Critical Review of the Statistical Analysis Performed by D. Viano
RE: Heco v. Midstate Dodge LLC and Johnson Controls, Inc.

Prepared by

Norma Faris Hubele, Ph. D.
Statistical Consultant

REFRAC SYSTEMS
7201 W. OAKLAND STREET
CHANDLER, AZ 85226-2434

April 5, 2013

Executive Summary

The intent of this document is to provide a statistical evaluation of the testimony, exhibits, reports and other statistical materials generated by Dr. David Viano (Viano) in support of his opinions in Heco v. Midstate Dodge LLC and Johnson Controls, Inc. My opinions and criticisms of Viano's statistical results are based on my 22+ years of experience as a full professor in statistics, authorship of numerous articles and a best-selling statistical textbook, plus my extensive work with the databases used by Viano. The following summary highlights only some of my key criticisms. The attached report expands and enumerates additional statistical problems found in Viano's analysis and testimony.

Viano analyzed the National Accident Sampling System Crashworthiness Data System (NASS CDS), an electronic database consisting of information from a statistical sample of vehicle crashes on our nation's roadways, to support his opinions about the safety of the Dodge Neon and JCI seatbacks in rear impacts.

Instead of using specifics about this case in his national estimates, Viano used estimates based on data from a broad spectrum of vehicles (from sporty coupes to large pick-up trucks with bench seating) in all types of crashes (including rollovers) and all occupants (ages 13 and older) in all seating positions.[1] Viano did not limit the vehicles to those comparable to the Dodge Neon or occupants seated in JCI-constructed seats (*Viano Deposition*, 50:1-25; 51:1-2, March 14, 2013).[2]

Viano testified that he would not generate an estimate when "I get down to one and twos and tens [number of data points in the NASS CDS] (*Id*., 25:11-12). In contrast to his own testimony, Viano used only 6 data points in the NASS CDS to generate his estimate of severely injured occupants in rear impacts with a delta V of 20-25 mph.[3] In fact, there is a repetitive history of Viano using exceedingly small samples in his analysis of rear impacts resulting in extremely unreliable estimates.

The unreliability of his estimates has far reaching implications when analyzed properly. According to his unreliable estimates of injury risks in rear impact, there is no statistically significant difference between the injury risk at less than 10 mph delta V and the injury risk at more than 45 mph delta V. When questioned about this statistical result, Viano called it a "statistical anomaly not a reality." (*Id*., 100:1-12). In fact, the statistical result is a consequence of extremely small samples used to generate very unreliable estimates (*Id*., 100:13-14).

As will be discussed in detail in this report, Viano's lack of statistical expertise is evident in his selection of the *wrong methodology* to estimate the standard error of his estimates from the NASS CDS data. As a consequence, he

- Incorrectly computed the standard error of his national estimates and overstated their reliability,
- Incorrectly computed the standard error of his risk rates and overstated their reliability,
- Incorrectly computed the standard error of his exposure estimates and overstated their reliability.

[1] Viano testified that Exhibit 5 only included front outboard occupants when, in fact, Exhibit 5 includes all occupants (*Id*., 83:22-23).

[2] In fact, Viano testified that he "would be reluctant to do it [construct national estimates] on a vehicle level" (*Id*., 67:23-24).

[3] The number of data points used to compute an estimate is called the sample size. The sample size of 6 is found in the detailed computer printouts of Exhibit 7. In contrast, Viano testified that he did not bring information about the sample size to his deposition (*Id*, 110:16-23). The delta V 20-25 mph is the agreed upon delta V of the Heco vehicle rear impact.

It is clear that he does not have sufficient statistical understanding to recognize the appropriate procedures for computing standard errors. In fact, Viano dismissed the importance of the statistical knowledge needed to properly analyze the NASS CDS data by claiming that he is merely doing "addition, subtraction and division" (*Id*., 45:15-16). As a result, he has repeatedly performed the wrong analysis, computed the wrong statistics and arrived at numerous misleading and incorrect conclusions.[4]

In addition to incorrectly computing the standard errors, he then proceeded to form non-standard 68% confidence intervals on his estimates.[5] When questioned about his use of non-standard confidence intervals, he responded that he "absolutely did not" use any confidence intervals on his materials (*Id*., 128: 1-6). Again, contrary to his testimony, he used 68% confidence intervals on Figure 15 (page 17) and Figure 1 of Appendix I (page 50) of his July 8, 2012 report (Exhibit 2). Furthermore, when he expressed his opinion in Opinion 56 in Exhibit 3, he used 68% confidence intervals.

