

# A New Accident Prediction Model for Highway-Rail Grade Crossings Using the USDOT Formula Variables

Jacob Mathew and Rahim F. Benekohal

*Civil and Environmental Engineering, University of Illinois Urbana-Champaign, Newmark Civil Engineering Building, 205 N. Mathews Ave, Urbana 61801, Illinois, USA*

**Abstract:** This paper presents the ZINDOT model, a methodology utilizing a zero-inflated negative binomial model with the variables used in the United States Department of Transportation (USDOT) accident prediction formula, to determine the expected accident count at a highway-rail grade crossing. The model developed contains separate formulas to estimate the crash prediction value depending on the warning device type installed at the crossing: crossings with gates, crossings with flashing lights and no gates, and crossings with crossbucks. The proposed methodology also accounts for the observed accident count at a crossing using the Empirical Bayes method. The ZINDOT model estimates were compared to the USDOT model estimates to rank the crossings based on the expected accident frequency. It is observed that the new model can identify crossings with a greater number of accidents with Gates and Flashing Lights and Crossbucks in both Illinois (data which were used to develop the model) and Texas (data which were used to validate the model). A practitioner already using the USDOT formulae to estimate expected accident count at a crossing could easily use the ZINDOT model as it employs the same variables used in the USDOT formula. This methodology could be used to rank highway-rail grade crossings for resource allocation and safety improvement.

**Key words:** Highway-rail grade crossing, accident prediction, USDOT formulae, zero inflated negative binomial, empirical Bayes.

## 1. Introduction

Accident prediction models are used to predict the expected accident frequency at railroad grade crossings. Such models are important since they are used to rank railroad crossings for safety improvements and to allocate resources. Mainstream estimation of safety risk at railroad grade crossings dates back to the early 1970s with the consolidation of the railroad grade crossing inventory by public agencies and private railroads. This allowed for the development of a systematic methodology to estimate safety risk at a crossing for prioritizing them for safety improvements.

One of the commonly used accident prediction models for railroad grade crossings is the United States Department of Transportation (USDOT) accident prediction formula [1]. This model was

developed in the 1980s and the formula has not been altered except for a normalizing constant. This model used non-linear regression approach to develop separate formulas for each of the warning device type installed at the crossing (gates, flashing lights with no gates, and crossbucks). Two main criticisms of this model are, age of the model and the lack of any study comparing the model estimates to the accident experience.

Researchers have used other methodologies to estimate the expected accident frequency such as the Poisson regression, Negative Binomial regression, Zero-inflated models [2-4]. These methodologies establish a relationship between the expected accident frequencies at a location as a function of some of the characteristics of the location. The estimates are further improved by considering the accident history of the crossing. The USDOT accident prediction formula employs a method that linearly combines the accident history at the crossing with the initial model estimate. Accident history is used in several hazard

---

**Corresponding author:** Jacob Mathew, Ph.D. candidate in transportation engineering, graduate research assistant, University of Illinois Urbana-Champaign.

models in the literature (Oregon, Utah, Detroit formulae) in an additive format [5]. Another method to include the accident history at the crossing to improve this initial estimate is Empirical Bayes approach [6]. Thus, two components that influence the final estimate of crossing safety are:

- (1) the format of the equations and the variable used in the accident prediction models;
- (2) the adjustment procedure used to account for the accident history at the crossing.

This study explored if a new model format but with the same variables as the USDOT model could provide a more accurate prediction result. Since the purpose of these models is to generate a priority ranking of crossings, a more accurate model is one which can generate a ranked list of crossings that identifies a greater number of crossings with accidents among its top crossings than the USDOT model. Using the same variables as the USDOT model in the new model makes it convenient for anyone who is already using the USDOT model as they would not have to collect additional data to use the new model. A new model format (Zero Inflated Negative Binomial or ZINB) using the USDOT variable is developed to estimate the initial expected number of accidents. The model is developed using the accident data from the state of Illinois between the years 2012-2016. Then, adjustments are made to the initial estimates using the Empirical Bayes method. The model developed using data from Illinois is also validated using data from Texas. Section 2 and Section 3 of this paper discuss the model formats and accident history adjustment procedures.

## **2. Model Formats**

### *2.1 USDOT Model Format*

The USDOT accident prediction formulas are given in the *Highway-Rail Grade Crossing Handbook* [7]. The development of these formulae is credited to Mengert [1] and is based on techniques applying nonlinear multiple regression techniques to crossing

characteristics. Further details about the model are given in the *Summary of DOT Rail-Highway Crossing Resource Allocation Procedure-Revised* [8].

As per the USDOT formula, the initial accident prediction value (a) for crossings is given in Table 1.

Please note that the coefficients in the initial accident prediction (a) in the USDOT formulae as mentioned in the *Highway-Rail Grade Crossing Handbook* [7] are different from the coefficients given in the *Summary of DOT Rail-Highway Resource Allocation Procedures-Revised* [8]. The FRA uses the coefficients mentioned in the Summary of DOT Rail-Highway Resource Allocation Procedures in their *Web Accident Prediction System* [9]. These values were used in this study to calculate the initial estimate using the USDOT model format. The initial accident prediction value (a) is adjusted based on the accident history using the method described in Section 3.1 of this paper.

### *2.2 Zero Inflated Negative Binomials*

Modeling of count data is usually done using the Poisson or the Negative binomial model. Park et al. [10] used Poisson regression on a stratified homogeneous dataset. Austin et al. [2] used a negative binomial regression model after determining that a Poisson model was inappropriate to use due to overdispersion in the data. An extension of standard Poisson and negative binomial regression is zero-inflated probability processes, such as the zero-inflated Poisson (ZIP) and zero-inflated negative binomial (ZINB) regression models. The assumption involved in the Zero-inflated model is that some crossings have a very low probability of accidents that they could be considered virtually safe [3]. The nature of the distribution of accident count data for grade crossings is unique in such a way that the number of crossings that observed zero accidents observed is very large. Only around 7% of the crossings had any accidents as shown in Table 2. This suggests a zero-inflated distribution of accidents across the crossings.

