

An Interactive Tool to Compare and Communicate Traffic Safety Risks: TrafficSTATS

Paul S. Fischbeck,* Barbara Gengler,** David Gerard,* and Randy S. Weinberg*
Carnegie Mellon University* and Multidimensionality**

November 2007

Paul S. Fischbeck
Social & Decision Sciences and Engineering & Public Policy
Center for the Study & Improvement of Regulation
Carnegie Mellon University
Pittsburgh, PA 15213
412 268 3240
Fischbeck@cmu.edu

Barbara Gengler
Multidimensionality, LLC
5445 Wilkins Ave
Pittsburgh, PA 15217
412 656 9014
bgengler@multidimensionality.com

David Gerard (*corresponding author*)
Engineering & Public Policy
Center for the Study & Improvement of Regulation
Carnegie Mellon University
Pittsburgh, PA 15213
412 268 1273
dgerard@cmu.edu

Randy S. Weinberg
Information Systems Program
Carnegie Mellon University
Pittsburgh, PA 15213
412 268 3228
rweinberg@cmu.edu

Forthcoming in *The Journal of the Transportation Research Forum*

An Interactive Tool to Compare and Communicate Traffic Safety Risks: TrafficSTATS

Abstract: TrafficSTATS (www.traffic-stats.us) is publicly-available, interactive, web-based query tool that provides estimates of passenger vehicle and other traffic safety risks. Using “cube” database technology, TrafficSTATS houses publicly-available government data on traffic fatalities from the Fatality Analysis Reporting System (FARS) and personal travel behavior from the National Household Travel Survey (NHTS) and calculates risk statistics in real time for user-specified queries. We describe the motivation for developing the tool, explain the technology developed to store the data and facilitate the queries, and provide a series of examples of the types of comparisons that can be made quickly and efficiently.

INTRODUCTION

Despite the availability of a number of excellent sources of government data,¹ information that allows for simple comparisons of passenger travel risks and risk tradeoffs is not readily available. Certainly, with some effort, experts can identify appropriate data sources and create risk and exposure measures (for example, a comparison of the relative risks of personal travel during the afternoon and the evening). However, policy makers, the media, and the general public often rely on anecdotal “statistics” that vary in their definitions, supporting evidence, and reliability. These problems are amplified when media reports fail to put problems in context or to provide appropriate baselines for comparison. As a result, there is a great divide between what risk analysts know about risks and what the media and the general public *think* they know about risks. The result is that some risks are perceived to be elevated (e.g., children riding on the school bus), while others that should warrant greater attention are neglected (e.g., children riding their bikes) (National Research Council 2002).

In this paper, the development of a tool to address these shortcomings is described – TrafficSTATS (Statistics on Travel Safety). TrafficSTATS was developed by the Center for the Study & Improvement of Regulation at Carnegie Mellon University with funding from the AAA Foundation for Traffic Safety and Carnegie Mellon University accessible at www.traffic-stats.us. It is a free, publicly-available, web-based query tool that provides the user with an interactive environment to identify and compare passenger vehicle and other traffic safety risks. Since its release in January 2007, the press has used this tool to convey travel risk information, including an article in the New York Times (Wald 2007), a widely-circulated Associated Press article (Borenstein 2007), and a front-page story in USA Today on elderly drivers (Davis and DeBarros 2007). Several state legislators have used the information to examine risks of motorcycles and school travel.

There were two central challenges to the development of the tool. The first is to provide an accurate and transparent characterization of information about passenger travel risks (specifically, fatality risks). Risk metrics are developed that are created using traffic fatalities from the Fatality Analysis Reporting System (FARS) and personal travel behavior from the National Household Travel Survey (NHTS). FARS and NHTS have a number of common fields, such as person's age, gender, person type (e.g., for drivers, passengers, pedestrians, bicyclists), time of day, day of week, and transportation mode (e.g., passenger car, SUV, bicycle, motorcycle). Based on these parameters, the user defines a query and the tool generates three risk metrics: deaths per person mile, deaths per trip, and deaths per minute traveled. In addition, the tool provides information about the reliability of the estimates by generating confidence intervals for each risk measure. The incorporation of uncertainty is a non-trivial matter given the available data. A simple query for 16 to 20-year old male drivers reveals 3.91 deaths per 100 million trips (with an associated 95% confidence interval of 3.55 to 4.35) compared with 1.99 female drivers

killed per 100 million miles driven (with a confidence interval of 1.76 to 2.29). These numbers illustrate that young male drivers are almost twice as likely to get killed as female drivers, and that both young males and young females compare poorly with the national average of 1.15 driver deaths per 100 million miles driven (1.13 to 1.17).

The second challenge is the effective presentation of the risk information in an easily understandable, interactive format that accommodates many types of risk comparison queries for a variety of potential end users. TrafficSTATS provides a level of detail and responsiveness that is not available from any other sources. Users can easily explore the relative risks of millions of different combinations of transportation modes, demographic variables, and vehicle types. As an added feature, the tool facilitates extremely fast queries of the underlying FARS and NHTS data. TrafficSTATS provides a single, centralized source for both general and specific traffic-safety risk information of interest to multiple stakeholders.

The second and third sections describe the motivation for the development of the tool, the data sources used and the resultant risk metrics, the method for calculating these risks, and an explanation of the means by which confidence bounds are put on these risk estimates. The next section describes the use and functionality of the TrafficSTATS query tool and the following section provides a brief description of the “cube” technology used to house the data and facilitate the queries. The next section provides three examples of the types of comparisons, showing both the breadth of the risk information that TrafficSTATS provides, as well as some limitations of inferences that can be made. Conclusions are in the last section.

PUBLIC PERCEPTIONS AND QUESTIONABLE RISK METRICS

Decision science researchers have shown repeatedly that given clear and comparable risk information, members of the public can make rational and reasoned risk trade-offs (Morgan,

Fischhoff, Bostrom, and Atman 2001; Morgan et al. 2001). For many risky decisions, however, this necessary information is simply not available. In some cases, the numbers have not been calculated correctly, and in others, the “correct” numbers are presented in a way that does a poor job of conveying information about risk tradeoffs.

These problems are certainly pervasive for the “facts” about travel safety, where there is often a divergence between perceived and actual risks. For example, a survey of 110 people selected from the general public found that a majority of the respondents did not appreciate the overwhelming safety advantages of school buses or the risks of walking and biking.² When asked why they thought that school buses were dangerous, many survey respondents commented that the news often had stories about school bus accidents. This is perhaps expected because fatal accidents involving a school bus are typically national news, whereas a car crash killing a teenage driver and his younger sibling is not (National Research Council 2002). The school transportation study helped to correct this understandable availability bias (Tversky and Kahneman 1973; Dawes 1988) by getting the facts out to the public. A limiting factor, however, is that this type of information might not be disseminated through the media or other outlets, providing little value for improving decision making.

Even when data are widely available and reported, however, conclusions regarding relative risks are often unreliable. Information about traffic risks is often presented as raw counts (number of deaths per year) or dubious ratios (deaths per registered vehicle), rather than a more appropriate risk measure. For example, using data on deaths per aircraft and deaths per vehicle would be a poor way to compare the relative risks of flying and driving. Evans, Frick, and Schwing (1990) quantify the risks. Commercial aviation is 50 times more dangerous using these metrics. A more useful measure for this comparison would be deaths per mile traveled or per

trip, where flying is roughly 10 times safer. A still more useful measure might be the risks associated with taking a 250-mile trip for a driver of average ability.

