ANALYSIS OF THE SAFETY IMPACT OF TRAIN HORN BANS AT HIGHWAY-RAIL GRADE CROSSINGS: AN UPDATE USING 1997-2001 DATA

Final Report

Prepared under Contract No. DTFR53-03-P-00005

Prepared for:

Federal Railroad Administration (FRA)
Brenda Moscoso
1120 Vermont Avenue, N.W.
Washington, DC 20590

Prepared by:

Paul L. Zador
Westat
1650 Research Boulevard
Rockville, Maryland 20850

August 13, 2003

TABLE OF CONTENTS

Chapter			Page
1	INTRODUCTION		1
2	METHODS		2
	2.1	Data sources and data preparation	2
		2.1.1 Data sources	2
		2.1.2 Predicting yearly accident rates by warning device class	3
		2.1.3 Grouping grade crossings on predicted accident risk	4
	2.2	Statistical models	5
		2.2.1 Predicting accident risk using FRA's basic prediction formula based on historical data	6
		2.2.2 Predicting accident risk using Poisson regression based on 1997-2001 accidents	6
		2.2.3 Poisson regression for aggregating safety effects across risk groups (GLM with Poisson link function)	7
		2.2.4 Poisson-Normal (non-linear) regression for aggregating safety effects across risk groups	7
		2.2.5 Model estimation and testing	8
3.	RESULTS		8
4	DISCUSSION		15
5	REFERENCES		17

List of Tables

Table		Page
A-1	Grade crossing counts by region and whistle ban status	3
B	Methods for estimating the overall safety impact of a whistle ban	6
R-1	Number of crossings and 1997-2001 accident frequencies by region and warning device class	9
S-1	Ban effect estimates by methods for predicting grade crossing accident risk, aggregating across risk groups and number of risk groups (U.S. except FL)	10

TABLE OF CONTENTS (CONTINUED)

List of Tables

Table		Page
S2	Ban effect estimates by methods for predicting grace crossing accident risk, Aggregating across risk groups and number of risk groups (U.S. except FL, Chicago area).	11
S3	Ban effect estimates by methods for predicting grade crossing accident risk, Aggregating across risk groups and number of risk groups (Chicago area)	12
S4	Ban effect estimates by methods for predicting grade crossing accident risk, Aggregating across risk groups and number of risk groups (Chicago area v. Continental U.S. (except FL)	13
S5	Ban effect estimates by methods for predicting grade crossing accident risk, Aggregating across risk groups and number of risk groups (Wisconsin)	14
S6	Ban effect estimates by methods for predicting grade crossing accident risk, Aggregating across risk groups and number of risk groups (Wisconsin v. Continental U.S. (except FL)	15

Analysis of the Safety Impact of Train Horn Bans at Highway-Rail Grade Crossings: an Update Using 1997-2001 Data

Paul Zador and Doug Duncan
August 13, 2003

1. Introduction

This report updates Westat's prior report to FRA (Zador, 2002) in terms of both data and statistical methods. Safety impact estimates in the prior report were based on 5-year accident data for 1992-1996. In this report, safety impact estimates are based on recent grade-crossing accidents during the 5-year period between 1997 and 2001. As in the prior report, this report also derives safety impact estimates by comparing observed accident frequencies to expected accident frequencies.

In the prior report, FRA's methods (U.S. DOT, FRA, 1987; U. S. DOT, FRA, 2000) were used, in a slightly modified form, to estimate the expected accident frequency at each grade crossing from the grade crossing's physical and operating characteristics. In Step 1 of this three step procedure, FRA's basic accident prediction formula (without adjustment for recent local accident history, U.S. DOT, FRA, 1987) was used to predict expected accidents. In Step 2, grade crossings were grouped on predicted accident risk into 10 groups with roughly comparable accident risks. In Step 3, general linear models (GLM, see McCullaghand Nelder, 1989) were used to estimate overall percent difference (across all risk groups but by warning device class) between grade crossings in the same accident risk group with and without the ban.

In the present report, we applied the method (Method 1a) just described to the more recent data but also performed several additional analyses of these data to assess the robustness of the safety impact estimates:

- Method 1b: Same as Method 1a, except that six risk groups were formed instead of ten.
- Method 2a: Same as Method 1a, except that Poisson-Normal models, rather than general linear models were used to estimate overall percent difference across risk groups by warning device class.
- Method 2b: Same as Method 2a, except that six risk groups were formed instead of ten.
- Method 3: Safety impacts are estimated from Poisson regression models, by warning device class, for 1997-2001 grade crossing accident frequencies. In each device class, the safety impact estimate is derived from the estimate for the coefficient of a whistle ban indicator variable. The models control for accident risk in the absence of the whistle ban using (1) predictors comprising the grade crossing characteristics included in FRA's basic prediction formula *and* (2) FRA's predicted accident frequency.

Method 4: Similar to Method 2a in that Poisson-Normal models are used to estimate overall percent difference across (N= 10) risk groups by warning device class but differs from it in the way the risk groups are formed. Risk predictions are based on Poisson regression models for 1997-2001 grade crossing accident frequencies—as in Method 3—except that the whistle ban indicator function is not in the models

Section 2 describes the data used and the statistical models used to estimate safety impact. Section 3 presents the results and Section 4 Summaries and conclusions.

