The Effect of Infrastructure and Demographic Change on Traffic-related Fatalities and Crashes: A Case Study of Illinois County-level Data

Robert B. Noland
Lyoong Oh

Centre for Transport Studies
Dept. of Civil & Environmental Engineering
Imperial College London
London, SW7 2BU
Ph: 44(0)20-7594-6036
Fx: 44(0)20-7594-6102
Email: r.noland@ic.ac.uk
Email: l.oh@ic.ac.uk

Revised: Feb. 24, 2003

Abstract

This paper presents analyses of data from the Highway Safety Information System (HSIS) for the State of Illinois. Our analyses focuses on whether various changes in road network infrastructure and geometric design can be associated with changes in road fatalities and reported accidents. We also evaluate models that control for demographic changes. County-level time-series data is used and fixed effect negative binomial models are estimated. Results cannot confirm the hypothesis that changes in road infrastructure and geometric design have been beneficial for safety. Increases in the number of lanes appears to be associated with both increased traffic-related accidents and fatalities. Increased lane widths appears to be associated with increased fatalities. Increases in outside shoulder width appear to be associated with a decrease in accidents. Inclusion of demographic results does not significantly change these results but does capture much of the residual time trend in the models. Potentially mis-leading results are found when the time trend is not included. In this case a negative association between vertical curvature and both accidents and fatalities. No statistical association with changes in safety is found for median widths, inside shoulder widths, and horizontal and vertical curvature.

Keywords: transport safety, road infrastructure, geometric design, negative binomial model, Highway Safety Information System

Introduction

Reductions in traffic casualties, especially fatalities, have traditionally been linked to three general areas of transport policy. These include efforts to change risk-taking behavior, such as drunk driving, regulations to improve the safety of vehicles, and efforts to build and design safer road infrastructure. Recent research by Noland (in press) has explicitly challenged whether new and improved road designs actually have produced fatality reductions in the US. Instead, Noland finds that increased seat-belt usage, demographic change, and improvements in medical care seem to be more associated with fatality reductions over time than various improvements to the road network.

This study examines this issue in more detail. Noland (in press) used state-level data on various infrastructure variables in his analysis. Specifically, he examined lane widths, road classes, number of lanes and total lane miles, the latter representing a proxy for new and better-designed roads. In this study we expand upon this by examining more detailed infrastructure variables, including median widths, shoulder widths, and both horizontal and vertical curvature. This is made possible by use of the Highway Safety Information System (HSIS) data, which was obtained for the state of Illinois. This data is considered more accurate than the Federal Highway Administration (FHWA) Highway Statistics (US DOT, 1998) data used in Noland (in press) and has the advantage of providing more detail on various infrastructure elements.

As Elvik (2002) points out, many safety studies do not control for confounding factors that may also affect safety. Elvik (2002) shows how this can often seriously inflate the estimated safety benefit of various infrastructure improvements. Our analysis includes demographic effects and a time trend to control for potential confounding factors.

Statistical models are estimated with both total fatalities and total crashes as the dependent variable. Demographic data is included in the analyses, especially age cohort data.

County level data is used over four years and for 102 counties in Illinois. Results tend to support the hypothesis that most infrastructure variables (with some exceptions) have no effect on fatalities and crashes when demographic changes are controlled for. Larger outside shoulder widths appear to reduce crashes, while increased lane widths appear to be associated with increased fatalities. Increases in the number of lanes appears to be associated with increased accidents and somewhat with increased fatalities. Most other effects are not statistically significant, which disputes many of the established findings in the traffic engineering literature.

This paper is organized as follows. First, we review some of the recent literature in this area focusing on more recent studies that have used appropriate statistical techniques. This is followed by a discussion of the data and methods used. Results are then discussed followed by conclusions.

Literature Review

Most previous statistical analyses of road infrastructure effects have had significant shortcomings. In addition to normally not controlling for other effects over time, many older studies have estimated mis-specified model structures. In particular, linear regression models that assume a normally distributed error structure were frequently used. These do not adequately account for the non-negative count structure of accident and fatality data (Jovanis & Chang, 1986; Miaou et al., 1993). In addition, accident rates are frequently modelled, but much evidence suggests that rates (such as accidents per capita or per VMT) may not be linear (Olmstead, 2002). Finally, most estimates tend to combine fatal and injury accident data, a logical step when data is sparse, but potentially mis-leading as the infrastructure factors associated with fatalities could be quite different than those associated with injuries (Noland & Quddus, 2002).

