVs (Table 1).

Table 1 Selected body type from NASS-CDS

Select body type from NASS-CDS

NASS-CDS Code	Body type	SUBARU Category
02	Car	2-Door Sedan/Hardtop/Coupe
04		4-Door Sedan/Hardtop
05		4-/5-Door Hatchbak
06		Station Wangon
14	Utility	Compact Utility
15		Large Utility

An additional inclusion criterion for the current study was the principal direction of force (PDOF) was known. To model the effect from PDOF as a continuous curve, we employed the statistical analysis called functional data analysis, where we used cyclic basis splines to model the PDOF curve. The occurrence of two or more significant impacts on a single vehicle was a potentially important explanatory variable in predicting injury outcomes. Vehicles that experienced multiple impacts and vehicles that experienced a single impact were identified from the NASS-CDS. Vehicles that sustained a non-horizontal event, such as a rollover, were excluded. As written in Kononen's paper, the number of occupants does not matter because it is the vehicle level at which vehicles involved in a crash are most likely to contain a seriously injured occupant. The occupant and injury parameters were based on the following criteria. Belt status was available as "all occupants belted" vs. "at least one occupant unbelted" and did not consider the belt status of rear-seat passengers because current Event Data Recorders (EDRs) might not sense presence of rear-seat occupants. We chose the presence of an older occupant if anyone in the vehicle was 55 or older [8], and "under 55" if all occupants were under 55. Similarly, the presence of a female occupant was at the vehicle level if any occupant was female and "all male" if all occupants were male. Injury Severity Score (ISS), [2] defined as sum of the squares of the Abbreviated Injury Scale (AIS) [12] of the three most significantly injured body regions, were available. The ISP model was designed to predict ISS at the crash involved vehicles. We chose highest ISS among the any of occupants. Following these selection criteria, the total number of crash case involving vehicle was 7,351. There was a weighting factor in the NASS-CDS, and it estimated the total number of similar crashes in the entire U.S. To be statistically sound, we used weighted data. NHTSA has limited resources to cover the entire nationwide crash data and can only investigate limited cases. Therefore, they purposely investigate more severe cases and created a bias towards more severe cases. The total number of weighted data was 1,645,762 (median weights = 75).

Functional Data Analysis

Kononen et al. [8] developed the algorithm which uses crash and occupant factors to predict whether anyone in the vehicle has significant probability of sustaining a severe injury. This algorithm was later updated to an algorithm [13] based on NASS-CDS data through calendar year 2013. In order to model the effect from PDOF as a continuous curve, we employed the statistical analysis called functional data analysis [10], where we used cyclic basis splines to model the PDOF curve. The number of basis splines (so called degree of freedom) was important here. The more splines one can use, the finer result of the PDOF curve one can obtain. In comparison, Kononen's model where we used four direction of impact (Front, Left, Right, Rear) to summarize impact directions, the degree of freedom (DOF) was three. In addition, the study of the effect of a right-front passenger [10] with respect to injury risk showed that the effect of PDOF on injury risk changed significantly when a right-front passenger was present, and this trend was more pronounced in the case of side impact.

In this study, ISP algorithms based on functional data analysis which consider the effect of right-front passenger were developed using a logistic regression model to predict the probability of sustaining ISS 15 + injury.

1. ISP-f1R: Functional data analysis using PDOF and select knots based on quantile.
2. ISP-f2R: Functional data analysis using PDOF and select knots based on direction of impact.

The only difference between the ISP-f1R and ISP-f2R model were the knots, which selected basically result in "shift" of basis curves to make them focus on certain regions. Traditionally, we choose knots where you have more data points i.e. according to quantiles of data. In ISP-f2R, we forced the knots at 90, 180, 270 degrees, since we know those are potential transition points.

The stepwise procedure was used to select a subset of more relevant features to construct the final ISP algorithm. The concept of using a statistical approach was to predict a 20% probability of ISS 15 or greater and the importance of each variable defined within a predictive model. Logistic regression models were fitted to investigate ISS 15 or greater injury with different configurations of the variables which was defined Table 2. In this study, the ISP algorithm using the NASS-CDS, was developed as the basis to predict injury. The logistic regression equation was as follows:

$$P(ISS\ 15+) = \frac{1}{\left(1 + exp(-z)\right)} \tag{1}$$

$$z = \beta_0 + \sum_{i=1}^{n} \beta_i x_i \tag{2}$$

where *P (ISS 15+)* is the probability that the ISS is 15 or greater (considered to be a serious injury level), β_0 was the intercept, β_i were the coefficients of the predictors x_i, and n is the number of predictors. These predictors were shown in Table 2. It is necessary to have a cutoff value for *P (ISS 15+)*. The CDC Expert panel's recommendations for development of an injury risk algorithm specified a probability cutoff value of 0.2 using Injury Severity Score (ISS). In another word, probability of 0.2 was defined as a serious injury at the crashes. In this study, this cutoff value was used with *P (ISS 15+)*.