2. Methods

2.1. Data sources and data preparation

2.1.1 Data sources

Westat received data files from the Federal Railroad Administration (FRA) with information for the 5-year period between January 1, 1997 and December 31, 2001 on the geometric and traffic characteristics of public at-grade highway-rail grade crossings and on accidents that occurred at these grade crossings. Westat also received data files that identified grade-crossings that were inactive during the study period and other data files that specified the whistle ban status of active grade crossings. Following FRA practice, grade crossings were grouped into three broad classes for analytic purposes: grade crossings with a passive warning device, grade crossings with flashing lights and grade crossings with gates. During the study period, 6,744 out of 146,774 crossings had a change in its warning device class. Detailed information on warning device class changes was used to calculate the proportion of time by warning device for each grade crossing. As described in Appendix A, Westat combined these data into an analysis file with one record per active grade crossing that included all relevant information on geometric and traffic characteristics, warning device class and class change, whistle ban status, and accident frequency.

Following FRA's guidance, a total of six *regional* data sets were identified for estimating the safety impact of a whistle ban by comparing crossings in the region with a ban to crossings without a ban. Table A1 shows grade crossing counts by whistle ban status in each region.

Table A1. Grade crossing counts by region and whistle ban status.

Region	Crossing with	
	No Whistle Ban	Whistle Ban
Continental U.S. except Florida	Continental U.S. except Florida, (N=144,734)	Continental U.S. except Florida, (N=1,788)
Continental U.S. except Florida and Chicago area	Continental U.S. except Florida and Chicago area, (N=143,888)	Continental U.S. except Florida and Chicago area, (N=1,429)
Chicago area	Chicago area, (N=846)	Chicago area, (N=359)
Chicago area v. Continental U.S. except Florida	Continental U.S. except Florida, (N=144,734)	Chicago area, (N=359)
Wisconsin	Wisconsin, (N=3,403)	Wisconsin, (N= 500)
Wisconsin v. Continental U.S. except Florida	Continental U.S. except Florida (N=144,734)	Wisconsin, (N= 500)

2.1.2 Predicting yearly accident rates by warning device class

A grade crossing's expected (or predicted) yearly accident rate, also referred to as its expected accident risk, depends on the crossing's physical and operating characteristics. FRA's traditional approach is to predict expected accident risk using the basic accident prediction formulas (U.S. DOT, FRA, 1987). There are three formulas, one for each of the warning device classes. These formulas were used in this report the same way as we used them in Westat's prior report (Zador, 2002), except for one modification. For crossings which had one or more changes in warning device class, predicted accident rate (AP) was calculated as a weighted sum of the crossing's predicted accidents rates assuming it had a passive warning device, it had flashing lights, and it had gates. These predicted rates (AP1, AP2, and AP3) were weighted by the proportions of days that the crossing actually had each of these warning devices. Thus,

$$AP = P*AP1 + F*AP2 + G*AP3,$$

where:

AP = final predicted accident rate,

AP1, AP2, and AP3 = predicted accident rates for the crossing with a passive warning device, with flashing lights, and with gates, respectively.

P, F, and G = proportions of days for the crossing with a passive warning device, with flashing lights, and with gates, respectively, (Note that P+F+G=1).

We note that the numeric coefficients (not shown here) that were used for calculating the expected accident rate from *current* grade crossing characteristics had been estimated by FRA staff using historical accident statistics, and not from current data on recent accidents.

We also developed a second approach that used accident frequencies during the study period to provide a set of alternative expected accident rate estimates. With this alternative approach, statistical models were fitted, by warning device class, to accident frequencies during 1997-2001. Statistical assumptions, variables used for modeling, and other details are described in Section 2.2. We note that when employing this alternative approach to estimate expected accident rates, we assigned grade crossings with multiple warning device classes to the warning device class with the largest proportion of days, that is to the class with the largest value among the three proportions: P, F and G.

To recap, the difference between the use of the historical accident statistics (via the FRA Accident Prediction Formulas and therefore using data that was as old as 5 years before the study period, or 1992) and the use of current data on collisions that occurred during the study period (1997-2001) was that with the former method, data on recent collisions could not affect the statistical relationships between collisions at a crossing and it's geometric and traffic characteristics but with the latter, it could affect those statistical relationships. The two methods were similar in that both were used to estimate the expected number of collisions from geometric and traffic characteristics by device class, but the predictions from more recent data reflected the current statistical relationships better than the predictions from the historical data.

2.1.3 Grouping grade crossings on predicted accident risk

When estimating the safety effect of a whistle ban, other factors that may affect accident risk need to be controlled. One approach for doing this is to group grade crossings on predicted accident risk so that grade crossings in each group have roughly the same predicted accident risk, and then compare grade crossings with the ban to grade crossings without the ban in the same accident risk group.