In addition, many studies have focussed on individual links within a road network, thereby missing potential system-wide effects. For example, improved perceptions of safety in one location may lead to increased risk taking on other parts of the network. Boyle & Wright (1984) termed this behavioral effect "blackspot migration". Link level analysis also makes controlling for other policy instruments and exogenous change problematic.

McGee et al. (1995) and Transportation Research Board (1987) were two major studies that were aimed at developing statistical "accident reduction factors". That is, they estimate models with infrastructure and geometric design variables with the purpose of developing specific coefficient estimates to use in cost-benefit analyses. Their models, however, do not control for confounding effects and do not consider system-wide impacts. Many also fail to distinguish between the severity of different crash types which is crucial information needed for cost-benefit analysis. In addition, studies that report coefficient estimates of road safety factors normally do not show the confidence intervals surrounding these estimates, which can be quite large (Noland & Karlaftis, 2003).

Those studies that have attempted to account for many of these statistical issues have had unanticipated findings. Fridstrom & Ingebrigsten (1991) estimate a model using monthly data on traffic accidents for 18 counties in Norway. They find that extensions and improvements to the national road network do not have the expected effect of improving safety. Karlaftis & Tarko (1998) analyze county level data from the state of Indiana and find that increased road mileage is associated with increased accidents. Both these studies use aggregate cross-sectional time-series data and a negative binomial regression as is done in the analysis presented here.

Milton & Mannering (1998) also examine various geometric design elements. They find that increasing the number of lanes on a given road segment, leads to more accidents and that in Eastern Washington, narrower "substandard" lane widths (of less than 3.5 metres or

11.5 ft) reduce accident frequency. They also found that horizontal curvature does not by itself increase accidents but is dependent upon whether large straight sections preceded the curves. While this latter result supports the hypothesis that reducing horizontal curvature reduces accidents, it does suggest that roads with extensive curvature may not necessarily be less safe than straighter roads. Milton & Mannering (1998) do not control for any time series or demographic effects in their study.

Shankar et al. (1995) estimated a series of negative binomial regression models in a study of the Interstate 91 corridor in Washington State. They found that when curves are spaced further apart (i.e., fewer curves per mile) more severe overturn accidents increase. This same study also found that highway segments that have curves with lower design speeds result in fewer accidents relative to those with higher design speeds; though the presence of snowfall tended to increase accidents on those segments with curves of lower design speeds. Shankar et al. (1995) found that those accidents attributable to curves of lower design speeds tended to be less severe than those associated with curves of higher design speeds.

Abdel-Aty & Radwan (2000) found that 'improvements' in geometric design variables reduce accidents. These included the degree of horizontal curvature and shoulder, lane and median widths. They estimated a negative binomial regression model with road segment data from an arterial highway in Florida. One problem with this study (other than the lack of control for time and demographic effects) is that it does not control for repeated observations (that is, multiple sampling of accidents from each segment). Shankar et al. (1995) do control for this by including section-specific constants in their models. This could perhaps account for the very different results shown by these two studies.

Vogt and Bared (1998) evaluate changes in design parameters for two lane rural roads using HSIS data. Using a population of highway segments for two states (Washington and Minnesota) they derive detailed statistical models linking design elements to both total crashes and more serious crashes involving a fatality or injury (however, not disaggregating between these two). They find that increasing lane widths and less horizontal curvature reduces total crashes. While using time-series data they do not appear to control for time in their model, nor other factors that may change over time. They acknowledge the limitations of their model and that various key variables may be omitted. The lack of controlling for time series effects, as well as cross-sectional effects, is likely to bias the results of this study.

Ivan et al. (2000) using data from Connecticut found that link segments with larger shoulder widths have more single-vehicle crashes. As will be shown, this contradicts the result found in this paper. Ivan et al. also found that those links with lower volume/capacity ratios had more crashes, even when controlling for the time of day of the crash. That is, those links with lower levels of congestion may have more crashes, a finding which is consistent with Zhou & Sisiokipu (1997) and other recent work documenting the safety benefits of congestion. While Ivan et al. use count models for analyzing their data, they do not disaggregate by severity level and have no way of controlling for time series or demographic effects.

Many studies that find unexpected or unconventional results tend to dismiss these results as aberrations within their dataset and have not examined the issue in further detail. In addition, one would expect that many studies that challenge accepted hypotheses would suffer from publication bias; that is, they would have a more difficult time passing the requirements of peer review.

The results of many of these studies lead us to conclude that the impact of various infrastructure and geometric design elements on safety are inconclusive. Most studies using more sophisticated statistical techniques either find no association or an unexpected association from infrastructure changes assumed to be beneficial. In most cases, they do not control for other more important effects that change over time (such as demographics, seat-belt usage, and alcohol consumption).

