Comparing the Application of the Product of Baseline Models and Accident Modification Factors and Models with Covariates: Predicted Mean Values and Variance

Dominique Lord*
Assistant Professor
Zachry Department of Civil Engineering
Texas A&M University
3136 TAMU
College Station, TX 77843-3136
Tel. (979) 458-3949
Fax. (979) 845-6481
Email: d-lord@tamu.edu

Pei-Fen Kuo
Research Assistant
Zachry Department of Civil Engineering
Texas A&M University
3136 TAMU
College Station, TX 77843-3136
Tel. (979) 862-3446
Fax. (979) 845-6481
Email: p-kuo@ttimail.tamu.edu

Srinivas R. Geedipally
Engineering Research Associate
Texas Transportation Institute
Texas A&M University
3136 TAMU
College Station, TX 77843-3136
Tel. (979) 862-3446
Fax. (979) 845-6481
Email: srinivas8@tamu.edu

Mar 4, 2009

Word Count: 4,324 + 2,500 (6 Tables + 4 Figures) = 6,824 Words

Submitted for the Presentation and Publication at the Transportation Research Board
*Corresponding Author
ABSTRACT

The objective of this research is to compare the application of full models (or models with several covariates) and baseline models, estimated using data meeting specific nominal conditions, combined with Accident Modification Factors (AMFs) for predicting motor vehicle crashes. The analysis focuses on the predicted values and associated inferences for both types of model. Over the last few years, some researchers have questioned the approach of multiplying baseline models with AMFs. In order to compare the models, full and baseline models are estimated using data collected on rural four-lane highways in Texas. AMFs describing the safety effects related to lane and shoulder widths as well as the number of horizontal curves per mile were extracted from previous work. Two scenarios describing different AMF values and their associated uncertainty were evaluated. The results of this study show that the product of baseline models and AMFs produced a much larger variance, hence wider 95% predicted CI, than the one calculated using the full models. This was consistent for the two scenarios evaluated and for all levels of uncertainty associated with the AMFs. The study concludes that the full model should be used over the product of the baseline model and AMFs when the study objective includes variance as part of the decision-making process.
INTRODUCTION

The Highway Safety Manual (HSM), which is near completion, is a document that will serve as a tool to help practitioners make planning, design, and operations decisions based on safety (see 1, for additional information). The document will serve the same role for safety analysis that the Highway Capacity Manual (HCM) (2) serves for traffic-operations analyses. The purpose of the HSM is to provide the best factual information and tools in a useful and widely accepted form to facilitate explicit consideration of safety in the decision making process.

The predictive methodology described in the HSM consists of first developing baseline models using data that meet specific nominal conditions, such as 12-ft lane and 8-ft shoulder widths for segments or no turning lanes for intersections. These conditions usually reflect design or operational variables most commonly used by state transportation agencies (defined as state DOTs). Consequently, baseline models typically only include traffic flow as covariates (e.g., \( \mu = \beta_0 F_1^{\beta_1} F_2^{\beta_2} \)). The second component of the methodology consists of multiplying the output of such models by AMFs to capture changes in geometric design and operational characteristics (1). An important assumption about using this methodology is that the AMFs are considered independent, which may not always be true in practice. The formulation of the predictive methodology is given by the following:

\[
\mu_{\text{final}} = \mu_{\text{baseline}} \times AMF_1 \times \ldots \times AMF_n
\]  

Where,

- \( \mu_{\text{final}} \) = final predicted number of crashes per unit of time;
- \( \mu_{\text{baseline}} \) = baseline predicted number of crashes per unit of time (via a regression model); and,
- \( AMF_1 \times \ldots \times AMF_n \) = accident modification factors assumed to be independent.

The use of baseline models and AMFs are often preferred over models that include several covariates, sometimes referred to as full models, by transportation safety analysts because the baseline models can be easily re-calibrated when they are developed in one jurisdiction and applied to another (3, 4). AMFs can also be useful when the transportation agency does not have the variables available in its databases.