We formed risk groups by sorting grade crossings on predicted accident risk in ascending order and located predicted accident risk cut-points so that each partition between consecutive cut-points included the same number of *observed accidents* (and *not* the same number of grade crossings). This step was performed by warning device class with 5 and 9 cut-points, yielding, respectively, 6 and 10 risk-homogeneous groups delimited by cut-point pairs. Analyses were then performed with 6 and 10 risk groups. (Note, the natural risk-range endpoints, 0 and 1, were treated as extra cut points.) In the prior study, all analyses used 10 risk groups. The additional analyses with 6 risk groups were performed because with 9 cut-points, some risk groups had very low accident rate counts, and such low counts can result in statistical instability. Reducing the number of risk groups helps reducing statistical instability by increasing the number of grade crossings in every risk group.

Balancing risk group size on accident count, rather than on the number of grade crossings, assured that the total number of accidents was reasonably high even in the risk group with the lowest accident rate. When using general linear models to estimate whistle ban safety effects from accident counts across risk groups, increasing the minimum number of accidents across risk groups tends to improve the procedures overall statistical stability (see p. 4 in Zador, 2002). We note that when using alternative analyses based on Poisson-Normal models for accident rate variation across risk groups (see Section 2.2), this is not so. For these analyses it was more appropriate to balance risk group size on grade crossing counts. This is because the assumption for a Poisson-Normal model is that a grade crossing accident count, conditional on crossing characteristics, is Poisson-distributed. This implies that the distribution of the sum of counts aggregated for a group of grade crossings is likewise Poisson-Normal. Under the Poisson-Normal model, the departure from the Poisson assumption is explicitly accounted for by Normally distributed within-group error terms. The Poisson-Normal framework is fully compatible with grouping units (i.e., grade crossings), but it would have had to be modified if outcomes (i.e., accident counts) had been allowed to directly affect the grouping procedure, and moreover, it would not have improved statistical stability.

2.2. Statistical Models

All but one of the methods are closely related to the method that was developed by the Federal Railroad Administration (U.S. DOT, FRA, 2000). The common feature of these methods is that they include three steps:

Step 1. Predict expected grade crossing accident risk from grade crossing characteristics.
Step 2. Group grade crossings on predicted accident risk so that within each risk group, grade crossings have roughly the same predicted accident risk.
Step 3. In each risk group, estimate the accident rate difference between grade crossings with and without a whistle ban, aggregate these differences across risk groups, and relate aggregated accident rate differences between grade crossings with and without a ban to the percent effect of having a whistle ban.

As shown in the Table B, varying Steps 1- Step 3, gives us various methods for estimating whistle ban safety effects:

Table B. Methods for estimating the overall safety impact of a whistle ban

Methods for: Predicting accident risk	Methods for: Estimating overall whistle ban safety effects	N of risk groups	Methods
FRA's basic prediction formula based on historical data	Poisson regression for aggregating across risk groups (GLM with Poisson link function)	10	1.a
		6	1.b
	Poisson-Normal (non-linear) regression	10	2.a
		6	2.b
Poisson regression based on 1997-2001 accidents	From regression coefficient	NA	3
	Poisson-Normal (non-linear) regression	10	4

2.2.1 Predicting accident risk using FRA's basic prediction formula based on historical data.

Our use of FRA's basic formula to predict accident risks from grade-crossing characteristics was summarized in a prior report (p. 4 in Zador, 2002), and this will not be reproduced here.

2.2.2 Predicting accident risk using Poisson regression based on 1997-2001 accidents.

For grade crossing k, we equate the logarithm of the expected number $E[N_k]$ of accidents to a linear function of grade crossing characteristics

$$\log E[N_k] = g(x) = \sum_{l=0}^{L} \beta_l x_{kl} \ . \qquad \text{(A1)}$$

where x_{kl} = (x_{kl} , j = 0, 1,..., L) represents an L-vector of grade crossing characteristics, and the β's are the regression parameters. According to the Poisson assumption for the (conditional) distribution of accidents specifies accident probabilities, we have that,

$$\Pr[N_k = n \mid g(x_k)] = e^{-g(x_k)} \frac{g(x_k)^n}{n!} \ . \qquad \text{(A2)}$$

Under this (or some other appropriate distributional) assumption, the regression coefficient in (A1) can be estimated from observed accident frequencies. The validity of the Poisson assumption, and of model A1, can also be tested.

For the purpose of this report we included four types of variables in the vector of grade crossing characteristics: (1) the six variables in the basic FRA prediction formula (number of vehicles per day, number of trains per day, number of tracks, a 0/1 flag for paved highways, maximum timetable speed, and number of highway lanes), (2) the (log-scale) probability (lprob) for having at least one accident per year and square of this probability, lprob*lprob (lprob was estimated from FRA's accident prediction formula),

(3) a variable for the combined effects daily train/vehicle traffic volumes (ADTTVAR = number of vehicles per day * number of trains per day), and (4) terms for the interactions between lprob and each one of the traffic variables, including ADTTVAR.

Models specified in (A.1) were estimated by warning device class from accident data for 1997-2001 and model parameter estimates were used to generate the predicted accident rates. (For the purpose of modeling, a grade crossing with more than one warning device class during the study period was assigned to the device class it belonged to for the longest time.)