A forward and backward stepwise variable selection method was used to select the optimal model in the logistic regression analysis. The Akaike's Information Criteria (AIC) was used as the criterion for selecting (adding or deleting) one variable in each step, and the best model was the combination of variables with the lowest value. Note that AIC is an information criterion that expresses the relationship (trade-off) between the goodness of fit of the data and the complexity of the model, and has the effect of suppressing overfitting when the model has many variables (overfitting). The performance of the obtained model was indicated by the sensitivity and specificity, and the model was evaluated based on the Area Under Curve (AUC), which was calculated from the area surrounded by the Receiver Operating Characteristic (ROC) curve, which was expressed based on these relationships. Statistical analysis was performed using R 4.1.1 with the Survey, pbs, and MASS packages.

Table 2 Predictor variables used in models that are able to be obtained using EDR (Event Data Recorder) in a crash, or by an operator communicating with the vehicle belong to Subaru category.

Variables	Categories
Delta V (mph)	Change in velocity (logarithm of Delta V)
Direction of impact	PDOF (0 to 350 degree in 10-degree increment)
Vehicle Type	Car, Utility (MY 2000 and new car)
Number of Events	Single, Multiple
Belt Use	Belted, Unbelted
Right-front Passenger	Presence of right-front passenger
Age	At least one greater than 55 years
Gender	At least one female

Validation

Five-fold Cross Validation [6] was used to validate the model prediction performance. As shown in Figure 1, D1, D2, D3, D4, and D5 are the five equal parts of NASS-CDS (1999-2015). The model is created using D1, D2, D3, and D4, and the performance is verified using D5. This process is repeated for D1, D2, D3, D5; D1, D2, D4, D5; D1, D3, D4, D5; D2, D3, D4, D5; and for each of D4, D3, D2 and D1. The model was evaluated by taking the arithmetic mean of the values obtained from each validation result (D1, D2, D3, D4, D5) as shown Figure 1.

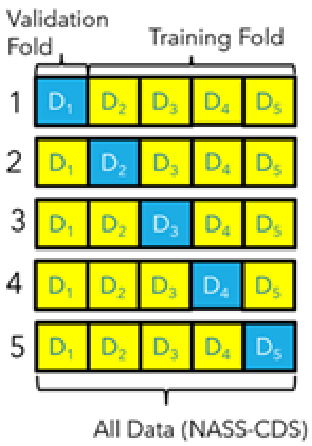


Figure. 1 Five-fold cross-validation.

III. RESULTS

Comparison of injury severity score

Table 3 shows the relationship between each predictor and severity (ISS 15+ vs. ISS 15-) obtained from the NASS-CDS (1999-2015). The numbers in parentheses indicate the mean value for delta V and the standard deviation for the others. For both vehicles, the proportion of frontal collisions was high, and in particular, the proportion of side impact from the left and right was high for ISS 15+. In terms of vehicle body type, there is no significant difference between the two groups, indicating that the proportion of Car (passenger cars) was high. For multiple events, the ISS 15+ tended to have a higher percentage. The percentage of wearing seat belts was higher for ISS 15-. On the other hand, the proportion of right-front passengers was higher in ISS 15+ than in ISS 15-, and the proportion of elderly was higher in ISS 15+ than in ISS 15-. There was no significant difference in gender between the two groups.

Table 3 Means and standard deviations of key variables stratified by ISS 15+ vs. ISS 15-. (SD).

Variables	ISS 15+	ISS 15-
Delta V (mph)	29.0 (0.5)	20.4 (0.1)
Direction of impact	Front: 55.3% (3.0%) Left: 23.0% (2.9%) Right: 17.8% (2.1%) Rear: 3.9% (2.2%)	Front: 72.1% (1.2%) Left: 7.1% (0.8%) Right: 9.6% (0.8%) Rear: 11.3% (0.5%)
Vehicle type	Car: 84.5% (1.4%) Utility: 15.5% (1.4%)	Car: 79.7% (1.1%) Utility: 20.3% (1.1%)
Number of Multiple Events	65.8% (2.4%)	54.1% (2.0%)
Seatbelt usage	63.3% (2.5%)	84.1% (2.4%)
Presence of right-front passenger	38.5% (4.6%)	21.6% (1.4%)
Presence of passengers older than 55 years old	39.3% (3.9%)	17.5% (1.0%)
Presence of female passengers	64.8% (3.2%)	60.6% (1.6%)