In the context of passenger vehicles, a recent *New York Times* article reported that SUVs are more dangerous than passenger cars because the associated deaths per registered SUV is 10% higher than that per registered passenger car (Hakim 2004). Yet, empirical evidence and survey data show that SUVs are driven more miles per year (Goh, Fischbeck, and Gerard forthcoming; FHWA 2004b) and typically carry more people per trip (FHWA 2004b) than the average passenger car. Therefore, the average SUV generates more exposure for its passengers (more passengers, more trips, more miles) than the average passenger car, making the deaths per registered vehicle metric of questionable validity.³

CALCULATING RISKS

A central issue in the calculation of travel risks is an appropriate and accurate measure of both the hazard and the level of exposure. It is not possible to understand the underlying risks by simply looking at fatality information. Just because a particular travel mode has more fatalities associated with it, does not mean that is necessarily riskier; the risk depends on how much that mode is used. For example, 3,779 people died on motorcycles and 18,819 died in passenger cars in 2004.⁴ However, the fact that there are nearly five times more deaths in cars does not mean that cars are a riskier travel mode. An appropriate risk comparison needs to incorporate exposure information. In this instance, passenger cars account for far more miles traveled than motorcycles, and consequently there are 1.05 fatalities per 100 million miles traveled in passenger cars compared with 32.61 in motorcycles.⁵ On a per-mile basis, motorcycles are over 30 times riskier.

These risk calculations require two sources of data: the number of fatalities as the hazard, and the number of miles traveled, trips taken, minutes traveled as the exposure measures. The

Fatality Analysis Reporting System (FARS) is the standard source of data for vehicle crashes in the United States that involve a fatality.⁶ To be included in FARS, a crash must involve a motor vehicle traveling on a traffic way customarily open to the public, and result in the death of a person (either an occupant of a vehicle or a non-motorist) within 30 days of the crash. The FARS file contains descriptions of each fatal crash reported. Each case has more than 100 coded data elements that characterize the crash, the vehicles, and the people involved. FARS is maintained by the National Highway Traffic Safety Administration (NHTSA) within the Department of Transportation (DOT) and is updated annually. Aside from the omission of suicides (or suspected suicides), it is a comprehensive accounting of U.S. traffic fatalities. We draw on a subset of FARS data from 1999 to 2004 to construct the risk measures used in this study.⁷

The National Household Travel Survey (NHTS) provides information on travel characteristics. NHTS is the nation's inventory of daily and long-distance travel. The survey includes demographic characteristics of people, vehicles, and detailed information on daily and longer-distance travel for all purposes by all modes.⁸ NHTS survey data are collected from a sample of U.S. households and expanded to provide national estimates of trips and miles by travel mode, trip purpose, and many household attributes. As with survey there are potential issues with accuracy. Itsubo and Hato (2006) used GPS tracking data to verify reported trips and found discrepancies. Regardless, NHTS is the best and only exposure data available at the national level (Beck, Dellinger, and O'Neil 2007).

Given the population of deaths from FARS and the exposure information from NHTS, the risk calculation is straightforward: the risk is the number of fatalities from FARS divided by the total number of miles traveled (or total trips, or total minutes in the car) using sample data from NHTS.

Figure 1: Screen Capture of TrafficSTATS Query Comparing Travel Risks of Males and Females

The screenshot shows the TrafficSTATS website interface. The main content area displays a table titled "Estimated Fatalities per 100 Million" with the following data:

Gender	Passenger Miles	Trips	Travel Minutes	Fatalities 1999 - 2004	See Confidence Interval Report
Male	1.30	14.51	0.70	149,066	>>
Female	0.73	6.55	0.36	70,840	>>
Other/Missing/Unknown	6.21	40.57	3.30	32	>>
All Genders	1.04	10.43	0.54	219,938	>>

Below the table is a "Parameter Summary" section listing the query filters:

- Age: All Age Groups
- Day of Week: All Days of Week
- Hour: All Hours
- Month: All Months
- Person Type: Vehicle Occupant
- Region: All Census Descriptions
- Transportation Mode: All Transportation Modes

The report was generated at 7/31/2006 11:27:38 PM.

TrafficSTATS QUERIES AND FUNCTIONALITY

TrafficSTATS contains both the risk calculations, as well as the underlying data sources. Figure 1 provides a screenshot of the front page of the site. In the left-hand frame, the user can select risk information, FARS, or NHTS. The available risk information is determined by the availability of the underlying data. Specifically, risk calculations are available for cases where FARS and NHTS have common fields: age, gender, time of day, day of the week (e.g., weekday

and weekend), season / month of the year, region of the country (grouped by state), person type (vehicle occupants and non-motorists), and transportation mode (e.g., personally owned vehicles, buses, walking, biking). In many cases, there is a tremendous amount of detail in the overlapping fields. Transportation mode, for example, facilitates risk calculations for personally owned vehicles generally, as well as specific information on cars, SUVs, pick up trucks, vans, and motorcycles.

The risk queries in TrafficSTATS are straightforward. The user begins by clicking a comparison criterion from the left-hand menu, and the tool will generate the most general results for this criterion. For example, by selecting “gender,” TrafficSTATS returns the estimated fatalities per 100 million passenger miles traveled, per 100 million passenger trips, and per 100 million minutes traveled (see Figure 1). For example, the estimated risk for females killed in passenger vehicles is calculated as the total number of females killed divided by the exposure measure – the estimated number of miles traveled by females. The calculations of “median” (or best estimated) risks are 0.73 deaths per 100 million miles traveled, 6.55 deaths per 100 million trips, and 0.36 deaths per 100 million minutes. In comparison, males have 1.30 deaths per 100 million miles traveled, 14.51 deaths per 100 million trips, and 0.70 deaths per 100 million minutes. This shows that males have much higher risks than females (78% greater risk for miles traveled, 122% greater risk for trips, and 94% greater risk for minutes traveled).

Several additional features ease the interpretation and use of the query results. First, in addition to the risk measures, there is a column that shows the total number of deaths for 1999 to 2004 as a means for putting the magnitude of these deaths in perspective. For the most general cases, the sum of the rows in this column gives the total number of passenger deaths for the five-year period. As shown at the bottom of Figure 1, TrafficSTATS provides a parameter summary

for each query as a means to track each query. These results can be exported to a number of file formats, including Excel, HTML, and Acrobat.

Because of the uncertainty inherent in the NHTS survey data, these risk estimates are not known precisely; we can only be confident that the values fall within a range. So, the user can click on the icon in the final column to generate lower and upper confidence bounds for each risk estimate. The miles-traveled risk measure for females has a median (best guess) value of 0.73, a lower bound of 0.72, and an upper bound of 0.75. This means that the process used to construct the interval insures that 95 out of 100 intervals so calculated will contain the true parameter. If this confidence interval includes 0, then the upper and lower bounds are not displayed and the median value shown must be viewed as being very uncertain. Figure 2 shows this display.