2.2.3 Poisson regression for aggregating safety effects across risk groups (GLM with Poisson link function)

As described in the prior report (p. 4., Zador, 2002), the idea here is to sum (within device class) accident frequencies by risk group and whistle ban and, *within each risk group,* compare the combined frequency of accidents at grade crossings with a ban to the combined frequency of accidents without a ban. It is valid to perform this comparison using Poisson regression if, conditional on accident risk group membership, the Poisson assumption applies to the accident frequency pairs in each risk group.

2.2.4 Poisson-Normal (non-linear) regression for aggregating safety effects across risk groups

Poisson-Normal regression differs from Poisson regression in the inclusion of a normal random term to account for 'over-dispersion' often observed when the model fit is not perfect (McCullagh and Nelder, 1989). The term over-dispersion refers to more variance than is compatible with the Poisson distribution, once the Poisson mean is given. Over-dispersion arises when the specification of the regression model for expected accident frequency is imperfect for reasons that can not be determined, or remedied. For example, some predictor variables included in the model may contain reporting errors, there may be grade crossing characteristics not represented by any measured predictor variable in the model, the mathematical form of predictor variables may be incorrect, and predictor variables may interact with one another in unknown ways that affect accident frequencies. A Poisson-Normal regression model for expected accident frequency accounts for the combined effects of these unspecified disturbances by explicitly including a random error term.

Model specification for Poisson-Normal regression is similar to the specification of the Poisson regression model above. Denote by *NA(g, B), NC(g,B)* and *APRED(g, B),* respectively, the accident count, the grade crossing count, and the mean predicted 5-year accident frequency in risk group *g* with whistle ban status specified by the binary flag *B* (= 1/0, with/without a ban). Generalizing A1 and A2 above, we specify a hierarchical model for accident counts,

$$NA(g,B) \mid \lambda(g,B) \sim Poisson(\lambda(g,B)) , \quad \text{(A1')}$$
$$\log[\lambda(g,B)] = \log[(c \times B + \beta) \times APRED(g,B) \times NC(g,B)] + e(g,B) , \quad \text{(A2')}$$
$$e(g,B) \sim Normal(0, \sigma^2) . \quad \text{(A3')}$$

Exponentiating equation A2', the Poisson parameter is seen in this setup to be proportional to the expected number of accidents, which is $APRED(g,B) \times NC(g,B)$. For grade crossings without a ban, B = 0, and the corresponding proportionality factor of the expected number of accidents is $\beta e^{e(g,B)}$ (=F1). For grade crossings with the ban, the proportionality factor becomes $((c \times B + \beta)e^{e(g,B)}$ (=F2). Except for random effects represented by exponential multipliers, the Poisson means of grade crossings with and without a ban, but with roughly the same predicted accident risk, have a ratio that equals 1 + c/β (=F1/F2), so that (100c/β)% equals the estimated effect of a whistle ban after controlling for predicted accident risk, and also allowing for imperfect model fit as that is represented by a normal adjustment factor with a variance of σ^2.

2.2.5 Model estimation and testing

SAS procedures were used to estimate and test the statistical models described in Sections 2.2.3 and 2.2.4 (SAS Institute, 1996). SAS procedure GENMOD was used to estimate Poisson regression models. SAS procedure NLMIXED was used to estimate Poisson -Normal models (see also Wolfinger and O'Connel, 1993; Wolfinger and Lin, 1997). We note that the Poisson-Normal models included only three parameters that needed to be estimated, they were c, β, and σ^2.

3. Results

Table R1 summarizes the numbers of crossings and 1997-2001 accident frequencies by region and warning device class. As these tables show, there were only 21 crossings with a whistle ban in the Chicago area with a passive warning device and the same number had flashing lights. Correspondingly, accident frequencies were also extremely low, with 1 and 2 accidents, respectively, during the study period. Selected detail tables are given in Appendix 2 by accident risk group for each row in the attached 18 tables entitled '1997-2001 grade crossing accident rates by accident risk group and whistle ban'. The detail tables confirm that the data are extremely sparse, with several of the risk groups containing no observations.

Table R1. Number of Crossings and 1997-2001 Accident Frequencies by Region and Warning Device Class.

Region	Warning Device	Number of Crossings		5-year Accidents	
		No Ban	Ban	No Ban	Ban
Continental US Except Florida	Passive	90,228	533	5,708	59
	Flashing Lights	25,511	344	2,695	59
	Gates	28,995	911	2,007	156
Continental US Except Florida and Chicago Area	Passive	89,893	512	5,689	58
	Flashing Lights	25,270	323	2,653	57
	Gates	28,725	594	1,964	93
Chicago Area	Passive	335	21	19	1
	Flashing Lights	241	21	42	2
	Gates	270	317	43	63
Chicago Area v. Continental U.S. Except Florida	Passive	90,228	21	5,708	1
	Flashing Lights	25,511	21	2,695	2
	Gates	28,995	317	2,007	63
Wisconsin	Passive	2,098	205	208	25
	Flashing Lights	972	155	75	29
	Gates	333	140	19	27
Wisconsin v. Continental U.S. Except Florida	Passive	90,228	205	5,708	25
	Flashing Lights	25,511	155	2,695	29
	Gates	28,995	140	2,007	27

Tables S1-S6 combine (percent) ban effect estimates for the six regions (see Table A.1 in Section 2) by method (see Table B in Section 2) and warning device class. Detail tables containing confidence limits, selected plots, and model fit statistics are included in Appendix A2.