There is another federal database widely used and recognized for analyzing vehicle crashes. The federal government compiles field data on all fatal crashes on the nation's roadways. This Fatality Analysis Reporting System (FARS) is based on the police accident report and other data collected in each of the states and US territories. While Viano found only 5 occupants in rear-impacted Dodge Neons with severe or worse injuries in the NASS CDS, I found 233 occupants of Dodge Neons in the FARS database who were either killed or sustained incapacitating injuries in rear impacts. Viano testified that he did not look at the injury and fatality data in FARS because the "amount of information available is very sparse" (*Id.*, 30:19-23). However, the comparison of the number of occupant injury records in rear impacts contained in the two databases, 5 versus 233, illustrates that Viano's focus on the numbers in the NASS CDS is not only misleading but also not representative of actual occupant injuries that are occurring in Dodge Neons.

In my opinion, Viano demonstrates that he lacks the statistical expertise to properly analyze the NASS CDS and to apply sound statistical principles in his analysis and interpretation of his statistical results. His analysis also lacks specificity about the performance of the Dodge Neon and the JCI seatback in rear impacts. As a result, I believe that his attempt to use the NASS CDS to support his opinions about the safety of the Dodge Neon and the JCI seatback in rear impacts is extremely flawed.

---

[4]The erroneous analyses detailed in this report are also manifest in Viano's publications.

[5] It is standard statistical practice to use 90, 95 and 99% confidence intervals, as sited by the National Institute of Standards and Technology, Engineering Statistics handbook (http://www.itl.nist.gov/div898/handbook/eda/section3/eda352.htm) and American Society for Testing and Materials (http://www.astm.org/SNEWS/JA 2011/datapoints ja11.htm)

Introduction

Per the request of Attorneys James L. Gilbert and Robert Langdon, I have performed a statistical evaluation of the testimony, exhibits, reports and other statistical materials generated by Dr. David Viano (Viano) in support of his opinions in Heco v. Midstate Dodge LLC and Johnson Controls, Inc.

In performing my review for this case, I have studied the following materials:

a. A 67-page letter/report authored by Viano and dated July 8, 2012.
b. A 21-page letter/supplemental report authored by Viano and dated October 21, 2012.
c. The transcript and exhibits from the deposition of Viano on September 11, 2012.
d. The transcript and exhibits from the deposition of Viano on March 14, 2013.
e. Numerous articles authored by Viano, as cited in his references.

My review and criticisms of Viano's statistical results are based on my 22+ years of experience as a full professor in statistics, authorship of numerous articles and a best-selling statistical textbook, plus 22+ years of extensive work with the databases used by Viano. My vita is attached in Appendix A.

Viano analyzed accident data from the National Automotive Sampling System Crashworthiness Data System (NASS CDS) which is collected, compiled and made available by the National Highway Transportation Safety Administration (NHTSA). This database of sampled accidents consists of electronic records of detailed information concerning approximately 5,000 crashes collected annually.[6] Details about the NASS CDS and some statistical terminology are given in the attached Appendices B and C, respectively.[7]

Briefly, the NASS CDS is used to estimate the number of crashes, vehicles, occupants, pedestrians or cyclists involved in tow-away crashes on our national road system. As a sample, each individual record in the NASS CDS is intended to represent a national subset of crashes. Taken as a whole, the NASS CDS can be used to generate or estimate national totals, averages or rates. The NASS CDS is a statistically-designed sample.

