FIELD ASSESSMENT OF GM/ONSTAR OCCUPANT-BASED INJURY SEVERITY PREDICTION

Susan Owen
General Motors
USA

Stewart Wang, Peng Zhang
University of Michigan
USA

Paper Number 23-0123

ABSTRACT

Injury Severity Prediction (ISP) models provide emergency responders with a rapid assessment of potential serious injury for occupants in vehicle accidents around the globe. ISP models predict the need for high level trauma care so that appropriate Emergency Medical Services (EMS) can be dispatched as quickly as possible to improve patient outcomes. In 2020, OnStar implemented its first occupant-based ISP models which predict outcomes for specific seating locations.[1] Models were developed and validated with NHTSA NASS CDS [2] and CISS [3] data. This paper seeks to assess model performance in the field using vehicle-based crash data and real-world occupant outcomes. This study leverages data from a sample of over 1,500 Michigan Advanced Automatic Crash Notification (AACN) events involving over 1,700 front row occupants to assess model performance. Vehicles include model years 2013 to 2020 and span several segments, including passenger cars, SUVs, and light trucks. AACN telemetry data and ISP-predicted outcomes are compared to actual Injury Severity Scores (ISS) for transported occupants. For non-transport cases, police reported injury severities (KABCO scores) are also examined. Measures of sensitivity, specificity, and likelihood ratio are calculated. False negative cases are used to understand model limitations. A range of threshold values used to assess "high" injury risk are also explored to highlight potential tradeoffs. Statistical analyses show that front row occupant models predict ISS 15+ injuries with high levels of accuracy. Metrics compare occupant-based model performance to prior vehicle-based ISP formulations. This study demonstrates that models based on government-sampled data sets are producing reliable results in the field.

INTRODUCTION

For over 20 years, Automatic Crash Notification (ACN) and Advanced Automatic Crash Notification (AACN) have leveraged vehicle telemetry to assist Emergency Medical Services (EMS) in providing rapid response to vehicle accidents. Injury Severity Prediction (ISP) models support this effort by providing a rapid assessment of potential serious occupant injury. Specifically, such models predict the need for high level trauma care so that appropriate resources can be dispatched as quickly as possible to improve patient outcomes.

Recently, the National Expert Panel on Field Triage released an update to the U.S. field triage guidelines.[4] The guidelines are intended to identify research-based criteria for determining when a seriously injured person should be transported to a trauma center, where "serious injury" is aligned with having an Injury Severity Score (ISS) greater than 15. In addition to updating the guideline based on current research findings, the Expert Panel also focused on reducing the rate of under-triage cases (defined as incidents where a seriously injured person is transported to a non-trauma medical center for treatment). The Panel chose a positive likelihood ratio value of 2 as a threshold for identifying research-based factors that merited inclusion in the guidelines. "Vehicle telemetry data consistent with severe injury" [4] was retained as a factor in the updated guidelines, though the authors noted a need for studies focused on evaluating the use of telemetry for field triage.

In 2020, GM/OnStar implemented its first occupant-based ISP models which predict outcomes for specific seating locations.[1] Models were developed and validated with NHTSA NASS CDS [2] and CISS [3] data to predict the occurrence of serious injury outcomes in line with the Field Triage guidelines (i.e., models produce a "High" injury severity prediction if there is a 20% or higher probability of resulting in an ISS of 15+). This paper describes efforts to assess model performance in the field using vehicle-based crash data and real-world occupant outcomes. It also considers potential trade-offs between over- and under-triage based on reducing the threshold for triggering a "High" prediction.

METHODS

U.S. crash investigations documented in CDS and CISS data sets provide detailed information on passenger vehicle collisions, including data on both crash dynamics and occupant outcomes. While these data sets are valuable for building and validating ISP models, it is important to further assess model performance on current field vehicle populations. Such testing ensures performance levels are maintained for actual delta-V inputs (rather than estimated values reported in CDS and CISS) and a broad spectrum of vehicle- and collision types (including late model year vehicles not represented in the latest NHTSA data sets due to lags in data reporting).

To conduct a large-scale field assessment of occupant-based ISP models, a broad set of Michigan AACN cases with calculated ISP scores were matched to available state UD-10 police reports, EMS data, and medical records. Researchers from the University of Michigan Hospital summarized medical outcomes, identifying potential co-morbidities and other factors beyond the scope of ISP models. Due to a small number of second-row occupants with severe injury (less than 5 cases), analysis focused on first-row occupants only. When possible, ISP predicted injury severity was compared to actual ISS score to assess model performance using a confusion matrix and associated statistical metrics of sensitivity, specificity, and positive likelihood ratio.

