

What's Speed Got To Do With It?

2010-01-0526

Published
04/12/2010G. Nusholtz and L. DiDomenico
Chrysler Group LLC

Copyright © 2010 SAE International

ABSTRACT

The statistical analysis of vehicle crash accident data is generally problematic. Data from commonly used sources is almost never without error and complete. Consequently, many analyses are contaminated with modeling and system identification errors. In some cases the effect of influential factors such as crash severity (the most significant component being speed) driver behavior prior to the crash, etc. on vehicle and occupant outcome is not adequately addressed.

The speed that the vehicle is traveling at the initiation of a crash is a significant contributor to occupant risk. Not incorporating it may make an accident analysis irrelevant; however, despite its importance this information is not included in many of the commonly used crash data bases, such as the Fatality Analysis Reporting System (FARS). Missing speed information can result in potential errors propagating throughout the analysis, unless a method is developed to account for the missing information.

In this paper, a method is presented to estimate the effect of missing speed information on the statistical effect of different variables (age, vehicle mass, ...) on injury or fatality. The method developed uses historical data, such as the crash velocity distributions and the injury as a function of velocity - obtained from the National Accident Sampling System/Crash Data System (NASS/ CDS)- to estimate the modeling errors or the confounding effect expected from not considering the velocity factor in accident analysis. The method can be used to correct the level of statistical significance or to adjust the potential effect of explanatory crash variables, obtained from an accident analysis, when there is missing velocity information.

Examples are provided where the suggested analytical procedure is used to estimate the confidence intervals for previously published risk of fatality based on crash data as well as to estimate a factor effect adjustment due to the missing speed data.

Keywords

Velocity, Crash Speed, Injury Risk, Average Risk, Risk of Fatality

INTRODUCTION

The determination of the risk of injury for an occupant involved in a vehicle crash is a difficult task. The complexity of crash analysis is aggravated by the possible lack of initial conditions and information about influential and relevant factors that could limit the ability to determine statistical relevance or may cause an identification of factors as significant, when they are not. For example, the critical factor speed is often missing; speed plays an essential role in explaining a vehicle's crash severity. However, this information is not straightforward to obtain, even impossible to get in some cases.

The missing information on crash speed could lead to the identification of crash surrogate variables as significant accident factors when in fact they are not explanatory. In particular, the missing crash speed consideration could raise the importance of unessential or minor factors to that of major factors, when they do not play a foundational role in a crash scenario. For example FARS does not provide crash speed information (i.e. delta-v). Thus in some FARS based analyses a control on the crash severity -in a two car frontal impacts- is performed using the other car driver fatality outcome. However, this methodology will implicitly raise the importance of unessential factors such as the other driver's medical conditions, the other driver's age or the other driver's seat belt use. Another example is provided by statistical analyses in which vehicle intrusion is used as a crash speed

surrogate. Vehicle intrusion is definitively related to crash speed but it is not interchangeable with it: using the vehicle intrusion will partly correct for the crash speed effect and will introduce errors of its own. A potential way to account for the missing speed information is to perform a “parallel” analysis that, using a-priori knowledge of the estimated crash speed range for the scenario under investigation, can be used to estimate possible speed effects on the occupant risk. This information can then be used to adjust the estimate of the variable under investigation to account for the speed effect.

In this study, we will present an analytical approach as well as a simulation procedure that can be used to estimate an approximate velocity effect in data sets where the crash speed is not directly available. Included is a definition of the probabilistic model along with the evaluation of the theoretical expected risk and its variance. Also included is a simulation procedure and examples are provided to supplement the theoretical approach for more complex speed distribution functions.

DEFINITION OF THE PROBABILITY MODEL

It is assumed that the crash speed, v , is a random variable with gamma probability density function (pdf) $f(v)$:

$$f(v) = \frac{1}{\Gamma(\alpha)\beta^\alpha} e^{-\frac{v}{\beta}} v^{\alpha-1} \quad \text{Eq.1}$$

with parameters α and β , and $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$.