Over the development of the Manual (more than 10 years), some researchers have questioned the use of baseline models combined with AMFs over the application of full models for predicting crashes. For some, flow-only models may be significantly influenced by the omitted variables bias, a predisposition that can affect the prediction of crashes and the standard errors associated with the estimation of coefficients (5). The lead author of this paper (e.g., 6) and others have indicated that adding several AMFs to baseline models could significantly increase the variance associated with the final prediction, which could make the final prediction less reliable. Given these potential problems, it may be preferable to use full models to predict crashes over the proposed predictive methodology in the forthcoming HSM. So far, nobody has investigated which one of these two types of model should be used over the other.
The objective of this research is to compare the application of full models and baseline models combined with AMFs. The analysis focuses on the predicted values and associated inferences when both types of model are used to predict motor vehicle crashes. The prediction of crashes will be one of the methodologies used in the HSM (Part II). In order to compare the models, full and baseline models were estimated using data collected on rural four-lane highways in Texas. Models were estimated for all crash severities KABCO (K=killed or fatal, A=incapacitating injuries, B=non-incapacitating injuries, C=possible injuries, O=property damage only or PDO) and serious injuries only (KAB) crashes. The data were originally assembled for NCHRP 17-29 titled “Methodology to Predict the Safety Performance of Rural Multilane Highways” (7). The AMFs used in this work were also estimated from that work: lane width, shoulder width and horizontal curve density on each segment. Then, the two modeling approaches were compared as a function of the predicted values, the 95% predicted confidence intervals (CIs) and different hypothetical variances used for characterizing the uncertainty associated with each AMF. Two scenarios with different AMF values were evaluated, one based on a base set of AMFs for lane width, shoulder width and curve density and the second with different values of (and thus AMFs for) shoulder width and curve density.

The paper is divided into six sections. The first section described background material about the modeling process, the computation of the 95% CIs, and the estimation of the variance for the product of baseline models and AMFs. The second section describes the characteristics of the data used for estimating the models. The third section presents the modeling results. The fourth section describes the comparison analysis. The fifth section provides important discussion points. The last section summarizes the key findings of this research and outlines ideas for further work.

BACKGROUND

This section describes background material about the modeling process, the computation of the 95% CIs, and the estimation of the variance for the product of baseline models and AMFs.

MODELING PROCESS

Despite its documented limitations (see, e.g., 8,9), the Poisson-gamma model is the most common type of model used by transportation safety analysts. This model is preferred over other mixed-Poisson models since the gamma distribution is the conjugate of the Poisson distribution (Note: because of the existing limitations with Poisson-gamma models, other researchers have proposed alternative models for analyzing crash data, see 10, and 11). The Poisson-gamma model has the following model structure (8): the number of crashes ‘\(Y_{it}\)’ for a particular ‘‘\(i\)’’ site and time period ‘‘\(t\)’’ when conditional on its mean ‘‘\(\mu_{it}\)’’ is Poisson distributed and independent over all sites and time periods

\[
Y_{it} \mid \mu_{it} \sim Po(\mu_{it}) \quad i = 1, 2, \ldots, I \text{ and } t = 1, 2, \ldots, T
\]
The mean of the Poisson is structured as:

$$\mu_{it} = f(X; \beta) \exp(e_{it}) \tag{2}$$

where,

- $f(.)$ is a function of the covariates ($X$);
- $\beta$ is a vector of unknown coefficients; and,
- $e_{it}$ is the model error independent of all the covariates.

With this characteristic, it can be shown that $Y_{it}$, conditional on $\mu_{it}$ and $\alpha$, is distributed as a Poisson-gamma random variable with a mean $\mu_{it}$ and a variance $\mu_{it} + \alpha \mu_{it}^2$, respectively. (Note: other variance functions exist for the Poisson-gamma model, but they are not covered here since they are seldom used in highway safety studies. The reader is referred to 12 and 13 for a description of alternative variance functions.) The probability density function (PDF) of the Poisson-gamma structure described above is given by the following equation:

$$f(y_{it}; \alpha, \mu_{it}) = \frac{\Gamma(y_{it} + \alpha^{-1})}{\Gamma(\alpha^{-1}) y_{it}!} \left( \frac{\alpha^{-1}}{\mu_{it} + \alpha^{-1}} \right)^{\alpha^{-1}} \left( \frac{\mu_{it}}{\mu_{it} + \alpha^{-1}} \right)^{y_{it}} \tag{3}$$

Where,

- $y_{it}$ = response variable for observation $i$ and time period $t$;
- $\mu_{it}$ = mean response for observation $i$ and time period $t$; and,
- $\alpha$ = dispersion parameter of the Poisson-gamma distribution.

