

UNIVERSITY OF MINNESOTA



CENTER FOR TRANSPORTATION STUDIES

**INTELLIGENT
TRANSPORTATION
SYSTEMS
INSTITUTE**

**A Case Control Study of Speed and
Crash Risk**

Technical Report 3

**Speed as a Risk Factor in Run-off
Road Crashes**

Final Report

Prepared by:

Gary A. Davis, Sujay Davuluri, Jian Ping Pei

Department of Civil Engineering
University of Minnesota

CTS 06-01C

Technical Report Documentation Page

1. Report No. CTS 06-01C	2.	3. Recipients Accession No.	
4. Title and Subtitle A Case Control Study of Speed and Crash Risk Technical Report 3 Speed as a Risk Factor in Run-off Road Crashes		5. Report Date April 2006	
		6.	
7. Author(s) Gary A. Davis, Sujay Davuluri, Jian Ping Pei		8. Performing Organization Report No.	
9. Performing Organization Name and Address Department of Civil Engineering University of Minnesota 500 Pillsbury Dr SE Minneapolis, MN 55455		10. Project/Task/Work Unit No. CTS Project Number 2001032	
		11. Contract (C) or Grant (G) No.	
12. Sponsoring Organization Name and Address Intelligent Transportation Systems Institute Center for Transportation Studies University of Minnesota 511 Washington Avenue SE, Suite 200 Minneapolis, MN 55455		13. Type of Report and Period Covered Final Report December 2000-February 2002	
		14. Sponsoring Agency Code	
15. Supplementary Notes http://www.cts.umn.edu/pdf/CTS-06-01C.pdf			
16. Abstract (Limit: 200 words) In the U.S.A., the imposition and subsequent repeal of the 55 mph speed limit has led to an increasingly energetic debate concerning the relationship between speed and the risk of being in a (fatal) crash. In addition, research done in the 1960s and 1970s suggested that crash risk is a U-shaped function of speed, with risk increasing as one travels both faster and slower than what is average on a road. Debate continues as to the causes of this relationship, and there is reason to suspect that it may be an artifact of measurement error and/or mixing of different crash types. This report first describes two case-control analyses of run-off road crashes, one using data collected in Adelaide, Australia and the other using data from Minnesota. In both analyses the speeds of the case vehicles were estimated using accident reconstruction techniques while the speeds of the controls were measured for vehicles traveling the crash site under similar conditions. Bayesian relative risk regression was used to relate speed to crash risk, and uncertainty in the case speeds was accounted for by treating these as additional unknowns with informative priors. Neither data set supported the existence of a U-shaped relationship, although crash risk clearly tended to increase as speed increased. The resulting logit model was then used to estimate the probability that a given speed could be considered a casual factor for each of the 10 Minnesota crashes.			
17. Document Analysis/Descriptors Ran off road accidents, risk taking, speed limits, speed, accident reconstruction, Bayes' theorem		18. Availability Statement No restrictions. Document available from: National Technical Information Services, Springfield, Virginia 22161	
19. Security Class (this report) Classified	20. Security Class (this page) Classified	21. No. of Pages 44	22. Price

**A Case Control Study of Speed and Crash Risk
Technical Report 3
Speed as a Risk Factor in Run-off Road Crashes**

Final Report

Prepared by:
Gary A. Davis
Sujay Davuluri
Jian Ping Pei

Department of Civil Engineering
University of Minnesota

April 2006

Intelligent Transportation Systems Institute
Center for Transportation Studies
University of Minnesota

CTS 06-01C

This report represents the results of research conducted by the authors and does not necessarily reflect the official views or policy of the Center for Transportation Studies or the University of Minnesota.

Acknowledgements

The authors wish to acknowledge those who made this research possible. Funding for the study was provided by the Intelligent Transportation Systems (ITS) Institute at the University of Minnesota's Center for Transportation Studies (CTS). The ITS Institute is a program funded by the United States Department of Transportation (USDOT) and administered by the USDOT's Research and Innovative Technologies Administration (RITA).

A Case Control Study of Speed and Crash Risk
Technical Report 3
Speed as a Risk Factor in Run-off Road Crashes

Gary A. Davis, Sujay Davuluri, Jian Ping Pei
University of Minnesota

Introduction

In March 1974 the Congress of the United States set the National Maximum Speed Limit (NMSL) at 55 mph. This was done as an energy conservation measure in response to the OPEC oil embargo of 1973, but following enactment of the NMSL traffic fatalities in the United States declined from 55,511 in 1973 to 46,402 in 1974 (TRB 1984, p. 1). Since at least some of this decline could plausibly be attributed reductions in highway speeds, safety advocates argued for retaining the NMSL after the oil shortages dissipated. Never the most popular of laws, the years following 1974 showed gradual but definite declines in compliance with the NMSL, and in 1982 Congress asked the National Academy of Sciences to investigate its costs and benefits. The committee charged with this task ultimately recommended retaining the NMSL, but this recommendation was by no means unanimous (TRB 1984).

Following publication of the committee's report attempts were made to use cross-sectional studies to identify a predictive relationship between speed and crash risk. One of the more provocative of these efforts was Lave's (1985) report that, when using the states of the union as observational units, variance in speeds appeared to correlate positively with fatal crash rate, even though average speed did not. This was seen as being consistent with Solomon's (1964) earlier finding that crash risk tended to be a U-shaped function of speed, with drivers travelling at speeds substantially above and below the average being at greater risk. A highly variable distribution of speeds would then tend to have more drivers travelling at high and low speeds, where crash risk was higher. A number of subsequent studies have also reported positive correlations between speed variability and crash risk, some also finding correlations between average speed and crash risk (e.g. Garber and Gadiraju 1988; McCarthy 1998).

In 1985 the U.S. Congress relaxed the NMSL for rural Interstate highways, and in 1995 repealed the NMSL entirely, returning the authority to set speed limits to the states. Following

these changes in the law several states raised speed limits on selected highways, and this opened up the possibility of supplementing the earlier cross-sectional studies with before-after investigations of the effect of speed limit changes. The general slant of this work appears to be that fatal crash rates increase following increases in speed limits (McCarthy 1998), but difficulties in controlling for possible confounding influences has meant that this interpretation remains controversial. For example, Lave and Elias (1994) argued that raising speed limits on rural interstates reduces overall fatal crash rates on a state's highway system, while Balkin and Ord "cast doubt on the blanket claim that higher speed limits and higher fatalities are directly related" (2001, p. 1).

To some extent, the current controversy regarding speed and crash risk appears to be due to the fact that many of the original, provocative findings admit alternative methodological explanations. For example, White and Wilson (1970) illustrated how U-shaped relationships between speed and crash risk, similar to those reported in Solomon (1964) and Cirillo (1968), could also be produced by plausible measurement errors in the estimates of the speeds of crash-involved vehicles, even when there is no underlying relationship between speed and crash risk. It has also been argued that Solomon's observed relationship is at least in part due to combining crashes where vehicles are making turns at intersections with other types of crashes (e.g. West and Dunn 1971). It turns out that positive correlations between speed variance and crash rates can arise when the relationship between individual risk and speed is increasing, decreasing, or U-shaped (Davis 2002a), and Balkin and Ord's inability to find, in several states, changes in fatal accidents following speed limit changes may be an artifact of the low power of their tests (Hauer 2004).

Elsewhere it has been argued (Davis 2002b, 2004a) that one way to begin resolving these issues is to treat individual crash events as being produced by (approximately) deterministic processes, but where knowledge of the details of any particular crash may be incomplete. Observed statistical regularities then result from aggregating particular types and frequencies of these individual mechanisms. In this view, a general statement like "Speeding causes crashes" is actually a claim that speeding causes (at least) a significant fraction of individual crashes. An inspiration for this approach is Snow's use of both clinical and statistical methods to identify the mechanism for transmission of cholera (Turner 1997). This approach draws on foundational treatments for causal inference developed by statisticians (e.g. Rubin 1974; Holland 1986) and

philosophers (e.g. Giere 1972), but its most important influences have been Judea Pearl's work on causality and artificial intelligence (Pearl 2000), and the use of accident reconstruction methods to study the causal effect of speed by the Road Accident Research Unit (RARU) at the University of Adelaide (McLean et al 1994; Kloeden et al, 1997, 2001, 2002). In what follows, we will show how Pearl's work can be used to give a foundation for the method used by Kloeden et al (2001, 2002), where logistic regression models were applied to case-control data in order to estimate the casual effect of hypothetical speed limit policies. Bayesian accident reconstruction methods will first be applied to individual crashes in order to compute posterior probabilities for each case vehicle's initial speed. The case vehicles will then be combined with speed measurements for vehicles not involved in crashes (controls), leading to a case-control problem with differential measurement error. The posteriors from the accident reconstructions will be used to characterize the uncertainty for each case vehicle's speed, and Markov Chain Monte Carlo (MCMC) methods will be used to compute posterior distributions for the logit model parameters. The approach will be used (1) to test whether or not a U-shaped relationship exists between speed and risk of run-off road crashes, and (2) to estimate the causal effect of strict adherence to speed limits in each of a set of fatal run-off road crashes occurring in Minnesota.

Causal Models of Crash Risk

Pearl's key notion is that of a structural model, consisting of a set of background variables, a set of endogenous variables, and for each endogenous variable, a structural equation describing how that variable changes in response to changes in the background or other endogenous variables. The dependencies in the model can be summarized using a directed acyclic graph (DAG), where the nodes of the graph represent the model's variables, while a directed arrow from node A to node B indicates that A is an argument in B's structural equation. Figure 1 shows a DAG for a generic crash, with v denoting the vehicle speed, u denoting other background variables, and y taking on the value 1 if the crash occurs, and 0 otherwise. (The meaning of the other variables will be taken up later.) The form of the structural equation relating y to v and u will of course depend on the type of crash under consideration. The variable e denotes the evidence available about the crash. If one were carrying out a reconstruction of this crash, one would begin with the evidence, then attempt to determine the structural equations relating e to v and u , and finally attempt to estimate values for v and/or u . Uncertainty

concerning the crash can be captured by specifying a probability distribution for the values taken on by the background variables, producing what Pearl calls a probabilistic causal model. The Bayesian approach to accident reconstruction then involves using the evidence to compute posterior distributions for the background variables (Davis 1999, 2004b).

Now imagine that we know (somehow) that the vehicle's speed prior to the crash was v_1 , and we are interested in determining whether or not this was a cause of the crash. Baker has defined a "causal factor" as a circumstances "contributing to a result without which that result would not have occurred" (1975, p. 274). This suggests that for the event $v=v_1$ to count as a causal factor for a crash there must exist some other plausible speed v_2 such that, other things equal, had the vehicle been travelling at speed v_2 instead of v_1 the crash would not have occurred. That is, we compare what happened to what would have happened in a counterfactual situation. Counterfactual tests of this sort are in fact a staple of accident reconstruction, and have been called "avoidance analyses" (Limpert 1989) or "sequence of events analyses" (Hicks 1989). Pearl's structural model approach can then be used to implement counterfactual tests by setting the speed to some other value v_2 , keeping all other background variables at their original values, and then solving the structural equation for y . For the majority of crashes however precise knowledge concerning the values taken on by the background variables will not be available, and so the conclusion as to whether or not $v=v_1$ was a causal factor will be to some extent uncertain. Pearl defines a probabilistic version of the notion of causal factor using the idea of probability of necessity (PN). If we let $y_{v=v_2}$ denote the value taken on by y when v is set to v_2 , probability of necessity is then defined as

$$PN(v_1, v_2) = P[y_{v=v_2}=0 \mid y=1 \ \& \ v=v_1] = \int (1-y(v_2, u)) dF(u \mid y=1, v=v_1, e) \quad (1)$$

(e.g. Pearl 2000, pp. 206, 286). In most cases, the original speed will also not be known with certainty, and a more useful measure of causal effect is what Pearl has called probability of disablement, but which we will call probability of avoidance

$$PA(v_2) = \int PN(v_1, v_2) dF(v_1 \mid y=1, e) \quad (2)$$

In essence, computing probabilities of necessity and avoidance involves first using Bayes theorem to compute posterior probabilities for the model's background variables, then setting the speed to a target value v_2 , and finally computing the probability that $y=0$ using this posterior distribution. When a complete structural model can be specified, the Twin Network method described in Balke and Pearl (1994) can be used to carry this out, by performing Bayesian updating on an augmented network where nodes have been added to reflect the counterfactual situation. Figure 1 illustrates these with nodes v^* and y^* standing for the counterfactual situation where the speed v is set to v^* . An application of this approach to vehicle/pedestrian and two-vehicle intersection crashes has been presented in (Davis 2004b).