**Table 1 Equations for crossing characteristics factors for USDOT formula.**

		EI	DT	MS	MT	HP	HL
	Formula constant	Exposure index factor	Day through trains factor	Maximum speed factor	Main tracks factor	Highway paved factor	Highway lanes factor
Crossing with gates	0.0005745	$\left(\frac{Aadt * TotalTrn + 0.2}{0.2}\right)^{0.2942}$	$\left(\frac{DayThru + 0.2}{0.2}\right)^{0.1781}$	1	$e^{0.1512 * MainTrk}$	1	$e^{0.1420 * (TrafficLn - 1)}$
Crossing with flashing lights and no gates	0.0003351	$\left(\frac{Aadt * TotalTrn + 0.2}{0.2}\right)^{0.4106}$	$\left(\frac{DayThru + 0.2}{0.2}\right)^{0.1311}$	1	$e^{0.1917 * MainTrk}$	1	$e^{0.1826 * (TrafficLn - 1)}$
Crossings with crossbuck	0.0006938	$\left(\frac{Aadt * TotalTrn + 0.2}{0.2}\right)^{0.2942}$	$\left(\frac{DayThru + 0.2}{0.2}\right)^{0.1781}$	$e^{0.0077 * MaxTtspd}$	1	$e^{-0.5966 * (HwyPved - 1)}$	1

where:

*Aadt* is the annual average daily traffic at the crossing;

*TotalTrn* is total number of trains using the crossing;

*DayThru* is number of daytime thru trains at the crossing;

*MaxTtSpd* is the maximum timetable train speed at the crossing;

*MainTrk* is the number of main tracks at the crossing;

*HwyPved* is a binary variable indicating if the highway is paved (= 1) or not (= 0);

*TrafficLn* is the number of highway lanes at the crossing.

**Table 2** Number of accidents at gated crossings in Illinois.

Accident count (2012-2016)	Number of crossings	Percentage of crossings
0	2,555	92.74
1	178	6.46
2	12	0.44
3	8	0.29
4	2	0.07
Sum	2,755	100

The ZINB model is a two-part model. The first part is the zero-inflation part that is based on the assumption that the possibility of accidents in certain locations is so close to zero that these locations would have zero accident counts. The second part is the count part of the model which gives the distribution of the accident frequency as a count (Negative Binomial) process. The variables that were used in the USDOT formula were used in this model.

Other arguments in favor of using a zero-inflated model are:

(a) Previous research in this area [11, 12] compared three different count models (Poisson, Negative Binomial (NB) and Zero Inflated Negative Binomial Models (ZINB)). Their research suggested a better goodness-of-fit for the ZINB model as compared to the Poisson or the NB models.

(b) The ZINB model gives parametric expressions for the expected accident counts and the variance of the expected accident counts making the ZINB model mathematically convenient. These values are required for the Empirical Bayes method used in this paper.

(c) ZINB models have been shown to give good results in the estimate of accident frequency in several other studies as well [3, 13, 14].

Mathematically, the accident counts ( $a$ ) would have the following probability distribution per the ZINB models.

$$a = \begin{cases} 0 & \text{with probability } \varphi \\ g(a|x) & \text{with probability } 1 - \varphi \end{cases} \quad (1)$$

This yields a probability mass function for individuals with zero counts as described by Eq. (2)

and for individuals with counts greater than zero as described in Eq. (3).

$$P(a = 0|x) = \varphi + (1 - \varphi)g(0|x) \quad (2)$$

$$P(a > 0|x) = (1 - \varphi)g(a|x) \quad (3)$$

where:

$a$  is the number of accidents at the crossing with characteristics  $x$ ;

$x$  is the vector of variables input to the model for crossing;

$g(a)$  generates the count from a negative binomial process for crossing.

The estimate for the expected value for  $a$  and its variance is given in Eqs. (4) and (5) below.

$$E[a] = \mu(1 - p) \quad (4)$$

$$V[a] = \mu(1 - p)(1 + \mu(p + \alpha)) \quad (5)$$

where:

$\mu$  is the mean of the negative binomial process described by  $g(a|x)$ ;

$\alpha$  is the over dispersion parameter of the negative binomial model;

$p$  is the probability of the entity being in the "always 0" case in the finite mixture model.

Details about parameter estimation for ZINB models using maximum likelihood estimation are given in Ref. [15].

### 3. Accident History Adjustment Procedures

In this section, two different methods of accident history adjustment to estimate the expected accident count at a railroad grade crossing are discussed. The two methods are the DOT method and the EB method.

#### 3.1 DOT Method

The USDOT accident prediction formulae give an accident prediction value based on a three-step computation. The first step involves the computation of an initial accident prediction based on the crossing parameters (described earlier). In the second step, the adjusted accident prediction value is related to the initial accident prediction value and the accident history at the crossing as:

$$B = \frac{T_0}{T_0 + T} * a + \frac{T}{T_0 + T} * \left(\frac{N}{T}\right) \quad (6)$$

$$T_0 = \frac{1}{0.05 + a} \quad (7)$$

where  $B$  is the adjusted accident prediction value and  $N$  is the number of observed accidents in  $T$  years and “ $a$ ” is the initial accident prediction value (accident prediction value before accident history adjustment). “ $a$ ” for a crossing is related to its crossing traits. The USDOT formulae use normalizing constants which are updated every few years in its third step. Some observations that can be made about the USDOT accident history adjustment include:

(1) The adjusted accident prediction value ( $B$ ) is related to the initial accident prediction value ( $a$ ) and the average number of accidents observed over  $T$  years ( $N/T$ ).

(2)  $B$  is a weighted average of  $a$  and  $N/T$  and therefore would lie between  $a$  and  $N/T$ .

(3) If no accident is recorded in  $T$  years, the expected number of accidents is equal to normalization constant times the initial accident prediction value. The recommended value for  $T$  is 5 years.