Note that if $\alpha \to 0$, the variance equals the mean and this model reverts back to the standard Poisson regression model.

The term $\alpha$ is usually defined as the "dispersion parameter" of the Poisson-gamma distribution or model (Note: in some published documents, the variable $\alpha$ has also been defined as the “over-dispersion parameter”). This term has traditionally been assumed to be fixed and a unique value applied to the entire dataset in the study. As described above, the dispersion parameter plays an important role in safety analyses, including the computation of the weight factor for the EB method and the estimation of confidence intervals around the gamma mean and the predicted values of models applied to a different dataset than the ones employed in the estimation process.

In this work, the models were estimated using varying dispersion parameters. Poisson-gamma models with a varying dispersion parameter use the same PDF shown in Equation (3) and estimate the same number of crashes for each observation, similar to those produced from the traditional Poisson-gamma model. However, instead of estimating a fixed dispersion parameter, these models use a varying dispersion parameter that can be estimated using the following expression (14-16):
\[ \alpha_{ii} = \exp(Z_{ii} \times \delta_i) \]  
\hspace{1cm} (4)

where,

- \( \alpha_{ii} \) = dispersion parameter for the site \( i \)
- \( Z_{ii} \) = a vector of secondary covariates (are not necessarily the same as the covariates used in estimating the mean function \( \hat{\mu}_{ii} \)),
- \( \delta_i \) = regression coefficients corresponding to covariates \( Z_{ii} \).

With Equation (4), the model can be used for estimating a different dispersion parameter according to the sites’ attributes (i.e., covariates). If there are no significant secondary covariates for explaining the systematic dispersion structure, the dispersion parameters will only contain a fixed value (i.e., constant term), resulting in a traditional Poisson-gamma regression model.

### COMPUTATION OF CONFIDENCE INTERVALS

Confidence intervals are important to assess the uncertainty associated point estimates of predictive models. Wood (17) proposed a method for estimating the confidence intervals for the mean response (\( \mu \)), for the gamma mean (\( m \)), and the predicted response (\( y \)) at a new site having similar characteristics as the sites used in the original dataset from which the model was developed. For Poisson-gamma models, the 95% confidence intervals for \( \mu \), \( m \) and \( y \) are described in Table 1. In this table, \( \eta \) is the logarithm of the estimated mean response \( \mu \), while \( \phi (\alpha = 1/\phi) \) is the inverse dispersion parameter of the Poisson-gamma model, as described above. In this paper, the confidence intervals were only estimated for the predicted value (\( y \)), since the technique proposed in the HSM applies only to predicted values. Furthermore, the confidence intervals were estimated using the varying dispersion parameter, \( \alpha = \frac{1}{e^{\phi L}} \), for the full models (18). It should be pointed out that NB models estimated using a varying dispersion parameter predicts narrower confidence intervals than one estimated using a fixed dispersion parameter (19).

### COMPUTING THE PRODUCT OF RANDOM VARIABLES

Using the theory related to the multiplication of independent random variables (20, 21), Lord (6) proposed a method to estimate the variance and associated confidence intervals for the product of baseline models and AMFs. This theory on which this method is based states that the equations presented below will be exact independent of the type of distribution to which each random variable belongs. For the purpose of this description, \( z \) will be defined as the product of independent random variables:

\[ z = x_1 x_2 x_3 \cdots \]  
\hspace{1cm} (5)

where,
\[ z = \text{the product of independent random variables; and,} \]
\[ x's = \text{random variable taken from any kind of distribution.} \]

It should be pointed out that the mean and variance estimates are defined as \( E[x] = \lambda \) and \( E[(x-\lambda)^2] = \nu \) (second central moment), respectively.

**Mean of a Product**

The mean of the product is the direct application of Equation (5):

\[
z = x_1x_2x_3\cdots
\]
\[
E[z] = E[x_1]E[x_2]E[x_3]\cdots
\]
\[
\lambda_z = \lambda_{x_1}\lambda_{x_2}\lambda_{x_3}\cdots
\]

The mean or average of the product is simply the product of the mean value of the random variables.

