Accident Analysis and Prevention 43 (2011) 112–122

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Identification and validation of a logistic regression model for predicting serious injuries associated with motor vehicle crashes

Douglas W. Kononen[a,*], Carol A.C. Flannagan[b], Stewart C. Wang[c]

[a] General Motors Corporation, Vehicle Structure and Safety Integration, 30200 Mound Road, Warren, MI 48090, United States
[b] University of Michigan Transportation Research Institute, United States
[c] University of Michigan Medical Center, United States

ARTICLE INFO

Article history:
Received 1 April 2010
Received in revised form 26 July 2010
Accepted 29 July 2010

Keywords:
Logistic regression
Clinical prediction modeling
Motor vehicle crash injuries

ABSTRACT

A multivariate logistic regression model, based upon National Automotive Sampling System Crashworthiness Data System (NASS-CDS) data for calendar years 1999–2008, was developed to predict the probability that a crash-involved vehicle will contain one or more occupants with serious or incapacitating injuries. These vehicles were defined as containing at least one occupant coded with an Injury Severity Score (ISS) of greater than or equal to 15, in planar, non-rollover crash events involving Model Year 2000 and newer cars, light trucks, and vans. The target injury outcome measure was developed by the Centers for Disease Control and Prevention (CDC)-led National Expert Panel on Field Triage in their recent revision of the Field Triage Decision Scheme (American College of Surgeons, 2006). The parameters to be used for crash injury prediction were subsequently specified by the National Expert Panel. Model input parameters included: crash direction (front, left, right, and rear), change in velocity (delta-V), multiple vs. single impacts, belt use, presence of at least one older occupant (≥55 years old), presence of at least one female in the vehicle, and vehicle type (car, pickup truck, van, and sport utility). The model was developed using predictor variables that may be readily available, post-crash, from OnStar®-like telematics systems. Model sensitivity and specificity were 40% and 98%, respectively, using a probability cutpoint of 0.20. The area under the receiver operator characteristic (ROC) curve for the final model was 0.84. Delta-V (mph), seat belt use and crash direction were the most important predictors of serious injury. Due to the complexity of factors associated with rollover-related injuries, a separate screening algorithm is needed to model injuries associated with this crash mode.

1. Background

Models for identifying and predicting the potential severity of occupant injuries associated with highway crashes can be used to both direct appropriate first responder resources to the crash scene and provide critical information to emergency trauma centers to facilitate appropriate preparations for receipt of transported seriously injured occupants (Bahouth et al., 2004). Modern telematics communications systems available to vehicle owners (e.g., OnStar®) can provide immediate information about the nature and severity of a motor vehicle collision. This information can serve as input to predictive models designed to classify a crash as a high injury probability event or a low injury probability event. Some of the information immediately available to telematics systems from Event Data Recorders (EDRs) in modern vehicles may include: principal direction of impact force (PDOF), the total change in vehicle velocity during the event (delta-V), seat belt use status of occupants and frontal or side airbag deployment. This information, when coupled with vehicle information derived from the vehicle identification number (VIN) (e.g., vehicle type and weight) and voice contact with vehicle occupants, when available, can serve as the basis for targeted allocations of first responder services to ensure that appropriately equipped and trained Emergency Management Services (EMS) are dispatched to the crash scene.

Since automatic collision notification (ACN) is relatively new, only a few studies have been published. Augenstein et al. (2007) discuss some of the predictors of crash outcome that can be measured by the EDR, including crash direction and delta-V. They illustrate the relationship between these variables and crash outcome and discuss the benefit of both ACN and the potential for more extensive crash data from EDR reports. Rauscher et al. (2009) report on the first field experience with an ACN system that transmits geographic coordinates of the crashed vehicle. This report focuses primarily on the types of voice contact (or lack thereof) that are made after notification, but it includes some information about outcome. These data

* Corresponding author. Tel.: +1 586 899 5510.
E-mail addresses: douglas.w.kononen@gm.com, kononend@aol.com (D.W. Kononen).

0001-4575/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.aap.2010.07.018

provide a starting point for learning about the potential benefits of ACN, especially when no voice contact can be made.

In 2008, the Centers for Disease Control (CDC) published the recommendations of National Expert Panel on Field Triage. In the new Field Triage Decision Scheme (Sasser et al., 2009), "vehicle telematics" was added to the list of crash-related information to consider in determining whether to transport an occupant to a hospital. At that time, the specifics of how to use vehicle telematics were not addressed, but the expert panel made several recommendations related to future use of such information.

In this paper, we present an algorithm for predicting risk of serious injury in a vehicle as a function of crash parameters that can be obtained from the EDR or from voice communication with the vehicle. The algorithm was developed for OnStar®, and in some instances, modeling choices were made based on the capabilities of the OnStar® system. However, the model as a whole is applicable to all vehicle makes and models. Moreover, the algorithm follows the recommendations of the expert panel and is based on publically available national crash data.

The objective of this modeling effort was to develop a statistical model to predict the risk of serious injury using outcome and predictor variables identified by the CDC National Expert Panel on Field Triage based on an analysis of planar motor vehicle crashes recorded in the 1999–2008 National Automotvie Sampling System (NASS-CDS) database files for model year 2000 and newer vehicles. The model will be incorporated into the existing OnStar® response-center system to help identify crashes that have high potential to result in severe injuries.