For some crashes, although it may be possible to estimate a vehicle's initial speed it is not possible to construct a complete structural model, and so apply Balke and Pearl's Twin Network method. For example, several of the crashes described in Kloeden et al (1997) involved vehicles which ran off the road and collided with fixed objects. Using measurements of the vehicle's deformation and the skidmarks left by its tires, estimates of the vehicle's initial speed could be computed, but modeling the driver's initial loss of control was not possible. Pearl (2000) and Tian and Pearl (2000) have addressed the problem of making causal inferences when structural knowledge is limited, but where statistical information on associations between model variables is available. For our purposes, their most useful result is that when the underlying structural model is monotonic (i.e. $v_1 \geq v_2$ implies $y(v_1, u) \geq y(v_2, u)$ for all u), and v is an exogenous variable, then the probability of necessity can be expressed in terms of conditional probabilities via

$$PN(v_1, v_2) = (P[y=1|v=v_1] - P[y=1|v=v_2]) / P[y=1|v=v_1] \quad (3)$$

If the conditional probabilities appearing in equation (3) can be estimated, then estimates of probabilities of necessity and avoidance can also be computed. A variant of this approach was described in the reports by Kloeden et al (2001, 2002), where logistic regression models of the form

$$P[y=1 | v] = \exp(b_0 + b_1 v + b_2 v^2) / (1 + \exp(b_0 + b_1 v + b_2 v^2)) \quad (4)$$

were fit to data from case-control studies. Substituting estimates of the conditional probabilities obtained from (4) into (3) produced estimates of probability of necessity, which the authors in turn used to predict the effects of various hypothetical speed management scenarios. As noted above however, in practice there will be uncertainty concerning the vehicle's actual speed (v_1), and when statistical methods have been used to fit a model such as (4), there will also be uncertainty concerning the values taken on by the model's parameters. Both these sources of uncertainty should be accounted for when computing probabilities of necessity and avoidance.

As noted above, Kloeden et al. (2001, 2002) used the logit model to relate speed to crash risk, and before proceeding it may be helpful to indicate how such a model might arise. Let X denote the distance an arbitrary driver travels until being involved in a run-off road crash. Although the actual crash event will be deterministic, our ignorance of particular details means it will not be possible to identify in advance which drivers will be involved in crashes. Arguably, the simplest model which allows crash probability to depend on distance traveled and on speed is a proportional hazards model, with hazard function

$$h(v) = \lambda \exp(g(v)) \quad (5)$$

where $g(v)$ is some function of speed, v . The probability of being in a crash while traversing a section of road of length x is then

$$P[X \leq x | v] = 1 - P[X > x | v] = 1 - e^{-(\lambda x) \exp(g(v))} \quad (6)$$

If run-off road crashes are rare (i.e. $0 < \lambda x \ll 1$) (6) can be approximated as

$$(\lambda x) \exp(g(v)) e^{-(\lambda x) \exp(g(v))} \quad (7)$$

which is the probability that the value 1 is taken on by a Poisson random variable Y with expected value

$$E[Y] = (\lambda x) \exp(g(v)) \quad (8)$$

If we then condition on $Y = 0$ or 1 (so that no one can crash more than once), we get

$$P[Y = 1|v] \approx \left(\frac{\exp(\log(Ix) + g(v))}{1 + \exp(\log(Ix) + g(v))} \right) \quad (9)$$

Letting $b_0 = \log(\lambda x)$ then leads to a general form for equation (4), with the prospective crash probability as a logit function of speed.

Case-Control Analyses

Two sources provided the data on run-off road crashes used in this study. The first was the case-control study conducted by the Road Accident Research Unit (RARU) at the University of Adelaide (Kloeden et al, 1997). The RARU reported data for 151 case vehicles involved in serious or fatal crashes on roads with 60 km/h speed limits. For each case vehicle four control vehicles were then selected by randomly sampling vehicles using the crash site at times when conditions were similar to those for when the crash occurred. Control speeds were measured using laser speed guns, while the case speeds were estimated using accident reconstruction techniques. 14 of the 151 cases were single vehicle run-off road crashes and of these, eight involved collisions with objects where it was possible to measure the deformation (crush) suffered by the vehicles, while two produced measurable yaw-marks near the point where the driver lost control of the vehicle.

As has been argued elsewhere (Davis 2004b), accident reconstructions are subject to nontrivial uncertainties, and the probability calculus can be used as a logic for uncertain reasoning. Our general approach to estimating case vehicle speeds for the RARU data was to develop probabilistic versions of the deterministic methods used by the RARU researchers. This was done by supplementing their measurements with training data, and then treating the case vehicle speeds as missing values to be estimated. For the fixed-object crashes, the training sample was the 19 staged collisions performed by the National Highway Traffic Safety Administration (NHTSA), and reported in Nystrom and Kost (1992). For the yaw-mark crashes, the training sample was 40 measured speeds and yaw radii tabulated in Semon (1995).

First, for the fixed-object crashes, the following variant of Nystrom and Kost's (1992) model was used to relate measured crush to impact speed

$$c = (v - v_0) / (b_0 + b_1 * w) + \epsilon \quad (10)$$

Where

c = measured crush

v = impact speed

v_0 = highest impact speed producing no crush (taken to be 5 mph)

w = vehicle weight

b_0, b_1 = coefficients to be estimated

ϵ = error.

The error term ϵ allows for differences between measured and predicted crush, and was assumed to be normally distributed with mean equal to 0 and unknown variance σ^2 . Since six of the fixed object crash vehicles left measurable skidmarks prior to collision, it was also necessary to account for speed lost while skidding. Treating the measured skidmark as an error prone observation, its expected value was computed using

$$\bar{s} = \frac{Lv_t^2 - v^2}{2\mu g} \quad (11)$$

where

v_t = denotes the vehicles initial speed

L = fraction of kinetic energy not lost between the initiation of braking and the point where the skidmark begins (taken by the RARU to be 0.8),

μ = coefficient of tire/pavement friction.

Garrot and Guenther (1982) conducted an extensive comparison of measured versus theoretical skidmarks, the differences for which showed a co-efficient of variation approximately equal to 0.11. Following the methods used in Davis (2004b), the measured skid mark was assumed to have a log normal distribution, with the mean equal to the natural log of the theoretical length, and a normal variance of 0.01. This gives a coefficient of variation for the measurement error equal to approximately 0.1.

In addition to the likelihood functions for the measured crush and skid lengths, Bayesian analysis requires prior distributions for all unknown quantities. For estimating the speeds of the fixed-object crash vehicles, the following prior distributions were used:

$$\begin{aligned}
 b_0 &\sim \text{Normal}(0, 10^6) \\
 b_1 &\sim \text{Normal}(0, 10^6) \\
 \sigma^2 &\sim \text{Inverse Gamma}(0.001, 0.001) \\
 v \text{ and } v_t &\sim \text{Normal}(\alpha, \beta), \\
 \alpha &\sim \text{Normal}(40 \text{ mph}, 10^6) \\
 \beta &\sim \text{Inverse Gamma}(0.001, 0.001), \\
 \mu &\sim \text{Uniform}(0.45, 1.0).
 \end{aligned}$$

With the exception of μ , all these are commonly-used "uninformative" priors. For μ the lower bound characterizes a dry, travel polished asphalt pavement while the upper bound characterized a dry, new concrete pavement (Fricke 1990).

Compared to the fixed-object crash model, the yaw-mark model is simpler, but is still based on the principle of imputing unknown speeds. Treating the radius of the yaw-mark as an error-prone measurement caused by the speed, the standard critical speed formula leads to

$$r = \frac{v^2}{\mu g} + \epsilon \quad (12)$$

where

r = measured yaw radius,

v = vehicle's speed

μ = friction coefficient

ϵ = measurement error.

The error term ϵ is assumed to be normally distributed with mean equal to zero, and unknown variance σ^2 . As stated earlier, the 40 experimental tests having information on observed speed and radius of curvature were used as a training data set for estimating the value of μ . The two

RARU cases were then treated as similar to the 40 tests but with missing speeds. The following priors were used:

$$v \sim \text{Normal}(\alpha, \beta),$$

$$\alpha \sim \text{Normal}(50, 10^6),$$

$$\beta \sim \text{Inverse Gamma}(0.001, 0.001),$$

$$\mu \sim \text{Uniform}(0.45, 1).$$

Posterior distributions for the case vehicle speeds were then computed using the Markov Chain Monte Carlo program WinBUGS (Spiegelhalter et al. 2000), and details of the WinBUGS models have been given by Davis and Davuluri (2002). Table 1 summarizes the case and control data for the ten RARU crashes.

The second data set was taken from a set of 46 fatal crashes occurring on Minnesota state highways between January 1, 1997 and June 30, 2000. These were all fatal crashes reported during this time period which happened near one of the locations where the Minnesota Dept. of Transportation collected automatic vehicle speed data, and for which crash investigation data could be obtained from the Minnesota State Patrol. The automatic speed data were used to produce control speeds by randomly sampling from the speed measurements taken during an hour when conditions were judged to be similar to those present when the crash occurred. Of the 46, 22 involved loss of control and running off the road, with 9 resulting in collisions with other vehicles, 10 resulting in rollover, and three resulting in collisions with fixed objects. For 10 of these it was possible to use accident reconstruction methods to estimate initial speeds. For two, initial speeds could be estimated from measured yaw-marks using the method described above, while for five the tripped rollover model described in Cooperrider et al. (1990) and Martinez and Schlueter (1996) was adapted to estimate initial speeds. For three additional crashes straightforward application of either the yaw-mark mark method or the rollover model was not possible, but special features of these crashes still permitted estimates of initial speeds. Table 2 summarizes the case and control data for the 10 Minnesota run-off road crashes.

As indicated in the introduction, one of the controversial issues in the debate on speed versus crash risk concerns whether or not crash risk is a U-shaped function of speed, with vehicles travelling at atypically low and high speeds having increased crash risk. Whether or not

this is the case is especially important if we hope to use equation (3) to estimate probabilities of necessity and avoidance, since a U-shaped relationship can render the monotonicity assumption problematic. If we accept that the role of speed may vary for different types of crashes, depending on the operative processes and circumstances, then appropriate tests for the possibility of a U-shaped relationship should be carried out using data disaggregated by crash type. Otherwise, there is the possibility of obscuring the speed effect by combining processes where speed is and is not causal, or of producing an apparent U-shaped relationship by mixing situations where high speed is causal with other situations where low speed is causal. Disaggregating by type of crash reduces sample sizes however, but as argued earlier, a simple proportional hazards model relating speed, distance traveled and crash risk leads to a prospective logit model of the form

$$P[\text{crash} | v] = \exp(b_0 + g(v, \mathbf{b})) / (1 + \exp(b_0 + g(v, \mathbf{b}))) \quad (13)$$

The parameter b_0 can be taken as summarizing those features shared by the cases and controls at a given location, while the function $g(v, \mathbf{b})$ describes how crash risk varies with speed and a vector of parameters \mathbf{b} . Assuming first that both the case and the control speeds are known without error, the fact that the cases and controls are matched by location means that a matched case-control approach can be used, which leads to a likelihood contribution from site k of the form

$$P[y_{k,0}=1, y_{k,j}=0, j=1, \dots, m] = \exp(g(v_{k,0}, \mathbf{b})) / (\sum_{j=0}^m \exp(g(v_{k,j}, \mathbf{b}))) \quad (14)$$

(e.g. Hosmer and Lemeshow 2000). Here $v_{k,0}$ denotes the case vehicle speed at site k , while $v_{k,j}$ $j > 0$ denotes the corresponding speeds for the control vehicles. The likelihood function obtained as the product of (14) over all cases would then provide the basis for either a Bayesian or a frequentist approach to estimation. If the case speeds are not known exactly then they are additional quantities to be estimated. If the prior distribution for the case speed at location k is $f(v_{k,0})$, Bayesian estimation is in principle straightforward. Alternatively, we could treat $v_{k,0}$ as a latent variable generated in a hierarchical model, and then compute a marginal likelihood by integrating it out of the joint likelihood

$$P[y_{k,0}=1, y_{k,j}=0, j=1\dots m] = \int [\exp(g(v_{k,0}, \mathbf{b})) / \int \exp(g(v_{k,j}, \mathbf{b}))] f(v_{k,0}) dv_{k,0} \quad (15)$$

Frequentist methods could then be applied to the resulting marginal likelihood function.