Mahalel & Szternfeld (1986) proposed a theory as to why various improvements in the driving environment could adversely affect safety. They observed that many design changes tend to make the driving task easier and less taxing on the driver. This could reduce the level of concentration needed to maintain the same level of safety. If driver perceptions then lead to an underestimation of the difficulties associated with the driving task the net result could be an increase in accidents. Most road improvements also allow greater speeds which could off-set the safety improvements associated with the road. This is essentially risk compensating behavior as originally described by Peltzman (1975). This theory would lead one to conclude that infrastructure changes will have no effect on safety outcomes, if risk reductions are off-set by increases in risk taking.

This paper does not attempt to propose alternative theories but focuses on empirical analyses of HSIS data from Illinois. As will be seen, we cannot show, in most cases, that infrastructure changes are beneficial from our analyses of the Illinois data. The next section describes the data that was analysed and the process of preparing it for analysis.

Illinois road and socio-demographic data

The State of Illinois is composed of 102 counties populated by 12.4 million people (in year 2000) over an area of 55,584 square miles. The rate of population increase was 8.6% over the past decade and the corresponding mean density is estimated at 223.4 persons per square mile, though this varies considerably between urbanized and rural areas. In 1987 there were 49830 reported accidents and 785 fatalities on state-maintained roads in the HSIS database. In 1990 these had decreased to 43394 reported accidents and 715 fatalities.

HSIS data was made available by the US Federal Highway Administration (FHWA) for the years 1987 to 1994 for the State of Illinois. This contains four main datasets: accident fatality/injury statistics, occupants involved, vehicle information and most significantly for the current analysis, the characteristics of the road infrastructure. These data have been provided to the FHWA by participating States and the roads covered by the HSIS scheme consist of the State System Mileage which is roughly 12.5% of Illinois State total mileage. HSIS data is considered to be of very high quality and greater care is taken in the collection of infrastructure information as opposed to that collected for the Highway Performance Monitoring System.

The HSIS for Illinois is based on approximately 15,000 state mile posts which were used as measuring points for identifying accidents as well as infrastructure characteristics of the network. Various different road categories are included in the data and are shown in Figure 1. As can be seen, rural two-lane roads account for the vast majority of state-maintained road mileage in the state. The various infrastructure characteristics were aggregated to the county-level according to the description in Table 1. This includes total lane mileage, the average number of lanes, average lane widths, average median widths, mean inside and outside shoulder widths, and measures of horizontal and vertical curvature. Data on road characteristics covered only state-maintained infrastructure. Thus, we do not include characteristics associated with local road networks. The vast majority of traffic, however, is normally carried on the state-maintained system, and thus this restriction is not a problem in our analyses. Only those crashes that occurred on the state-maintained system are used in the analyses.

To calculate mean values for the road characteristics it was assumed that those characteristics were constant for each road section (as delineated by the mileposts in the data). For example, the mean lane width was estimated by calculating the section miles times the width of each section and averaged by the total county road miles. This applies to all mean dimensions such as mean road width, median width and shoulder width.

Horizontal curvature is measured by the radius of the distance between the road tangent and the center of the curve. A large radius implies a large circle, hence a flat curvature or straight road. Only those curves greater than 2.5 degrees (or radius of less than 2293 feet) are measured. The larger the angle of curvature the higher the severity of deflection is. The total number of mile posts whose deflection angles are larger than 2.5 degrees were counted for each mile and averaged over the entire county's road network. The actual average deflection angles (as opposed to radii) at the county level was used as a covariate in our model.

For vertical curves, the key measurement point is the approach and downside gradient and the distance separating them. The HSIS Illinois State guidelines suggest that these variables are only coded in those cases in which the field engineering staff considers the existing vertical curve to be "substandard" in nature. It is assumed to imply that the observations are taken where the measurements are below some specified standard for road design (not stated in the HSIS manual). The variable used is the number of such "substandard" vertical curves for each mile averaged over the entire county.

Road medians are also included in the database. The majority of road miles are rural and do not have median dividers. Median installations on these non-urban roads are observed to be minimal, consisting of less than 5% of the total road miles. Moreover, most of these were 'painted' medians (as opposed to barriers) and have no substantial width. Consequently, no median implies no inside shoulder. Our analyses of the data included models both with and without the median and inside shoulder data, but the latter resulted in the omission of 33 rural counties since they did not have any roads with medians.[1]

Traffic volume is represented by annual average daily traffic (AADT) collected by automatic traffic recorders (ATR). In addition to 49 permanent counters that operates 24 hours each day for 365 days, a series of short-term "coverage" surveys were carried out at the county level. AADT was estimated by averaging the roads over the entire county. In our analyses of the HSIS data we found various inconsistencies between years in the AADT data. We were unable to resolve the source of these inconsistencies and consequently omitted this variable from our analysis.