Although crashes must meet a deployment threshold to trigger an AACN event, they represent a range of crash severities. Many cases do not involve EMS transport and not all occupants receive medical treatment. For cases without noted transport, it is impossible to know if occupants elect to seek treatment on their own for sustained injuries, either through a hospital or other medical provider. If hospital records cannot be located, the only available data on potential injury is captured in police-reported KABCO scores. Previous research has shown that KABCO scores can vary dramatically from medically assessed ISS values.[5] For purposes of this study, cases are thus separated into two groups: “matched” cases which include hospital-based occupant ISS scores, and “unmatched” cases which only have KABCO score assessments of occupant injury. For the unmatched set, KABCO scores of “K” and “A” are used to represent cases with severe injury. While it would be cleaner to consider only cases with actual ISS scores, this would bias analysis results by ignoring most cases for which the ISP models are relied upon to accurately identify potential severe injury.

In addition to validating the field performance of occupant-based ISP models, the current study compares their performance to the prior vehicle-based ISP model.[6] In this model, a “High” risk of severe injury is assessed at the vehicle level based on aggregated information about all occupants. In the field, OnStar Emergency Advisors in contact with vehicle occupants determine whether there are any female occupants or any occupants over age 55 and can recalculate ISP findings accordingly. For the current study, age and gender are used to evaluate these parameters when available. The resulting prediction of whether a vehicle is likely to have any occupant with ISS greater than 15 is compared to the maximum first row occupant injury level. Performance metrics for the vehicle-level predictions are compared to those for occupant-based predictions to evaluate any potential loss of predictive power in shifting to more granular models.

Finally, analysis to quantify the tradeoff between over- and under-triage is conducted by lowering the threshold for classifying an occupant as having a “High” probability of severe injury. Shifting from a 20% threshold to a 5% threshold in 5% increments, model sensitivity and specificity are re-evaluated at each level. The resulting plot reveals the increase in over-triaged occupants (who may be sent unnecessarily to a high level trauma center) associated with capturing previously under-triaged occupants.

RESULTS

AACN Cases

The data set for this analysis included a large sample of vehicles that experienced non-rollover AACN events in Michigan between June 2014 and June 2021. 604 vehicles had AACN data matched to police reports, EMS, and hospital records for at least one occupant, while 965 vehicles were in the “unmatched” set that included only AACN and police report data. Vehicles spanned model years 2013 to 2020, with the majority (99%) from 2014 to 2019 and about half from model years 2016 and 2017. Aside from removing predicted rollover cases (for which ISP is not calculated due to the complexity of non-horizontal forces), no filters were applied to AACN cases based on crash type or severity. A variety of vehicle types and sizes were included in the data set, spanning passenger cars (15%), SUVs (53%), and light trucks (32%).

Model Results

Quantitative summaries Table 1 summarizes model performance for vehicle drivers (left front seat position) in matched cases in a confusion matrix based on predicted outcome and actual ISS values. The sensitivity for these cases is 67%, specificity is 89%, and positive likelihood ratio is 6.

Table 1.
Confusion matrix for left front occupants in matched cases

Predicted 20% probability of ISS 15+

Actual ISS 15+	Yes	No	Row totals
Yes	**6**	**3**	**9**
No	**57**	**477**	**534**
Column totals	**63**	**480**	**543**

Table 2 summarizes right front passenger model performance for matched cases in a confusion matrix based on predicted outcome and actual ISS values. The sensitivity for these cases is 50%, specificity is 88%, and positive likelihood ratio is 4.

Table 2.
Confusion matrix for right front occupants in matched cases

Predicted 20% probability of ISS 15+

Actual ISS 15+	Yes	No	Row totals
Yes	**1**	**1**	**2**
No	**13**	**97**	**110**
Column totals	**14**	**98**	**112**

Table 3 provides driver model performance for non-matched cases based on police-reported KABCO scores. Assuming that "KA" values align with ISS values of 15+, these cases have a sensitivity of 46%, specificity of 98%, and positive likelihood ratio of 19.

Table 3.
Confusion matrix for left front occupants in unmatched cases

Predicted 20% probability of ISS 15+

KABCO in "KA"	Yes	No	Row totals
Yes	**5**	**6**	**11**
No	**23**	**935**	**958**
Column totals	**28**	**941**	**969**

Table 4 provides right front passenger model performance for non-matched cases based on police-reported KABCO scores. These cases have a sensitivity of 100%, specificity of 96%, and positive likelihood ratio of 26.