In the simplest case, a test as to whether or not the risk function is U-shaped can be carried out by comparing a quadratic form for the function $g(\cdot)$

$$g(v, \mathbf{b}) = b_1(v-v_0) + b_2(v-v_0)^2 \quad (16)$$

to a linear form

$$g(v, \mathbf{b}) = b_1(v-v_0). \quad (17)$$

Looking first at the Minnesota crashes, Bayesian estimation of the linear model (17) was accomplished using the Markov Chain Monte Carlo routine WinBUGS, with v_0 being set to the average speed for the control population at each site. When we attempted to estimate the quadratic model however, the MCMC routine was unstable, producing chains with poor mixing properties, and the simulated values for b_1 and b_2 tended to be highly correlated with each other. At least one reason for this can be seen in Figure 2, which shows a contour plot of the marginal log-likelihood as a function of b_1 and b_2 . The narrow ridge-shape of the log-likelihood function indicates that the data tend to be uninformative about b_2 over a range of values including zero, while the relatively flat region in the upper right-hand corner suggests that the log-likelihood function is not universally log-concave. For the relative risk function to be U-shaped however b_2 must be positive so additional MCMC runs, with b_2 constrained to be non-negative, were conducted. Table 3 displays posterior estimation summaries for the linear and constrained quadratic models as fit to the Minnesota run-off road data.

The results in Table 3 indicate that the linear and constrained quadratic models provide roughly equivalent fits to the data. The posterior deviances have similar distributions, and the deviance information criteria (DIC) (Spiegelhalter et al. 2002) are approximately equal. As a check, maximum likelihood applied to the marginal likelihood functions was also used to fit the linear and quadratic models, producing maximum likelihood estimates $b_1=0.102$, $b_2=0.005$ for

the quadratic model and $b_1=0.159$ for the linear model. A likelihood ratio test of the hypothesis $b_2=0$ yielded $X^2=0.521$, $p>0.53$.

Since a convex parabola and a straight line have different implications for the relationship between speed and crash risk, the rough equivalence of these two models may seem contradictory. The contradiction is resolved by looking at the point of minimum risk, which for the quadratic models occurs at a speed equal to $v_0-b_1/(2b_2)$. Substituting the posterior means for b_1 and b_2 into this expression reveals that for the quadratic model minimum risk occurs at a speed between 10 and 11 mph *below* the average for the controls. Since most (97 out of 100) of the control speeds in Table 2 are above this value, what appears to be happening is that the quadratic model achieves parity with the linear model simply by being monotonically increasing over the range of the available data. More particularly, the results in Table 3 do not appear to support the earlier findings, that minimum risk tends to occur near the mean or median of the control speeds. Similar effects were described in the RARU reports by Kloeden et al. (2001, 2002).

For the RARU data, the Markov Chain Monte Carlo simulations showed poor mixing even when attempting to fit the linear model, and again studying a plot of the marginal log-likelihood function is informative. Figure 3 shows this, and the distinctive feature here is how the log-likelihood flattens out for higher values of b_1 , indicating that the data contain no information about how large b_1 is. The reason for this is found by inspecting Table 1, where it can be seen that most controls speeds are below the posterior mean speed for their corresponding cases, and none are greater than two standard deviations above this mean. To work around this problem, it was decided to supplement the controls speeds with those obtained for all other controls in the RARU study. We considered this acceptable since, unlike the Minnesota data, the RARU sites tended to be similar in terms of road and weather conditions. This produced an unmatched case control study with ten cases and 604 controls. As with the Minnesota data, WinBUGS was used to compute Bayesian estimates of the parameters for the linear and quadratic models, along with measures of goodness of fit, and these are displayed in Table 4. Again, the linear and quadratic models appear to fit the data about equally well, and again as a check maximum likelihood estimation was applied to the marginal likelihood. This produced estimates $b_0=-8.3$, $b_1=0.63$, and $b_2=-0.006$ for the quadratic model and $b_0=-7.7$, $b_1=0.49$ for the linear model. A likelihood ratio test of the hypothesis $H_0: b_2=0$ yielded an estimated $X^2=0.19$, $p=0.338$.

To summarize, for both the Minnesota crashes and the RARU's crashes it appears that at least over a typical range of speeds, risk of being in a serious or fatal run-off road crash increases as speed increases. If in fact there are situations where low speeds are dangerous these likely involve processes or conditions that are different from those characterizing the crashes in our samples.

Individual Probabilities of Avoidance

The second issue identified in the Introduction concerns the relationship between speed and fatal crashes, and in particular the relationship to the 55 mph National Maximum Speed Limit. The approach taken here was to compute, for each of the 10 Minnesota cases, the probability of avoidance as a function of speed using equation (3) and the linear version of the logit model. Before proceeding however we ought to consider whether or not this model is at least reasonably consistent with our data. Traditionally, a model's goodness-of-fit has been assessed by computing what the model would predict for the observed data and comparing these predictions to what was actually observed. The most common device for making these comparisons has been the computed residuals, which are the differences between the predictions and the observations. For matched case-control studies the conditional likelihood function (14) can be interpreted as giving the probability that the observed case would be selected as the case when compared to the controls, and the Pearson residual for case k is

$$s_k = (1-p_{k0})/p_{k0} \quad (18)$$

(Bedrick and Hill 1996). Alternatively, we could use the deviance residual (Spiegelhalter et al. 2002)

$$d_k = -2\log(p_{k0}) \quad (19)$$

which is the contribution made by case k to the overall deviance. Gelman et al (2004) recommend that, for Bayesian analyses, goodness-of-fit assessments can be carried out by comparing a "test statistic" to the posterior predictive distribution for that statistic. This distribution can be simulated by first drawing from the posterior distribution of the model's

parameters, then generating a simulated data replication using these parameters values, and finally computing the value of the test statistic using the replicated data. For the Minnesota data we used two test statistics, both of which reflect the ability of the linear logit model to correctly select the actual case from the set of cases and controls.. The first is the difference between the deviance residual computed for the actual case and that computed using the replicated case selection. This will equal zero if the actual case was selected as the case in all the replications, and differ from zero otherwise. The second measure was simply the proportion of replications where each case or control was selected as the 'case.' Since the case speeds are not known with certainty it would be unreasonable to expect the actual case to be selected in all replications, but if the model is reasonable there should be a tendency to select 'most' of the actual cases, and to not ignore any actual case. Using WinBUGS, 15,000 posterior replications were generated for each set of the 10 cases with matched controls, by running three Markov chains for 50,000 iterations and saving every 10th iteration for the analysis. Table 5 summarizes these results.

The first column of Table 5 gives the crash number, while the next three columns give quantiles for the predictive distribution of the difference between observed and replicated deviance residuals. If the model is reasonable one would expect the value of zero to be contained between the 2.5 percentile and the 97.5 percentile, and this happens with all 10 case/control sets. The final five columns of Table 5 give the fraction of replications where the actual case was selected as the replicated case, and also the fractions for any controls which were selected more frequently than the actual case. For five of the case/control sets, 1,2, 5, 6, and 9, the actual case was selected most often, while for the remaining case/control sets the actual case did no worse than being fifth out of eleven. These results appear to make sense, since Table 2 shows that for crashes 1,2,5,6, and 9 the control speeds tended to fall in the left-hand tails of the priors for the case speeds, while for the remaining crashes the control speeds are more evenly scattered over the support of the priors. Overall though, it seems reasonable to conclude the our logit model is consistent with these data.

For the linear logit model, the probability of necessity for lowering speed from v_1 to speed v_2 can, when crashes tend to be rare, be approximated as

$$PN(v_1, v_2, b_1) = 1 - P[y=1|v=v_2] / P[y=1|v=v_1] \approx 1 - \exp(b_1(v_2 - v_1)) \quad (20)$$

For each of the 10 Minnesota crashes, probabilities of avoidance were computed using Markov Chain Monte Carlo simulation by treating v_2 as fixed, evaluating equation (20) for each simulated value of v_1 and b_1 , and then averaging over the iterations. Carrying this out over a range of values for v_2 then gave probability of avoidance as a function of target speed. Figures 4-13 display these results for the 10 crashes, along with the prior distributions used for the case speeds, and the distribution of speeds in the populations from which the control speeds were sampled.

Figure 4 displays results for a run-off road crash which occurred on a rural Interstate highway during snowy conditions. The posted speed limit was 70 mph, but the average speed during the time prior to the crash was about 57 mph, with most drivers travelling at speeds lower than the posted limit. The posterior mean for the case vehicle was about 70 mph, and the driver was probably travelling faster than most other drivers. Looking at the probability of avoidance (PA) curve, if the crash-involved driver had been travelling at 30 mph it is highly probable that a crash with a fatal outcome would have been avoided, and this probability declines as speed increases. At 55 mph the probability of avoidance equals about 0.80, while at the posted speed limit of 70 mph it equals about 0.23. If the crash-involved driver had been travelling at the average speed for the control population (57 mph) the probability of avoidance equals about 0.75. So for this case we would be inclined to conclude that it is not clear whether or not the driver was exceeding the 70 mph speed limit, but that the speed of the crash-involved vehicle was probably less than the speed limit plus the 5 mph tolerance most people accept. Since the probability of avoidance for the posted limit is only about 0.23 it does not appear that exceeding the speed limit should be considered a cause of this crash. However, if we take the speeds of the control population as indicative of what was reasonable and prudent for the prevailing conditions, it appears that the crash-involved driver was probably travelling too fast for conditions and, since the probability of avoidance for the average control speed equals 0.78, we can conclude that excess speed was probably a cause of the fatal outcome.

As another example, consider crash number 7, the results of which are displayed in Figure 10. Here the driver attempted to avoid a deer crossing a rural Interstate highway but lost control of his car, rolled over in the median and was killed. Looking at figure 10, the posterior mean for the crash-involved vehicle's speed was about 75 mph, and although the posted speed limit was 70 mph it does not appear that this driver's speed was atypical when compared to the

speeds of other drivers using the road at about the same time. The probability of avoidance for an initial speed equal to the posted limit was about 0.46, that for an initial speed equal to the average control speed (74.5 mph) was about 0.28, but the probability of avoidance for an initial speed of 55 mph was about 0.89. So for this crash it does not appear that the driver was going excessively fast either compared to the posted limit or to what other drivers were doing. But in a hypothetical world where the old 55 mph speed limit was strictly obeyed it is probable that, other things equal, this crash, or at least its fatal outcome, would not have occurred.