Table S1 presents the percent difference between grade crossings with a whistle ban and grade crossings without a ban for the continental Unites States outside of Florida. All estimates are positive, regardless of method of estimation and warning device class, confirming that nationwide, whistle bans were associated with an increase in grade crossing accidents. The increase associated with banning train horns was statistically significant at the 5% level for grade crossings with gates with all but one of the methods. With that method, the increase was almost statistically significant at the 5% level. The number of risk groups and the method of aggregating across risk groups had little effect on the size of the estimates that ranged from a 24.9% increase to a 30.5% increase as long as they were based on the FRA method for estimating accident risk from historical data. The estimated adverse effect of whistle ban was larger when estimated based on accident risk estimates derived by Poisson regression from 1997-2001 accident frequencies. These estimates were 36.2 percent with using the risk group methodology and 43.4% based direct calculations using a model parameter estimate.

For grade crossings with a passive warning device, estimates based on FRA accident risk estimates ranged between 35.3% and 36.7% but were not statistically significant at the 5% level. Whistle ban effect estimates based on predicted accident risk estimates derived from recent accident frequencies were 70.3% using risk groups and 74.9% percent when calculated directly from model parameter for whistle ban. Both were statistically significant at least at the 5% level.

For grade crossings with a flashing lights, estimates based on FRA accident risk estimates ranged between 10.6% and 12.3% but were not statistically significant at the 5% level. Whistle ban effect estimates based on accident risk estimates derived from recent accident frequencies were 21.5% using risk groups and 21.7% percent when calculated directly from model parameter for whistle ban. However, neither were statistically significant at the 5% level.

Table S1. Ban effect estimates by methods for predicting grade crossing accident risk, aggregating across risk groups and number of risk groups.

Region: Continental U. S. (Except Florida)

Method for predicting accident risk	N of risk groups	Method for aggregating across risk groups	Warning device class		
			Passive	Flashing lights	Gates
FRA, historic data	6	Poisson Regression	35.5	11.3	28.0*
		Poisson-Normal	35.3	10.6	24.9
	10	Poisson Regression	36.7	12.3	30.5**
		Poisson-Normal	36.5	11.9	27.4*
Poisson Regression, 1997-2001 data	10	Poisson-Normal	70.3*	21.5	36.2*
	Ban effect estimated directly from regression coefficient		71.6***	21.7	43.4***

* statistically significant at less than 0.05; ** less than 0.01; *** less than 0.001

Table S2 presents the percent difference between grade crossings with a whistle ban and grade crossings without a ban for the continental Unites States outside of Florida and the Chicago area. Removing the Chicago area from the study strengthened the estimated effect of a whistle in general, but the pattern of results was not substantially changed. We note that for grade crossings with flashing lights, removing the Chicago area from the comparison increased the direct estimate for the whistle ban's adverse effect from a 21.7% increase to a 30.9% increase when estimating the effect directly from the Poisson regression coefficient (1997-2001 data). While this later estimate was not statistically significant at the conventional 5% level, it was nearly so, (Prob > |t| = 0.08).

Table S2. Ban effect estimates by methods for predicting grade crossing accident risk, aggregating across risk groups and number of risk groups.

Region: Continental U. S. (Except Florida and Chicago area)

Method for predicting accident risk	N of risk groups	Method for aggregating across risk groups	Warning device class		
			Passive	Flashing lights	Gates
FRA, historic data	6	Poisson Regression	38.2	19.0	49.7**
		Poisson-Normal	38.1	18.3	47.8*
	10	Poisson Regression	38.8	20.0	50.9**
		Poisson-Normal	38.6	19.6	49.1*
Poisson Regression, 1997-2001 data	10	Poisson-Normal	73.5*	30.7	63.3**
	Ban effect estimated directly from regression coefficient		74.9***	30.9	66.8***

* statistically significant at less than 0.05; ** less than 0.01; *** less than 0.001

Table S3 presents the percent difference between grade crossings with a whistle ban and grade crossings without a ban for the Chicago area. As noted earlier, the estimates for grade crossings with a passive warning device or with flashing lights are based on very few grade crossings and are not reliable -- they are excluded from Table S3. At grade crossings with gates, crossings with a ban had a lower rate of accidents than crossings without a ban, but none of the estimates were statistically significant at the 5% level.