Table 4.
Confusion matrix for right front occupants in unmatched cases

Predicted 20% probability of ISS 15+

KABCO in "KA"	Yes	No	Row totals
Yes	**2**	**0**	**2**
No	**5**	**125**	**130**
Column totals	**7**	**125**	**132**

When matched and unmatched cases are taken together, the driver model evaluates to a sensitivity of 55%, specificity of 95%, and a positive likelihood ratio of 10. The right front passenger model evaluates to a sensitivity of 75%, specificity of 93%, and a positive likelihood ratio of 10.

False negative cases Four front passengers in the matched data set were under-triaged, assessed with a "not high" ISP but found to have severe (ISS 15+) injury outcomes. Further examination revealed confounding factors beyond the scope of ISP models in three of the four cases:

1. A 60 year old driver experienced an apparent pre-crash medical event and had confounding co-morbidities that resulted in an ISS of 29 in a crash which evaluated the probability of high injury severity to be only 0.4%.
2. A 42 year old belted driver was struck by a falling object that came through the windshield, resulting in an ISS of 19. The associated crash was assessed to have probability of high injury severity of 0.5%.
3. A fatal 86 year old belted driver experienced an apparent pre-crash medical event and had confounding co-morbidities that resulted in an ISS of 17. The associated crash was assessed to have an ISP of 19.4%, just below the ISP "High" threshold.

In the unmatched data there appear to be 6 under-triaged drivers based on police-reported "KA" injury scores, four of which were listed as not being transported by EMS. Without further information, it is impossible to know if the assigned KABCO scores accurately reflected occupant injuries.

Comparison to vehicle-based ISP Table 5 provides a confusion matrix for the 1553 vehicle-based ISP assessments. Using this method, the sensitivity for these cases is 48%, specificity is 97%, and positive likelihood ratio is 15.

Table 5.
Confusion matrix for vehicle-based ISP applied to combined matched/unmatched cases

Predicted 20% probability of ISS 15+

Any occupant ISS 15+ or "KA"	Yes	No	Row totals
Yes	**12**	**13**	**25**
No	**50**	**1478**	**1528**
Column totals	**62**	**1491**	**1553**

Over- Vs. Under-Triage Tradeoff

As described above, analysis was conducted to explore over- vs. under-triage tradeoffs by setting the threshold for "High" ISP evaluation below the current 20% level. Figure 1 shows the impact of dropping the threshold from 20% to 5% in 5-point increments. As expected, reducing the threshold probability causes more occupants to be classified as "High" risk, and thus reduces the amount of under-triage while also increasing the amount of over-triage. The positive likelihood ratio falls from 10 at the 20% threshold to just under 4 at the 5% threshold. While moving the threshold from 20% to 5% allowed the algorithm to drop the under-triage (false negative) group from 10 occupants to 5, it did so at substantial cost – more than tripling the over-triage (false positive) group from 100 occupants to 379.