### 3.2 Empirical Bayes Method

The Empirical Bayes approach is a way to use two “inputs” to estimate the safety at an entity (railroad grade crossing). The first of the two “inputs” are an initial estimate of the expected number of accidents at the railroad grade crossing. This is estimated based on a reference population which shares the same traits as the crossing in consideration. By defining traits of the railroad grade crossing (like annual average daily traffic (AADT), total number of trains, maximum timetable train speed, etc.), we can define the reference population. “A reference population of entities is the group of entities that share the same set of traits as the entity in the safety of which we have an interest” [16]. The initial estimate of the expected number of accidents could be calculated based on the observed number of accidents at the grade crossings in

the reference population. The second “input” is the accident history recorded (number of accidents observed) at the railroad grade crossing. These are combined as follows [17].

$$B = k * E[a] + (1 - k) * N \quad (8)$$

$$k = \frac{1}{1 + \frac{Var[a]}{E[a]}} \quad (9)$$

where  $E[a]$  is the estimate of the expected number of accidents based on the reference population,  $Var[a]$  is the variance of this initial estimate.  $N$  is the number of accidents observed at the crossing. The duration for which the number of accidents ( $N$ ) is observed at the crossing is equal to the duration of accident counts used to estimate  $E[a]$ .  $E[a]$  can be estimated for a crossing using the multivariate regression method and would depend on the traits of the crossing.

A few observations based on Eq. (8) can be made.

(1) The adjusted accident prediction value is related to the number of observed accidents in the before period and the initial estimate of the expected accident count based on the crossing parameters.

(2) The duration of the before period is the same as the duration of accidents counts used in the estimation of  $E[a]$ .

(3) The adjusted accident prediction value depends on the variance of the initial estimate of the expected accident count  $Var[a]$ .

(4) If  $Var[a]$  is 0, the adjusted accident prediction value is equal to the estimated value for the accident count at the crossing. This means that, if the variance of the estimated expected accident count is 0, the expected number of accidents could be solely predicted based on the crossing parameters.

(5) If  $Var[a]$  is very high, the adjusted accident prediction value for the crossing is influenced more by its accident history observed than the initial estimate based on crossing characteristics.

The Empirical Bayes method has also been used in before-after studies to estimate the effectiveness of safety improvements [18, 19].

### 3.3 Comparison of the Two Accident History Adjustment Methods

Some of the similarities and differences identified between the two methods are listed below.

#### 3.3.1 Similarities

The two different methods of accident history adjustment are a weighted average of the initial estimate of the expected accident count (initial accident prediction value) which is related to the traits at a crossing, and the number of accidents observed at a crossing. Both the models can be written as

$$B = w * a' + (1 - w) * N' \quad (10)$$

where  $w$  is a constant,  $a'$  is related to the initial estimate of the expected number of accidents at the crossing and  $N'$  is related to the accident history at the crossing. The weight ( $w$ ) used to find the adjusted estimate of the expected number of accidents is related to the initial estimate of the number of accidents in both the methods.

#### 3.3.2 Differences

In the DOT method,  $w = T_o/(T_o + T)$  and  $T_o$  is related to “ $a$ ”, i.e., the weight used in the accident history adjustment is related to the initial estimate of the safety and the duration for which accident history was observed. This is different from the Empirical Bayes method in which  $w = 1/(1 + (Var[a]/E[a]))$ , depends on the initial estimate based on the reference population and also the variance of the initial estimate.

## 4. Data

The data used in this study were obtained from the databases maintained by the Federal Railroad Administration [20]. Three separate databases available in this website are relevant for this study: The Highway Rail Accident database, the Grade Crossing Inventory database, and the Grade Crossing inventory History database.

Highway Rail Accident database contains information about “any impact, regardless of severity, between a railroad on-track equipment consist and any

user of a public or private crossing site”. All grade crossing collisions are reported to the FRA regardless of the monetary value of damage caused. The database contains a variety of information including data about the type of highway vehicle involved, speed of the train at collision, and environmental factors such as time of day and weather conditions.

Grade Crossing Inventory database includes information reported to the FRA by each state DOT about the condition of each crossing. This includes information about the highway (i.e. annual average daily traffic (AADT), number of traffic lanes, posted highway speed) and the rail line (i.e. timetable speed, daily number of trains).

Grade Crossing Inventory History database includes data about the changes to the crossing inventory database. This was used to filter out the crossings which had a change in its warning device type during the analysis period.

The Grade Crossing Inventory database for Illinois had 26,089 records. This list contains crossings which are on both public and private highways, some crossings with old data that may not have been updated, crossings with missing or incomplete data, etc. Therefore, this database was filtered based on the filters given in Table 3 to obtain a meaningful dataset (Table A-1) for at grade crossings on public roads. The researchers recommend the application of such filters before using the dataset for any analysis so that the analysis is done on a meaningful dataset.

During the filtering, it was also ensured that no crossing had any variables with missing entries in the dataset. Furthermore, the warning device in the crossing was compared to the warning device of the

**Table 3 Number of crossings and accidents in Illinois (2012-2016).**

Warning device type	Number of crossings	Number of accidents (2012-2016)
Gates	2,755	234
Flashing lights	960	42
Crossbucks	1,382	52

corresponding crossing as given in the grade crossing inventory history dataset. In this way, crossings that had a change in its warning device were also eliminated from the study.

Using the crossing ID as the key, the filtered grade crossing inventory database and the grade crossing accident database were combined. Three separate databases were created, one for each warning device category. The number of crossings in Illinois (after the filters) and the number of accidents observed in the 5-year span (2012-2016) is given in Table 3.

### 5. Equations for the Fitted Models for Each Device Type

The developed models use a Zero Inflated Negative Binomial format and the variables used in the model are the same variables as used in the USDOT accident prediction formulae. For that reason, the models described in the paper are called the ZINDOT models

(Zero Inflated Negative Binomial Models with USDOT formula variables). Models are fitted for all three warning device types and are described below. The final calculation from the ZINDOT also includes adjustment to account for the accident history at the crossing.