Table 6 summarizes the results for the 10 Minnesota crashes. The first thing to note is that the role of speeding as a causal factor ranges from about 0.09 for crash 8 to about 0.87 for crash 4. A similar range occurs when we replace the posted limit with the average control speed, but when we consider the effect of travelling at 55 mph it appears that at least the fatal outcome of most these crashes could have been avoided. In Davis (2002) it was shown how the expected crash reductions could be estimated by summing the probabilities of avoidance for a set of individual crashes. Summing over the rows of Table 6 we find that strict adherence to posted limits would have avoided about 4.9 out of the 10 fatalities, while strict adherence to a 55 mph limit would have avoided about 8.3 out of 10 fatalities.

Summary and Conclusion

In the Introduction we identified two salient issues with regard to the role of speed in road crashes, one concerning the existence of a U-shaped relationship between speed crash risk, and the other concerning the relationship between the 55 mph National Maximum Speed Limit and the occurrence of fatal crashes. Despite extensive research a clear resolution of these issues has yet to be achieved. The view adopted here is that at least some of the current confusion may result from (1) aggregating crashes which are caused by fundamentally different processes and (2) failure to account for uncertainty in an analysis. In this report, we have shown how pioneering work conducted at the Road Accident Research Unit can be combined with recent advances in computation for Bayesian statistical models in order to apply case control methods to studies with relatively small numbers of cases. Applying the method to two case-control samples with 10 run-off road crashes each, we found that these data did not support the existence of a U-shaped relationship between speed and crash risk, although risk tended to increase as a function of speed. Applying Bayesian methods to compute the probability that a fatal outcome

would have been avoided as a function of counterfactual initial speeds, we found that strict adherence to posted speed limits would have prevented about 5 out of the 10 fatal crashes we investigated, while strict adherence to a 55 mph limit would have prevented most of them.

One implication of this study appears to be that, as common sense tells us, high speed in and of itself is not sufficient to cause a fatal crash. For most of the 10 fatal Minnesota crashes other drivers were observed traveling the same road under the same conditions, as fast or faster than the crash-involved drivers, without being involved in a fatal crash. A reasonable interpretation would be that some type of triggering event, which places the driver in a crash-avoiding situation, also appears necessary. This is consistent with Hauer's pyramid (1997, p. 19), which distinguishes normal driving from conflict situations, and from those situations resulting in crashes. Study of the Minnesota crash reports revealed events such as the appearance of a deer in the driver's path, the merging of a slower-moving vehicle into the driver's lane, driver distraction resulting in the need to avoid a rear-ending collision, and loss of control following the driver's turning to interact with a passenger in the back seat.

The presence or absence of a crash-avoiding situation is allowed for in Figure 1 by the variable z , which in the simplest situations takes on the value 1 if a crash-avoidance situation arises and 0 otherwise. In Figure 1 z is depicted as being exogenous, that is independent of the values of v and u , and for some types of crashes this may be plausible. For example it is not unreasonable to assume that whether or not a deer attempts to cross in front of a vehicle is approximately independent of what the driver has done. For other crash types however this independence assumption may not be justified. For instance, there is some reason to suspect that gap selection errors by left-turning drivers can increase as the speed of the oncoming vehicle increases. Unfortunately, our ability to model how such crash avoidance situations arise does not appear to be as well-developed as our ability to model what happens once the crash the crash sequence has started.

Table 1. Posterior Means and Standard Deviations for Case Speeds, and Control Speeds, for 10 RARU Run-off Road Crashes. All Speeds are in mph.

Crash #	Case Speeds		Control Speeds			
	Mean	Stan. Dev.	1	2	3	4
1	43.4	4.1	34	35	38	38
2	41.5	3.9	27	30	34	38
3	48.3	4.3	30	34	35	39
4	53.2	3.3	32	34	37	42
5	54.8	4.5	37	43	38	30
6	86.9	9.0	37	37	38	38
7	41.5	2.4	33	34	38	43
8	38.7	2.3	31	40	40	42
9	59.2	3.4	39	39	42	43
10	62.5	3.2	34	38	39	43

Table 2. Posterior Means and Standard Deviations for Case Speeds, and Control Speeds, for 10 Minnesota Run-off Road Crashes. All Speeds are in mph.

Crash #	Case Speeds		Control Speeds									
	Mean	St.Dev	1	2	3	4	5	6	7	8	9	10
1	69.8	5.6	58	58	66	50	56	65	60	70	72	54
2	71.0	5.0	57	53	38	55	65	51	59	61	55	56
3	71.4	4.4	69	75	82	77	75	82	78	70	77	71
4	81.4	2.1	69	75	89	70	75	77	87	65	77	71
5	59.8	3.6	52	60	56	55	56	56	58	53	54	62
6	80.8	2.1	69	78	71	82	80	70	69	70	70	78
7	74.5	5.0	85	78	85	71	83	72	85	71	63	79
8	67.7	4.0	69	68	65	56	72	75	75	69	70	70
9	80.4	1.7	77	81	78	77	72	75	74	80	70	76
10	73.3	4.0	73	73	71	75	80	82	78	70	87	77

Table 3. Bayesian Estimation Results for Linear and Constrained Quadratic Models: Minnesota Data

Variable	Linear Model			(Constrained) Quadratic Model		
	Mean	2.5%-ile	97.5%-ile	Mean	2.5%-ile	97.7%-ile
b ₁	0.19	0.03	0.46	-.014	-0.013	0.381
b ₂	--	--	--	0.006	0.0003	0.016
Deviance	40.3	29.5	47.9	40.6	230.8	48.5
DIC	41.1			40.3		

Table 4. Bayesian Estimation Results for Linear and Quadratic Models: RARU Data

Variable	Linear Model			Quadratic Model		
	Mean	2.5%-ile	97.5%-ile	Mean	2.5%-ile	97.7%-ile
b ₀	-7.35	-10.64	-5.29	-7.66	-11.72	-5.36
b ₁	0.54	0.33	0.84	0.55	0.15	1.11
b ₂	--	--	--	0.018	-0.018	0.025
deviance	36.33	26.0	48.0	36.3	25.6	48.5
DIC	40.0			39.6		

Table 5. Goodnes of Fit Measure for Minnesota Case-Control Study

Crash Number	Quantiles for Residual Differences			Probability of Being the Case				
	2.5%	50%	97.5%					
1	-2.3	0	5.8	.35				
2	0	0	7.0	.55				
3	-6.7	-1.2	2.5	.24	.22	.10	.08	
4	-5.6	-1.3	3.7	.39	.25	.11		
5	-1.9	0.3	4.8	.23				
6	-1.3	0.3	4.9	.23				
7	-6.2	-1.2	3.1	.21	.19	.18	.13	.10
8	-4.3	-0.2	3.2	.19	.19	.12		
9	-0.9	0.5	4.0	.18				
10	-8.5	-1.7	2.2	.39	.15	.072	.070	

Table 6. Probability of Avoidance Summary for Minnesota Case-Control Study

Crash Number	Posted Limit	Mean Case Speed	Mean Control Speed	Posted Limit PA	Control Mean PA	55 mph PA
1	70	69.8	57.7	0.23	0.75	0.80
2	65	71.0	56.2	0.51	0.82	0.84
3	70	71.4	74.2	0.28	0.17	0.85
4	65	81.4	73.0	0.87	0.64	0.95
5	55	69.2	58.3	0.47	0.27	0.47
6	70	80.8	74.2	0.80	0.54	0.94
7	70	74.5	74.5	0.46	0.28	0.89
8	70	67.7	77.4	0.09	0.02	0.77
9	70	80.4	74.4	0.79	0.50	0.94
10	70	73.3	75.8	0.40	0.17	0.88

Figure 1. Directed Acyclic Graph Model for a 'Generic' Crash.

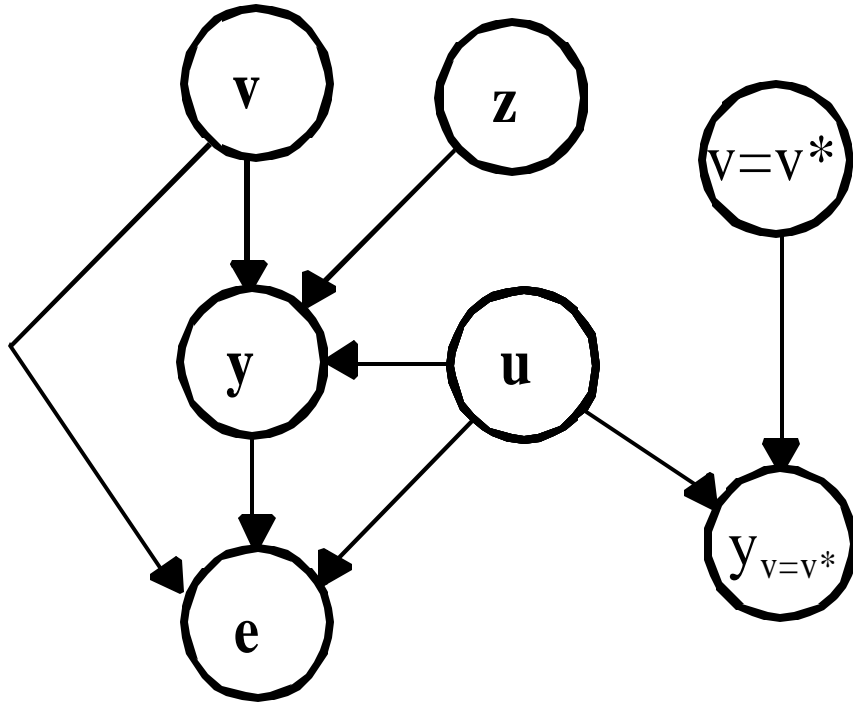
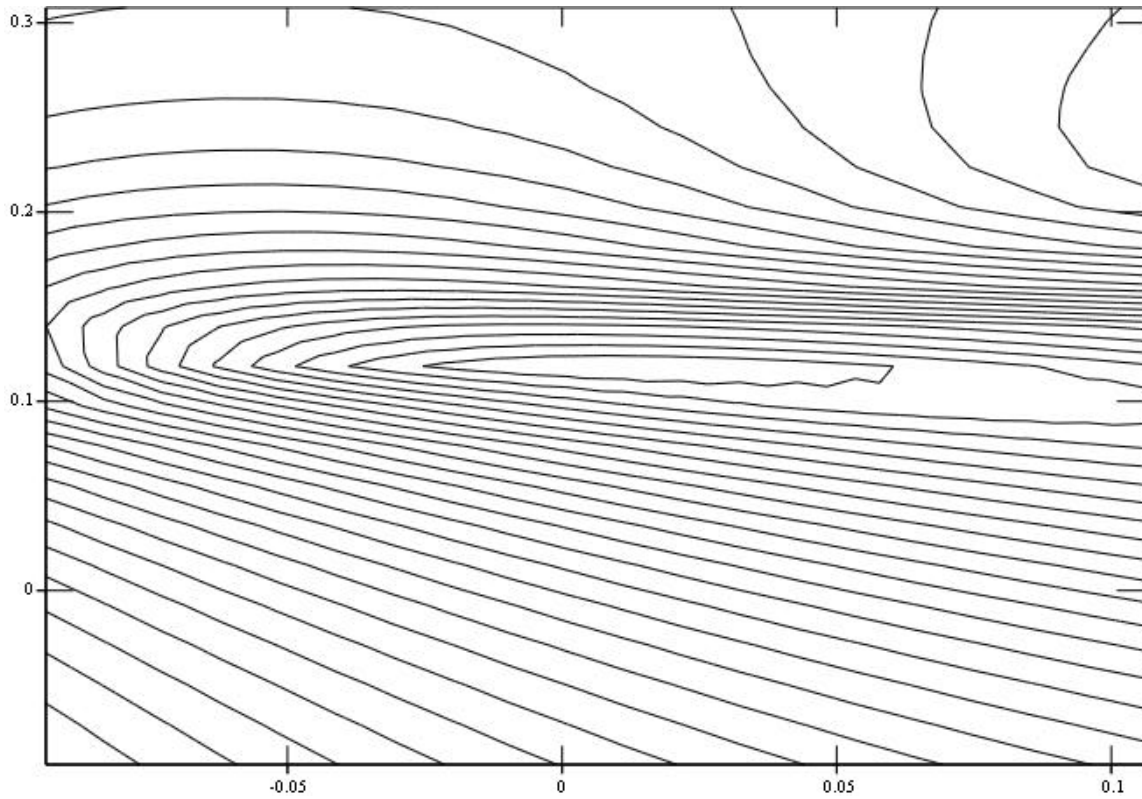


Figure 2. Contour Plot of Log Likelihood Function of Quadratic Model Fit to the Minnesota Data. b_1 is on the y-axis and b_2 is on the x-axis.