Logistic regression analysis using the target vehicles defined in the previous chapter was performed to create the ISP, and the contribution to each predictor was determined. Table 4-5 shows the results of the logistics regression model (ISP-f1R, ISP-f2R) using the stepwise variable selection method including the coefficient, standard error, and p-value for each variable. The most significant variable in the analysis was delta V. The PDOF was approximated by four cyclic curves (C1, C2, C3, C4) obtained from the functional data analysis. By cyclic, it meant values at 0 and 360 were equal, since they correspond to the same direction. Number of curves (4 here) was the degree of the freedom. We basically used the combination of these four curves to approximate the PDOF effect. Those coefficients were determined by "fitting" the model and we therefore tried to find the best combination which was close to the truth. The belt use was significantly different in both models, indicating that wearing a seat belt was safer than not wearing a seat belt. Among the factors related to the occupants, age was the most important factor, with a significant difference of p=0.009 and p=0.010 in both models, respectively. On the other hand, gender was not selected as a predictor, and the selection of variables by AIC was a necessary factor to improve the accuracy of the model, although some variables were not significantly different for some items.

Figure 2-3 shows the results for the ISP-f1R case with only the driver's seat occupant and ISP-f2R case with the right-front passenger in the vehicle, by crash direction (PDOF: 0 deg. (Front), 90 deg. (Right), 180 deg. (Rear), 270 deg. (Left)). The dotted line represents the confidence interval. Other variables were fixed at all occupants belted, car, single impact, and no older occupants. Figure 2(a) shows the ISP-f1R case with only the driver's seat occupant with the highest risk for left-side impact (red), followed by right-side (yellow), rear (black), and front impact (blue). Compared to Kononen's model, the risk curves were higher overall (Frontal, Left, Right, Rear), and the risk curve for rear impact was higher than that for frontal impact. Figure 2(b) shows the ISP-f1R case with a right-front passenger. Compared to the driver-only case in Figure 2(a), all risk curves (Frontal, Right, and Rear) were higher when there was a passenger on the right-front passenger, except for the left side impact. Figure 3(a) shows the ISP-f2R case with only the driver's seat occupant, with the highest risk for left-side impact (red), followed by right-side (yellow), front (blue), and rear impact (black). In the case of the driver seat occupant alone, the risk curve for rear impact tended to be close to that of frontal impact. On the other hand, Figure 3(b), which considered the effect of the right-front passenger, shows a higher risk curves compared to Figure 3(a), except for the left-side impact.

Table 4. Estimation of coefficients and their standard

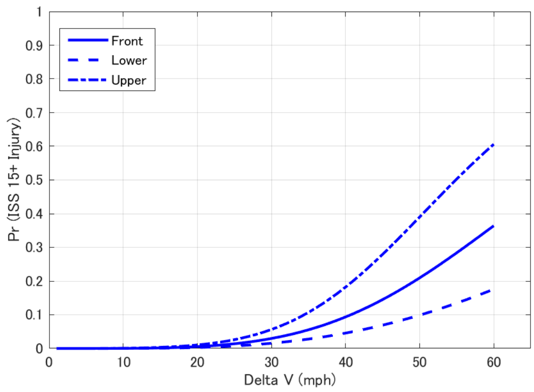
Table 5. Estimation of coefficients and their standard

error and p-value for variables from the ISP-f1R model.

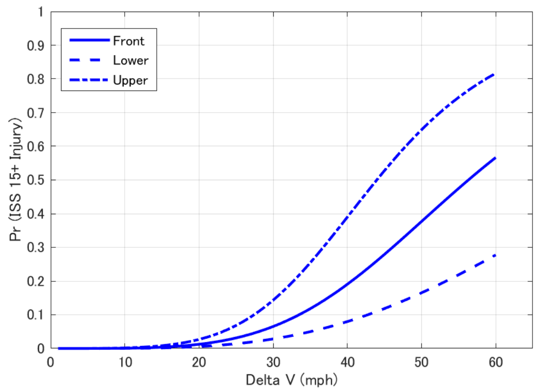
Parameters		Estimate	Std. Error	t value	Pr(>|t|)
Intercept		-13.842	1.094	-12.647	0.006
ln Delta-V (mph)		4.229	0.329	12.866	0.006
PDOF	C1	-5.642	1.510	-3.736	0.065
	C2	1.969	0.576	3.421	0.076
	C3	-2.989	0.317	-9.437	0.011
	C4	-2.588	0.327	-7.914	0.016
Belt use	Belted	-1.283	0.196	-6.538	0.023
	Unbelted	0.000			
Vehicle type	Utility	-0.499	0.150	-3.327	0.080
	Car	0.000			
Number of events	Multiple	0.351	0.166	2.117	0.169
	Single	0.000			
Presence of older occupants	55 or greater	1.517	0.142	10.695	0.009
	under 55	0.000			
Presence of right-front passenger	C1	2.222	1.201	1.850	0.205
	C2	-1.138	0.574	-1.982	0.186
	C3	1.322	0.767	1.724	0.227
	C4	0.495	0.245	2.019	0.181

error and p-value for variables from the ISP-f2R model.