The expression for the ZINDOT model is given by Eq. (11).

$$ZINDOT = k * zinb + (1 - k) * A \quad (11)$$

$$zinb = (1 - p)\mu \quad (12)$$

The expressions for  $p$ ,  $\mu$  (and therefore  $zinb$ ), and  $k$  are dependent on the type of warning device at the crossing. The expressions for crossings with Gates, Flashing Lights without Gates and Crossbucks are given below.

#### 5.1 Model for Crossings with Gates

The equation for the ZINDOT model for crossings with gates is given by:

$$p = \frac{1}{1 + e^{180.671 - 16.933 * \log(Aadt * TotalTrn) - 232.296 * MainTrk + 7.258 * DayThru + 7.562 * TrafficLn}} \quad (13)$$

$$\mu = e^{-7.234796 + 0.357974 * \log(Aadt) + 0.407312 * MainTrk + 0.003097 * DayThru + 0.096585 * TrafficLn} \quad (14)$$

$$k = \frac{1}{2 + \mu * (1.74 - p)} \quad (15)$$

#### 5.2 Model for Crossings with Flashing Lights and no Gates

The equations for the ZINDOT model for crossings with flashing lights and no gates are given by:

$$p = \frac{1}{1 + e^{17.3191 - 2.2290 * \log(Aadt * TotalTrn) - 2.3569 * MainTrk + 0.3392 * DayThru + 1.2132 * TrafficLn}} \quad (16)$$

$$\mu = e^{-3.467 - 0.0057 * \log(Aadt) - 0.1660 * MainTrk + 0.1240 * DayThru + 0.3594 * TrafficLn} \quad (17)$$

$$k = \frac{1}{2 + \mu * (1.0413 - p)} \quad (18)$$

#### 5.3 Model for Crossings with Crossbucks

The equation for the ZINDOT model for crossings with crossbucks is given by:

$$p = \frac{1}{1 + e^{13.454 - 1.482 * \log(Aadt * TotalTrn) + 0.246 * DayThru - 0.138 * MaxTtSpd + 0.2885 * HwyPved}} \quad (19)$$

$$\mu = e^{-1.4097 + 0.0098 * \log(Aadt * TotalTrn) + 0.0654 * DayThru - 0.0338 * MaxTtSpd + 0.1981 * HwyPved} \quad (20)$$

$$k = \frac{1}{2 + \mu * (1.00012 - p)} \quad (21)$$

Appendix B to this manuscript includes further details on the model including the standard error of the estimates and the p-values.

## 6. Comparing the ZINDOT Model to the USDOT Model

The ZINDOT model and the USDOT model were compared on its ability to select high-risk crossings. A crossing which has a higher predicted value than another crossing is considered a higher risk crossing as per the model. A model which selects crossings which had or are likely to have a higher number of accidents is considered to be a “more accurate” model (among the two models compared).

The number of accidents observed at the top crossings selected by each of the models is compared to each other. For both models, a crossing with a higher predicted accident is ranked higher than a crossing with a lower predicted accident. The ranked crossings based on the accident prediction values are compared based on the number of accidents observed at those crossings in the years 2012-2016. This comparison is done separately for crossings based on its warning device type. Two sets of data are used in this study: Illinois data and Texas data. Illinois data are used for developing the ZINDOT models and Texas data are used for validation of the models.

### 6.1 Crossings with Gates

The comparison is done for two different states: Illinois shown in Fig. 1 and Texas, shown in Fig. 2.

An observation that could be made about the ZINDOT model is that the list of top crossing generated using this model is almost consistently better than or at least equal to the list generated based on USDOT formula in terms of accident counts observed (orange curve tends to be on or above blue curve in Fig. 2). At the top 50 crossings as selected by the ZINDOT model, there were 84 accidents observed.

This is 2 more than the number of accidents observed at the top 50 crossings as selected by the USDOT model (82 accidents). The new model could identify a greater number of crossings that had accidents than the USDOT formula. This model developed using data from Illinois was tested for crossings in Texas. The crossings ranked based on ZINDOT model are seen to have a greater number of accidents than the crossings ranked USDOT model. This result is consistent with the observation made for top crossings in Illinois.

Fig. 2 shows the comparison for gated crossings in Texas. The orange curve (representing ZINDOT) is consistently above the blue curve (representing USDOT model) in the comparison for Texas as well. At the top 50 crossings, the crossings selected based on the ZINEBS model had 141 accidents while the crossings selected based on the USDOT model had only 123 accidents (which is 18 accidents less than the ZINDOT model).

### 6.2 Crossings with Flashing Lights and No Gates

The comparison for crossings with Flashing Light is shown in Fig. 3 for Illinois and Fig. 4 for Texas.

Comparing the number of accidents observed at the top-ranking crossings in Illinois, it can be seen that the ZINDOT model is almost consistently more accurate at ranking crossings than using USDOT model. This observation is repeated in crossings with Flashing Lights in Texas as well.

### 6.3 Crossings with Crossbucks

A similar comparison among crossings with crossbucks is made. In the dataset Illinois, as shown in Fig. 5, the ZINDOT model is able to identify crossings with a greater number of accidents in the future than the USDOT model. There were 52 accidents in the top 50 crossbuck locations identified using the ZINDOT model as compared to the 47 accidents in the top 50 crossbuck locations identified using the USDOT model.