L

Figure 3. Matched Case Control Log likelihood as a Function of b_1 for Linear Model Fit to RARU Data.

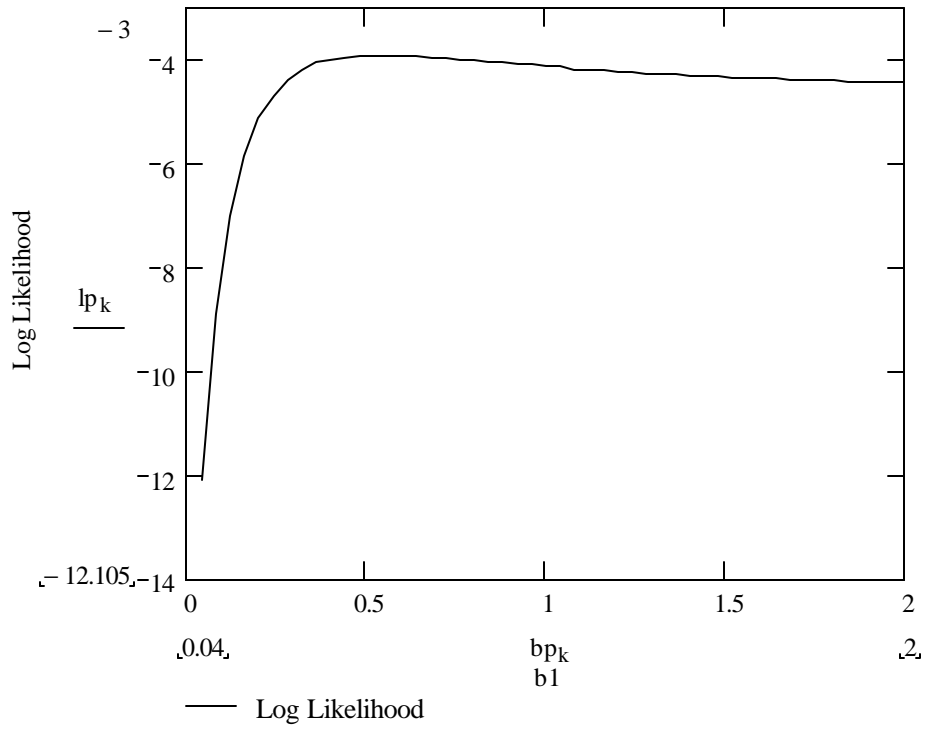


Figure 4. Case Speed Posterior, Distribution of Control Speed Population, and Probability of Avoidance for Minnesota Crash #1. Interstate Highway under Snowy Conditions.

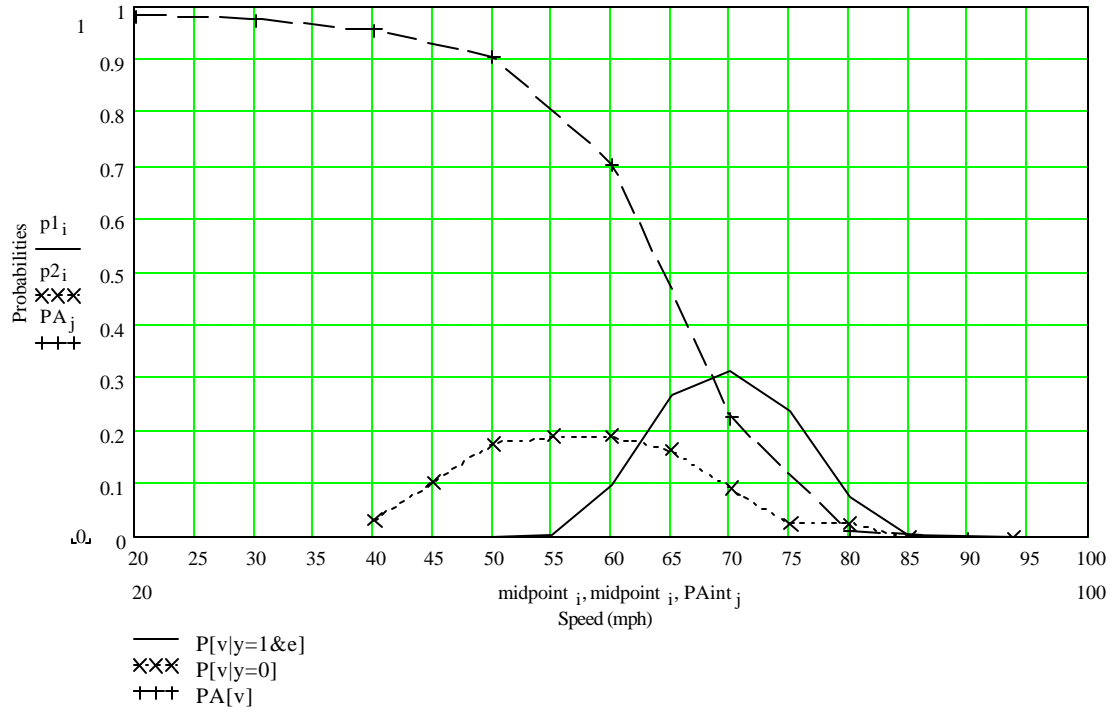


Figure 5. Case Speed Posterior, Distribution of Control Speed Population, and Probability of Avoidance for Minnesota Crash #2. Divided Highway under Snowy Conditions.

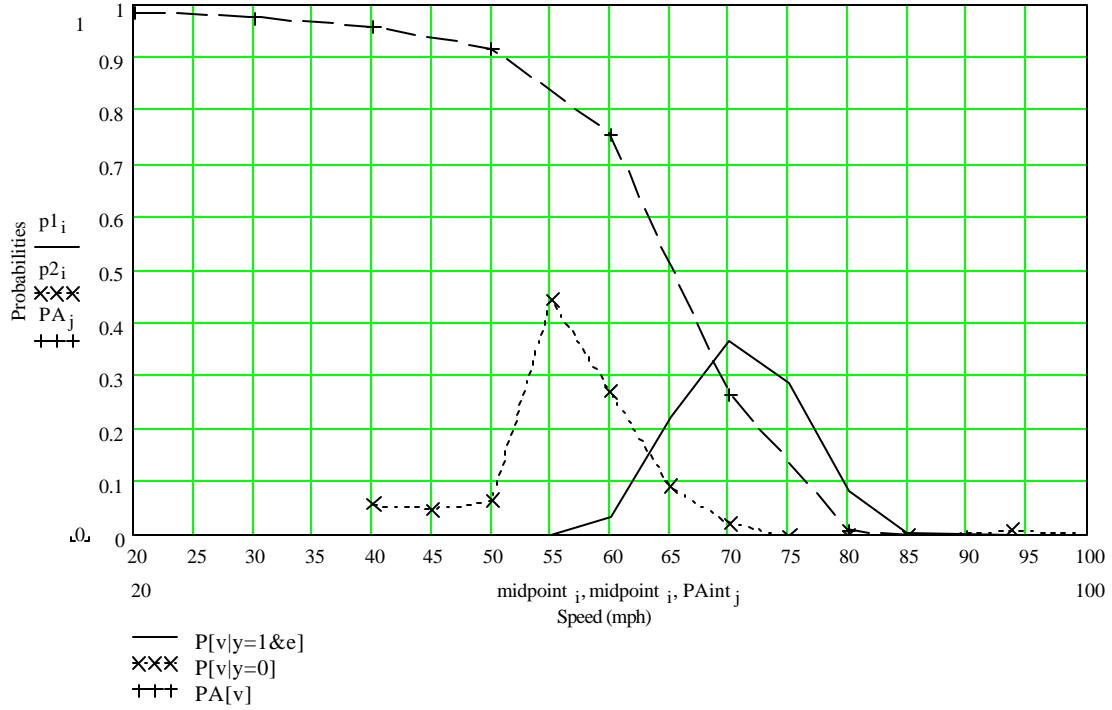


Figure 6. Case Speed Posterior, Distribution of Control Speed Population, and Probability of Avoidance for Minnesota Crash #3. Interstate Highway under Wet Conditions.

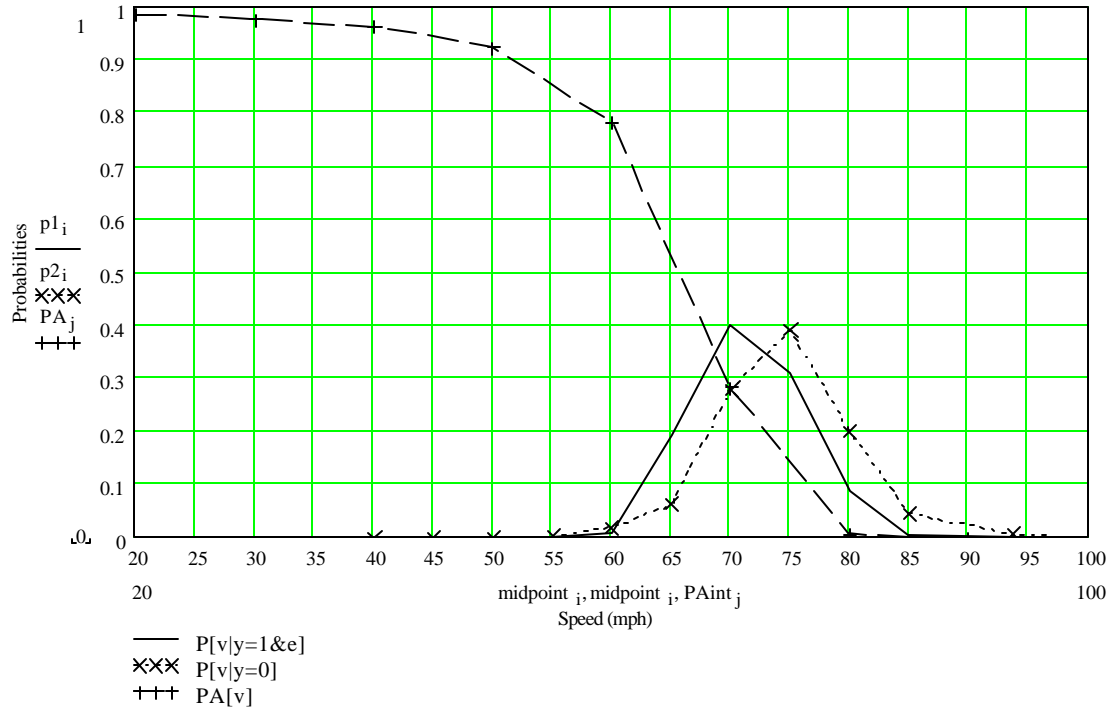


Figure 7. Case Speed Posterior, Distribution of Control Speed Population, and Probability of Avoidance for Minnesota Crash #4. Interstate Highway under Dry Conditions.