Statistics is a branch of mathematics. While it is possible to perform statistical analysis involving only theory and mathematical concepts, here I am focused on the evaluation of Viano's materials with respect to the necessary and appropriate use of the statistical theory and concepts for sound data analysis of the information derived from the NASS CDS. By his own admission, when questioned about doing statistical analysis on the NASS CDS, Viano minimized the importance of statistical knowledge:

> "I'd say it is pure mathematics, addition, subtraction and division.
> It doesn't require a Ph. D. in statistics." (*Viano Deposition*, 45:15-16, March 14, 2013)

As will be shown, Viano repeatedly demonstrated his lack of statistical expertise throughout his materials and testimony. He computed statistical estimates based on extremely small samples sizes which made his estimates very unreliable. In addition, he applied the wrong methodology to the NASS

[6] In 2010 there were nearly 5.5 million police-reported crashes on the nation's roadways, as reported in NHTSA's *Traffic Safety Facts*, 2010.
[7] More documentation about the NASS CDS can be found in Exhibit 3 of the March 14, 2013 deposition.

CDS which, in turn, made his estimates appear more reliable. He adopted non-standard statistical confidence intervals to arrive at weak statistical conclusions.These problems with his statistical analysis, as well as others, make his statistical-based opinions in this case unqualified, misleading and deceptive.

Gross Analysis of Injury Rates

I begin my analysis with a discussion of the data used by Viano to support his Opinion 56 of his October 21, 2012 (Exhibit 3, p. 5) report. In particular, he used data from Table 6 (Exhibit 4) of that report which he compiled from the NASS CDS database for the years 1993-2007. At his March 14, 2013 deposition, Viano testified "Because of the upgrades of computers and software, it's no longer possible to generate data from 1993 using the SAS program" (*Id*., 9:1-4).[8] Consequently, he changed the years contained in his Table 6 to 1994-2010 (Exhibit 5). The remainder of this analysis focuses on Exhibit 5 since Viano's testimony at his March 14, 2013 deposition references this changed material.

For convenient reference, I have reproduced Exhibit 5 as Table A in this report. I also reproduce Viano's opinion from his October report with changes to his statistics reflecting his changed numbers from those in Exhibit 4 to the new numbers in Exhibit 5, shown here in brackets. In his October 21, 2012 report (Exhibit 3), Viano's 56th opinion on page 5 states:

> Table 6 summarizes the 1994-2007[9] [1994-2010] NASS-CDS weighted data for crashes by severity (delta V). There were 29,055,072 [35,819,408] cases with 49% [48%] having a determined crash delta V. There were a number of reasons delta V was not determined, e.g., rollovers with multiple impacts, accidents with very severe intrusion, underbody or roof impacts.
>
> There were 291,013 ± 8,109 [356,164 ± 9,174] occupants with MAIS 4+F injury and 5,437 ± 932 [6,719 ± 983] occupants with MAIS 4+F injury in rear impacts (1.87%)[1.89%]. Overall, the risk for MAIS 4+F injury was 1.00% ± 0.028% [0.99% ± 0.026%]. The risk was 0.58% ± 0.031% [0.60% ± 0.029%] in frontal crashes, 1.53% ± 0.09% [1.5% ± 0.075%] in side impacts and 0.30% ± 0.05% [0.29% ± 0.043%] in rear impacts. The risk was lowest in rear impacts.

When Viano was questioned about the data in Exhibit 4 (the original Table 6), he demonstrated that he did not know the contents of his own Exhibit or his stated opinion. He incorrectly testified that 14,129,503 occupants or 49% did *not* have a delta V calculated when, in fact, the exact opposite is true (*Id*., 34:12-25; 35:1-7).

In order to discuss the updated data in Exhibit 5, I will briefly describe each of the blocks of data.[10] Exhibit 5 contains statistically-based national estimates for the calendar years 1994-2010. In the first block of data in Table A Viano gave national estimates of the number of occupants in crashes for the various intervals of delta V together with his estimate of the associated standard error "se" for all impact

---

[8] Using the same software, Statistical Analysis Software (SAS), I had no such problem re-creating the data in Viano's original Table 6.

[9] He referenced the wrong years; they were actually 1993-2007 in his July report.

[10] When asked about the formulas needed to compute the numbers in his Exhibits 4 and 5, Viano provided the wrong equations in Exhibit 21. Please see Appendix D for a more detailed discussion.

(including rollovers), frontal, side and rear impacts. The first entry in this block: 6,642,324 occupants were in all impacts with a delta V of less than 10 mph. That estimate has a “se” of 200,338 occupants.