Calculation of the Confidence Interval

Although fatalities data is based on the entire population of deaths involving motor vehicles,⁹ exposure measures are based on survey data comprising approximately 160,000 people. As a result, members of the sample group represent others in the country that share similar demographic variables. For example, a teenage boy living in an urban community in a southern state who is part of the survey represents many other teenage boys with similar backgrounds that are not in the survey. The determination of how many other non-survey people that a person in the survey represents is a complicated statistical calculation. There is assumed to be many-to-one mapping from people in the country to people in the survey.

Figure 2: Screen Capture of TrafficSTATS Query Output Comparing Median Travel Risk of Males and Females with Confidence Intervals

Use the browser's BACK button to return to calling report.

Gender	Passenger Miles			Trips			Travel Minutes			Fatalities 1999 - 2004
	Low	Median	High	Low	Median	High	Low	Median	High	
Male	1.27	1.30	1.34	14.28	14.51	14.74	0.69	0.70	0.71	149,066
Female	0.72	0.73	0.75	6.45	6.55	6.65	0.35	0.36	0.37	70,840
Other/Missing/Unknown	3.25	6.21	67.79		40.57		1.66	3.30	769.21	32
All Genders	1.02	1.04	1.06	10.31	10.43	10.54	0.53	0.54	0.54	219,938

Low, Median, and High correspond to the following percentiles: 2.5%, 50%, 97.5%

Parameter Summary

Age: All Age Groups
Day of Week: All Days of Week
Hour: All Hours
Month: All Months
Person Type: Vehicle Occupant
Region: All Census Descriptions
Transportation Mode: All Transportation Modes
Report Generated: 7/31/2006 11:29:57 PM

In NHTS, the “multiplier” used to map survey respondents to general demographic characteristics is called the replicate weight. Since there were approximately 290 million people in the country at the time of the most recent survey, each survey participant represents, on average, approximately 1,800 other people. In practice, however, the replicate weights range from hundreds to ten of thousands.

Because this extrapolation from a relatively small sample size to the entire country can not be done exactly, there is uncertainty as to what the exact replicate weights are for each of the survey’s participants. This uncertainty is measured in terms of a standard error on the replicate

weight estimates. For each individual in the survey, a confidence interval on his/her replicate weights can be calculated by assuming a normal distribution. Generally, the standard errors of the replicate weights are about 10% of the weight itself.

The calculation of the risk metric requires that the number of fatalities in the combination of interest be divided by these estimates. An alternate way to express this is that the total number of fatalities must be multiplied by the reciprocal of the estimated exposure metric. We assume the NHTS exposure measures for miles, trips, and minutes are normally-distributed random variables. The reciprocal of a normal distribution has a median equal to the reciprocal of the median and the 5th and 95th percentiles will be the reciprocal of the 95th and 5th percentiles, respectively (the percentile values reverse because of the reciprocal calculation). As an artifact of the reciprocal calculation, the confidence interval is not symmetric in many cases, with a slightly longer tail towards the higher end. Nevertheless, TrafficSTATS displays only the median values for the denominator.¹⁰

Because of these steps, it is difficult to complete the calculation so that results can be displayed interactively. Therefore, the task of calculating the best estimate is separated from calculating the confidence interval for the estimate. The default display on the web page is just the median value. By selecting the double arrow icon on the right side of the output table, confidence intervals can be calculated for each row or for the entire table. The response time for a single row is tens of seconds, where the calculation for an entire table with many rows could take several minutes.

TECHNOLOGY UNDERLYING TrafficSTATS

Typically, the FARS and NHTS data sets are stored and queried through statistical or relational database reporting tools, such as SAS or an SQL (Structured Query Language) system.

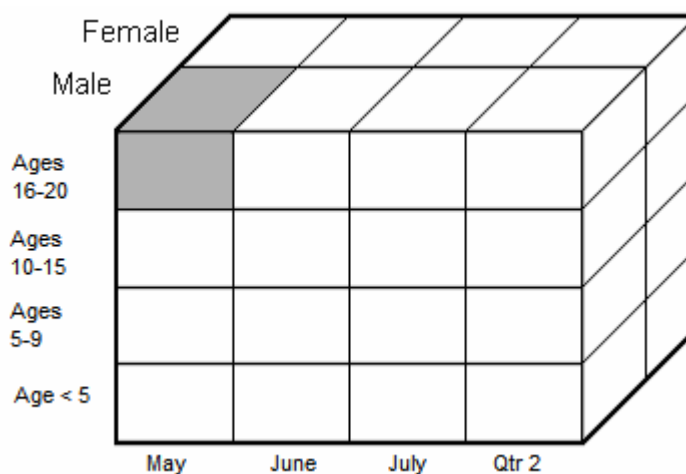
For rapid analysis, ease of use, and complex cross-queries, however, traditional SQL or SAS database query tools would be technically unsuitable and unacceptably slow for most users.

Multidimensional database technology enables users to execute complex ad hoc queries quickly and reduces the need for users to deal directly with the complexities of the underlying record and data structures, variable names and semantics, coding conventions, and specialized query languages. TrafficSTATS uses Microsoft SQL Server 2000, Analysis Services and Reporting Services to enable multidimensional queries on the FARS and NHTS datasets, as well as queries between FARS and NHTS along common dimensions.

Conceptually, multidimensional databases can be envisioned as a set of “cubes,” whose edges represent dimensions of interest and whose intersections represent “facts.” Facts are summaries, or aggregations, of information derived from the underlying datasets. Queries are executed by selecting and aggregating the facts for dimensions and ranges of interest. Since the facts have been pre-calculated, queries on a multidimensional database are uniformly fast. For example, the three dimensional cube shown in Figure 3 consists of dimensions for Month, Age and Gender. At any intersection, the relevant pre-calculated facts (e.g., number of fatal crashes, number of fatalities) are stored. As a result, any fact within the cube, such as the number of fatalities for 16-20 year old males during May, is quickly and easily retrieved. More complex queries on the multidimensional database can be executed by combining and expanding input variables from the cube in interesting ways. For instance, one may query and aggregate the underlying facts for teen-agers in April, May and June by selecting the range of values along the cube's Ages dimension corresponding to *Ages 10-15* and values along the Time dimension corresponding to *Quarter 2*.

Similar to a “view” in a relational database, multidimensional databases can be combined into “virtual cubes” to allow cross-queries on combinations of measures. In Traffic STATS, the

FARS and NHTS data sets are joined into a virtual cube along various common dimensions: Age, Gender, Day of Week, Month, Hour, Person Type, Region and Transportation Mode. Queries among these dimensions extract facts from both FARS and NHTS cubes and can therefore calculate a variety of risk ratios. In contrast, confidence intervals are not pre-stored but calculated upon request based on the dimensions selected and, therefore, take longer to process. TrafficSTATS is comprised of three cubes: a FARS cube, an NHTS cube and a third virtual cube that joins the FARS and NHTS cubes, as summarized in Table 1. As is evident, the intersection of FARS and NHTS provides the dimensions for the virtual cube. The corresponding risk calculations are constrained by the availability of data, and therefore the database with the more general dimensions. For example, FARS and NHTS each report data by the hour, day, and month of the trip, making it possible to calculate risks, for example, for an individual driving on Monday at 8 a.m. in September. However FARS reports by state whereas NHTS reports by region. Therefore, we cannot produce a risk calculation for individual states, but can fit the data to create regional estimates, based on the NHTS data. Because NHTS does not survey individuals on their propensity to drink alcohol, the virtual cube contains no information on alcohol-related fatalities.