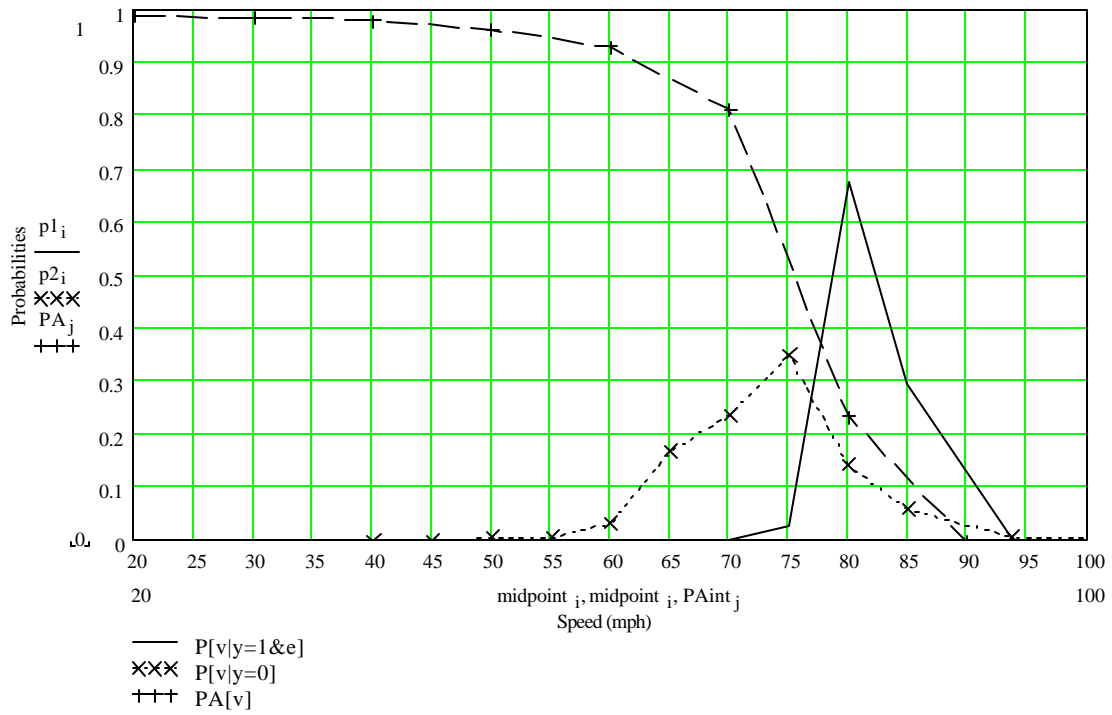


Figure 8. Case Speed Posterior, Distribution of Control Speed Population, and Probability of Avoidance for Minnesota Crash #5. 2-lane Highway under Dry Conditions.

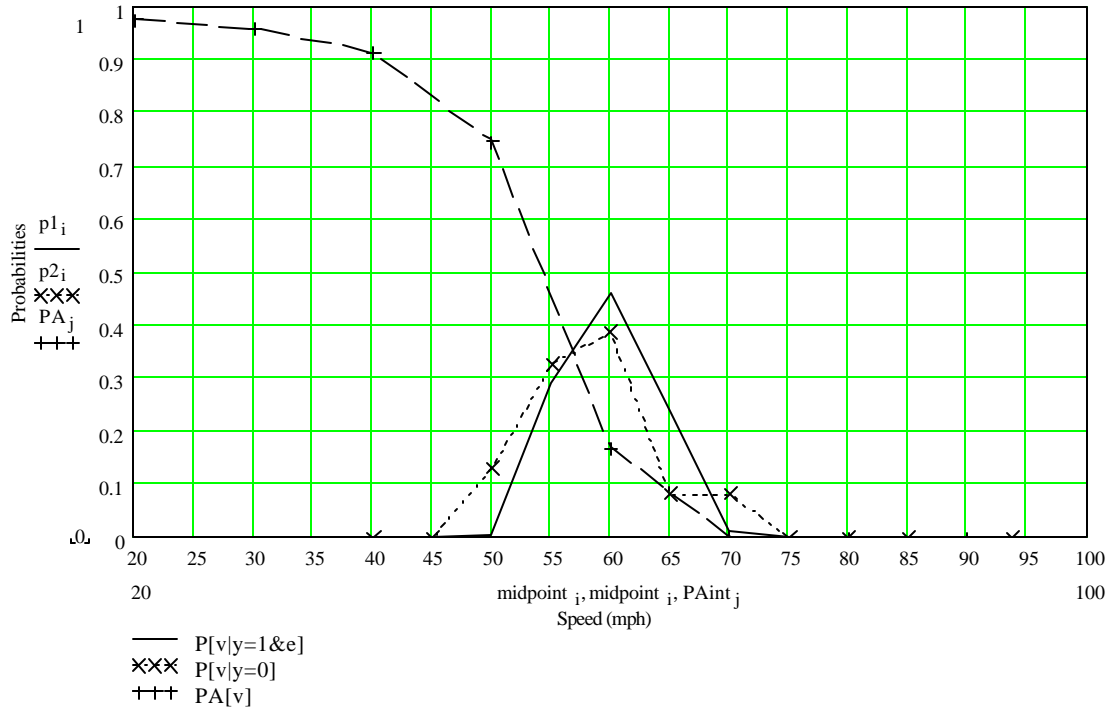


Figure 9. Case Speed Posterior, Distribution of Control Speed Population, and Probability of Avoidance for Minnesota Crash #6. Interstate Highway under dry Conditions.

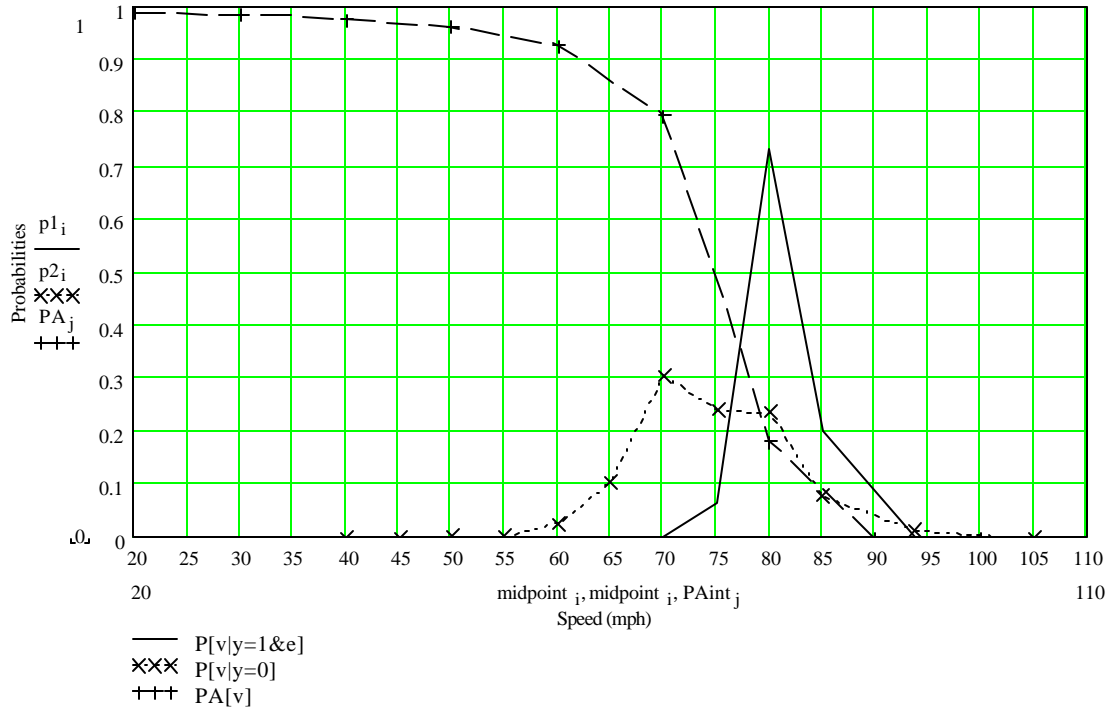


Figure 10. Case Speed Posterior, Distribution of Control Speed Population, and Probability of Avoidance for Minnesota Crash #7. Interstate Highway under Dry Conditions.

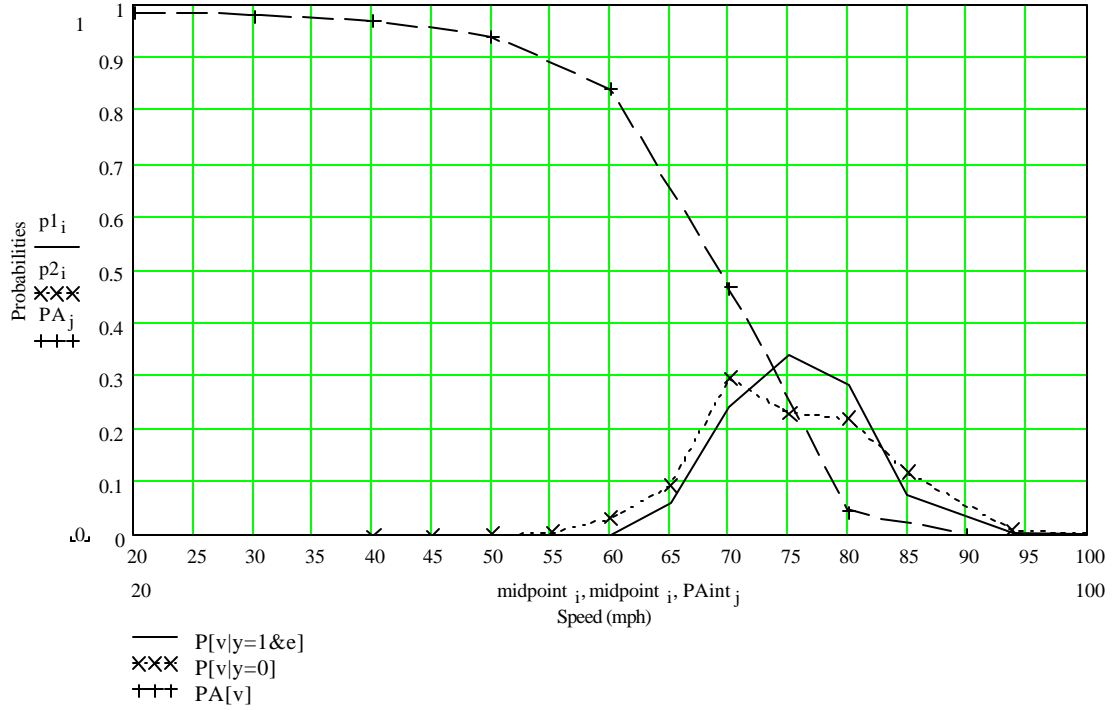


Figure 11. Case Speed Posterior, Distribution of Control Speed Population, and Probability of Avoidance for Minnesota Crash #8. Interstate Highway under Dry Conditions.