We performed a Monte Carlo simulation to investigate sample size requirements for achieving stable whistle ban effect estimates, for details of the simulation, (see the Appendix on Monte Carlo Simulation). As the simulation results show, grade crossing sets with sample sizes equal to the number of Chicago area grade crossings are not large enough to derive stable whistle ban impact estimates for any of the warning device classes. The instability is particularly striking for grade crossings with passive devices and for grade crossings with flashing lights. The hypothetical whistle ban effect estimates ranged from a low of a 100 % accident reduction to a high of 111% accident increase for grade crossings with a passive device, and the 25 simulated estimates had an average accident increase of 29.2%. Doubling the simulation sample sizes had little effect on the instability of these estimates. The comparable hypothetical results were similarly unstable for grade crossings with flashing lights.

Table S3. Ban effect estimates by methods for predicting grade crossing accident risk, aggregating across risk groups and number of risk groups.

Region: Chicago Area

Method for predicting accident risk	N of risk groups	Method for aggregating across risk groups	Warning device class		
			Passive	Flashing lights	Gates
FRA, historic data	6	Poisson Regression	-[1]	-	-26.4
		Poisson-Normal	-	-	-29.0
	10	Poisson Regression	-	-	-28.3
		Poisson-Normal	-	-	-29.1
Poisson Regression, 1997-2001 data	10	Poisson-Normal	-	-	- 8.6
	Ban effect estimated directly from regression coefficient		-		-16.5

[1] - Indicates unreliable estimates that were excluded because only few grade crossings were equipped with a passive warning device or flashing lights in the Chicago area

* statistically significant at less than 0.05; ** less than 0.01; *** less than 0.001

Table S4 presents the percent difference between grade crossings with a whistle ban in the Chicago area grade crossings without a ban in the continental United States outside of Florida. As noted earlier, the Chicago area estimates for grade crossings with a passive warning device or with flashing lights are based on very few grade crossings and are not reliable, (see discussion above, and the Appendix on Monte Carlo Simulation). At grade crossings with gates, crossings with a ban had a higher rate of accidents than crossings without a ban, but none of the estimates were statistically significant at the conventional level of 5%.

Table S4. Ban effect estimates by methods for predicting grade crossing accident risk, aggregating across risk groups and number of risk groups.

Region: Chicago Area v. Continental U. S. (Except Florida)

Method for predicting accident risk	N of risk groups	Method for aggregating across risk groups	Warning device class		
			Passive	Flashing lights	Gates
FRA, historic data	6	Poisson Regression	-[1]	-	5.8
		Poisson-Normal	-	-	1.5
	10	Poisson Regression	-	-	6.9
		Poisson-Normal	-	-	3.0
Poisson Regression, 1997-2001 data	10	Poisson-Normal	-	-	13.5
	Ban effect estimated directly from regression coefficient		-	-	17.3

[1] - Indicates unreliable estimates that were excluded because only few grade crossings were equipped with a passive warning device or flashing lights in the Chicago area

* statistically significant at less than 0.05; ** less than 0.01; *** less than 0.001

Table S5 presents the percent difference between grade crossings with a whistle ban and grade crossings without a ban for the Wisconsin area. The data for the Wisconsin area was sparse, especially for gated grade crossings. There were only 19 collisions at gated grade crossing accidents without a ban, and 27 at gated grade crossing accidents with a ban. For these data, ban effects could not be estimated directly using only Wisconsin grade crossing data, because the statistical procedure used in Method 3 did not converge (i.e. not all the estimates could be calculated). Also, the estimates obtained using the other procedures were unstable; they varied by over a factor of 2 from 98.7 to 218.6, and had correspondingly wide confidence limits. All estimates for crossings with passive devices and flashing lights were positive, indicating that grade crossings with a ban had a higher rate of accidents in the Wisconsin area than grade crossings without a ban. (For gates, the effect estimates were also positive, and large, but as noted earlier, these estimates are unreliable.) For grade crossings with flashing lights, the effects estimates obtained with some, but not all, methods were significant at the 5% level. They were not significant with any of the methods for grade crossings with a passive device.

Table S5. Ban effect estimates by methods for predicting grade crossing accident risk, aggregating across risk groups and number of risk groups.

Region: Wisconsin

Method for predicting accident risk	N of risk groups	Method for aggregating across risk groups	Warning device class		
			Passive	Flashing lights	Gates
FRA, historic data	6	Poisson Regression	11.9	33.0	98.7*
		Poisson-Normal	7.9	47.8	138.9
	10	Poisson Regression	12.6	33.7	116.7*
		Poisson-Normal	7.5	46.3	138.9
Poisson Regression, 1997-2001 data	10	Poisson-Normal	25.2	65.7	218.6
	Ban effect estimated directly from regression coefficient		32.2	74.8*	-

* statistically significant at less than 0.05; ** less than 0.01; *** less than 0.001

Table S6 presents the percent difference between grade crossings with a whistle ban in the Wisconsin area and grade crossings without a ban in the continental United States outside of Florida. The general pattern of results in Table S6 is similar to the pattern of results in Table S5 but with less statistical unstability because of the increased size of the comparison groups. On the one hand, these results are still based on only the 140-205 grade crossings in Wisconsin that had the ban, but on the other hand, the comparison groups are large, and include 25,000 or more grade crossings. For grade crossings with a passive device, method 3 estimated a statistically significant increase of 94.7%. The estimate based on method 4 was a statistically significant 92.9% increase. Methods 1 and 2 yielded lower, and not significant, estimates. For grade crossings with flashing lights, the estimated increases were 11%-12%, and none were statistically significant. For grade crossings with gates, the method 3 ban effect estimate was a statistically significant increase of 82.5%. The other estimates ranged from 72% to 79%; three were significant, two not. However, the 95% confidence intervals were very wide for all estimates, indicating that the numeric values of these estimates are quite imprecise.