In the next block of data, Viano gave his corresponding national estimates and “se” values for occupants who suffered severe injuries.[11] The first entry in this block: 7,782 occupants were in all impacts with a delta V of less than 10 mph and suffered severe injuries. That estimate has a “se” of 1,875 occupants.[12]

For his “Risk” rates in the third block of data, Viano performed a division operation. For each of the categories of impacts and delta V, he divided the severely injured occupant estimates by the corresponding total number of occupant estimates to create a risk rate. He performed the same division operation to obtain his “se” for his risk rate. The first entry in this block:
0.12% =(7,782/6,642,324)*100%. That estimate has a “se” of 0.028%=(1,875/6,642,324)*100%.[13]

In the fourth block of data, Viano performed a division operation again. He used the first block of data and computed the distribution of occupants across the various delta V categories. He also obtained his standard error values by dividing the numbers in the first block by the total estimate in the category.The first entry in this block: 38.8%=(6,642,324/17,119,408)*100%. That estimate has a “se” of 1.2%=(200,338/17,119,408)*100%.[14]

As footnoted below, there are important mistakes in Exhibit 5 which will be discussed in detail in a later section of this report. For the present, I focus on the general problem of the lack of specificity of this table and Viano’s associated opinion, in relation to this case.

The entire Exhibit 5 was generated using all the vehicles in all the crashes in the NASS CDS with very few restrictions. Viano did not limit the vehicles to those comparable to the Dodge Neon or occupants seated in JCI-constructed seats (*Id*., 50:1-25; 51:1-2).[15] The vehicles included 49 types of bodies, as listed in Appendix E. Furthermore, all the occupants (ages 13 and older) and seating positions were included in his compilation. This fact is found by closely examining Viano’s Exhibit 6, the computer program used to generate the data in Exhibit 5. Contrary to the contents of his own exhibits, Viano testified incorrectly that only front outboard occupants were included Exhibit 5 (*Id*., 83:22-23).

In the end, he used this broad cross-section of occupants to arrive at a summary opinion that the severe injury risk was lowest in rear impacts. While he cited the estimates in Exhibit 5 to support his opinion, he did not perform a rigorous statistical analysis. In fact, when the mistakes in his computation of the standard errors are corrected and a standard confidence interval is properly applied to the corrected estimates, I demonstrate that there is no statistically significant difference between severe injury rates in frontal and rear impacts at the 95% confidence level using this NASS CDS data. Viano has no statistical foundation for his Opinion 56 expressed in his supplemental report of October 21, 2012.

[11]I label occupants as “severely injured” for ease of understanding whereas Viano labels these same occupants as “MAIS 4+F.” These occupants suffered severe, critical or fatal injuries.
[12]The “se” values in the first and second block of data are incorrect due to Viano’s lack of understanding of the NASS CDS statistical design, as will be discussed later in this report.
[13]Viano obtained this “se” value by incorrectly performing the simple arithmetic operation of division. In fact, the correct “se” requires calculus and sophisticated statistical analysis software, as discussed later in this report.
[14] Sophisticated statistical analysis software, not division, is required to compute the correct standard errors.
[15] Viano testified that he “would be reluctant to do it [form national estimates] on a vehicle level” (*Id*., 67:23-24).