The justification for the selection of the gamma pdf is described in [1]. The parameters depend on the particular crash scenario under examination. For example, in a frontal impact scenario Nusholtz et al. found that $\alpha=5.37$ and $\beta=2.38$ could reasonably approximate the field data [1]. A gamma pdf with parameters $\alpha=5.37$ and $\beta=2.38$ is plotted in Figure 1. From the figure it can be observed that in a frontal impact there is a probability of 0.90 that the crash speed will be inside the interval (5.25, 22.99) mph.

Nusholtz et al. suggested the gamma function because it gave the best fit to the crash severity data among the standard other functions that they checked the data against. For example, Nusholtz et al. did not consider a normal distribution because it does not have as high goodness of fit as the gamma distribution and it will assign positive probability to negative velocity values.

The risk (of either injury or fatality) function usually can be assumed to be an exponential function of the speed (either crash speed or travel speed) [1],[2], and [8]:

$$R(\text{injury}|v) = C v^\eta \quad \text{Eq.2}$$

where η and C are constants whose value depends on the particular occupant risk and v is the speed that depending on the specific study under examination it will be either the travel speed or the crash speed.

Possible values for η and C , suggested in the literature, for frontal impact, when v is the crash speed are:

- $\eta = 4.21$ and $C = 4.00 \times 10^{-8}$ for the risk of serious/fatal injury [1] Eq.2.1

- $\eta = 4.00$ and $C = 4.16 \times 10^{-8}$ for the risk of fatality [2] Eq.2.2

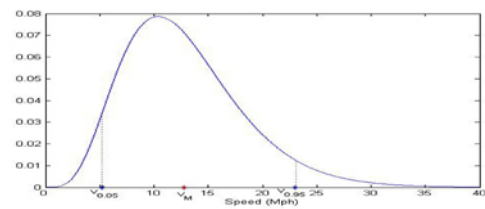


Figure 1. Crash speed (Mph) probability density function (pdf) in a frontal impact [1]. V_M is the mean crash speed ($V_M = 12.78 \text{ Mph}$). $V_{0.95}$ indicates the 95% percentile, i.e., that with probability 0.95 the crash speed for a vehicle involved in a frontal impact is lower than $V_{0.95}$ ($V_{0.95}=22.99 \text{ Mph}$). Similarly, $V_{0.05}$ is the 5% percentile, i.e., with probability 0.05 the crash speed for a vehicle involved in a frontal impact is lower than $V_{0.05}$ ($V_{0.05}=5.25 \text{ Mph}$).

AVERAGE RISK

Assuming that the crash speed is a random variable with pdf defined in Eq.1 and the risk is expressed by Eq.2 then the average risk is:

$$\begin{aligned} E(Risk) &= \int_0^\infty R(\text{injury}|v) f(v) dv = \int_0^\infty C v^\eta \frac{1}{\Gamma(\alpha)\beta^\alpha} e^{-\frac{v}{\beta}} v^{\alpha-1} dv = \\ &= C \frac{\Gamma(\eta+\alpha)}{\Gamma(\alpha)} \beta^\eta \int_0^\infty \frac{1}{\Gamma(\alpha+\eta)\beta^{\alpha+\eta}} e^{-\frac{v}{\beta}} v^{\alpha+\eta-1} dv \end{aligned} \quad \text{Eq.3}$$

Thus the overall expected risk can be found by evaluating the integral as:

$$E(Risk) = C \frac{\Gamma(\eta + \alpha)}{\Gamma(\alpha)} \beta^\eta \quad Eq.4$$

For example, for the frontal crash scenario, with parameter values $\alpha = 5.37$, $\beta = 2.38$, $\eta = 4.21$ and $C = 4.00 \times 10^{-8}$ then the expected risk of serious/fatal injuries is:

$$E(Risk) = (4 \times 10^{-8}) \frac{\Gamma(4.21+5.37)}{\Gamma(5.37)} 2.38^{4.21} = 0.0052 \quad Eq.5$$

This means that the expected risk of either serious or fatal injury is about 0.5% when a driver is involved in a crash.