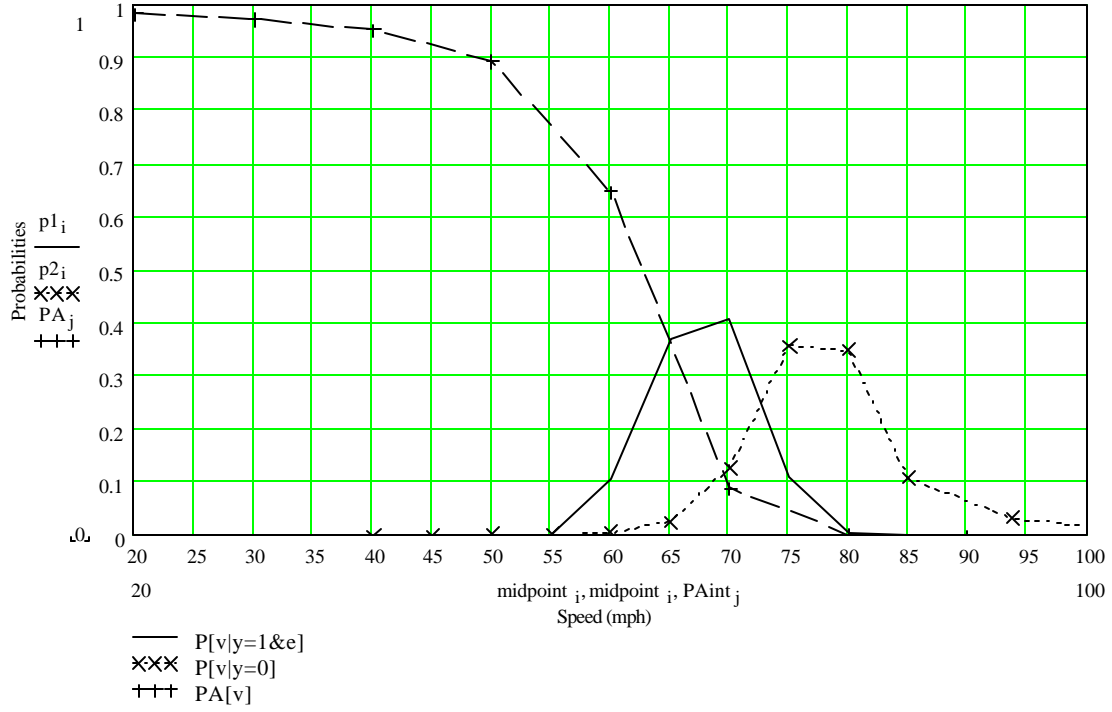


Figure 12. Case Speed Posterior, Distribution of Control Speed Population, and Probability of Avoidance for Minnesota Crash #9. Interstate Highway under Dry Conditions.

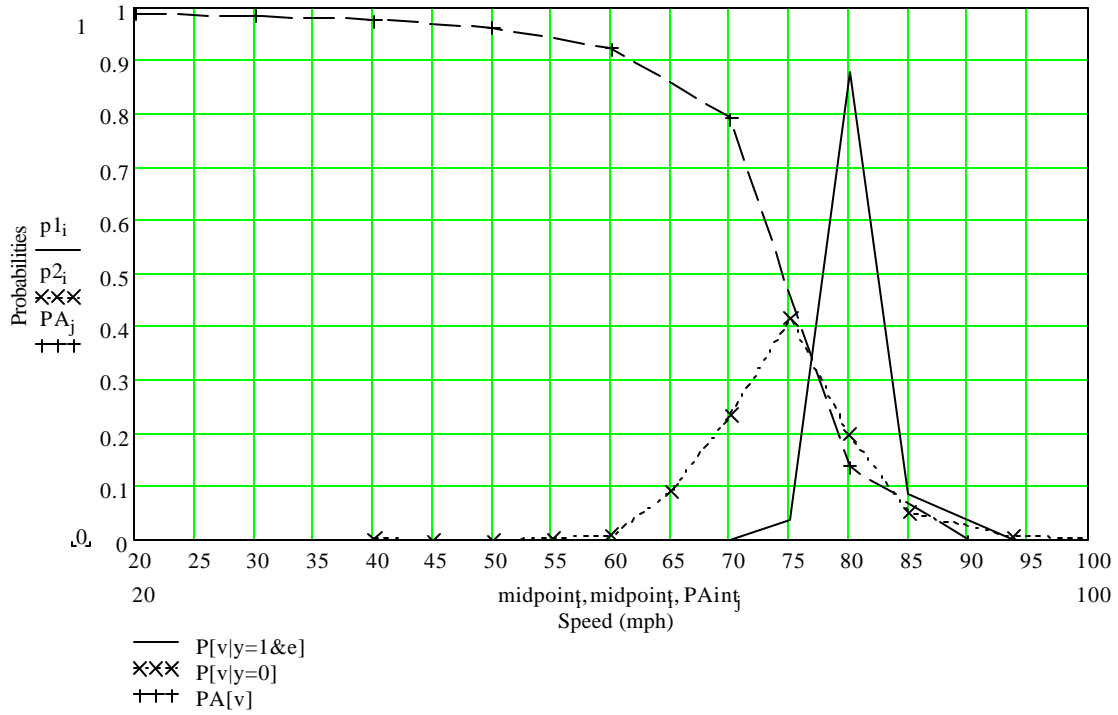
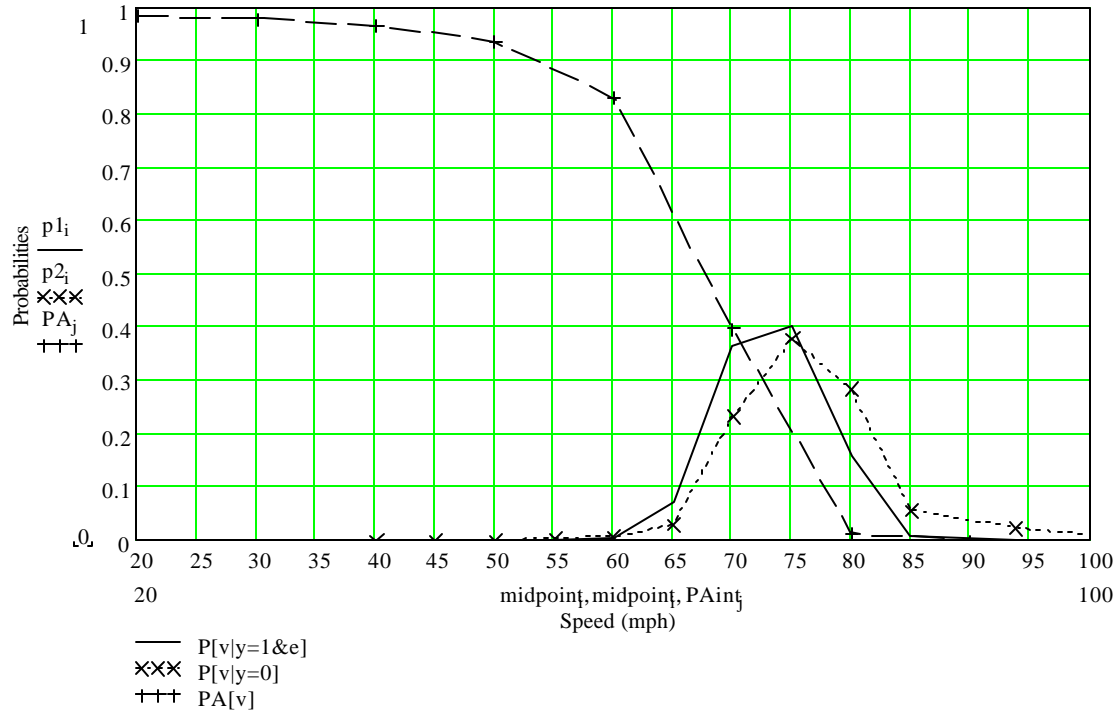


Figure 13. Case Speed Posterior, Distribution of Control Speed Population, and Probability of Avoidance for Minnesota Crash #10. Interstate Highway under Dry Conditions.



References

- Baker, S. 1975. Traffic Accident Investigation Manual. Evanston, IL, Northwestern University Traffic Institute.
- Balke, A., and Pearl, J. 1994. Probabilistic evaluation of counterfactual queries. In Proceedings of the 12th National Conference on Artificial Intelligence, MIT Press, Menlo Park, CA, 230-237.
- Balkin, S., and Ord, J. 2001. Assessing the impact of speed-limit increases on fatal interstate crashes. *Journal of Transportation and Statistics*, 4, 1-12.
- Cirillo, J. 1968. Interstate system accident research: study II. *Public Roads*, 35, 71-75.
- Cooperrider, N., Thomas, t., and Hammoud, S. 1990. Testing and analysis of vehicle rollover behavior. SAE Technical Paper 900366, SAE, Inc. Warrendale, PA.
- Davis, G., 2004a, Possible aggregation biases in road safety research and a mechanism approach to accident modeling, Technical Report 1 to ITS Institute, University of Minnesota.
- Davis, G., 2004b. Bayesian reconstruction of traffic accidents and causal effect of speed in intersection and pedestrian accidents. Technical Report 2 to ITS Institute, University of Minnesota.
- Davis, G. 2002a. Is the claim that 'variance kills' and ecological fallacy? *Accident Analysis and Prevention*, 34, 343-346.
- Davis, G. 2002b. Towards a unified approach to causal analysis in traffic safety using structural causal models. In M. Taylor (ed.) *Transportation and Traffic Theory in the 21st Century*, Pergamon, Amsterdam, 247-266.
- Davis, G. 1999. Using graphical Markov models and Gibbs sampling to reconstruct vehicle/pedestrian accidents. Proceedings of the Conference Traffic Safety on Two Continents. Linkoping, Sweden, Swedish national Road and Transport Institute.
- Davis, G., and Davuluri, S. 2002. Development and Application of Case-Control Methods to Traffic Safety Problems. Bureau of Transportation Statistics, Washington, DC.
- Fricke, L. 1990. Traffic Accident Reconstruction. Evanston, IL, Traffic Institute, Northwestern University.
- Garber, N., and Gadiraju R., 1988. Speed Dispersion and Its Influence on Accidents. AAA Foundation for Highway Safety.
- Garrott, W., and Guenther, D. 1982. Determination of precrash parameters from skid mark analysis. *Transportation Research Record*, 893, 38-36.
- Gelman, A., Carlin, J., Stern, H., and Rubin, D. 2004. *Bayesian Data Analysis*, 2nd edition. Boca Raton, Chapman-Hall.
- Giere, R., 1980. Causal systems and statistical hypotheses. In L. Cohen and M. Hesse (eds.) *Applications of Inductive Logic*. Oxford University Press, Oxford, UK.
- Hauer, E., 1997. *Observational Before-After Studies in Road Safety*. New York, Elsevier.
- Hauer, E., 2004. Harm done by tests of significance. Presentation at 83rd Annual Meeting of Transportation Research Board, Washington, DC.
- Hicks, J. 1989. Traffic accident reconstruction. In K. Carper (ed.) *forensic Engineering*, New York, Elsevier, 101-129.
- Holland, P., 1986. Statistics and causal inference. *J. of American Statistical Association* 81, 945-960.
- Hosmer and Lemeshow 2000. *Applied Logistic Regression*, 2nd edition. Wiley, New York.

- Kloeden, C., McLean, A., Moore, V., and Ponte, G., 1997. Traveling Speed and the Risk of Crash Involvement. NHMRC Road Accident Research Unit, University of Adelaide, Adelaide, Australia.
- Kloeden, C., Ponte, G., and McLean, J. 2001. Travelling Speed and the Risk of Crash Involvement on Rural Roads. Road Accident Research Unit, Adelaide, University of Adelaide.
- Kloeden, C., McLean, J. and Glonek, G. 2002. Reanalysis of Travelling Speed and the Risk of Crash Involvement in Adelaide South Australia, Road Accident Research Unit, University of Adelaide. Adelaide, Australia.
- Lave, C., 1985. Speeding, coordination, and the 55 mph limit. *American Economic Review* 75, 1159-1164.
- Lave, C., and Elias, P. 1994. Did the 65 mph speed limit save lives? *Accident Analysis and Prevention* 26, 49-62.
- Limpert, R. 1989. *Motor Vehicle Accident Reconstruction and Cause Analysis*, 3rd edition, Charlottesville, VA, Michie Co.
- Martinez, J., and Schlueter, R. A primer on the reconstruction and presentation of rollover accidents. SAE Technical paper 960647, SAE Inc., Warrendale, PA.
- McCarthy, P., 1998. Effect of speed limits on speed distributions and highway safety: A survey of the literature. In *Managing Speed, Special Report 254*, Transportation Research Board, Washington, DC, 277-358.
- McLean, J., Anderson, R., Farmer, M., Lee, B., and Brooks, C. 1994. Vehicle Travel Speeds and the Incidence of Fatal Pedestrian Crashes, Road Accident Research Unit, Adelaide, University of Adelaide.
- Nystrom, G., and Kost, G. 1992. Application of NHTSA crash database to pole impact prediction SAE Technical Paper 920605, SAE Inc. Warrendale, PA.
- Pearl, J., 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, UK.
- Rubin, D., 1974. Estimating causal effects of treatments in randomized and nonrandomized studies. *J. of Educational Psychology* 66, 688-701.
- Semon, M. 1995. Determination of speed from yaw marks. In T. Bohan and A. Damask (eds.) *Forensic Accident Investigation: motor Vehicles*. Charlottesville, VA, Michie Butterworth, 149-198.
- Solomon, D. 1964 *Accidents on Main Rural Highways Related to Speed, Driver and Vehicle*. Bureau of Public Roads, Washington, Dc.
- Spiegelhalter, D., Best, N., Carlin, B., and vander Linde, A., 2002 Bayesian measures of model complexity and fit. *Journal of Royal Statistical Society B*. 64, 583-616.
- Spiegelhalter, D., Tomas, A., and Best, N. 2000. *WinBUGS Version 1.3 User Manual*. MRC Biostatistics Unit, Cambridge University, Cambridge, UK.
- Tian, J. and Pearl, J., 2000. Probabilities of causation: Bounds and identification. *Annals of Mathematics and Artificial Intelligence*, 28, 287-313.
- Turner, S., 1997. "Net effects": A short history. In V. McKim and S. Turner (eds.) *Causality in Crisis*. University of Notre Dame Press, Nortre Dame, IN, 23-46.
- TRB 1984. 55 A Decade of Experience, Special Report 204, Transportation Research Board, Washington, DC.
- West, L. and Dunn, J. 1971. Accidents, speed deviation and speed limits. *Traffic Engineering*. July, 52-55.