Income data was collected from the Bureau of Economic Analysis (BEA, US Department of Commerce, http://www.bea.doc.gov/) and the demographic data from the US Census Bureau (US Department of Commerce, http://www.census.gov/). The magnitudes of the counties within the State vary significantly. For example, the county population ranges from just over 4000 (Pope) to 5 million (Cook). Age cohorts for each county were calculated.[2]

HSIS data for Illinois was available for the years 1987-1992 and for 1994. Unfortunately, we found that data for years after 1990 appeared to have some definitional changes. We were unable to determine what those changes were and hence could not adjust the data to be consistent with earlier years. Therefore, our analyses includes only the years 1987-1990. The statistical analyses methods used are described in the next section followed by a discussion of results.

Methodology

As discussed previously, accident data consists of counts and thus violates the normality assumptions of standard linear regression models. For this reason we estimate a negative binomial model which is a generalization of the Poisson regression model. Since our data is a mix of time-series and cross-sectional data we utilize the method derived by

---

[1] These models found that mean median width and mean inside shoulder width were not statistically significant variables. These results are not presented for brevity.

[2] We also estimated models without Cook county (metropolitan Chicago). This was due to recommendations that data from Chicago may not be of the same quality as the rest of the state and also because of the large

Hausman et al. (1984) for estimating negative binomial model with panel data. This method has the advantage of factoring out the overdispersion parameters and accounting for heterogeneity in the data.

To account for the fixed individual effects in the negative binomial model, we rewrite the Poisson parameter as,

$$?_{it} \ ? \ ?_{it} ?_{i} \qquad (1)$$

where,

$$?_{it} \ ? \ E(n_{it}) \ ? \ \exp(\boldsymbol{ß} \ \boldsymbol{X}_{it}) \qquad i \ ? \ 1, 2, \ldots. N \qquad t \ ? \ 1, 2, \ldots\ldots.. T \qquad (2)$$

in which $?_{it}$ is the Poisson parameter indicating expected numbers of casualties in an observation unit *i*, in a given time period t, n_{it} is the number of observed casualties in an observation unit *i*, during a given time period t, $\boldsymbol{X}_{it}$ is a vector of covariates which describe the characteristics of an observation unit *i*, during a given time period t, $\boldsymbol{ß}$ is a vector of estimable coefficients representing the effects of the covariates, and $?_{i}$ is the individual-specific fixed effect. To derive the joint probability of the fixed effect negative binomial model, it is necessary to find a convenient distribution for the sum of events for a given individual, $?_{t} n_{it}$. A detailed derivation can be found in Hausman et al. (1984). The resulting joint probability of the i^{th} individual conditional on total years is

$$\Pr(n_{i1},\ldots\ldots,n_{iT} \mid ?_{t} n_{it}) \ ? \ \left(?_{t} \ \frac{?(?_{it} \ ? \ n_{it})}{?(?_{it})?(n_{it}) \ ? \ 1} \right) ? \left(\frac{?(?_{t} ?_{it})?(?_{t} n_{it} \ ? \ 1)}{?(?_{t} ?_{it} \ ? \ ?_{t} n_{it})} \right) \qquad (3)$$

which includes $\boldsymbol{ß}$ via $?_{it}$ but does not include $?_{i}$ and the overdispersion parameter k. From this the likelihood function can be derived as,

variation in the population of the county. Results were not substantively different from results including all counties.

$$L(\sum_t n_{it} \mid n_{it},......,n_{iT}) = \prod_t \left[\prod_t \frac{\Gamma(\lambda_{it} + n_{it})}{\Gamma(\lambda_{it})\Gamma(n_{it}) + 1}\right] \cdot \left[\frac{\Gamma(\sum_t \lambda_{it})\Gamma(\sum_t n_{it} + 1)}{\Gamma(\sum_t \lambda_{it} + \sum_t n_{it})}\right] \qquad (4)$$

This method appears quite robust for analyses of accident data as shown by Noland & Karlaftis (2002). Use of ordinary least squares with an exposure variable, such as fatalities per capita, as the dependent variable, appears to reduce the robustness of coefficient estimates.

The results of models estimated with this technique are discussed in the next section.