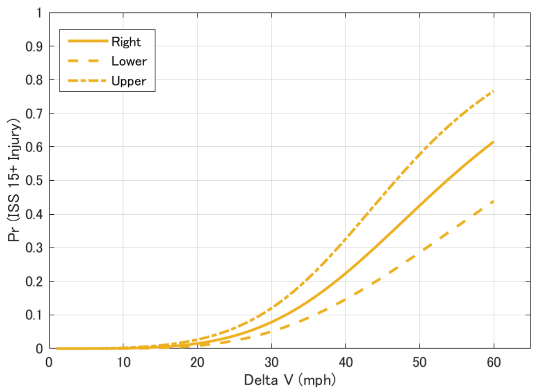
Parameters		Estimate	Std. Error	t value	Pr(>|t|)
Intercept		-12.754	1.016	-12.554	0.006
ln Delta-V (mph)		4.171	0.313	13.308	0.006
PDOF	C1	-6.076	1.566	-3.881	0.060
	C2	1.347	0.427	3.154	0.088
	C3	-3.238	0.281	-11.518	0.007
	C4	-3.908	0.373	-10.481	0.009
Belt use	Belted	-1.256	0.192	-6.541	0.023
	Unbelted	0.000			
Vehicle type	Utility	-0.472	0.153	-3.095	0.090
	Car	0.000			
Number of events	Multiple	0.356	0.183	1.945	0.191
	Single	0.000			
Presence of older occupants	55 or greater	1.534	0.157	9.787	0.010
	under 55	0.000			
Presence of right-front passenger	C1	1.026	1.292	0.794	0.510
	C2	-0.220	0.787	-0.280	0.806
	C3	0.519	0.677	0.767	0.523
	C4	1.123	0.375	2.994	0.096



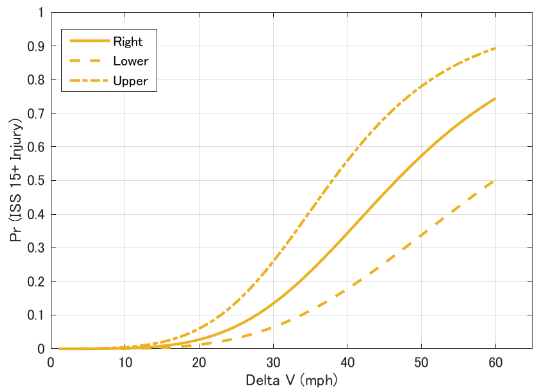
PODF=0 deg. (Front)



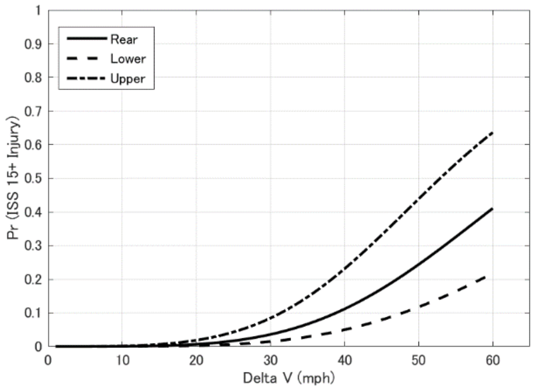
PODF=0 deg. (Front)



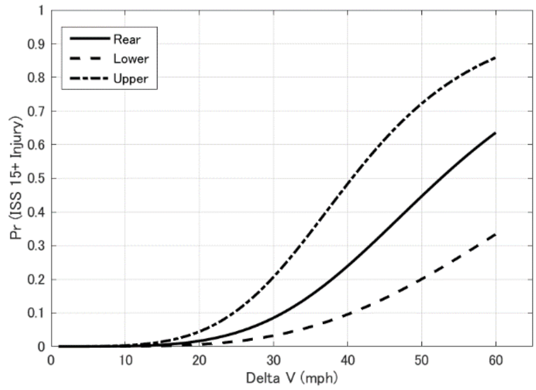
PODF=90 deg. (Right)



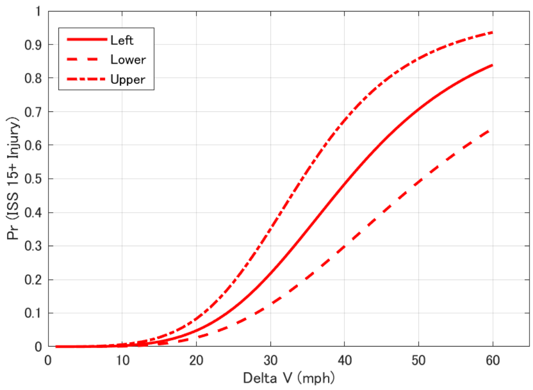
PODF=90 deg. Right



PODF=180 deg. (Rear)