Table S6. Ban effect estimates by methods for predicting grade crossing accident risk, aggregating across risk groups and number of risk groups.

Region: Wisconsin v. Continental U. S. (except Florida)

Method for predicting accident risk	N of risk groups	Method for aggregating across risk groups	Warning device class		
			Passive	Flashing lights	Gates
FRA, historic data	6	Poisson Regression	60.1	12.0	74.2**
		Poisson-Normal	60.0	11.1	72.7
	10	Poisson Regression	60.3	12.7	72.8**
		Poisson-Normal	60.1	12.3	72.0
Poisson Regression, 1997-2001 data	10	Poisson-Normal	92.9*	10.9	79.1*
	Ban effect estimated directly from regression coefficient		94.7**	10.9	82.5**

* statistically significant at less than 0.05; ** less than 0.01; *** less than 0.001

Discussion

Expected accident frequency estimates for grade crossings with and without a ban were calculated by warning device class using six different methods. The estimates were then compared within warning device class to obtain six sets of estimates for ban effects on expected accident frequency. Four of the methods (1.a, 1.b, 2.a and 2.b) used FRA's basic prediction formula to estimate expected accident frequencies. Two of the methods (3 and 4), re-estimated the dependence of expected accident frequency on grade crossing characteristics from accident data for 1997-2001 and grade crossing characteristics in 2001, and used these estimates instead of FRA's basic prediction formula. The expected accident rate predictions based on FRA's basic prediction formula worked well even though the formula had been estimated from historical data. However, expected accident rate predictions based on recent data -- as in models 3 and 4 -- were always superior to predictions based on historical data.

Model quality is assessed in terms of the correlation between observed and model-predicted accident frequencies: good models result in high correlation, poor models in low correlation. Table 2 compares the correlations between observed and model-predicted accident frequencies for models based on historical data to similar correlations for models based on recent data. As the table shows, the correlation between observed and predicted accident frequency was always higher for models based on recent data (e.g. Model 4) than for models based on historical data (e.g. models 1.a-2.b). In comparisons for the continental U. S. (outside Florida and the Chicago area), the correlations based on recent-data models exceeded the correlations based on historical-data models by 31.5% for grade crossings with passive warning devices, by 11.8% for grade crossings with flashing lights, and by 6.8% for gated grade crossings. For just the Chicago area, the correlations for recent-data models exceeded the correlations for historical-data models even more, by 178.9%, 365.9%, and 35.4%, respectively.

Controlling the effect of grade crossing characteristics on expected accident frequency is critically important to obtaining valid whistle ban effect estimates on expected accident frequency. We selected as our final model the statistical procedure that accomplished this best, and chose the direct estimates based on model 3. We note that model 4 estimates for whistle ban effects were very similar, both numerically and conceptually. The two models only differ in that in model 4, grade crossings were grouped on expected accident frequency before estimating ban effects, but in model 3, they were not. We opted for model 3 estimates because grouping data tends to reduce both precision and statistical significance.

Relative to the gated crossings without a ban, the estimates for the Chicago area ban crossings shown in Table S3 were below the comparable estimates for ban crossings for the rest of the nation (except Florida and Chicago) shown in Table S2. The 95% confidence limits (-52.5%, +46.6%) for the Chicago area estimate of a -16.5% reduction for the ban grade crossings was not statistically significant at the conventional 5% level, as evidenced by the fact that the value 0, associated with no ban effect, was bracketed by the confidence limits. In contrast, the 95% confidence limits of (32.4%, 110.2%) for the estimated 66.8% increase at ban grade crossings in the rest of the nation was statistically significant, as evidenced by the fact that the confidence limits did not bracket 0, the value associated with no effect. Since the confidence limits for each of the point estimates excluded the point estimates for the other, the ban effect in the Chicago area is different from the ban effect in the rest of the nation.