Results

Our initial objective in analyzing the data was to examine the association between the infrastructure and geometric design variables and total reported accidents and fatalities. The standard hypothesis is that newer and improved infrastructure leads to fewer accidents and fatalities and we test these hypotheses for the variables in the dataset. We also test whether inclusion of variables to correct for changes over time have an effect. This is done by including fixed effect year dummy variables. Finally, the effect of various demographic and socio-economic variables are examined and the robustness of the results for the infrastructure variables when these other effects are included is examined. Results for the estimated fixed-effect negative binomial models are shown in Table 2 for total reported accidents and Table 3 for total fatalities. All models have logarithmic independent variables to minimize heteroskedasticity in the data.

For the accident models (Table 2) our results show that most of the infrastructure variables are not statistically significant. Two exceptions are the mean number of lanes and mean outside shoulder width. Increased numbers of lanes is associated with increases in accidents. However, this variable is not significant when time series effects are not controlled for. This is an important result and demonstrates the potential problem of not explicitly controlling for time series in the model. Noland (in press) found a similar effect from the number of lanes but with differences in effects depending on the functional class of road. Our data was not sufficient to examine this effect for each road category. This is an interesting result as it questions the policy of increasing many two-lane rural roads to four-lane roads for safety reasons. The fatality models (Table 3) find a positive association with number of lanes, but below the 95% level of statistical significance.

Increases in outside shoulder width seem to be associated with fewer accidents. Note also that increases in lane width has no statistically significant effect, suggesting that when increasing the width of right-of-ways it is more beneficial to devote the new space to creating a sufficient outside shoulder. This result, however, is quite different in the fatality models (Table 3). In these models we find a small positive (but not significant) association of increased outside shoulder width with increased fatalities, but a strong statistically significant positive association with increased lane widths.

As mentioned previously, we also examined median width and inside shoulder widths. This necessitated dropping 33 counties from the analysis since all their roads were two-lane roads with no medians and no inside shoulder. These variables were not found to be statistically significant.

The accident models also show a negative association with vertical curvature, although this is statistically insignificant and the result disappears when time series effects are properly controlled for. The lane mile variable in the fatality models shows a negative association, but this is not statistically significant.

The coefficients for the fixed effect dummy variables for each year are all statistically significant and negative. This means that some unmeasured factor is reducing total accidents over time. This is also seen in the fatality models (Table 3), though with lower levels of statistical significance.

When our demographic variables are included in the analyses we find that they pick up much of the residual effect from the year dummy variables, which are no longer statistically significant. This implies that inclusion of the demographic variables captures much of the unmeasured changes over time. Results for the infrastructure variables are generally the same when demographic variables are included.

Among the demographic variables, population is positively associated with increased accidents and fatalities but only at the 90% level of significance. Per capita income was not statistically significant. Most studies have found a correlation between accidents and economic activity, measured as income, (Noland, in press), but we do not find this effect at the county level.

The age cohort variables show that increases in the 15-24 age bracket is weakly associated with more fatalities but not with more accidents. The 65 and over age group also shows a weak association with more fatalities, but not accidents.

The results on the age cohort variables were generally disappointing and did not clearly give the anticipated results. Despite this we did find that the higher risk age categories (younger people and elderly people) seems to be associated with more fatalities but not more accidents. One reason for the relatively weak statistical significance on these variables may be the spatial unit (counties) used in the analyses. While the demographic data was an indicator of the age distribution of the county population, it may not completely represent those driving in the county, especially for smaller counties. Therefore, as a control for demographic factors a larger spatial unit for age cohorts may be preferable. Noland (in press) found a strong effect on age cohort variables when using states as the unit of spatial analyses.

To examine this in more detail, we estimated models with age cohort data aggregated across all neighboring counties. That is, we calculated the population data for a given county by including all counties with which it shared a border.[3] This did not improve the modelling results and actually increased the collinearity between some of the age cohort variables (these results are not shown).

Despite these relatively weak results on the age cohort variables, we still find that the time trend is no longer statistically significant when demographic variables are included, suggesting that these variables are capturing much of the residual effect from changes over time.

Multi-collinearity in the independent variables does not appear to be a problem. The highest level of correlation is between two of the age cohort variables, population between 25-34 and population over 65, with a correlation coefficient of –0.74. However, omission of the age cohort variable for population between 25-34 does not affect the overall results. In addition, the lane mileage variable is moderately correlated with number of lanes, number of horizontal curves, and number of vertical curves. Omission of the lane mile variable does not substantively affect the coefficients on these other variables. These results are omitted for brevity.

One of the important results of the analyses is the non-significance of many factors that are assumed to be beneficial for safety. This includes increased lane mileage, increased median and inside shoulder widths (not shown in these results), and reductions in horizontal and vertical curvature. While it is possible that this result is due to insufficient data, the analyses includes over 400 data points, so this is unlikely. Therefore, we conclude that our results do not support the hypothesis that improved geometric design and infrastructure reduces reported accidents and fatalities, for these specific measures.