Crash Type	**<10**	**10-15**	**15-20**	**20-25**	**25-30**	**30-35**	**35-40**	**40-45**	**45+**	**Total w/o unk**	**Total w/ unk**
	MAIS 0+F all occupants										
All	6,642,324	6,056,384	2,704,624	1,092,585	374,223	131,450	61,906	31,115	25,097	17,119,408	35,819,408
se	200,388	150,857	104,445	59,674	19,907	9,312	4,736	4,031	3,328	270,784	405,996
Front	3,186,814	3,548,324	1,654,714	714,444	227,287	82,938	38,544	18,639	17,645	9,498,348	15,502,645
se	134,098	113,867	93,168	53,479	16,792	6,879	3,790	2,234	2,905	300,631	278,703
Side	2,184,612	1,223,780	537,899	186,745	63,792	19,599	12,077	6,410	1,805	4,236,718	6,900,948
se	127,754	53,403	35,401	18,261	5,543	3,012	1,776	1,496	482	140,663	177,651
Rear	463,408	729,808	278,610	99,652	48,205	12,389	4,252	4,175	2,479	1,642,977	2,294,564
se	61,744	76,024	23,321	15,409	7,793	3,358	1,589	2,923	1,273	99,254	115,795
	MAIS 4+F all occupants										
All	7,782	25,000	21,428	26,903	18,198	13,473	9,803	10,250	9,518	142,356	356,164
se	1,875	4,913	1,637	3,072	1,455	1,333	1,131	1,619	1,153	6,453	9,174
Front	3,615	5,135	5,674	8,474	5,166	5,632	4,454	5,868	5,872	49,888	92,243
se	1,633	2,072	770	1,457	741	897	702	1,243	887	3,371	4,535
Side	1,991	13,104	11,910	14,868	10,369	6,241	3,954	3,079	1,167	66,683	103,481
se	787	3,160	1,178	2,644	1,143	880	753	857	355	4,378	5,151
Rear	50	606	914	160	228	457	128	277	761	3,581	6,719
se	50	329	576	75	106	124	56	133	325	667	983
	Risk MAIS 4+F all occupants										
All	0.12%	0.41%	0.79%	2.46%	4.86%	10.2%	15.8%	32.9%	37.9%	0.83%	0.99%
se	0.028%	0.081%	0.061%	0.28%	0.39%	1.0%	1.8%	5.2%	4.6%	0.038%	0.026%
Front	0.11%	0.14%	0.34%	1.19%	2.27%	6.79%	11.6%	31.5%	33.3%	0.53%	0.60%
se	0.051%	0.058%	0.047%	0.20%	0.33%	1.1%	1.8%	6.7%	5.0%	0.036%	0.029%
Side	0.091%	1.07%	2.21%	8.0%	16.3%	31.8%	32.7%	48.0%	64.6%	1.6%	1.5%
se	0.036%	0.26%	0.22%	1.4%	1.8%	4.5%	6.2%	13.4%	19.7%	0.10%	0.075%
Rear	0.011%	0.083%	0.33%	0.16%	0.47%	3.7%	3.0%	6.6%	30.7%	0.22%	0.29%
se	0.011%	0.045%	0.21%	0.075%	0.22%	1.0%	1.3%	3.2%	13.12%	0.041%	0.043%
	Exposure MAIS 0+F all occupants										
All	38.8%	35.4%	15.8%	6.4%	2.2%	0.77%	0.36%	0.18%	0.15%	100.0%	
se	1.2%	0.88%	0.61%	0.35%	0.12%	0.054%	0.028%	0.024%	0.019%		
Front	33.6%	37.4%	17.4%	7.5%	2.4%	0.87%	0.41%	0.20%	0.19%	100.0%	
se	1.4%	1.2%	1.0%	0.56%	0.18%	0.072%	0.040%	0.024%	0.031%		
Side	51.6%	28.9%	12.7%	4.4%	1.5%	0.46%	0.29%	0.15%	0.043%	100.0%	
se	3.0%	1.3%	0.84%	0.43%	0.13%	0.071%	0.042%	0.034%	0.011%		
Rear	28.2%	44.4%	17.0%	6.1%	2.9%	0.75%	0.26%	0.25%	0.15%	100.0%	
se	3.8%	4.6%	1.4%	0.94%	0.47%	0.20%	0.10%	0.18%	0.077%		

Table A. Reproduction of Exhibit 5: "Delta V (mph) based on NASS-CDS 1994-2010 (MY 1994+), Age 13-104"

Missing Delta V Data

Over 200 pieces of information are routinely collected on a crash included in the statistically-designed sample for the NASS CDS. Some of the collected information is then used by NHTSA to construct additional information, such as the delta V of the vehicle. It is not always possible for NHTSA to compute the delta V due to various reasons such as complexity of the crash and missing data. As a consequence, there are large percentages of occupants in vehicles in NASS CDS who do not have associated delta V data.

From the data in Exhibit 5, I computed these percentages of occupants who are missing delta V data. Those percentages are shown in Table B.