As we noted, ban effect estimates for the Chicago area exhibited a great deal of variability. To explore this further, we conducted a Monte Carlo simulation to assess the inherent accuracy limitations on ban effect estimates due to sample size limitations. The study involved repeatedly estimating hypothetical ban effects from data that were explicitly constructed with no ban effect but otherwise resembled the grade crossing inventory and accident data for the Chicago area. The results showed that grade crossing sets with sample sizes equal to the number of Chicago area grade crossings are not large enough to derive stable ban effect estimates for any of the warning device classes. The lack of stability was particularly striking for grade crossings with passive devices, and for grade crossings with flashing lights. In 25 repetitions of the experiment, the hypothetical whistle ban effect estimates ranged from a low of a 100 % accident reduction to a high of 111% accident increase for grade crossings with a passive device. Doubling the simulation sample sizes had little effect on this lack of estimate stability. The comparable hypothetical results were similarly unstable for grade crossings with flashing lights. Although somewhat better, the simulated accident effect estimates for gated grade crossings were also unstable, with an average of a 3.2% adverse effect and a full range that extended from a 36.2% reduction to a 45.8% increase. After doubling the sample size, the full range still extended from a 33.8% reduction to a 38.4% increase. Based on this simulation, it is clear that even doubling the study period would not provide adequate power for assessing ban effects in the Chicago area. Extending the study period to 20-25 years might provide adequate sample sizes for gated crossings, but only under the assumption that other circumstances affecting grade crossing accidents remain constant.

Based on past experience, the assumption of little or no change over such long periods is hardly justified.

Because of data limitations, it is not possible to reliably estimate ban effects for such comparatively small areas as the Chicago area. Although it was not possible to determine whether or not gated ban crossings had a lower or a higher accident rate than the rest of the gated crossings in the Chicago area, the results presented in this study did show that ban effect for gated crossings was different in the Chicago area than in the rest of the nation. Why this is so, was not determined in this study, and we can only speculate here.

One possible explanation for the difference is that crossing characteristics not used in any accident prediction model influenced decisions about whistle ban permits in the Chicago area differently than elsewhere. If so, the effect of these unspecified characteristics would not be accounted for in the analyses presented here. A second possible explanation is that the Chicago area has a combination of unique grade crossing characteristics and traffic patterns that are only imperfectly represented in statistical models derived from data both inside and outside the Chicago area. In fact, we found statistical evidence to support this possibility, but because of data limitations, the 'Chicago effect' could not be adequately accounted for by our statistical models. To the extent that accident predictions are inaccurate, ban effect estimates can also be inaccurate. A third possible explanation is that driver behavior is different in the Chicago area than elsewhere, perhaps because the exceptionally high exposure to gated crossings, high train traffic, or the high levels of recent public controversy about bans, focused driver attention to grade crossings.

References

Holland, P.W. (1986). Statistics and causal inference. *Journal of the American Statistical Association*, 81, 945-960.

McCullagh, P., and Nelder, J. A. (1989). *Generalized linear models*. (2nd ed.). New York: New York: Chapman and Hall.

SAS Institute, Inc. (1996). *SAS/STAT Software: changes and enhancements for release 6.12.*

U.S. Department of Transportation, Federal Railroad Administration. (1987). *Highway-rail crossing resource allocation procedure. user's guide*. (3rd ed.). DOT/FRA/OS-87/10 DOT-TSC-FRA-87-1.

U.S. Department of Transportation, Federal Railroad Administration. (2000). *Updated analysis of train whistle bans*. Washington, D.C.

Wolfinger, R. D., and Lin, X. (1997), "Two Taylor-series Approximation Methods for Nonlinear Mixed Models," *Computational Statistics and Data Analysis,* 25, 465 -490.

Wolfinger, R. D. and O'Connell, M. (1993), "Generalized Linear Mixed Models: a Pseudo-likelihood Approach," *Journal of Statistical Computation and Simulation,* 48, 233 -243.

Zador, P. L. (2000). *The effect of train whistle bans and collisions at public at-grade highway-crossings: an analysis of the FRA grade crossing accident inventory*.

Zador, P. L. (2002). Analysis of the Safety Impact of Train Horn Bans at Highway-Rail Grade Crossings. Final Report. Prepared under Contract No. DTFR53-00Q-00479. Westat, 1650 Research Boulevard, Rockville, MD 20850.

Tables

1997-2001 Grade Crossing Accident Rates By Predicted Accident Probability (PAP) Group and Whistle Ban

1997-2001 Grade Crossing Accident Rates By Accident Risk Group and Whistle Ban

Accident rate difference (and its lower and upper 95% confidence limits) between crossings with and without a train horn ban. Estimates based on Poisson-Normal model for risk group frequencies.

Plot 1. Observed (AVGACC) v. estimated (AVGAPP3) average accident rates by risk group for grade crossing with ban (1) and grade crossing without a ban (0).

Accident rate difference (and its lower and upper 95% confidence limits) between crossings with and without a train horn ban. Estimates based on Poisson-Normal model for risk group accident frequencies.

Accident rate difference (and its lower and upper 95% confidence limits) between crossings with a train horn ban and crossings without a train horn ban. Estimates based on General Linear Model with Poisson link function.

Accident rate difference (and its lower and upper 95% confidence limits) between crossings with a train horn ban and crossings without a train horn ban. Estimates based on General Linear Model with Poisson link function.

Estimated rate difference (and its lower and upper 95% confidence limits) between crossings with and without a train horn ban by region and warning device class. Estimates based on General Linear Models for 1997-2001 accident frequency.

Accident rate difference (and its lower and upper 95% confidence limits) between crossings with and without a train horn ban. Estimates based on Poisson-Normal model for risk group frequencies. Grade crossings accident risk groups were derived from general linear model for 1997-2001 accident frequencies.