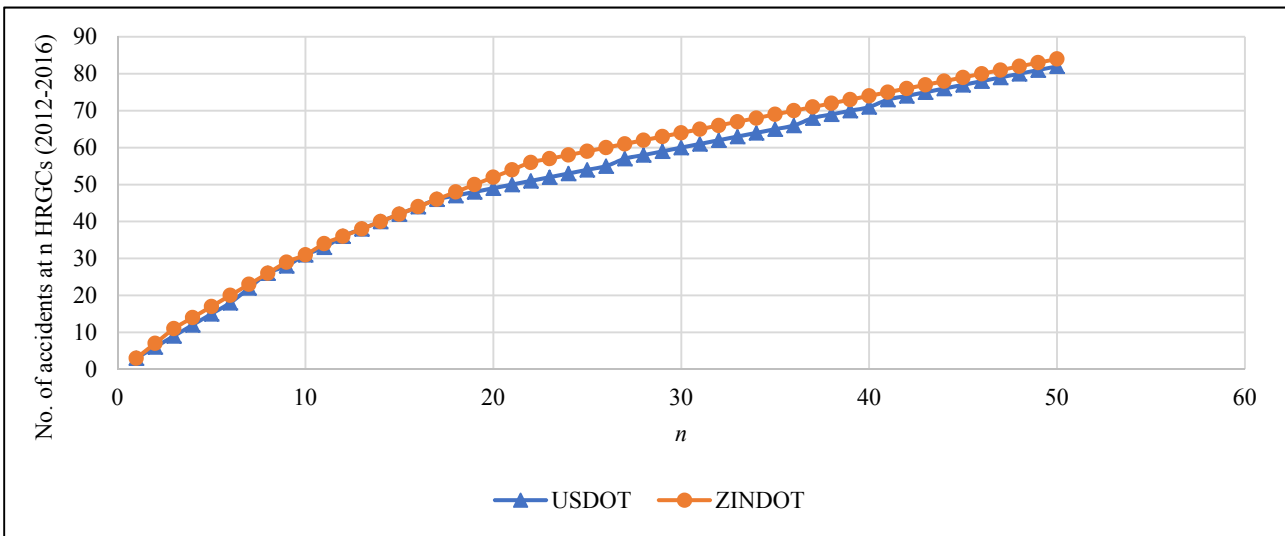


Fig. 1 Cumulative number of accidents at the top-ranking locations: crossings with gates in Illinois.

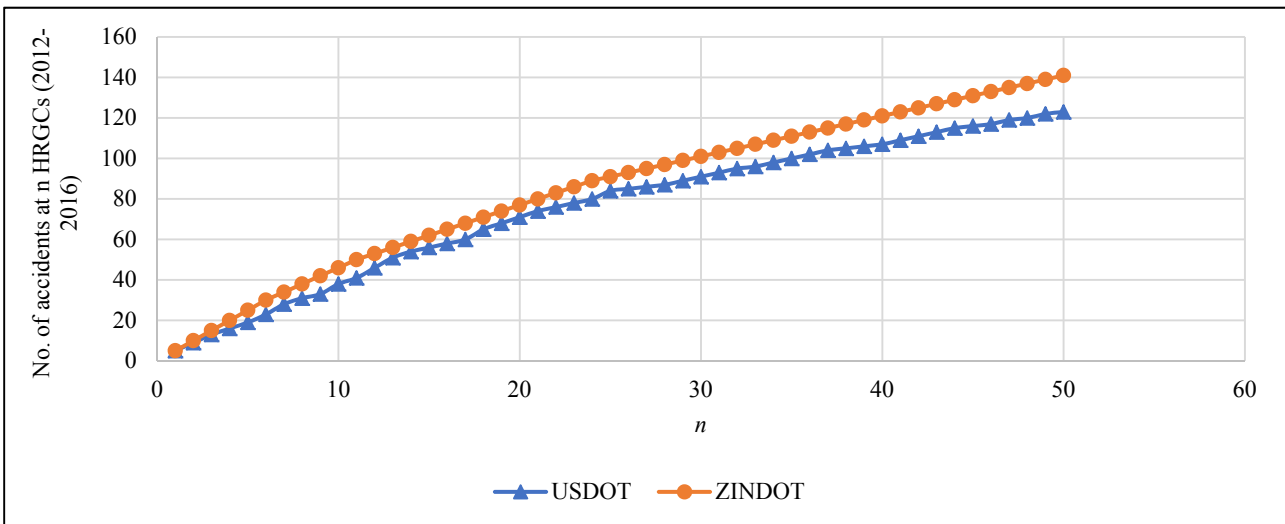


Fig. 2 Cumulative number of accidents at the top-ranking locations: crossings with gates in Texas.

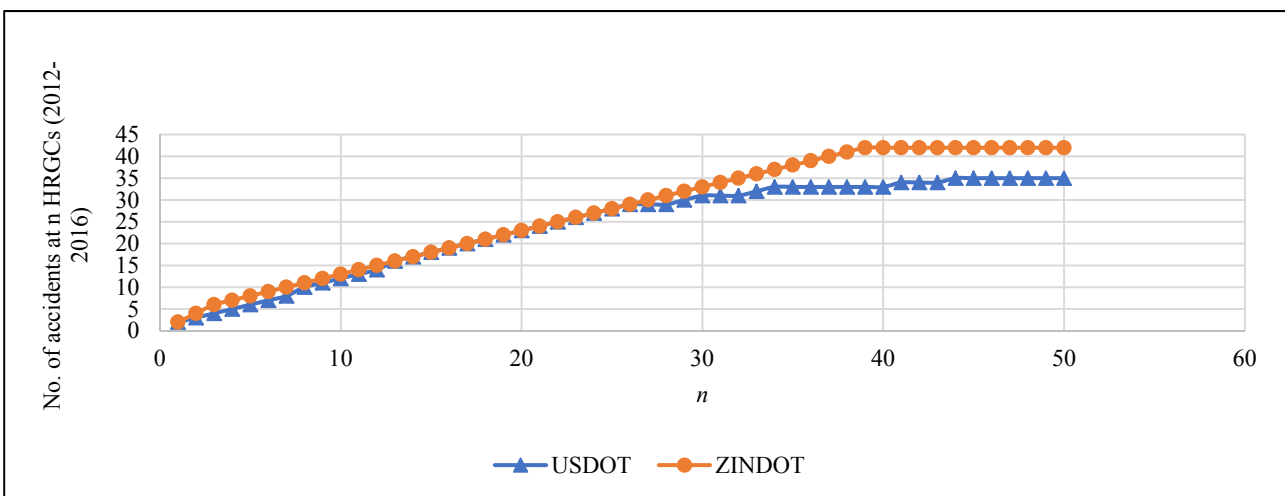


Fig. 3 Cumulative number of accidents at the top-ranking locations: crossings with flashing lights in Illinois.