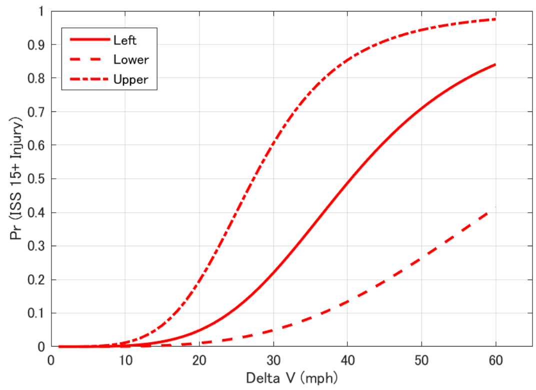


PODF=180 deg. (Rear)



PODF=270 deg. (Left)

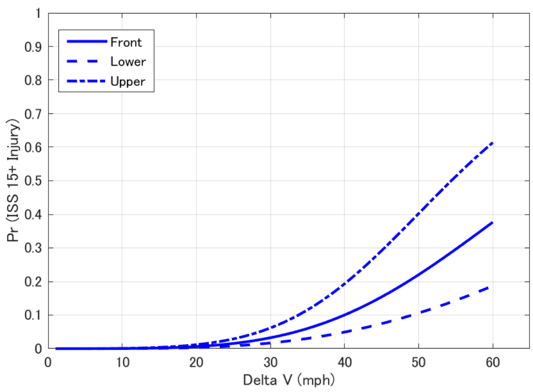
(a) Driver only



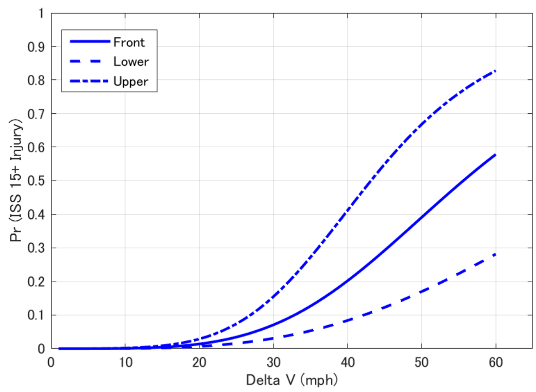
PODF=270 deg. (Left)

(b) Driver with a right-front passenger

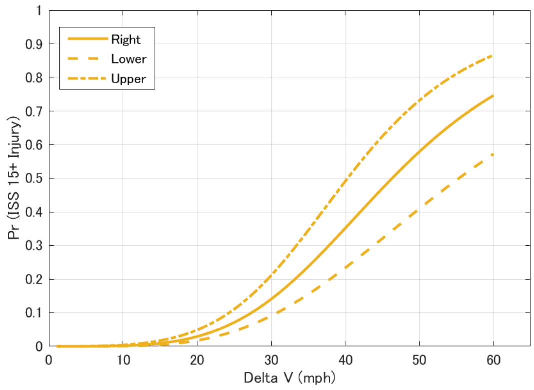
Figure 2 Result of risk prediction curve using ISP-f1R. The figure showed the injury risk curve for four impact directions: front (blue), left (red), right (yellow), and rear (black).



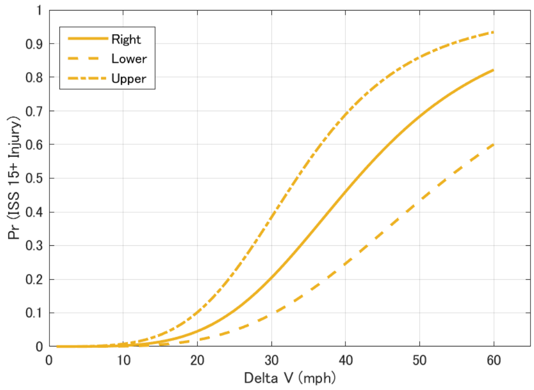
PODF=0 deg. (Front)



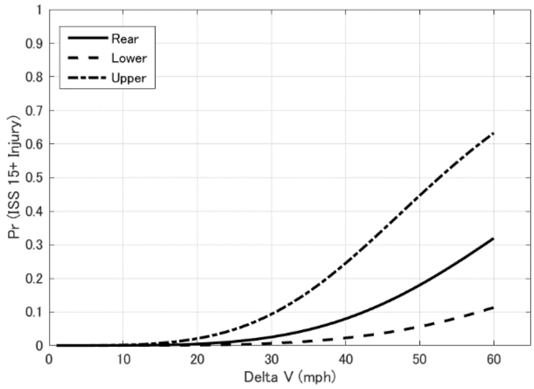
PODF=0 deg. (Front)