**Variance of a Product**

The variance of a product is obtained by taking the expectation square of \( z \):

\[
z^2 = x_1^2x_2^2x_3^2\cdots
\]
\[
E[z^2] = E[x_1^2]E[x_2^2]E[x_3^2]\cdots
\]
\[
(\lambda_z^2 + \nu_z) = (\lambda_{x_1}^2 + \nu_{x_1})(\lambda_{x_2}^2 + \nu_{x_2})(\lambda_{x_3}^2 + \nu_{x_3})\cdots
\]

Note: \( E[x^n] = E[(x-\lambda + \lambda)^n] \);

\[
\text{for } n = 2, E[x^2] = E[(x-\lambda + \lambda)^2] = E[(x-\lambda)^2] + E[2\lambda(x-\lambda)] + E[\lambda^2] = \nu + \lambda^2.
\]

The variance \( \nu_z \) is computed by first calculating the product on the right hand side and then subtracting the square of the mean \( \lambda_z^2 \) computed in Equation (7):

\[
\nu_z = (\lambda_{x_1}^2 + \nu_{x_1})(\lambda_{x_2}^2 + \nu_{x_2})(\lambda_{x_3}^2 + \nu_{x_3})\cdots - \lambda_z^2
\]

Note that if all \( x \)'s equal zero, the variance \( \nu_z \) will also equal zero:

\[
\nu_z = (\lambda_{x_1}^2)(\lambda_{x_2}^2)(\lambda_{x_3}^2)\cdots - \lambda_z^2
\]
\[
\nu_z = \lambda_z^2 - \lambda_z^2 = 0
\]
The reader is referred to Lord (6) for additional information about how apply Equations (5 – 8) in the product of a baseline model and AMFs. The paper also describes an example about how to apply the theory.

DATA CHARACTERISTICS

This section describes the characteristics of the data used for estimating the full and baseline models. The data were collected in Texas (1997-2001) and in Washington State (1999-2002). The data were obtained from the Department of Public Safety (DPS) and the Texas Department of Transportation (TxDOT) for Texas and from the Highway Safety Information System (HSIS) (www.hsis.org) managed by the University of North Carolina on behalf of the Federal Highway Administration for the State of Washington. 1,570 undivided segments were used for developing the full models and the summary statistics are shown in Table 2. Additional information about the characteristics of the data can be found in the final NCHRP 17-29 report (7).

Baseline models are developed using a given set of baseline conditions. The baseline conditions usually reflect the nominal conditions agencies most often used for designing segments and intersections. In this study, nominal baseline conditions used for segments were paved 11-12 ft lanes, 7-8 ft shoulder widths and no horizontal curves (i.e., only tangents). To increase the sample size (at the request of the NCHRP advisory panel), lane widths of 11ft and 12ft and shoulder widths of 7ft and 8ft were combined together. A total of 183 segments were used for developing baseline models and the summary statistics are shown in Table 3. It should be pointed out that baseline models can also be estimated by inputting the nominal value for each element other than flow in the full model. This method is more useful when the sample size of the data is limited, but has not been found to be better than estimating baseline models from data meeting nominal conditions (22).

MODELING RESULTS AND AMFS

This section describes the modeling results. The coefficients of the models were estimated using the PROC NLMIXED function in SAS (23). The functional forms used for the models were as follows:

Full Model: $\mu = \beta_0 LF^{\beta_1} e^{(\beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4)}$

Baseline Model: $\mu = \beta_0 LF^{\beta_1}$

Dispersion parameter: $\alpha = \frac{1}{e^{\gamma_0} L}$
Table 4 summarizes the modeling results for KABCO and KAB models. The model results show that increasing lane and shoulder widths decreases the number of crashes, while adding horizontal curves on segments increases the number of crashes.

Table 5 shows the characteristics of the variance-covariance matrix. The inverse of that matrix is needed for computing the predicted confidence intervals.

Table 6 summarizes the three AMFs used for the two scenarios. The AMFs were taken from the original NCHRP 17-29 report. Since the AMFs do not have standard errors, three hypothetical standard deviation values were used for each AMF: 0.05 (small variance), 0.10 (medium variance), 0.20 (large variance). The objective was to investigate how sensitive were the AMF values for computing the variance of the predicted number of crashes. It should be pointed out that for all three standard deviation values, the AMF could actually lead to an increase in the number of crashes, since the upper bound of the 95% confidence interval crosses the value 1.