### A New Accident Prediction Model for Highway-Rail Grade Crossings Using the USDOT Formula Variables

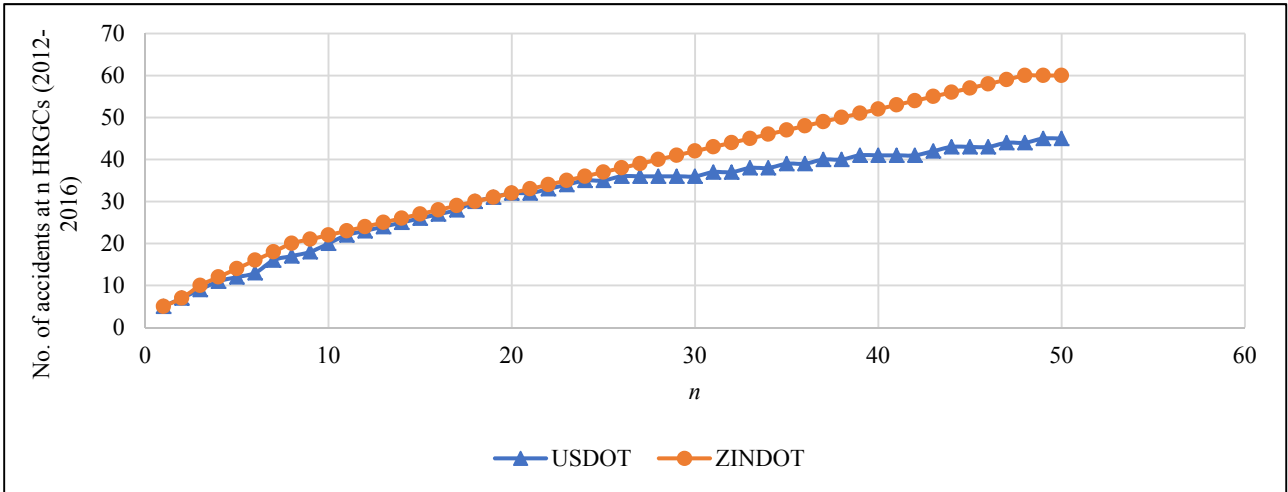


Fig. 4 Cumulative number of accidents at the top-ranking locations: crossings with flashing lights in Texas.

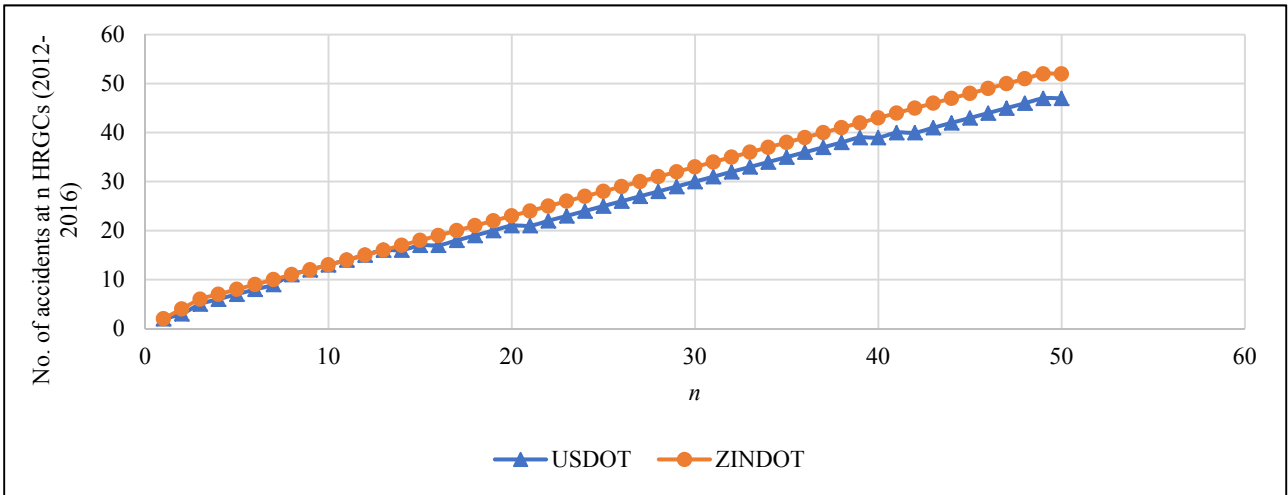


Fig. 5 Cumulative number of accidents at the top-ranking locations: crossings with crossbucks in Illinois.

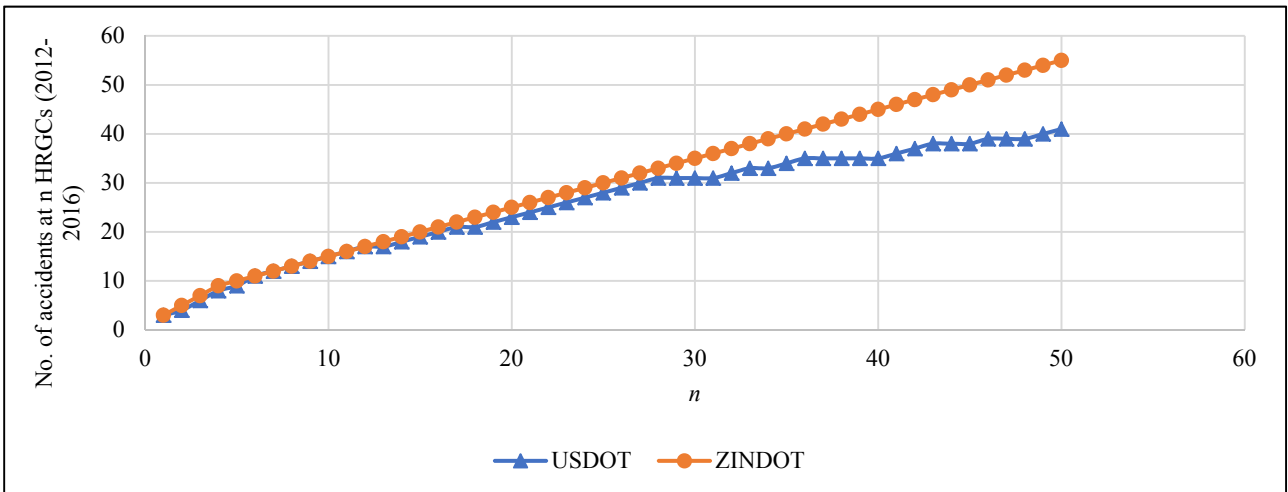


Fig. 6 Cumulative number of accidents at the top-ranking locations: crossings with crossbucks in Texas.