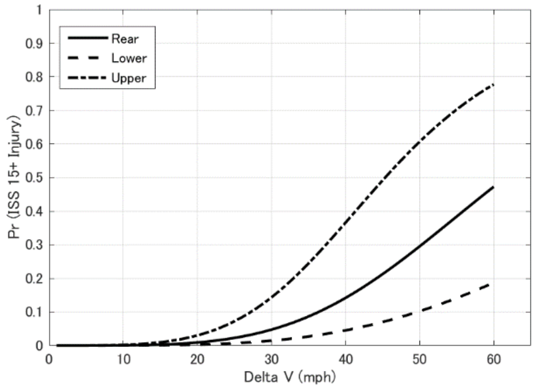
PODF=90 deg. (Right)



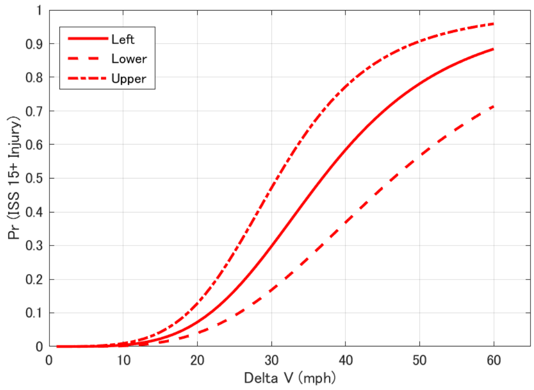
PODF=90 deg. (Right)



PODF=180 deg. (Rear)

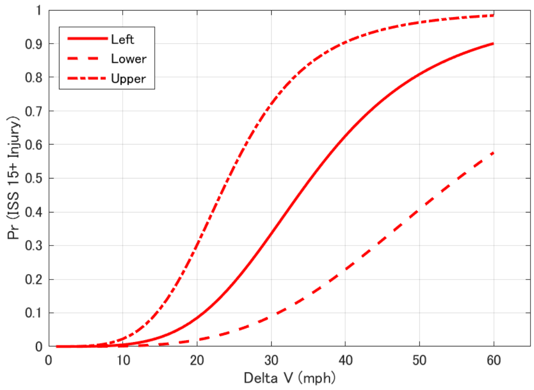


PODF=180 deg. (Rear)



PODF=270 deg. (Left)

(a) Driver only



PODF=270 deg. (Left)

(b) Driver with a right-front passenger

Figure 3 Result of risk prediction curve using ISP-f2R. The figure showed the injury risk curve for four impact directions: front (blue), left (red), right (yellow), and rear (black).

The results of the five-fold cross-validation using the ISP-f1R and ISP-f2R models were shown in Table 6-7, and the sensitivity and specificity of the ISP models when the cutoff values were varied from 0.10, 0.15, 0.20, 0.25, and 0.30. The area under the receiver operator characteristic curve (AUCs) was used as the metric to evaluate model performances, which was 0.847 for the ISP-f1R model and 0.856 for the ISP-f2R model for the cross-validation, respectively.

Table 6 Comparison of sensitivity and specificity with the five-fold cross-validation with the ISP-f1R model.

cutoffs	sensitivity	specificity
0.10	0.581	0.906
0.15	0.461	0.945
0.20	0.389	0.968
0.25	0.349	0.979
0.30	0.316	0.984

Table 7 Comparison of sensitivity and specificity with the five-fold cross-validation with the ISP-f2R model.

cutoffs	sensitivity	specificity
0.10	0.579	0.899
0.15	0.474	0.946
0.20	0.406	0.967
0.25	0.360	0.977
0.30	0.320	0.983

IV. DISCUSSION

Injury severity prediction models were developed using NASS-CDS (1999-2015). The model showed a higher risk curves than the risk curves proposed by Kononen et al. [8] for each direction of impact. The main reason for this difference was due to the “Body Type” selected from NASS-CDS (1999-2015). While Kononen used Car (Code: 0-9) and Utility (Code; 14-19) as body types, the current analysis uses Car (Code. 2-4, 6) and Utility (Code; 14-15) as belonging to the Subaru vehicle categories were selected. Subaru vehicles were relatively small or medium-sized, and the difference in vehicle weight was considered to have resulted in a higher risk curve. Generally, in a crash between vehicles, the heavier vehicle will push the lighter vehicle out of the way, resulting in a large impact force more on the people in the lighter vehicle [14]. Although the vehicle body structure absorbs energy and mitigates the impact to the occupants [15], the selection of a relatively light vehicle (small or medium-sized) for the vehicle may have resulted in a higher risk curve due to the weight of the vehicle.