Figure 3: A Three-Dimensional Cube**Table 1: TrafficSTATS Facts and Dimensions**

	FACTS	DIMENSIONS
FARS	Number of crashes Number of persons involved Number of fatalities	Age Gender Person type (e.g., driver, pedestrian) Hour, day, month, year of crash Vehicle body type Road function (e.g., rural interstate) First harmful event (e.g., rollover, impact with fixed object) State Injury severity
NHTS	Number of passenger miles traveled Number of trips taken Number of minutes traveled	Age Gender Person type (e.g., driver) Hour, day and month of trip Transportation Mode Trip length and purpose Region of country
Risk (NHTS-FARS Virtual cube)	Fatalities per 100 million miles traveled Fatalities per 100 million trips Fatalities per 100 million minutes traveled Lower- and Upper-bound Confidence Intervals for each of the above risk measures	Age Gender Person type (e.g., driver, pedestrian) Hour, day and month of trip Transportation mode Region of country

APPLICATIONS

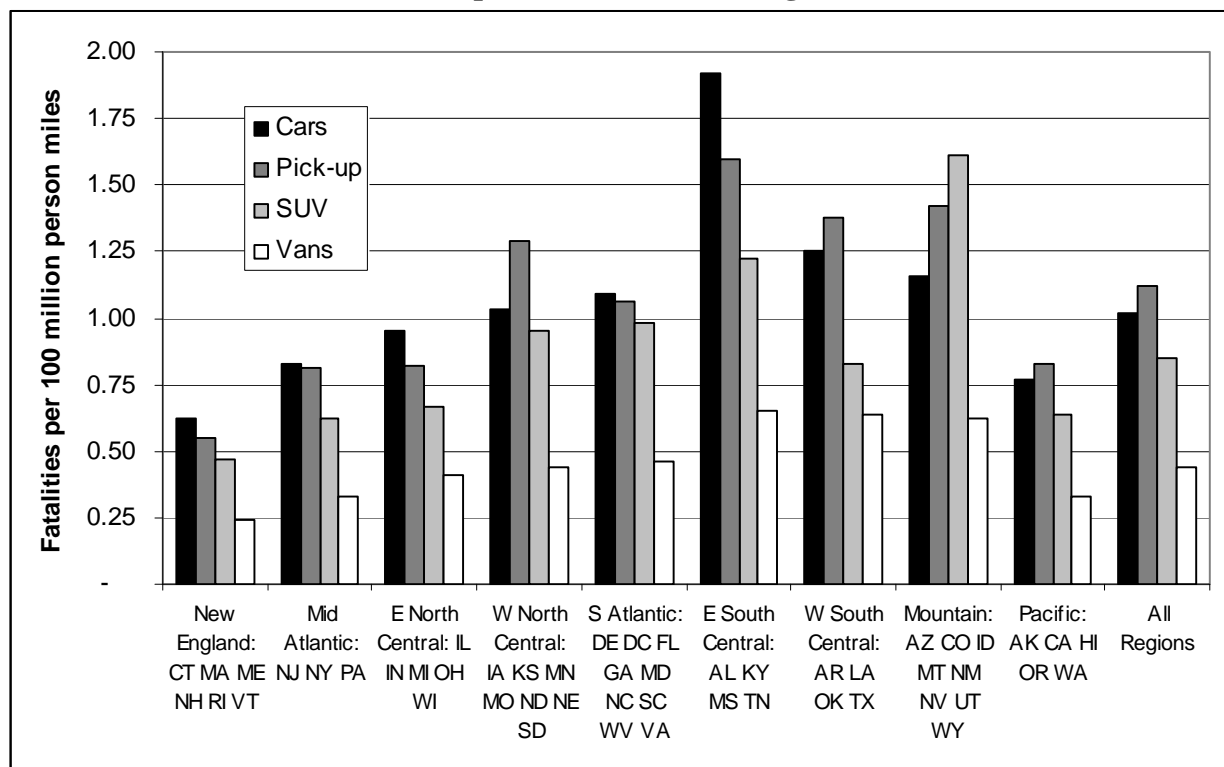
By greatly lowering the barriers between users and data, travel risk information can now be calculated over a wide variety of dimensions quickly. In this section, we demonstrate the power of the interface through three examples.

Example 1: Differences in Travel Risk by Geographic Region and Type of Vehicle

For this comparison, geographic region was selected as the primary comparison variable. Then four queries were submitted using different types of personally-owned vehicles (i.e., car, SUV, van, pick-up truck). After each query the results were exported to Excel. The four runs were combined into one spreadsheet for graphical analysis. The total time for completing this analysis was less than 15 minutes. Figure 4 provides a summary of the point estimates of the results.

Using TrafficSTATS, the national baseline risk for vehicle occupants is 1.04 fatalities per 100 million passenger miles traveled (for all hours, ages, days, vehicle types, etc...), with a confidence interval of 1.02 to 1.06. It is clear from Figure 4 that many of these point estimates deviate substantially from the baseline. The maximum risk value for vans, which are always the safest vehicle type, is approximately 40% lower at 0.64. Using the confidence interval functionality of TrafficSTATS, it is possible to investigate the uncertainty behind these estimates. The confidence interval for vans in the mountain states is 0.52 to 0.80,¹¹ demonstrating that the difference between vans and the national baseline is statistically significant. Moreover, inspection of confidence intervals for each vehicle type shows that vans in the mountain states are significantly safer than pick-ups, cars, and SUVs.

**Figure 4: Travel Risk for Different Vehicle Types Across Geographic Regions
(Fatalities per 100 Million Passenger Miles)**



The overall risk and the riskiest vehicle type varies across the country. New England, mid-Atlantic, and the Pacific regions have the lowest risks, while the east south central, west south central, and mountain regions have the highest. SUVs are the riskiest vehicle type only in the mountain region. In all other regions, SUVs are lower risk than cars and pick-ups. Over all, pick-up trucks present the highest fatality risks (right-most columns of the graph). At the national level, there is no overlap in the confidence intervals for vans (0.42-0.46), SUVs (0.78-0.86), cars (1.01-1.06), or pick-up trucks (1.07-1.16), demonstrating statistically significant differences in risk.

Example 2: SUV Travel Risk by Geographic Region and Yearly Season

For this comparison, geographic region was once again selected as the primary comparison variable. A query was run for each season (i.e., winter, spring, summer, and fall) while holding vehicle type at SUV. Again, the results from each query were exported to Excel and combined, and the analysis took less than 10 minutes to complete. Figure 5 shows the results.

Note that contrary to intuitive speculation, the greatest fatality risks are not associated with winter travel. In fact, winter driving in SUVs is never the riskiest and is safest over all (right-most columns). Summer risk for two regions (east south central and mountain) are dramatically higher than other seasons. In addition, summer driving is riskiest for six out of the nine regions and is riskiest overall. Risks in the mid-Atlantic and New England regions are low and fairly constant across seasons. Additional queries could be done to put confidence bounds on these estimates and to determine the risks of other vehicle types by season to see whether this particular pattern is unique to SUVs. Traffic STATS, unlike any other available data source, allows for this easy, interactive exploration of risk.