In Texas, as shown in Fig. 6, the observed accident count at crossings selected based on the ZINDOT model is higher than the observed accident count at crossings selected based on the USDOT model. At the top 50 crossings, there were 55 accidents in the crossings selected based on the ZINDOT model as compared to 41 accidents in the crossings selected based on the USDOT model.

## 7. Conclusions

One of the commonly used accident prediction models for railroad grade crossings is the United States Department of Transportation (USDOT) formula. This formula, however, has been developed over 40 years ago with the changes to the normalizing coefficients made once every few years based on the accident experiences. This paper presented ZINDOT model to estimate the expected accident frequency that used a different model format, more recent accident and inventory data, different accident history adjustment approach, but the same variables that were used in the USDOT accident prediction formula.

The ZINDOT model, like the USDOT model, is a multi-part model. In the first part of the ZINDOT model, the Zero Inflated Negative Binomial model is used to estimate the initial accident prediction value. This part of the ZINDOT model uses the same variables as the USDOT model, but a different model format. In the second part of the ZINDOT model, Empirical Bayes method is used to adjust for the location accident history. This part of the ZINDOT model is similar to the USDOT model as it uses a weighted average of the initial estimate and the accident history. The difference in the second part is that the calculated weights in the ZINDOT model are dependent both on the initial estimate and the variance of the initial estimate unlike the USDOT model, which are only dependent on the initial estimate.

The coefficient of the ZINDOT model was estimated by fitting the model using data from Illinois. The model was also validated using data from Texas.

The ZINDOT model estimates were compared to the USDOT model estimates to rank the crossings based on the expected accident frequency. It is observed that the new model can identify crossings with a greater number of accidents with Gates and Flashing Lights and Crossbucks in both Illinois and Texas.

Since the ZINDOT model utilizes the same variables that were used in the USDOT accident prediction formulae, using this new model would not require the collection of additional data, thereby making it easier for practitioners, already using the USDOT formula to use the ZINDOT model. The ZINDOT model can complement the USDOT model to identify high risk crossings for resource allocation and safety improvements.

## References

- [1] Mengert, P. 1979. *Rail-Highway Crossing Hazard Prediction Research Results*. FRA-RRS technical report.
- [2] Austin, R. D., and Carson, J. L. 2002. "An Alternative Accident Prediction Model for Highway-Rail Interfaces." *Accid. Anal. Prev.* 34 (1): 31-42.
- [3] Lee, J., Park, D., and Nam, D. 2005. "Analyzing the Relationship between Grade Crossing Elements and Accidents." *J. East. Asia Soc. Transp. Stud.* 6: 3658-68.
- [4] Lu, P., and Tolliver, D. 2016. "Accident Prediction Model for Public Highway-Rail Grade Crossings." *Accid. Anal. Prev.* 90: 73-81.
- [5] Farr, E. H. 1981. *Rail-Highway Crossing Resource Allocation Model*. U.S. Department of Transportation.
- [6] Hauer, E., Harwood, D. W., Council, F. M., and Griffith, M. S. 2002. "Estimating Safety by the Empirical Bayes Method: A Tutorial." *Transportation Research Record* 1784 (1): 126-31.
- [7] FHWA. 2007. *Railroad-Highway Grade Crossing Handbook Revised*. 2nd ed. [http://safety.fhwa.dot.gov/xings/com\\_roaduser/07010/sec05.htm](http://safety.fhwa.dot.gov/xings/com_roaduser/07010/sec05.htm).
- [8] Farr, E. H. 1987. *Summary of the Dot Rail-Highway Crossing Resource Allocation Procedure-Revised*. U.S. Department of Transportation.
- [9] Federal Railroad Administration. *FRA Web Accident Prediction System (WBAPS)*. Generate report. <https://safetydata.fra.dot.gov/webaps/>.
- [10] Park, Y., and Saccomanno, F. F. 2005. "Collision Frequency Analysis Using Tree-Based Stratification." 1908 (1): 121-9.
- [11] Medina, J. C., and Benekohal, R. F. 2015. "Macroscopic Models for Accident Prediction at Railroad Grade

## A New Accident Prediction Model for Highway-Rail Grade Crossings Using the USDOT Formula Variables

- Crossings.” *Transp. Res. Rec. J. Transp. Res. Board* (2476): 85-93.
- [12] Mathew, J., Benekohal, R. F., and Medina, J. C. 2019. *Accident Prediction Models Using Macro and Micro Scale Analysis: Dynamic Tree and Zero Inflated Negative Binomial Models with Empirical Bayes Accident History Adjustment*.
- [13] Shankar, V., Milton, J., and Mannering, F. 1997. “Modeling Accident Frequencies as Zero-Altered Probability Processes: An Empirical Inquiry.” *Accid. Anal. Prev.* 29 (6): 829-37.
- [14] Kumara, S. S. P., and Chin, H. C. 2003. “Modeling Accident Occurrence at Signalized Tee Intersections with Special Emphasis on Excess Zeros.” *Traffic Inj. Prev.* 4 (1): 53-7.
- [15] Cameron, A. C., and Trivedi, P. K. 2013. *Regression Analysis of Count Data*. Cambridge Univ. Press.
- [16] Hauer, E. 1997. *Observational Before/After Studies in Road Safety: Estimating the Effect of Highway and Traffic Engineering Measures on Road Safety*.
- [17] Hauer, E. 1986. “On the Estimation of the Expected of Accidents.” *Accident Analysis & Prevention* 18 (1): 1-12.
- [18] Persaud, B., and Lyon, C. 2007. “Empirical Bayes Before-After Safety Studies: Lessons Learned from Two Decades of Experience and Future Directions.” *Accid. Anal. Prev.* 39 (3): 546-55.
- [19] Sharma, S. L., and Datta, T. K. 2007. “Investigation of Regression-to-Mean Effect in Traffic Safety Evaluation Methodologies.” *Transp. Res. Rec. J. Transp. Res. Board* (1): 32-9.
- [20] FRA. 2019. “Federal Railroad Administration Office of Safety.” [https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on\\_the\\_fly\\_download.aspx](https://safetydata.fra.dot.gov/OfficeofSafety/publicsite/on_the_fly_download.aspx).