The next section presents the analysis results.

COMPARISON ANALYSIS

This section describes the results of the comparison analysis. The results are summarized for each scenario.

SCENARIO 1 – 12-ft lanes, 8-ft shoulders, 4 curves/mile

Figure 1 shows the comparison analysis for the KAB models. Figure 1a shows the 95% CI for the predicted value, while Figure 1b illustrates the absolute difference between the mean of the full model and the product between the baseline model and AMFs as a function of AADT. The figure shows that the product of the baseline model and AMFs predicts more crashes than that the full model for traffic volumes above 5,500 vehicles per day. Even though the difference appears to be large in Figure 1b), the true picture of the difference is seen in Figure 1a). This figure also illustrates that the 95% CI for the predicted values is much smaller for the full model, especially for vehicular traffic above 2,500 vehicles per day.

Figure 2 shows the comparison analysis for the KABCO models for the first scenario. Figure 2a shows the 95% CI for the predicted value, while Figure 2b illustrates the absolute difference between the mean of the full model and the product between the baseline model and AMFs as a function of AADT. Similar to KAB modeling output, Figure 2 shows that the product of the baseline model and AMFs predicts more crashes than that the full model for traffic volumes above 7,000 vehicles per day. Interestingly,
the 95% CI for the KABCO baseline model combined with AMFs is actually much smaller for traffic flow below 5,000 vehicles per day.

SCENARIO 2 – 12-ft lanes, 12-ft shoulder width, and 1 curve/mile

Figure 3 shows the comparison analysis for the KAB models for the second scenario. Figure 3a shows the 95% CI for the predicted value; Figure 3b illustrates the absolute difference between the mean of the full model and the product between the baseline model and AMFs as a function of AADT. Again, the figure shows that the product of the baseline model and AMFs predicts more crashes than that the full model for traffic volumes above 5,000 vehicles per day. This figure also illustrates that the 95% CI for the predicted values is much smaller for the full the model for the entire range of vehicular traffic.

Figure 4 shows the comparison analysis for the KABCO models for the second scenario. This figure shows that the product of the baseline model and AMFs predicts much more crashes than that the full model for traffic volumes above 5,000 vehicles per day. This figure also illustrates that the 95% CI for the predicted values is much smaller for the full the model, especially for vehicular traffic above 3,000 vehicles per day.

DISCUSSION

The comparison analysis detailed above provided very interesting results. First, the results show that the 95% CIs for the predicted values have, in general, much wider boundaries for the product of the baseline models and AMFs than those produced by the full models. This difference gets larger as the traffic flow increases. The difference in the width of the boundaries could be as high as 122%. However, it should be noted that for low traffic volumes (less than 5,000 vehicle per day), the boundaries for the full models were larger than those produced by the baseline combined with AMFs. This is explained by the low predicted values of the baseline models used in Equation 8.

To follow on the first point, the results indicate that the variance produced by the uncertainty associated with the AMFs significantly increases the variance produced from the product of these factors with the baseline models. This was noted for all three standard deviation values used in the two scenarios: 0.05, 0.10 and 0.20. To illustrate this characteristic, an example is shown below. The example used the KABCO models applied in the first scenario for 15,000 vehicles per day and the standard error equal to 0.10 for each AMF. Using the equations described above, the variance for the two types of models can be computed as follows:

\[ \nu_{\text{Full}} = 4.71 \text{ (from Table 1)} \]

\[ \nu_{\text{Baseline}} = 3.67 \text{ (from Table 1)} \]
Baseline Model + AMFs: 

\[ v_{\text{baseline+AMFs}} = \left( 2.62^2 + 3.67 \right) \left( 1^2 + 0.1^2 \right) \left( 1.11^2 + 0.1^2 \right) \]

\[ \left( 1.31^2 + 0.1^2 \right) - 3.82^2 = 8.30 \]

(from Equation 8)

This example shows that the variance associated with the full model is slightly higher than the one estimated from the baseline model, as expected. By including the AMFs, the final variance is increased by about 76% (note: the full model also includes the same variables as the AMFs; yet, the final variance is much smaller). Thus, every new AMF added to the baseline model has an important influence on the final variance of the product, confirming the results provided by Lord (6).