White, S. and Wilson, A. 1970. Some effects of measurement errors in estimating involvement rate as a function of deviation from mean traffic speed. *Journal of Safety Research*, 2, 67-72.

Appendix
Examples of WinBUGS Code Used for this Report

Model for Estimating Initial Speed for Minnesota Rollover Crash #7

```
Model
# tripped rollover reconstruction for case 99105543
{
# Speed at start of post-trip phase
mu.roll ~ dunif(mu.roll1,mu.roll2)
#d.roll ~ dunif(droll_low, droll_up)
#v3 <-sqrt(2*mu.roll*g*d.roll)
v3 ~ dnorm(0,.001)
d.roll.bar <-log((v3*v3)/(2*mu.roll*g))
d.roll ~ dlnorm(d.roll.bar,100)

# Speed at start of trip phase
t.trip ~ dunif(t.trip1,t.trip2)
M <- M_veh+w1
w1~ dunif(wm1,wm2)
t2 <- pow(t.trip,2)
A <- 1/(M*g)
B <- TRACK/(2*H)
C <- 1/(M*g*H)
D <- M*pow(TRACK,2)/4
C1 <- 2*C*(IX+D)*(sqrt(pow(B,2)+1)-1)
FT1 <- A*B*t2
FT2 <- 4*pow(A,2)*t2*C1
FT3 <- 2*A*A*t2
FT <- (FT1 + sqrt(pow(FT1,2)+FT2))/FT3
a.trip <- FT/M
v2 <- v3+a.trip*t.trip

# Speed at start of pre-trip phase
mu.pave ~ dunif(mu.pave1,mu.pave2)
mu.grass ~ dunif(mu.grass1,mu.grass2)

for ( i in 1:N) {
d[i] ~ dunif(d_low[i], d_up[i])
alpha[i] ~ dunif(alpha_low[i], alpha_up[i])
}
d.trip <- (pow(v2,2)-pow(v3,2))/(2*a.trip)
tripstart <- sum(d[])-d.trip
vcrit <- sqrt(2*FT*d.trip/M)
dsum[1] <- d[1]
for ( i in 2:N) {
```

```

dsum[i] <- dsum[i-1]+d[i] }

for ( i in 1:N) {
F[i] <- (wheel[i]*mu.pave+(4-wheel[i])*mu.grass)*M*g*sin(alpha[i])/4}
W[1]<-F[1]*d[1]
for ( i in 2:N) {
W[i] <- F[i]*(d[i]*step(tripstart-dsum[ i]) + (tripstart-dsum[i-1])*step(tripstart-dsum[i-1])*step(dsum[i]-tripstart))
}
K <- (M/2)*pow(v2,2)+sum(W[])
v1 <- sqrt(2*K/M)

```

```

# Post-processing
v1.mph <- v1*2.237
v2.mph <- v2*2.237
v3.mph <- v3*2.237
vcrit.mph <- vcrit*2.237
speed.tau <- 1/(speed.sig.mph*speed.sig.mph)
speed.atr~dnorm(speed.bar.mph,speed.tau)
faster <- step(v1.mph-speed.atr)
}

```

```

Data list(N=16,
g=9.807,M_veh=1773,TRACK=1.65,H=0.713,IX=847.285,mu.roll1=.3,
mu.roll2=.6,mu.pave1=.55,mu.pave2=.9,mu.grass1=.35,mu.grass2=.55,t.trip1=.4,
t.trip2=.6, d.roll=62.179,speed.bar.mph=74.565,speed.sig.mph=7.07,
wm1=57.15,wm2=98.43)

```

alpha_low[]	alpha_up[]	d_low[]	d_up[]	wheel[]
0.074	0.083	4.359	5.578	4
0.074	0.083	5.486	6.706	4
0.087	0.096	5.425	6.645	4
0.103	0.115	6.706	7.925	4
0.120	0.133	7.010	8.230	4
0.145	0.161	5.182	6.401	4
0.170	0.188	3.627	4.846	4
0.244	0.271	5.852	7.071	4
0.252	0.280	1.067	2.286	3
0.529	0.588	3.597	4.816	3
0.852	0.946	2.743	3.962	2
0.881	0.978	3.048	4.267	1
0.959	1.065	2.377	3.597	1
0.939	1.042	1.585	2.804	0
1.096	1.217	2.408	3.627	0
1.331	1.479	0.975	2.195	0

```

Inits list(mu.roll=.45,mu.grass=.45,mu.pave=.7,t.trip=.5,v3=23)

```

Fit Linear Model and Compute Probabilities of Avoidance for Minnesota Data

```
model
# matched case-control analysis for 10 Minnesota run-off road crashes
# uses new bounds for u_roll, and reanalysis of 99417595 and 00404813
# uses 41 test data for yaw model estimates
# uses 'veh 1&2' stats for vehicle 2 from 00602583
# uses yaw speed estimates for 99104860
# uses site-specific means for centering
# linear or quadratic model
# uses slip-corrected speed for 99605587
# attempts to estimate mean critical speed from beta0
{
  # PRIORS
  beta1 ~ dnorm(0,1.0E-06)
  beta2 ~ dnorm(0,1.0E-06)

  # LIKELIHOOD
  for (i in 1 : I) {
    # METHOD 3 fit standard Poisson regressions relative to baseline
    Y[i,1]~dpois(mu[i,1])
    log(mu[i,1]) <- beta0[i]+beta1*(V[i]-V.bar[i])
    tau[i]<-1/(X[i,J+1]*X[i,J+1])
    V[i]~dnorm(X[i,1],tau[i])
    log(mup[i,1]) <- beta1*(V[i]-V.bar[i])
    p[i,1] <- mup[i,1]/sum(mup[i,1:J])

    for (j in 2:J) {
      Y[i, j] ~ dpois(mu[i, j])
      log(mu[i, j]) <- beta0[i] + beta1*(X[i,j]-V.bar[i])
      log(mup[i,j]) <- beta1*(X[i,j]-V.bar[i])
      p[i,j] <- mup[i,j]/sum(mup[i,1:J])
    }
    beta0[i] ~ dnorm(0,1.0E-06)
    d[i] <- -2*log(p[i,1])
    s.star[i] <- (1-p[i,1])/p[i,1]
  }

  nonzero1 <- step(beta1)
  D <- sum(d[1:I])

  # Compute 'replications' for testing goodness of fit

  for (i in 1: I) {
```

```

yrep[i,1:J] ~ dmulti(preparep[i,1:J],1)
log(murep[i,1]) <- beta1*(V[i]-V.bar[i])
preparep[i,1] <- murep[i,1]/sum(murep[i,1:J])
srep[i,1]<-yrep[i,1]*(1-preparep[i,1])/preparep[i,1]
drep[i,1] <- -2*yrep[i,1]*log(preparep[i,1])
for (j in 2:J) {
log(murep[i,j]) <- beta1*(X[i,j]-V.bar[i])
preparep[i,j] <- murep[i,j]/sum(murep[i,1:J])
drep[i,j] <- -2*yrep[i,j]*log(preparep[i,j])
srep[i,j] <- yrep[i,j]*(1-preparep[i,j])/preparep[i,j] }
drow[i] <- sum(drep[i,1:J])
srow[i] <- sum(srep[i,1:J])

}

# Goodness of fit measures
for (i in 1:I) {
p_pear[i] <- step(srow[i]-s.star[i])
s_diff[i] <- srow[i]-s.star[i]
p_dev[i] <- step(drow[i]-d[i])
d_diff[i] <- drow[i]-d[i]
}

# Compute probability of avoidance functions
for (i in 1:I) {Vt[i]~dnorm(X[i,1],tau[i])
for (n in 1:N) {
PN[i,n] <- step(Vt[i]-V2[n])*(1-exp(beta1*(V2[n]-Vt[i])))
count[i,n] <- step(Vt[i]-V2[n]) }}
}

```

Data list(I=10,J=11,
V.bar=c(57.7, 56.2, 74.2, 73.0, 58.3, 74.2, 74.5, 77.4, 74.4, 75.8),
N=9,V2=c(20,30,40,50,60,70,80,90,100))

X[,1]	X[,2]	X[,3]	X[,4]	X[,5]	X[,6]	X[,7]	X[,8]	X[,9]	X[,10]	X[,11]	X[,12]
69.77	58.371	58.095	66.417	50.197	55.557	65.440	60.495	69.959	71.608	53.898	5.583
70.95	56.614	53.237	38.095	54.754	64.810	50.557	58.879	61.222	54.626	55.999	4.97
71.42	69.020	74.754	82.484	76.688	74.626	82.333	78.115	69.780	76.715	70.501	4.354
81.41	69.020	74.754	88.831	69.925	75.086	77.333	86.673	64.780	76.715	70.501	2.06
59.80	52.000	60.000	56.000	55.000	56.000	56.000	58.000	53.000	54.000	62.000	3.6
80.77	69.020	77.785	71.398	81.688	80.086	69.633	68.898	70.248	70.005	78.201	2.12
74.47	84.684	77.988	85.329	70.602	82.897	71.867	84.963	71.237	62.639	79.338	5.02
67.65	69.020	68.331	65.197	55.557	71.688	74.626	74.633	68.898	70.248	70.005	4.016
80.44	77.240	80.708	78.375	76.570	72.176	74.979	74.463	80.214	69.953	76.346	1.659
73.30	72.957	72.785	71.398	75.440	79.626	82.333	78.115	69.780	86.711	76.531	3.964

Y[,1]	Y[,2]	Y[,3]	Y[,4]	Y[,5]	Y[,6]	Y[,7]	Y[,8]	Y[,9]	Y[,10]	Y[,11]
1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0

```

Inits list(beta1=0.14,beta2=0,V=c(70,70,70,70,70,70,70,70,70,70),
beta0=c(-20,-20,-20,-20,-20,-20,-20,-20,-20,-20))
list(beta1=.25,beta2=.3,V=c(80,80,80,80,80,80,80,80,80,80),
beta0=c(-5,-5,-5,-5,-5,-5,-5,-5,-5,-5))
list(beta1=.006,beta2=.05,V=c(60,60,60,60,60,60,60,60,60,60),
beta0=c(-10,-10,-10,-10,-10,-10,-10,-10,-10,-10))

```