### Appendix A: List of Filters Used on Grade Crossing Inventory Dataset

**Table A-1** Filters used on grade crossing inventory dataset.

Variable	Description	Filter	Description of filter	Number of crossings after filter
		None	Before any filters	26,089
<i>TypeXing</i>	Crossing type	Select “3”	Select public crossings only	17,054
<i>PosXing</i>	Crossing position	Select “1”	Select at grade crossing only	13,703
<i>ReasonID</i>	Reason for update	Remove 16	Remove closed crossings	7,925
<i>TotalTrain</i>	Total number of trains	> 0	Select crossings with 1 or more trains operating per day	7,183
<i>TotalTrack</i>	Total number of tracks	> 0	Select crossings with 1 or more tracks at the crossing	7,090
<i>TrafficLn</i>	Number of highway lanes	> 0	Select crossings with 1 or more highway lanes at the crossing	6,774
<i>Aadt</i>	Annual average daily traffic count	> 0	Select crossings with AADT > 0	6,763
<i>AadtYear</i>	Annual average daily traffic year	> 2,000	Select crossings with year of AADT > 2,000	6,673
<i>HwySpeed</i>	Posted highway speed limit	> 0	Select crossings with posted speed limit > 0	6,625
<i>MaxTtSpd</i>	Maximum timetable train speed	> 0 and ≤ 79	Select crossings with maximum timetable train speed between 0 and 79 mph	6,624
<i>WdCode</i>	Warning device code	Select “3”, “7”, “8” and “9”	Select crossings with crossbucks, flashing lights, four quad gates and all other gates	6,476
<i>XSurfaceID</i>	Crossing surface	Remove “17”, “19”, “20” and unknowns	Remove crossings with metal, unconsolidated, other or unknown crossing surfaces	5,883

### Appendix B: Details of Fitted Zero Inflated Negative Binomial Models

**Table B-1** Zero inflated negative binomial coefficients for model for gates.

**Table B-1a** Coefficients of the count part of the model.

	Estimate	Std. error	z value	Pr (>  z )
<i>Intercept</i>	-7.2348	0.4688	-15.4326	0.0000
<i>Log(Aadt * TotalTrain)</i>	0.3580	0.0536	6.6765	0.0000
<i>MainTrk</i>	0.4073	0.1415	2.8791	0.0040
<i>DayThru</i>	0.0031	0.0047	0.6577	0.5107
<i>TrafficLn</i>	0.0966	0.0778	1.2415	0.2144

**Table B-1b Coefficients of the zero-inflation part of the model.**

	Estimate	Std. error	z value	Pr (>  z )
<i>Intercept</i>	180.6707	188.0501	0.9608	0.3367
<i>log(Aadt * TotalTrain)</i>	-16.9334	17.3306	-0.9771	0.3285
<i>MainTrk</i>	-232.2965	250.6382	-0.9268	0.3540
<i>DayThru</i>	7.2582	7.8799	0.9211	0.3570
<i>TrafficLn</i>	7.5622	7.3975	1.0223	0.3067

**Table B-2 Zero inflated negative binomial coefficients for model for flashing lights and no gates.**

**Table B-2a Coefficients of the count part of the model.**

	Estimate	Std. error	z value	Pr (>  z )
<i>Intercept</i>	-3.4674	3.5507	-0.9765	0.3288
<i>log(Aadt * TotalTrain)</i>	-0.0057	0.3445	-0.0165	0.9869
<i>MainTrk</i>	-0.1660	0.5339	-0.3109	0.7559
<i>DayThru</i>	0.1240	0.0565	2.1949	0.0282
<i>TrafficLn</i>	0.3595	0.2632	1.3656	0.1721

**Table B-2b Coefficients of the zero-inflation part of the model.**

	Estimate	Std. error	z value	Pr (>  z )
<i>Intercept</i>	17.3191	6.6297	2.6124	0.0090
<i>log(Aadt * TotalTrain)</i>	-2.2290	0.9396	-2.3723	0.0177
<i>MainTrk</i>	-2.3569	2.1043	-1.1200	0.2627
<i>DayThru</i>	0.3392	0.1754	1.9339	0.0531
<i>TrafficLn</i>	1.2132	0.9882	1.2277	0.2196

**Table B-3 Zero inflated negative binomial coefficients for model for crossbucks.**

**Table B-3a Coefficients of the count part of the model.**

	Estimate	Std. error	z value	Pr (>  z )
<i>Intercept</i>	-1.4097	2.1661	-0.6508	0.5152
<i>log(Aadt * TotalTrain)</i>	0.0099	0.2131	0.0464	0.9630
<i>DayThru</i>	0.0654	0.0498	1.3124	0.1894
<i>MaxTtSpd</i>	-0.0338	0.0192	-1.7654	0.0775
<i>as.factor(HwyPved)2</i>	0.1982	0.5737	0.3455	0.7297

**Table B-3b Coefficients of the zero-inflation part of the model.**

	Estimate	Std. error	z value	Pr (>  z )
<i>Intercept</i>	13.4543	3.3962	3.9616	0.0001
<i>log(Aadt * TotalTrain)</i>	-1.4826	0.4134	-3.5859	0.0003
<i>DayThru</i>	0.2460	0.0958	2.5695	0.0102
<i>MaxTtSpd</i>	-0.1382	0.0388	-3.5644	0.0004
<i>as.factor(HwyPved)2</i>	0.2886	1.0726	0.2690	0.7879