[3] This was not done, however, for counties bordering other states.

Conclusions

This paper has analyzed HSIS data for the State of Illinois to evaluate the hypothesis that improved road infrastructure geometric design is beneficial for safety. Our results tend to reject this hypothesis in contrast with standard assumptions in the traffic safety literature. However, this result is not inconsistent with many recent studies that have used more sophisticated statistical techniques and also those that control for other confounding effects.

Our results suggest that some changes in infrastructure have actually led to increased accidents and fatalities. In particular, increases in the number of lanes appears to be associated with increased fatalities and accidents, and increases in lane widths are associated with increased fatalities. Increases in outside shoulder width appear to be associated with reduced accidents, but show a positive but statistically insignificant association with increased fatalities. Other factors, including median width, inside shoulder width, and horizontal and vertical curvature were not found to be statistically significant. We also found that our results did not change much when demographic variables were included, although these variables do appear to capture the residual time trend associated with reductions in both fatalities and reported accidents. Omission of any time trend in the model appears to give different results, for example on our variable for vertical curvature. This suggests the importance of controlling for time trend in this type of analyses.

Clearly, these results challenge some assumptions about the benefits of various road safety "improvements." Additional research is clearly needed to better understand these effects including additional empirical analyses of data. One conclusion is that care should be taken to control for changes in other factors over time. This study still omits two major potential effects, which are changes in seat-belt usage and changes in alcohol consumption. Additional research can clarify the behavioral responses to changes in infrastructure and how this affects safety.

Acknowledgments

This work was funded by the Engineering and Physical Sciences Research Council. We would like to thank Yusuf Mohamedshah and Li wan Chen of the LENDIS Corporation for supplying HSIS data, and Jean M. Dee from the Census Bureau for supplying Illinois demographic data. All errors and omissions are the responsibility of the authors.

References

Abdel-Aty, Mohamed A. and A. Essam Radwan, 2000, Modeling Traffic Accident Occurrence and Involvement, *Accident Analysis and Prevention*, 32: 633-642.

Boyle, A.J., and C.C. Wright, 1984, Accident 'Migration' After Remedial Treatment at Accident Blackspots", *Traffic Engineering and Control*, 25: 260-267.

Elvik, Rune, 2002, The Importance of Confounding in Observational Before-and-after Studies of Road Safety Measures, *Accident Analysis and Prevention*, 34: 631-635.

Fridstrom, Lasse and Siv Ingebrigsten, 1991, An Aggregate Accident Model Based on Pooled, Regional Time-Series Data, *Accident Analysis and Prevention*, 23: 363-378.

Hausman, J.C., Hall, B.H., Griliches, Z., 1984, Econometric models for count data with an application to the patents-R&D relationship. *Econometrica*, 52(4), 909-938.

Ivan, John N., Chunyan Wang & Nelson R. Bernardo, 2000, Explaining Two-lane Highway Crash Rates Using Land Use and Hourly Exposure, *Accident Analysis and Prevention*, 32: 787-795.

Jovanis, P. and Chang, H., 1986, Modeling the relationship of accidents to miles travelled, *Transportation Research Record* 1068, 42-51.

Karlaftis, Matthew G. & Andrzej P. Tarko, 1998, Heterogeneity Considerations in Accident Modeling, *Accident Analysis and Prevention*, 30: 425-433.

Mahalel, David and Zvi Szternfeld, 1986, Safety Improvements and Driver Perception, *Accident Analysis and Prevention*, 18(1): 37-42.

McGee, H.W., W.E. Hughes, and K. Daily, 1995, *Effect of Highway Standards on Safety*, National Cooperative Highway Research Program, Report 374, Transportation Research Board, National Research Council, National Academy Press: Washington, DC.

Miaou, Shaw-Pin and Harry Lum, 1993, Modeling Vehicle Accidents and Highway Geometric Design Relationships, *Accident Analysis and Prevention*, 25(6): 689-709.

Milton, John and Fred Mannering, 1998, The Relationship Among Highway Geometries, Traffic-related Elements and Motor-vehicle Accident Frequencies, *Transportation*, 25: 395-413.

Noland, Robert B., 2002, Traffic Fatalities and Injuries: The Effect of Changes in Infrastructure and Other Trends, forthcoming in *Accident Analysis and Prevention*.

Noland, Robert B., and Matthew G. Karlaftis, 2002, "Policy Recommendations from Accident Models: Are the Results Specification-Sensitive?", paper submitted to *Transportation Research E*.