The result of lower risk during the winter presents one of the limitations of the risk metrics, which is the reliance on fatality risks. It could be the case that the risk of a crash or an injury is higher during the winter, but the risk metric does not account for these possibilities.

Example 3: Rollover Risk vs. Non-rollover Risk Comparison Between SUVs and Cars

The third comparison looks at the often discussed rollover risk of SUVs. This is done by looking at rollover risk (measured in fatalities per 100 million passenger miles) by age and gender categories. In this case, age was selected as the primary comparison variable, and eight queries were run using all combinations of three double-category variables (male and female,

rollover and non-rollover, and car and SUV). As in the previous examples, query results were exported to Excel and combined. See Figures 6-8 for results.

Figure 5: Travel Risk for SUVs at Different Times of the Year across Geographic Regions (Fatalities per 100 Million Passenger Miles)

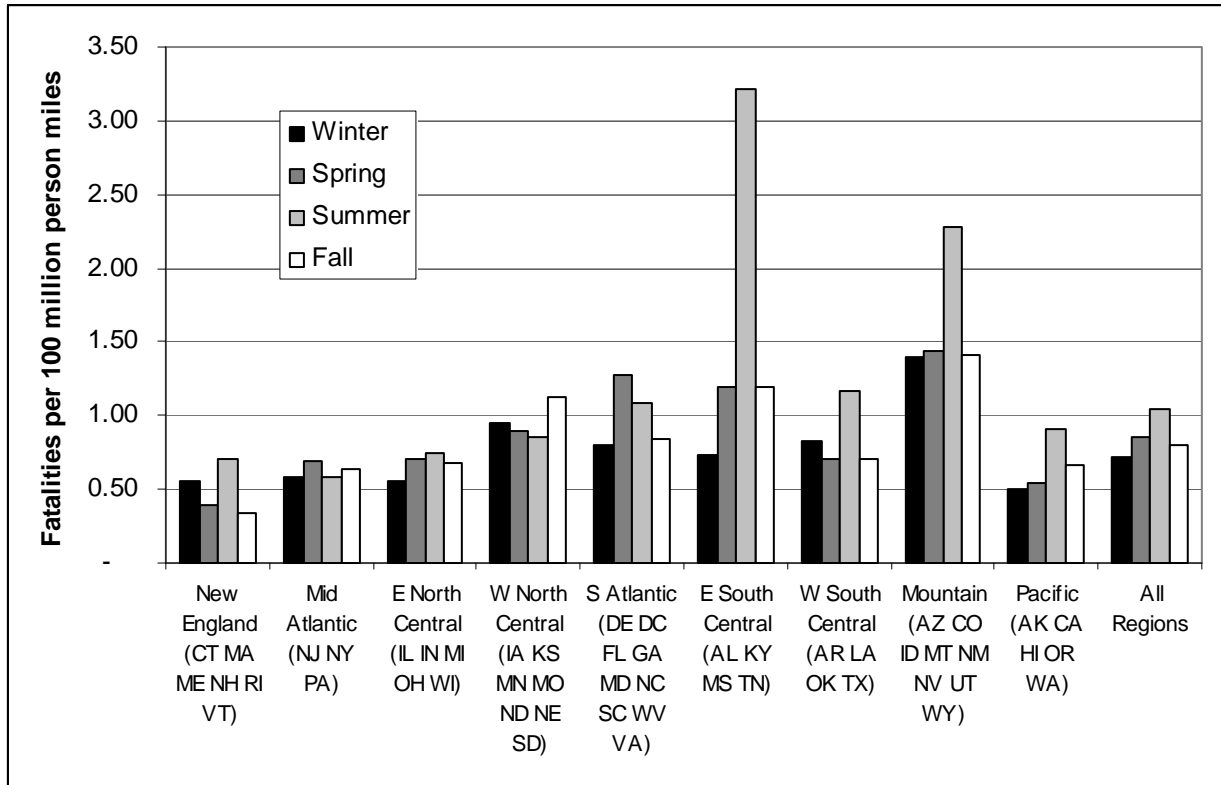
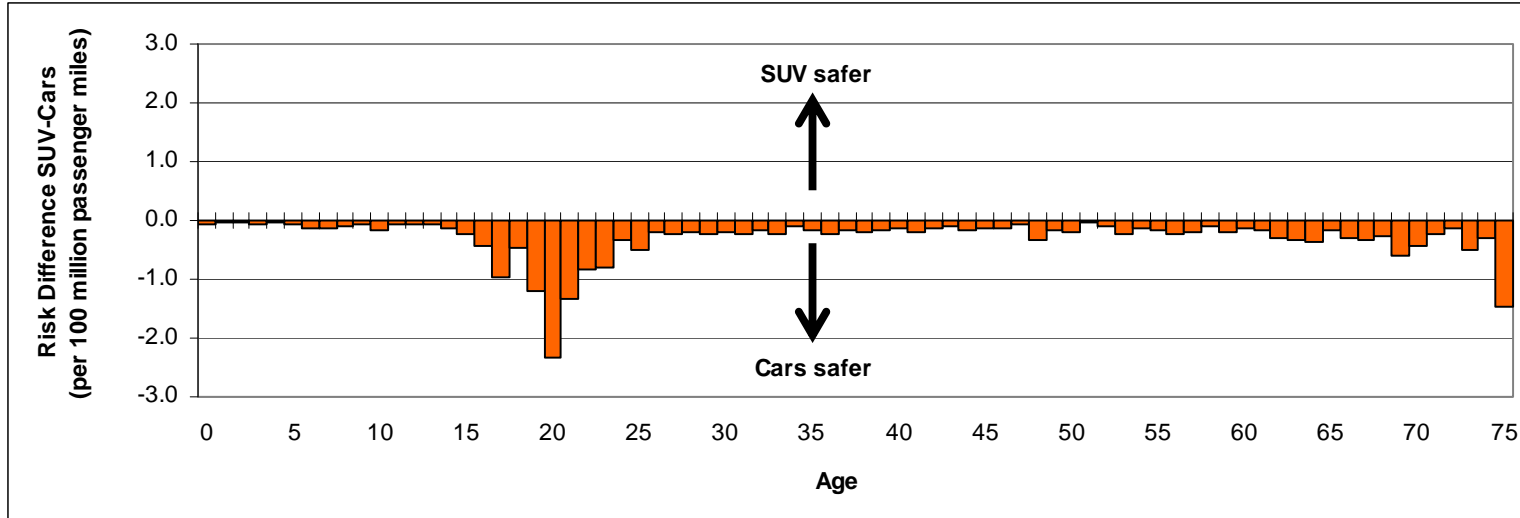


Figure 6: Differences in Rollover Risks by Age of Vehicle's Occupant (SUV-Cars)
(Measured in Fatalities per 100 Million Passenger Miles for (A) Males and (B) Females)