2. Data and methods

Data from the National Automotive Sampling System (NASS) Crashworthiness Data System (CDS), years 1999–2008, were used to develop and validate a multivariate logistic regression model of serious injury as a function of those predictor variables that may be readily transmitted from Event Data Recorder (EDR) modules to the OnStar® system. The NASS-CDS database is a complex stratified sample of crashes in the United States (National Highway Traffic Safety Administration, 2007). A NASS-CDS crash must: (1) be police reported, (2) involve a harmful event (property damage and/or personal injury) resulting from a crash and (3) involve at least one towed passenger car or light truck or van in transport on a trafficway.

From a practical perspective, this modeling effort was restricted to using only those variables that can be immediately and accurately obtained via telematic transmission (i.e., EDR data) supplemented with information obtained from voice communication with crash-involved vehicle occupants, if available. This constraint on potentially significant explanatory variables was imposed to reflect the current state of information available from current vehicle telematic systems.

Most predictive models of occupant injury outcome of which we are aware focus on the occupant level (e.g., Bahouth et al., 2004; Augenstein et al., 2007). However, in the scenarios being considered here, a vehicle's telematic system would make the initial contact and Emergency Medical Services (EMS) would respond to the vehicle/crash as a whole. Thus, the models in this paper are designed to predict which *vehicles* involved in a crash are likely to contain a seriously injured occupant, and all modeling is done at the vehicle level. Occupant information available to the EDR, as well as any gathered from voice contact, is coded to the vehicle level. Methods for this are discussed in the sections describing each predictor.

Although with many systems (including OnStar at this time), EDR-based occupant information may only be available for front-seat occupants, we included the maximum injury of any passenger in a vehicle. In this way, the model represents the current state of information available to the EDR, but predicts the complete range of vehicles and their occupants. As rear-seat information (occupancy and belt use) becomes available, prediction may improve. However, it is worth noting that the most severely injured occupant is a rear-seat occupant in only 3.5% of vehicles in NASS-CDS, so future information about rear-seat occupant presence and belt use may provide limited additional predictive value.

Published models of serious injury associated with motor vehicle crashes are often based upon an injury outcome criterion such as MAIS3+ (Bahouth et al., 2004; Farmer, 2003). For the purposes of this study, Injury Severity Score (ISS) was considered a better, more clinically reliable indicator of severe injury than indices based upon the score attributed to a single (presumably most severe) coded injury using the Abbreviated Injury Scale (AIS) injury coding system (e.g., MAIS3+). Because the ISS score, defined as the sum of the squares of the AIS severity level of the three most significant coded occupant injuries, captures a larger portion of an injured occupant's "harm profile", it is thought to provide a better, more realistic assessment of occupant harm than does a simpler univariate score such as MAIS3+. While there are many ways to dichotomize the severity of injuries, the National Expert Panel chose ISS of 15 as the partition during their revision of the Field Triage Decision Scheme (Sasser et al., 2009). The National Expert Panel also chose a 20% probability of ISS 15 or greater as the threshold for considering individual triage criterion for inclusion in the Field Triage Decision Scheme. Thus, the dependent measure for this study is the binary variable, ISS 15+, indicating whether any occupant of a vehicle experienced an injury of ISS 15+ or not.

The final sets of data exclusions were related to the nature of the vehicles that contain EDRs. Although EDRs have been introduced slowly by manufacturer and model, they were generally not available before model year 2000. In addition, we limited our sample to planar collisions (i.e., excluding rollovers and the rare crashes coded with the primary general area of damage as top or bottom) and passenger vehicles. Finally, cases with weights of 5000 and up were trimmed (excluded) to improve standard errors (Little et al., 1997; Potter, 1990). Previous experience using the NASS–CDS complex survey data indicate that cases with weights greater than 5000 are usually extreme outliers that often exert a large influence on resulting model parameter estimates and their standard errors. Cases with such large weighting factors are very rare and readily identified. Our analyses with and without these influential cases demonstrated that a single case with a weight of 5000 or greater can dramatically change some model parameter estimates and their standard errors. Therefore, these cases were excluded from subsequent analyses. The resulting dataset contained 14,673 vehicles. Of these, 1212 (8.3%) contained one or more occupants with ISS 15+ injuries.

The CDC National Expert Panel on Field Triage recommended the following variables as predictors of serious occupant injury risk (ISS 15+): delta-V, principal direction of force (front, left, right, rear), seat belt restraint use (yes vs. no), vehicle type (car, sport utility, pickup, passenger van) and multiple vs. single crash events. In addition, information about occupant age and gender may be obtained if verbal contact is made by an OnStar operator or EMS dispatcher.

The following section describes each predictor variable in detail.

2.1. Delta-V

Delta-V is the change in vehicle velocity associated with the primary direction of force of the crash event. In the NASS-CDS database, delta-V is defined as the difference between Impact Velocity and Separation Velocity and is calculated by a computer model (WinSmash) based upon detailed vehicle crush measurements obtained by NASS-CDS crash investigators. Typically, the