Noland, Robert B., and Mohammed A. Quddus, 2002, A Spatially Disaggregate Analyses of Road Casualties in England, paper submitted to the 82nd Annual Meeting of the Transportation Research Board and to *Regional Science and Urban Economics*.

Olmstead, Todd, 2001, Freeway Management Systems and Motor Vehicle Crashes: A Case Study of Phoenix, Arizona, *Accident Analysis and Prevention*, 33: 433-447.

Peltzman, S., 1975, The Effects of Automobile Safety Regulation, *Journal of Political Economy*, 83: 677-725.

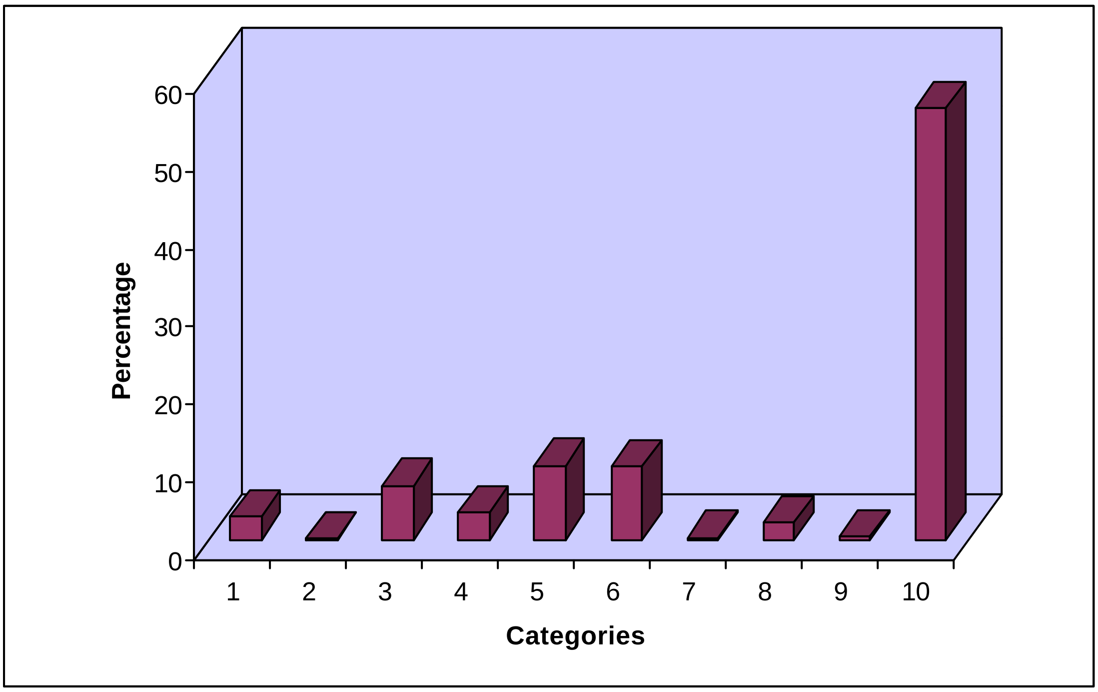
Shankar, Venkataraman, Fred Mannering, and Woodrow Barfield, 1995, Effect of Roadway Geometrics and Environmental Factors on Rural Freeway Accident Frequencies, *Accident Analysis and Prevention*, 27(3): 371-389.

Transportation Research Board, 1987, *Designing Safer Roads: Practices for Resurfacing, Restoration, and Rehabilitation*, Special Report 214, National Research Council, Washington, DC.

US DOT, 1998, *Highway Statistics 1997*, Federal Highway Administration, Office of Highway Information Management, FHWA-PL-98-020.

Vogt, Andrew and Joe Bared, 1998, Accident Models for Two-Lane Rural Segments and Intersections, paper presented at the 77th Annual Meeting of the Transportation Research Board.

Figure 1
Distribution of Road Categories in Illinois HSIS Data (1992)



1 = Urban freeways
2 = Urban freeways < 4 lanes
3 = Urban multilane divided non-freeways
4 = Urban multilane undivided non-freeways
5 = Urban 2 lane highways
6 = Rural freeways
7 = Rural freeways < 4 lanes
8 = Rural multilane divided non-freeways
9 = Rural undivided non-freeways
10=Rural 2 lane highway