(A) Males



(B) Females

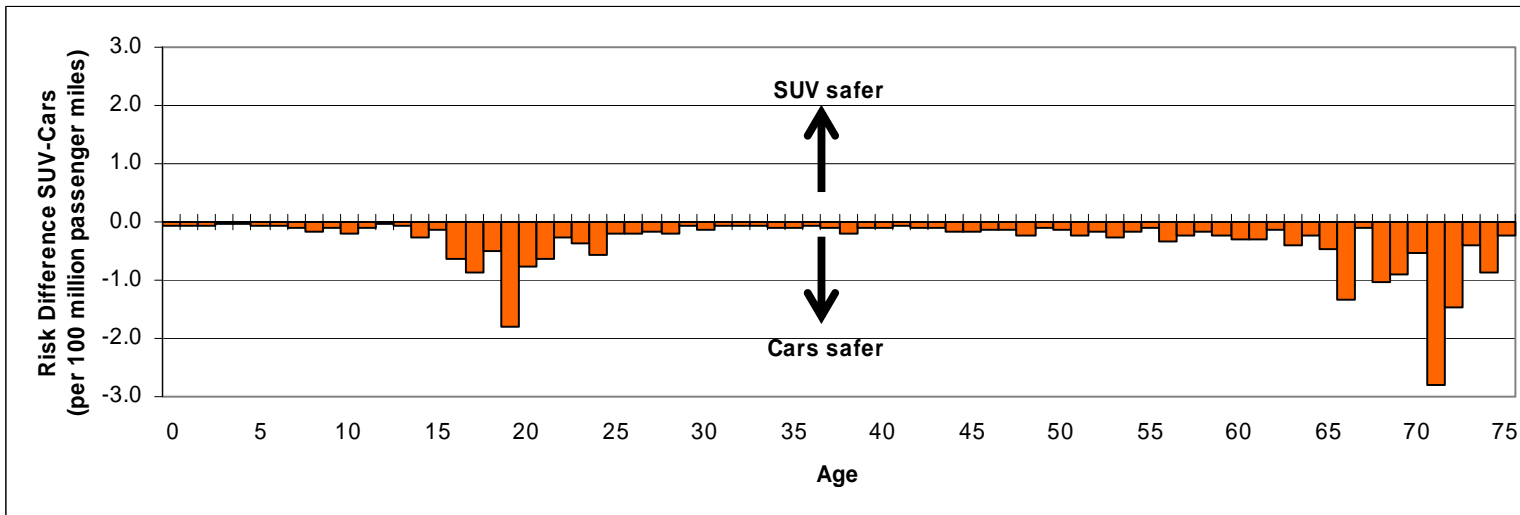
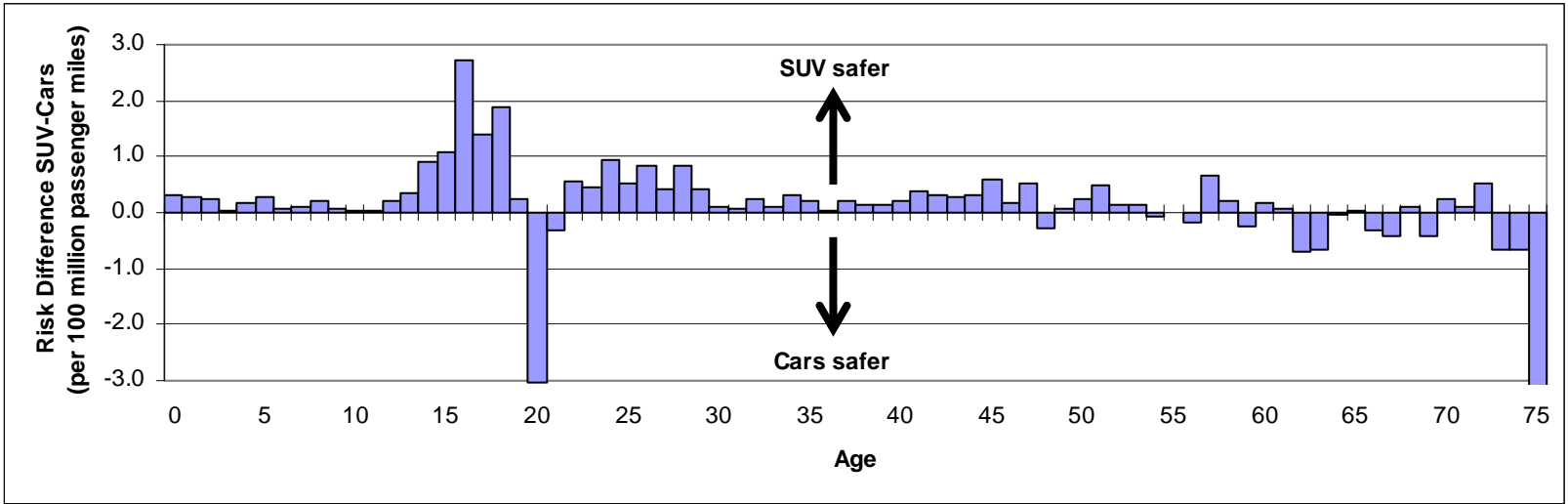


Figure 7: Differences in Non-rollover Crash Risks by Age of Vehicle's Occupant (SUV-Cars)
(Measured in Fatalities per 100 Million Passenger Miles for (A) Males and (B) Females)