Third, when the 95% CI is included in the analysis, the prediction produced by the product of the baseline model and AMFs is not statistically different from the one produced by the full model. This means that in terms of prediction, both approaches could be used, if the study objective only consists of estimating the mean value of the prediction. However, if one is interested using the model having the smallest variance for decision-making process, hence smaller 95% CIs, the full model should be used over the baseline model combined with AMFs. The lead author recently participated in a seminar for training practitioners about the applications of AMFs in safety analyses. Interestingly, the participants indicated that providing the best and worst case scenarios (i.e., 95% CI) was the most important criterion when AMFs are used as part of these analyses.

In summary, the study outcome shows that models with covariates are preferable over the application of baseline models and AMFs when uncertainty is part of the decision-making process. However, as discussed by Persaud et al. (3) and Lord and Bonneson (4), full models can be very difficult to recalibrate when they are estimated in a given jurisdiction are applied to another jurisdiction.

SUMMARY AND CONCLUSIONS

The objective of this research consisted in evaluating the application of full models and baseline models combined with AMFs for predicting motor vehicle crashes. The analysis focused on the predicted values and associated inferences for both types of model. The prediction of crashes will be one of the methodologies proposed in the forthcoming HSM. At this point in time, the recommended predictive methodology is to use baseline models combined with AMFs. In order to compare the models, full and baseline models were estimated using data collected on rural four-lane highways in Texas; these data were originally assembled for the NCHRP 17-29 research project. AMFs describing the safety effects of lane and shoulder widths as well as the number of horizontal curves per mile were also extracted from that work. Regression models were estimated for all and serious injuries crashes. Then, the two types of model were compared as a function of the predicted values, the 95% predicted CIs and different standard deviations to characterize the uncertainty associated with each AMF. Two scenarios with different values of AMFs were evaluated.

The results of this study show that the product of baseline models and AMFs produce much wider 95% predicted CIs than the ones calculated using the full models, except at low traffic flow volumes. This was consistent for the two scenarios evaluated.
and for all levels of uncertainty associated with the AMFs. The width of the lower and upper boundaries becomes increasingly larger as traffic flow increases. Considering the uncertainty associated predicted values, both types of model can be used to predict motor vehicle crashes. However, if the study objective includes minimizing the variance associated with the prediction, then the full model should be used over the product of the baseline model and AMFs, since the former produces much smaller variance. Further work includes looking at the effects of adding several AMFs to baseline models; in this work, only three were applied. A new method to re-calibrate full models should also be investigated, if full models become the proposed predictive methodology in a future edition of the HSM.

ACKNOWLEDGEMENTS

This paper benefited from the input of three anonymous reviewers. Their comments and suggestions are greatly acknowledged.

References


TABLE 1: 95% Confidence Intervals for $\mu$, $m$, $y$. (Wood, 2005)

TABLE 2. Summary Statistics for the Data Used for the Full Models

TABLE 3. Summary Statistics for the Data Used for the Baseline Models

TABLE 4: Estimated Coefficients for Full and Baseline Models

TABLE 5: Variance-Covariance Matrix for the Full and Baseline Models

FIGURE 1. Comparison Analysis between the Full Model and Baseline Model (KAB) for Scenario 1.

FIGURE 2. Comparison Analysis between the Full Model and Baseline Model (KABCO) for Scenario 1.

FIGURE 3. Comparison Analysis between the Full Model and Baseline Model (KAB) for Scenario 2.

FIGURE 4. Comparison Analysis between the Full Model and Baseline Model (KABCO) for Scenario 2.
### TABLE 1: 95% Confidence Intervals for $\mu$, $m$, $y$. (Wood, 2005)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu$</td>
<td>$\left[ \frac{\hat{\mu}}{e^{1.96\sqrt{\text{Var}(\hat{\eta})}}} \pm 1.96\sqrt{\text{Var}(\hat{\eta})} \right]$</td>
</tr>
<tr>
<td>$m$</td>
<td>$\left[ \max\left{ 0, \hat{\mu} - 1.96 \sqrt{\hat{\mu}^2 + \frac{\hat{\mu}^2 \text{Var}(\hat{\eta})}{\phi} + \hat{\mu}^2} \right}, \hat{\mu} + 1.96 \sqrt{\hat{\mu}^2 + \frac{\hat{\mu}^2 \text{Var}(\hat{\eta})}{\phi} + \hat{\mu}^2} \right]$</td>
</tr>
<tr>
<td>$y$</td>
<td>$\left[ 0, \hat{\mu} + \sqrt{19} \sqrt{\hat{\mu}^2 + \frac{\hat{\mu}^2 \text{Var}(\hat{\eta})}{\phi} + \hat{\mu}^2} \right]$</td>
</tr>
</tbody>
</table>