As shown in Figure 2, the ISP-f1R risk curve for rear impact was higher than that for front impact. This was due to the ISP-f1R's quantile-based knot positioning. As shown in Table 3, the data used in this analysis was biased toward frontal impact, and the number of rear impact was small, which might be a reason why the feature values could not be reflected accurately. In addition, about 70% of rear impacts occur at less than 15 mph[16], which may have been excluded from the scope of our analysis. Although rear impacts are considered unlikely to cause fatalities, they are extremely hazardous once they occur, depending on the severity of the collision and the severity of the injuries being investigated. Further collection of data on rear impacts is needed to investigate the relationship between Delta V and severity of injuries.

On the other hand, ISP-f2R with the knot position set to 90, 180, and 270 degrees produced the risk curve by capturing the rear impact feature, indicating that the risk of rear impact was the smallest compared to the other three directions. However, as can be seen from the ISP-f2R risk curves, there were limitations in developing the model using approximately 7,000 data points, and it was not possible to set the degree of freedom to more than 5. Ended up using DOF = 4, which was smaller than the DOF (=10) of Wang et al[10]. In addition, Kononen et al.[8] modeled by direction of impact (DOF=3), and the results of current analysis did not fully demonstrate the advantages of functional data analysis.

As shown in Figure 2(b) and Figure 3(b), a new severity risk term was added when a passenger was on board, as indicated by the logistic regression analysis results (Table 4-5). The effect of the right-front passenger was evident in the risk curves for frontal, right-side, and rear impact, with the exception of the left-side impact. This result indicated that ISP-f2R can distinguish differences in severity risk with and without a right-front passenger. As for the effect of the right-front passenger on the left side impact, it was found that the presence or absence of the passenger seat occupant did not make a significant difference, as shown by Ejima et al. [17], and the same results were obtained.

Five-fold cross-validation using NASS-CDS (1999-2015) showed that the ISP-f1R model performed with a sensitivity of 38.9% and specificity of 96.8% with a cutoff value of 0.2, and the accuracy of the model was evaluated with an AUC of 0.847. On the other hands, the ISP-f2R model performed with a sensitivity of 40.6% and specificity of 96.7% with a cutoff value of 0.2, and the accuracy of the model was evaluated with an AUC of 0.856. As shown in Table 8, the sensitivity and specificity of the model by Kononen et al. [8] were 39.6% and 98.3%, respectively, and the results were similar to those of Kononen’s model. The main problem we see from the Kononen's model is that the model is driver-centric, but it tries to predict the severe injury risk of the whole car. In the side-impact scenario, however, presence of additional right-front passenger would greatly change the injury risk of the while car. By incorporating the interaction between right-front passenger and impact direction, we can effectively address the issue above. The U.S. Centers for Disease Control and Prevention's

Specialty Council on Field Triage recommends an acceptable range of 0-5% and 25-50% for under-triage (1-Sensitivity) and over-triage (1-Specificity), respectively. Sensitivity and specificity were in a trade-off relationship, and it is important to balance sensitivity and specificity in setting cutoff values, and it is necessary to discuss appropriate triage decisions based on field data.

Table 8 Summary of the specificity, sensitivity and AUC of the models (cutoff value of 0.2)

	Sensitivity	Specificity	AUC
Kononen et al.[8]	0.396	0.983	0.843
ISP f1R	0.389	0.968	0.847
ISP f2R	0.406	0.967	0.856

Limitations

Delta V is an important variable among the information obtained from the vehicles in this algorithm and is calculated in a computer model (WinSmash) based upon the detailed vehicle crash measurements obtained by the crash investigators in the NASS-CDS database. On the other hand, Delta V obtained from telematics (EDR) is calculated from the lateral and longitudinal components of Delta V and is also used to calculate the PDOF. Differences can be observed between the two, and the Delta V from WinSmash is often underestimated compared to telematics, which should be taken into account when verifying accuracy using field data.

This algorithm was a model that selects vehicle categories and could predict severity at the vehicle level; as described in the paper by Kononen et al. [8], a model at the occupant level was needed to improve the accuracy of the injury severity prediction algorithm. At the scene of an actual traffic accident, it is assumed that telematics in the car will make the call and treat the patient in cooperation with the emergency services. To properly triage the injured, it is necessary to collect information on each passenger (Driver, Right-front passenger, and rear seated occupants). In the future, we plan to study the possibility of extending the algorithm to the severity prediction algorithm at the occupant level.

In addition, the NASS-CDS used in this analysis was based on data from 1999-2015, which may differ from current vehicles, and therefore validation with field data is necessary.