(A) Males



(B) Females

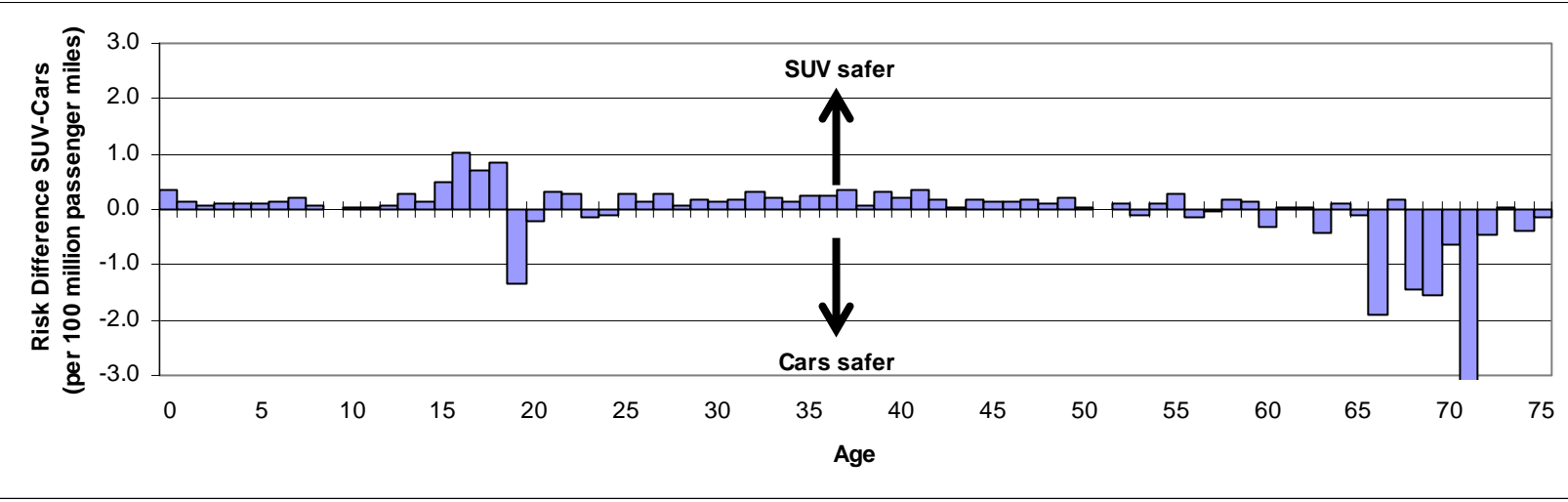
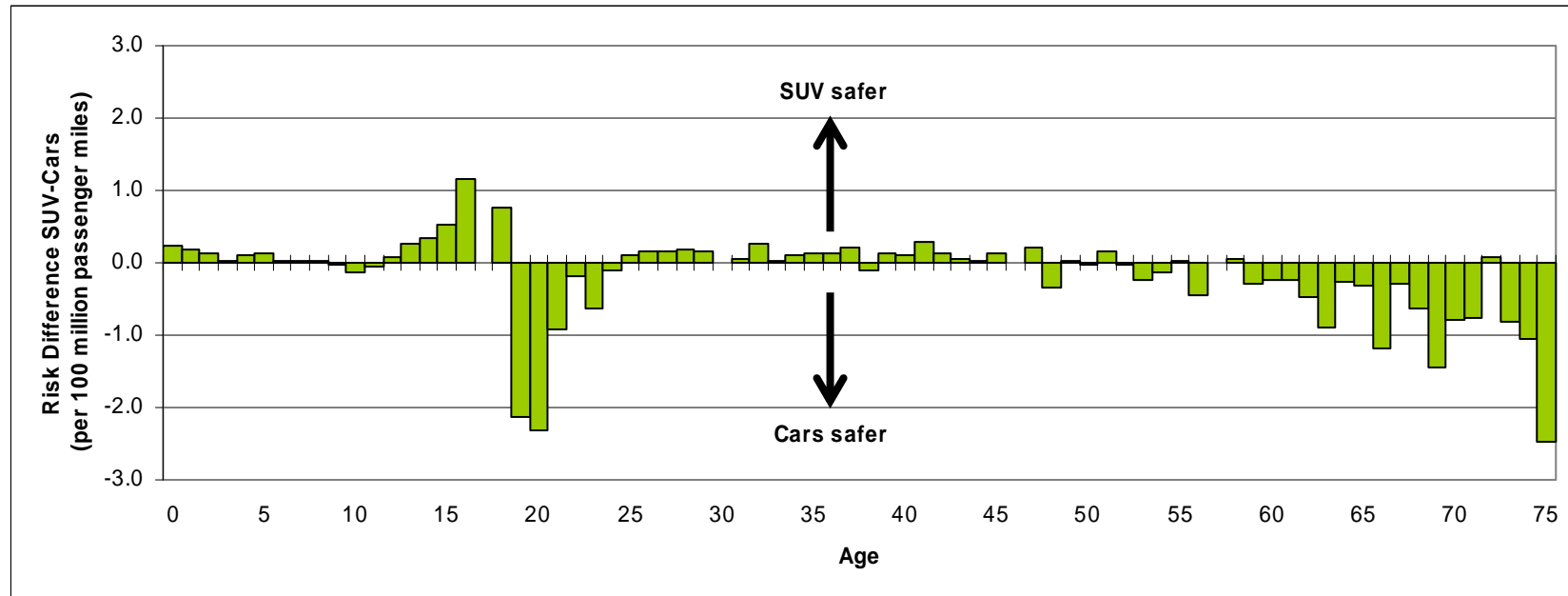


Figure 8: Differences in Travel Risks (Both Rollover and Non-rollover) by Age of Vehicle's Occupant (SUV-Cars)
(Measured in Fatalities per 100 Million Passenger Miles)



Note that for rollover risk, for every age and both genders, the SUV is riskier (Figure 6). There is a pronounced increase in the relative risk of SUVs for both genders during ages 19-22 and for people over age 65 (though the paucity of data prevents drawing conclusions about the older drivers). Almost the opposite occurs for risk from non-rollover crashes (Figure 7). In this case, SUVs are generally safer except for 20 year olds of both genders and for vehicle occupants over age 65. Figure 8 shows the overall risk for all vehicle occupants. SUVs are safer (or equally risky) for most ages except 19-23 and over 60 years old. Overall when using risk per passenger mile as a metric, SUVs are safer than cars. This does not account for the greater risk that they may be inflicting on other vehicles (White 2004).

CHOOSING A RISK METRIC

TrafficSTATS calculates risks for many transportation modes, from walking to driving to riding the bus, and each of the three metrics it generates (per mile, per trip, per minute) can provide useful insights depending on the application. For example, a TrafficSTATS query of passenger type risks shows that in terms of fatalities per mile, walking is 71% more dangerous than biking (19.48 to 13.65 fatalities per trip). On a per trip basis, however, biking is 57% riskier than walking (21.48 to 13.65). This reflects that bike trips are on average about 1.9 miles compared to 0.7 miles for walking trips. On a per-minute travelled basis, there is little difference between the two 0.93 for bikers and 0.90 for walking, a statistically insignificant difference.¹² In this context, it is useful to examine each of the three risk metrics to infer the relative risks.

The use of more than one metric is also useful when comparing vehicle travel. For example, when comparing men and women drivers, men have 78% higher risks per mile driven (1.39 to 0.78), but 105% higher risk per minute traveled (0.76 to 0.37) and 161% higher per trip

taken (15.96 to 6.10). The variance reflects that on average men take longer trips and drive faster, whereas women take more trips. As a result, there is not necessarily a singular correct metric to measure or to convey travel risks, and the objective of TrafficSTATS is to give users the information necessary to begin to compare traffic safety risks.

These measures, however, cannot account for behavioral aspects of driving. There is a large literature on how the inherent safety of the vehicle might cause drivers to be less cautious, “offsetting” the safety improvements.¹³ Moreover, alcohol is a major cause of vehicle fatalities, especially in the evenings, and this factor could make the interpretation of the raw risk calculations more difficult.

CONCLUSIONS

TrafficSTATS draws on multidimensional database technology to provide users with a straightforward, interactive tool that quickly provides travel-risk information to millions of possible user-specified queries. Although there are certainly limitations to both FARS and NHTS as data sources, these are two of the best, most comprehensive sources of data on traffic fatalities and personal travel behavior. Because there is no existing database of travel risks, this research holds the promise of achieving several goals. First, TrafficSTATS provides a single, centralized source that provides accessible, reliable, and understandable risk information to multiple stakeholders, including the media, safety advocates, policy makers, and the general public. The speed and ease of the query tool allows users to generate risk metrics and make comparisons that would have taken much longer by individually querying FARS and NHTS data. The system also provides (albeit somewhat less quickly) confidence bounds on the risk estimates. These benefits should inform individual decision making, traffic safety research, and regulatory policy. Second, the process by which the risk estimates are calculated will be transparent allowing for focused

discussions on real issues of risk trade-offs. Third, TrafficSTATS also provides users with access to rapid retrieval of large portions of the FARS and NHTS databases, which has utility for safety researchers and may also have broader interest. Finally, by using a form of data mining, researchers could uncover multidimensional insights previously not recognized by a study of either database individually. By systematically determining which risk values can be calculated (the same resolution/combination of risk category and dimension must occur in both FARS and NHTS), this approach could reveal previously unknown relationships among fatal motor vehicle crashes, demographics and travel behaviors. The uncovering of peculiar results could lead to new insights, or, alternatively, provide the impetus to improve accuracy of survey data collection.