**Note:**

$\text{Var}(\hat{\eta}) = XX^{-1}X^T$ where $X^{-1}$ is the variance-covariance matrix and $X$ is a matrix of observed values.

$\left\lfloor x \right\rfloor$ denotes the largest integer less or equal than $x$.

For computing any other % CI for $y$, $\sqrt{19}$ can be substituted with $\frac{1.0 - \tau}{\tau}$, where $\tau = 1 –$ percentile in %; For instance, for estimating the upper boundary 90% CI, the upper value becomes $\sqrt{\frac{1.0 - 0.1}{0.1}} = \sqrt{9} = 3$. 

TABLE 2. Summary Statistics for the Data Used for the Full Models

<table>
<thead>
<tr>
<th>Variables</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean (SD)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment length (Miles)</td>
<td>0.1</td>
<td>6.275</td>
<td>0.542 (0.66)</td>
<td>850.399</td>
</tr>
<tr>
<td>AADT (Vehicles/day)</td>
<td>42</td>
<td>33118</td>
<td>6926.0 (4532.5)</td>
<td>--</td>
</tr>
<tr>
<td>Lane Width (Feet)</td>
<td>9.75</td>
<td>17.5</td>
<td>12.69 (1.71)</td>
<td>--</td>
</tr>
<tr>
<td>Shoulder Width (Right + Left – both sides combined) (Feet)</td>
<td>0</td>
<td>40</td>
<td>9.84 (8.05)</td>
<td>--</td>
</tr>
<tr>
<td>Number of Curves</td>
<td>0</td>
<td>16</td>
<td>0.69 (1.30)</td>
<td>1089</td>
</tr>
<tr>
<td>Curve Density (Curves/Mile)</td>
<td>0</td>
<td>18.07</td>
<td>1.46 (2.41)</td>
<td>--</td>
</tr>
<tr>
<td>Total Crashes</td>
<td>0</td>
<td>97</td>
<td>2.85 (5.67)</td>
<td>4478</td>
</tr>
<tr>
<td>KAB Crashes</td>
<td>0</td>
<td>36</td>
<td>0.99 (2.33)</td>
<td>1561</td>
</tr>
</tbody>
</table>
### TABLE 3. Summary Statistics for the Data Used for the Baseline Models

<table>
<thead>
<tr>
<th>Variables</th>
<th>Min.</th>
<th>Max.</th>
<th>Mean (SD)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segment length (Miles)</td>
<td>0.1</td>
<td>3.169</td>
<td>0.349 (0.34)</td>
<td>63.85</td>
</tr>
<tr>
<td>AADT (Vehicles/day)</td>
<td>690</td>
<td>18772</td>
<td>5194.4 (3479.9)</td>
<td>--</td>
</tr>
<tr>
<td>Total Crashes</td>
<td>0</td>
<td>24</td>
<td>1.33 (2.90)</td>
<td>243</td>
</tr>
<tr>
<td>KAB Crashes</td>
<td>0</td>
<td>8</td>
<td>0.49 (1.11)</td>
<td>90</td>
</tr>
</tbody>
</table>
### TABLE 4: Estimated Coefficients for Full and Baseline Models