V. CONCLUSIONS

In this study, the Subaru vehicle categories were selected as vehicle body types from the NASS-CDS, and an algorithm for predicting occupant injury severity was developed by applying functional data analysis. According to the guidelines of the Centers for Disease Control and Prevention's Expert Committee on Field Triage, the ISP-f1R model showed 38.9% sensitivity and 96.8% specificity, and the ISP-f2R model showed 40.6% sensitivity and 96.7% specificity when cutoff value was 0.20. The area under the receiver operator characteristic curve (AUCs) was used as the metric to evaluate model performances of models, which was 0.847 for the ISP-f1R model and 0.856 for the ISP-f2R model, which was an improvement compared to Kononen's model. The main predictors were delta V, seat belt use, and age, and the influence of the right-front passenger during right-side impact was also significant. These models need to be continuously improved by collecting the crash data.

VI. REFERENCES

[1] MacKenzie, EJ., Rivara, FP., Jurkovich, GJ., et al. A national evaluation of the effect of trauma-center care on mortality. The New England Journal of Medicine, 354:366–378. 2006.

[2] Sasser, SM., Hunt, RC., Sullivent, EE., et al. Guidelines for field triage of injured patients: recommendations of the National Expert Panel on Field Triage., MMWR Morbidity and mortality weekly report, 58:1–35. 2009.

[3] Baker, SP., O'Neill, B., Haddon, WJr., Long, WB., The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *Journal of Trauma.*14:187–196. 1974.

[4] Bahouth, GT., Digges, KH., Bedewi, NE,. et al. Development of URGENCY2.1 for the prediction of crash injury severity. Top Emergency Medicine. 26:157–165. 2004.

[5] Ayoung-Chee, P,, Mack, CD., Kaufman, R., Bulger, E., Predicting severe injury using vehicle telemetry data. Journal Trauma Acute Care Surgery. 74, 190–194; discussion 4–5. 2013.

[6] Kusano, KD., Gabler, HC., Comparison and Validation of Injury Risk Classifiers for Advanced Automated Crash Notification Systems. Traffic Injury Prevention, 15, S126–S133, 2014.

[7] McMurry, TL., Cormier, JM., Daniel, T., et al. An omni-directional model of injury risk in planar crashes with application for autonomous vehicles. Traffic Injury Prevention, DOI: 10.1080/15389588.2021.1955108, 2021.

[8] Kononen, DW., Flannagan, CA., Wang, SC., Identification and validation of a logistic regression model for predicting serious injuries associated with motor vehicle crashes. Accident Analysis and Prevention, 43:112–122. 2011.

[9] National Highway Traffic Safety Administration. National Automotive Sampling System. Washington, DC: National Highway Traffic Safety Administration; Available at: https://www.nhtsa.gov/crash-data-systems/national-automotive-sampling-system. Accessed May 15, 2020.

[10] Wang, SC., Kohoyda-Inglis, CJ. et al., Second Generation Aacn Injury Severity Prediction Algorithm: Development and Real-World Validation. 25th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Detroit, United State. Paper Number: 17-0133-O, 2017.

[11] National Highway Traffic Safety Administration, National Automotive Sampling System (NASS) Crashworthiness Data System Analytical User's Manual 2007 File. National Center for Statistics and Analysis, NHTSA, U.S. Department of Transportation, Washington, DC.

[12] Association for the Advancement of Automotive Medicine. "Abbreviated Injury Scale (AIS)": Available at: https://www.aaam.org /abbreviated-injury-scale-ais/. Accessed May 1, 2020.

[13] Wang SC., Kohoyda-Inglis, CJ. et al. Results of First Field Test of Telemetry Based Injury Severity Prediction. 24th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Gothenburg, Sweden. Paper Number: 15-0388-O, 2015.

[14] The Insurance Institute for Highway Safety (IIHS). Vehicle size and weight. https://www.iihs.org/topics/vehicle-size-and-weight#how-size-and-weight-affect-safety. Accessed Oct. 12, 2021.

[15] Teoh, RE., Nolan, JM., Is Passenger Vehicle Incompatibility Still a Problem? Traffic Injury Prevention, 13:585–591, 2012.

[16] Parentreau, C., Why Rear Crash Safety Is Not That Simple, https://www.exponent.com/~/media/knowledge /thought-leadership/2020/01/why-rear-crash-safety-is-not-that-simple/tl_why-rear-crash-safety-is-not-that-simple.pdf. Accessed June 1, 2022.

[17] Ejima, S., Goto, T., Zhang, P., Cunningham, K., and Wang, S., Injury Severity Prediction Algorithm based on Select vehicle Category for Advaced Automatic Collision Notification, SAE Technical Paper 2022-01-0834, 2022, doi:10.4271/2022-01-0834.