RISK VARIANCE

Using the same procedure that was used to evaluate the average risk, the risk variance is given by:

$$Var(Risk) = E(Risk^2) - E(Risk)^2 = C^2 \frac{\Gamma(\alpha + \eta)^2}{\Gamma(\alpha)^2} \beta^{2\eta} \left(\frac{\Gamma(2\eta + \alpha)}{\Gamma(\eta + \alpha)^2} \Gamma(\alpha) - 1 \right) \quad Eq.6$$

For example, in a frontal crash scenario, the standard deviation (*Std*) of serious/fatal injury risk is:

$$Std(Risk) = \left(4 \times 10^{-8} \frac{\Gamma(4.21+5.37)}{\Gamma(5.37)} 2.38^{4.21} \right) \dots \sqrt{\left(\frac{\Gamma(2 \times 4.21 + 5.37) \Gamma(5.37)}{\Gamma(4.21 + 5.37)^2} - 1 \right)} = 0.0132 \quad Eq.6.1$$

In some cases, pole impacts, rear impacts, guardrail impact, roll-overs for example, the crash speed distribution could become very complex, and difficult to handle analytically. A simulation procedure allows estimation of the risk effect in situations where it is not possible to follow the analytical approach described above.

SIMULATION PROCEDURE

It is assumed that after a crash event has been observed there is only one of two possible occupant scenarios: either a driver with a serious/fatal injury or a driver with no/minor injury. Thus, the driver's condition after a crash -at speed v_i - can be modeled using independent distributed Bernoulli random variables, χ_i , with parameter $Risk(injury|v_i)$. The total number of seriously/fatally injured drivers in N crashes, F_N , is given by the sum of all the crash outcomes:

$$F_N = \sum_{i=1}^N \chi_i \quad Eq.7$$

Where $\chi_i = 1$ for serious/fatal injury and $\chi_i = 0$ for no/minor injury.

TOTAL FATALITY MEAN

The mean of the total number of seriously/fatally injured drivers can be evaluated as [3]:

$$E(F_N) = E \left[E \left[\sum_{i=1}^N \chi_i \mid v \right] \right] = E(N) E(Risk) \quad Eq.8$$

Where the operator $E[\cdot | v]$ is the conditional expectation given crash speed v [6] and N is independent of the crash speed. Eq.8 implies that the average number of serious/fatal injured drivers is given by the product of the average number of total crashes times the average risk of serious/fatal driver injury. When the total number of crashes, N , is a known constant then:

$$E(F_N) = N * E(Risk) \quad Eq.8.1$$

Then, the average number of serious/fatal injured drivers is given by the product of the total number of crashes, N , times the average risk of serious/fatal driver injury.

SIMULATION SCHEME

The simulation procedure consists of M repeat runs,

In each run, a random vector of size N of either crash speed or travel speed (depending on the model assumptions) is generated. Where N is the total number of crashes. A vector of serious/fatal injury risk is then calculated based on the assumed risk model. For example, in a frontal crash the crash speed is generated by sampling from a gamma pdf with parameters $\alpha=5.37$ and $\beta=2.38$ (Figure 1). Then, the risk of serious/fatal injury is evaluated using Eq.2.

A vector of binary serious/fatal outcome is obtained based on the serious/fatal injury risk from the previous step.

In particular, in each run of the above scheme an estimate is obtained for the average risk. With M runs, M estimates are obtained for the average risk. Therefore it is possible to obtain the median for the M estimates. An application of this procedure is presented in the following section.

The software built to execute the above fleet simulation steps was verified by confirming that it could reproduce solutions of Eq.4 and Eq.6 with error less than (1×10^{-6}) with 200 runs. In particular, the simulation has been performed for different values of η and C (see Eq.2) suggested in the literature ([1],[2], and [8]). Assuming a gamma pdf for the crash speed then from all the different values of η and C (see Eq.2) that have been used the simulation outputs show an empirical probability density function (pdf) for the risk that