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Full Model</th>
<th>Baseline Models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>KAB</td>
<td>KABCO</td>
</tr>
<tr>
<td>Ln ((\beta_0))</td>
<td>-6.8567 (0.5069)</td>
<td>-7.6559 (0.3574)</td>
</tr>
<tr>
<td>AADT ((\beta_1))</td>
<td>0.7832 (0.0547)</td>
<td>0.962 (0.0390)</td>
</tr>
<tr>
<td>Lane Width ((\beta_2))</td>
<td>-0.0835 (0.021)</td>
<td>-0.06528 (0.0146)</td>
</tr>
<tr>
<td>Total Shoulder Width ((\beta_3))</td>
<td>-0.01381 (0.004)</td>
<td>-0.0122 (0.003)</td>
</tr>
<tr>
<td>Curve Density ((\beta_4))</td>
<td>0.08127 (0.0177)</td>
<td>0.08546 (0.0129)</td>
</tr>
<tr>
<td>(\gamma_0)</td>
<td>1.2657 (0.1390)</td>
<td>1.512 (0.0781)</td>
</tr>
<tr>
<td>MAD(^b)</td>
<td>0.82</td>
<td>1.74</td>
</tr>
<tr>
<td>MSPE(^c)</td>
<td>2.31</td>
<td>11.20</td>
</tr>
</tbody>
</table>

\(^a\) Insignificant and the dispersion parameter tends towards zero; this is essentially a Poisson model.

\(^b\) Mean Absolute Deviance (MAD) = \(\frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|\), where \(n\) is the sample size, \(y_i\) and \(\hat{y}_i\) are the observed and estimated mean values at site \(i\), respectively.

\(^c\) Mean Squared Predicted Error (MSPE) = \(\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2\).
<table>
<thead>
<tr>
<th>TABLE 5: Variance-Covariance Matrix for the Full and Baseline Models</th>
<th>β0</th>
<th>β1</th>
<th>β2</th>
<th>β3</th>
<th>β4</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KAB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.2572</td>
<td>-0.02406</td>
<td>-0.00369</td>
<td>0.00029</td>
<td>0.00027</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.02406</td>
<td>0.002993</td>
<td>-0.00018</td>
<td>-0.00002</td>
<td>-0.00009</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.00369</td>
<td>-0.00018</td>
<td>0.00044</td>
<td>-0.00003</td>
<td>-1.70E-07</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.00029</td>
<td>-0.00002</td>
<td>-0.00003</td>
<td>0.000019</td>
<td>3.55E-06</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.00027</td>
<td>-0.00009</td>
<td>-1.70E-07</td>
<td>3.55E-06</td>
<td>0.000313</td>
</tr>
<tr>
<td>KABCO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>0.1278</td>
<td>-0.01213</td>
<td>-0.0017</td>
<td>0.000129</td>
<td>0.00007</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.01213</td>
<td>0.001523</td>
<td>-0.0001</td>
<td>-8.80E-06</td>
<td>-0.00004</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>-0.0017</td>
<td>-0.0001</td>
<td>0.000213</td>
<td>-0.00001</td>
<td>1.09E-06</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.000129</td>
<td>-8.80E-06</td>
<td>-0.00001</td>
<td>9.63E-06</td>
<td>1.91E-06</td>
</tr>
<tr>
<td>$\beta_4$</td>
<td>0.00007</td>
<td>-0.00004</td>
<td>1.09E-06</td>
<td>1.91E-06</td>
<td>0.000167</td>
</tr>
<tr>
<td><strong>Baseline Models</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KAB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>2.7832</td>
<td>-0.3135</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.3135</td>
<td>0.03544</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>KABCO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>1.7325</td>
<td>-0.1962</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>-0.1962</td>
<td>0.02231</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Scenario 1</td>
<td>Lane Width (12ft)</td>
<td>Total Shoulder Width (8ft)</td>
<td>Curve Density (4/mile)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------------------</td>
<td>---------------------------</td>
<td>------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.11</td>
<td></td>
<td>1.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scenario 2</td>
<td>Lane Width (12ft)</td>
<td>Total Shoulder Width (16ft)</td>
<td>Curve Density (0/mile)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
a) Comparison of Confidence Intervals

b) Comparison of mean values (Full – Baseline + AMFs)

FIGURE 1. Comparison Analysis between the Full Model and Baseline Model (KAB) for Scenario 1.
a) Comparison of Confidence Intervals

b) Comparison of mean values (Full – Baseline + AMFs)

FIGURE 2. Comparison Analysis between the Full Model and Baseline Model (KABCO) for Scenario 1.
FIGURE 3. Comparison Analysis between the Full Model and Baseline Model (KAB) for Scenario 2.
a) Comparison of Confidence Intervals

b) Comparison of mean values (Full – Baseline + AMFs)

FIGURE 4. Comparison Analysis between the Full Model and Baseline Model (KABCO) for Scenario 2.