BIOGRAPHIES

Paul S. Fischbeck, Ph.D., is a Professor of Engineering & Public Policy and of Social & Decision Sciences and the Director of the Center for the Study & Improvement of Regulation at Carnegie Mellon University. He has extensive experience as a risk analyst studying complex engineered systems in general (space shuttle and off-shore oil platforms) and transportation systems specifically (school transportation).

Barbara Gengler is a consultant in the area of Business Intelligence. She is the principal of Multidimensionality, providing solutions for data analysis and reporting using Business Intelligence technologies, including development of the multidimensional databases (cubes) and reporting tools used for the TrafficSTATS system.

David Gerard, Ph.D., is the Executive Director of the Center for the Study & Improvement of Regulation in the Department of Engineering & Public Policy and a teacher in the Engineering &

Technology Innovation Management program at Carnegie Mellon University. His expertise is in regulatory economics and policy, focusing on environmental and safety risks and technology policy.

Randy S. Weinberg, Ph.D., is a Teaching Professor of Information Systems, and director of the undergraduate major program in Information Systems at Carnegie Mellon University. He holds the Ph.D. in Operations Research from the University of Minnesota. He worked as a software developer for several major corporations prior to entering academia. His primary interests are in software development methodologies, software project management, software systems in nonprofit organizations and intelligent management support systems.

ACKNOWLEDGEMENTS

The project has been supported by from the AAA Foundation for Traffic Safety and from Carnegie Mellon University. We are thankful for excellent comments from Lindsay Griffin. A Carnegie Mellon student Information Systems team of Amarpal Singh Banger, Glen Bischoff, Kanishka Maheshwari, and Edward Mok completed portions of the database development, the query tool, and the preliminary web interface. Immanuel Alam provided excellent technical assistance in building and maintaining the server.

ENDNOTES

¹ The Bureau of Transportation Statistics website is a store of data links. See www.transtats.bts.gov/DataIndex.asp

² The survey was conducted as part of the NRC (2002) study, but is unpublished. Most people, however, did recognize that teenage drivers are an extraordinarily high-risk group.

³ This measure does not account for any differential risks that SUVs might *create* for passengers in other vehicles or for non-motorists (White 2004; Gayer 2004).

⁴ Numbers are from the FARS database.

⁵ Risks are calculated using the FARS and NHTS databases.

⁶ The FARS database and documentation is available on the web at <http://www-fars.nhtsa.dot.gov> (last accessed August 1, 2006). Users should refer to the FARS website for the complete dataset, details on the underlying data, a description of the collection methodology, and limitations of the data.

⁷ Traffic STATS also includes the underlying FARS data necessary to construct our risk metrics, but not all FARS fields are included on the site.

⁸ NHTS and attendant supporting materials are located at <http://nhts.ornl.gov> (last accessed August 1, 2006). See FHWA (2004a).

⁹ There are, however, some exclusions and caveats. For example, suicides are excluded, as well as fatalities that occur more than 30 days after the incident.

¹⁰ The mean of the reciprocal of a normal distribution can be calculated using a Taylor series expansion. However, based on preliminary inspections, the mean and median are close and differ by only fractions of a percent.

¹¹ Confidence intervals were calculated using the TrafficSTATS website.

¹² We conducted a simple difference-in-means t-test using the means and standard deviations extracted from the TrafficSTATS confidence interval functionality.

¹³ For a recent review of the literature and another empirical test of the offsetting-behavior hypothesis, see Cohen and Einav (2003).

REFERENCES

- Beck, Laurie F., Ann M. Dellinger, and Mary E. O'Neil. "Motor Vehicle Crash Injury Rates by Mode of Travel, United States: Using Exposure-Based Methods to Quantify Differences." *American Journal of Epidemiology* 166, (2007): 212 – 218.
- Borenstein, Seth "Risk of Death Higher for Male Drivers." *Washington Post*, January 18, 2007.
- Cohen, Alma and Liran Einav. "The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities." *The Review of Economics and Statistics* 85(4), (2003): 828-843.
- Dawes, R.M. *Rational Choice in an Uncertain World*. Harcourt, San Diego, CA, 1988 and 2001.
- Davis, Robert and Anthony DeBarros. "Older, Dangerous Drivers a Growing Problem," *USAToday*, May 2, 2007.
- Evans, Leonard, Michael C. Frick, Richard C. Schwing. "Is It Safer to Fly or Drive?" *Risk Analysis* 10(2), (1990): 239-246.

- U.S. Department of Transportation, Federal Highway Administration. *Highway Statistics*. Washington, DC. Report FHWA-PL-96-023-annual, 2004a.
<http://www.fhwa.dot.gov/policy/ohim/hs03/index.htm>
- U.S. Department of Transportation, Federal Highway Administration. *Summary of Travel Trends: 2001 Nationwide Household Transportation Survey*. Washington, DC. 2004b.
<http://nhts.ornl.gov/2001/pub/STT.pdf>.
- Gayer, Ted. "The Fatality Risks of Sport-Utility Vehicles, Vans, and Pickups Relative to Cars" *Journal of Risk and Uncertainty* 28(2), (2004): 103-133.
- Goh, Vincent, Paul S. Fischbeck, and David Gerard. "Identifying and Correcting Errors with Odometer Readings from I/M Data: A "Rollover" Problem for the Estimation of Emissions and Technical Change." Forthcoming in *Transportation Research Record: The Journal of the Transportation Research Board*.
- Hakim, Danny. "Safety Gap Grows Wider Between S.U.V.'s and Cars." *New York Times*, p. C1, Aug 17, 2004.
- Itsubo, Shinji and Eiji Hato. "Effectiveness of Household Travel Survey Using GPS-Equipped Cell Phones and Web Diary: Comparative Study with Paper-Based Travel Survey." Transportation Research Board Annual Meeting Paper #06-07012006, Washington DC, 2006.
- Morgan, M.G., B. Fischhoff, A. Bostrom, C. and Atman. *Risk Communication: The Mental Models Approach*. Cambridge University Press, New York, 2001.
- Morgan, K. M., M. L. DeKay, P. S. Fischbeck, M. G. Morgan, B. Fischhoff, and H. K. Florig. "A Deliberative Method for Ranking Risks (2): Evaluation of Validity and Agreement Among Risk Managers." *Risk Analysis* 21, (2001): 923-938.
- National Research Council. *The Relative Risks of School Travel: A National Perspective and Guidance for Local Community Risk Assessment on School Transportation Safety*. Special Report #269, Transportation Research Board, 2002.
- Tversky, A. and D. Kahneman. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5, (1973): 207-232.
- Wald, Mathew L. "Site Calculates Risk Factors for Travelers." *New York Times*, January 21, 2007.
- White, Michelle. "The 'Arms Race' on American Roads: The Effect of Sport Utility Vehicles and Pickup Trucks on Traffic Safety." *Journal of Law and Economics* 67(2), (2004): 333-356.