Table 1
Summary of variables

Category	Variable (Unit)	Notes
Key variables	County name	102 counties
	Calendar year	1987-1990
Accidents	Number of accidents	Incapacitating injury = injury other than fatal requiring hospitalization Non-incapacitating injury = injury evident to others at scene Possible injury = no visible injury but complaint of pain
Road functional classes		Class10 = Non-urban/urban interstate freeway; Class20 = Non-urban/urban major highway; Class30 = Non-urban major/minor local ; collector; Class40 = Urban arterial; Class50 = Urban local collector
Annual average daily traffic	AADT	Not used due to data inconsistencies
Road Infrastructure	Lane miles (miles)	
	Number of lanes	
	Lane width (ft)	
	Median width (ft)	
	Inside shoulder width (ft)	is0 =no shoulder, is1 = soft shoulder (earth, sod), is2 = hard shoulder (aggregate, surface treated, paved), is3 = curb
	Outside shoulder width (ft)	os0 =no shoulder, os1 = soft shoulder (earth, sod), os2 = hard shoulder (aggregate, surface treated, paved), os3 = curb
	Number of horizontal curves	Horizontal curves measure the severity of horizontal curve angles. Considered as a ‘curve’ if the radius is less than 2293 ft.
	Horizontal deflection angle (degrees)	County means of all curves.
	Number of vertical curves	
Demographics & socioeconomics	Population	
	Income per capita (US $)	Consumer Price index (CPI) applied to normalize to 1992 real values
	County age group as percent of total population	Age bands: [15-24], [25-34], [35-64], [65+]

Table 2
Conditional fixed-effects negative binomial (Years 1987-1990): Dependent variable = Number of accidents

	Dependent variable = Number of Accidents					
	Model without time correction		Model with year dummy variables		Model with demographic variables	
	Coef.	ß/S.E.	Coef.	ß/S.E.	Coef.	ß/S.E.
Dummy for year =1988			-0.044	-3.02	-0.028	-1.47
Dummy for year =1989			-0.078	-4.78	-0.035	-0.92
Dummy for year =1990			-0.103	-6.27	-0.050	-1.07
log (Total lane miles)	0.218	1.58	-0.019	-0.14	0.032	0.23
log (Mean number of lanes)	0.033	0.08	1.206	2.72	1.249	2.84
log (Mean lane width)	0.069	0.21	0.216	0.69	0.182	0.59
log (Mean outside shoulder width)	-0.722	-2.78	-0.737	-3.07	-0.797	-3.33
log (Number of horizontal curves per mile)	0.280	1.09	0.335	1.36	0.273	1.10
log (Mean horizontal deflection angle)	-0.170	-0.56	-0.144	-0.50	-0.185	-0.64
log (Number of vertical curves per mile)	-0.453	-1.62	-0.301	-1.13	-0.259	-0.97
log (Population)					0.012	1.65
log (Income per capita, 1992$)					-0.331	-1.24
log (Percent 15-24 years of age)					-0.004	-0.10
log (Percentt 25-34 years of age)					-0.145	-1.26
log (Percent 65+ years of age)					-0.037	-0.60
Constant	7.002	3.69	6.123	3.45	8.967	2.84
LL(0)	-1191.29		-1191.29		-1191.29	
LL(?)	-1182.74		-1164.12		-1161.06	
N	408		408		408	

Table 3
Conditional fixed-effects negative binomial (Years 1987-1990): Dependent variable = Number of fatalities

	Dependent variable = Number of Fatalities					
	Model without time correction		**Model with year dummy variables**		**Model with demographic variables**	
	Coef.	ß/S.E.	Coef.	ß/S.E.	Coef.	ß/S.E.
Dummy for year =1988			0.085	1.69	0.106	1.37
Dummy for year =1989			-0.093	-1.45	-0.066	-0.44
Dummy for year =1990			-0.162	-2.53	-0.142	-0.72
log (Total lane miles)	-0.132	-0.26	-0.845	-1.48	-0.821	-1.42
log (Mean number of lanes)	1.211	0.70	3.329	1.75	2.952	1.53
log (Mean lane width)	2.560	1.91	2.848	2.09	2.942	2.12
log (Mean outside shoulder width)	1.437	1.40	1.427	1.38	1.145	1.08
log (Number of horizontal curves per mile)	1.413	1.40	0.982	0.98	0.879	0.86
log (Mean horizontal deflection angle)	-0.300	-0.24	0.249	0.20	-0.082	-0.06
log (Number of vertical curves per mile)	-1.987	-2.01	-0.961	-0.95	-0.736	-0.71
log (Population)					0.059	1.92
log (Income per capita, 1992$)					0.016	0.02
log (Percent 15-24 years of age)					0.267	1.63
log (Percentt 25-34 years of age)					0.472	1.03
log (Percent 65+ years of age)					0.388	1.63
Constant	14.290	0.04	11.248	0.06	13.110	0.03
LL(0)	-543.82		-543.82		-543.82	
LL(?)	-538.44		-529.73		-525.53	
N	404		404		404	