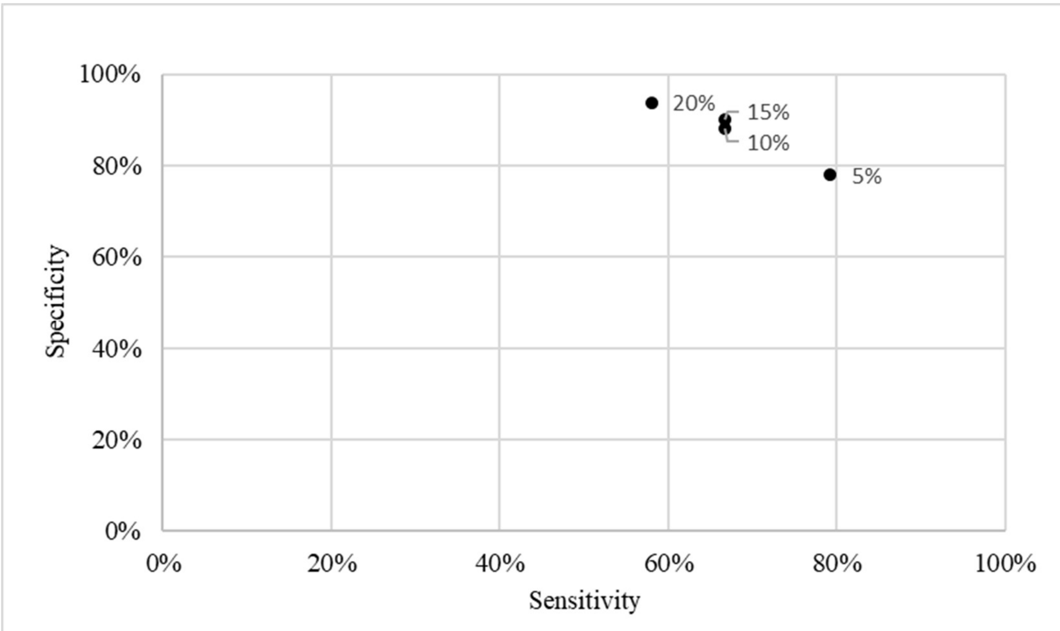

Figure 1. Performance tradeoff associated with lowering the threshold for front row HIGH ISP from 20% to 5%.

DISCUSSION AND LIMITATIONS

The analysis of cases in this study provides a field-based assessment of occupant-based ISP models for front row outboard occupants. Additional models for second row occupants have been fit, validated, and implemented, but limited populations of seriously injured second row occupants make it difficult to capture meaningful statistics on model performance. The cases were pulled from a broad sample of Michigan AACN crashes due to accessibility of police report, EMS, and medical outcome data. While the cases span several model years, vehicle platforms, and crash severities, they may not be representative of the overall US AACN crash population.

Based on this study, results demonstrate that occupant-based ISP model performance for front row passengers is strong. At the current 20% threshold, the positive likelihood ratio of 10 far exceeds the level specified in updated Field Triage Guidelines. Compared to the vehicle-based model, the front row occupant-based ISP models improve sensitivity (from 48% to 58%) with a slight degradation in specificity (from 97% to 94%). This performance is promising, given the potential value that occupant-based models provide – enabling OnStar to identify for EMS the number and location of severely injured occupants.

Work to monitor model performance in the U.S. is ongoing. Efforts to quantify 2nd row model performance continue as additional cases are collected and analyzed. Additional exploration of lowering the threshold for "High" probability of severe injury also continues. Tradeoffs need to consider the systemic costs of over- and under-triage. While the Expert Panel focused on potential improvements in occupant outcomes associated with reducing under-triage, recent NHTSA publications (e.g., [7]) highlight constraints in EMS resources that could be exacerbated by increasing levels of over-triage. Finally, research to quantify the effect of AACN/ISP on EMS response times and occupant outcomes is needed, but requires access to significant volumes of crash, EMS, and injury data not currently available in the US.

ACKNOWLEDGEMENTS

The authors acknowledge John Capp, Director of Vehicle Safety Technology, Strategy, & Regulations at GM, and Cathy Bishop, Senior Manager of Emergency Services at OnStar in their continued support for improved system safety, including post-crash emergency response.

REFERENCES

[1] Owen SH, Joyner JW, Zhang P, Wang SC. Occupant-based Injury Severity Prediction. Stapp Car Crash J. 2021 Nov; 65:17-28.
[2] Radja GA. National Automotive Sampling System – Crashworthiness Data System, 2015 Analytical User's Manual. Report No. DOT HS 812 321. 2016 Sept. Washington, DC: National Highway Traffic Safety Administration.
[3] National Center for Statistics and Analysis. Overview of the 2017 Crash Investigation Sampling System. Traffic Safety Facts Research Note. Report No. DOT HS 812 787. 2019, Sept. National Highway Traffic Safety Administration.
[4] Newgard CD, Fischer PE, Gestring M, Michaels HN, Jurkovich GJ, Lerner EB, Fallat ME, Delbridge TR, Brown JB, Bulger EM. National Guideline for the Field Triage of Injured Patients: Recommendations of the National Expert Panel on Field Triage, 2021. J Trauma Acute Care Surg. 2022 Aug; 93(2):e49-e60.
[5] Burdett B, Li Z, Bill AR, Noyce DA. Accuracy of Injury Severity Ratings on Police Crash Reports. Trans. Res. Rec. 2015 Jan; 2516(1):58-67.
[6] Wang S, Kohoyda-Inglis C, Ejima S, MacWilliams J, Zhang P, Stacey L, Melocchi A, Gorman D, Kral J, Joyner J. Second Generation AACN Injury Prediction Algorithm: Development and Real-World Validation. 25th International Conference on Enhanced Safety of Vehicles. 2017; 17-0133.
[7] Office of Behavioral Safety Research. Continuation of research on traffic safety during the COVID19 public health emergency: January – June 2021. Report No. DOT HS 813 210. 2021 Oct. National Highway Traffic Safety Administration.