Crash Type	All Occupants	MAIS 4+F Occupants
All	52%	60%
Front	39%	46%
Side	38%	36%
Rear	28%	47%

Table B. The Percentage of Occupants Estimated by Viano in Exhibit 5 With No Delta V Information

Fifty-two percent (52%) of occupants in all impact directions do not have a delta V. Sixty percent (60%) of the severely injured occupants in all impact directions do not have a delta V. For rear impacts, the corresponding percentages are 28% and 47%. With missing information on such large percentages of the data, between 28% and 60%, any statistical relationship between injury risk and delta V is unreliable.

Unreliable National Estimates Based on Extremely Small Sample Sizes

The number of data points used to compute an estimate is called the sample size. When asked about his sample sizes at the deposition, Viano testitfied he did not have that information available (*Id.*, 110:16-23). In fact, the sample size information *was* available to him in Exhibit 7 (computer output tables) which was part of his file at his deposition.

As described above, Viano used Exhibit 5 to examine the relationship between direction of impact, the delta V of the crash and the risk of severe injury of occupants, for all impact directions, but particularly for rear impacts. One statistical problem lies in the fact that there are only 69 occupants (sample size) in the NASS CDS with known delta V in rear impact, as shown in the "Frequency" column on page 10 of Exhibit 7. Viano divided the 69 occupants among the 9 categories of delta V. The resultant extremely small sample sizes are listed in Table C below.

The sample sizes listed in Table C correspond to the number of data points used to estimate the severely injured occupants in rear impacts in Exhibit 5. For example, Viano found only 1 occupant in the delta V category "<10" which he then used to estimate 50 severely injured occupants with a standard error of 50. Similarly, for the category of "20-25 mph" he had only 6 occupants to arrive at a national estimate of 160 severely injured occupants in rear impacts with a standard error of 75.

Delta V Ranges	<10 mph	10-15 mph	15-20 mph	20-25 mph	25-30 mph	30-35 mph	35-40 mph	40-45 mph	>45 mph	Total
Sample Sizes	1	7	8	6	7	15	8	6	11	69

Table C. The Number of Data Points, Known as the Sample Size, Used by Viano to Compute the Severe Injury Risk Rates and Standard Errors of His Estimates for Rear Impacts in Exhibit 5 (Table A).

Viano testified that he would not generate an estimate when "I get down to one and twos and tens" [sample size or number of data points, i.e., occupants, in the NASS CDS] (*Id.*, 25:11-12). The data in Table C, reproduced from his own computer printout, stand in sharp contrast to his testimony.

According to his unreliable estimates of injury risks in rear impact, there is no statistically significant difference between the injury risk at less than 10 mph delta V and the injury risk at more than 45 mph delta V. When questioned about this statistical result, Viano called it a "statistical anomaly not a reality." (*Id*., 100:1-12).

When questioned about using extremely small sample sizes to arrive at national estimates, he testified that these samples were "adequate" for his purposes despite "some variability" due to extremely small samples sizes(*Id.*, 103:7-25).

What Viano labeled as "some variability" is, in fact, excessive variability in the estimates. An extremely small sample size is a two-fold problem in the statistical analysis of the NASS CDS. First, the data are not sufficiently representative of all the types of injuries. Second, the associated large standard error signals a very unreliable estimate. Viano mis-stated this basic statistical concept when he testified at his deposition in September:

> If you are just reporting the data, standard errors don't have any meaning.
>
> (*Viano Deposition*, 162:7-8, September 11, 2012)

To detect the unreliable nature of an estimate, it is useful to express the standard error of an estimate as a percentage of the estimate. Table D shows the percentage error for the rear impact severely injured occupant estimates. For example, the estimate for severely injured occupants in rear impacts with delta V less than 10 has a standard error equal to the estimate, 50/50=100%. For the 20-25 mph category, the percent is 75/160=46%. Such high percentages, in excess of 25%, result in a very poor basis for drawing statistical conclusions. In other words, these percentages signal very unreliable estimates.[16]

[16] Such unreliable estimates have no meaning. For example, polling results at 56%+/-8% could mean that candidate is predicted to win or lose the election.

Delta V Ranges	<10 mph	10-15 mph	15-20 mph	20-25 mph	25-30 mph	30-35 mph	35-40 mph	40-45 mph	>45 mph
Viano's "se" as a Percent of His Estimate	100%	54%	63%	46%	46%	27%	44%	48%	43%

Table D. Viano's Standard Error Expressed as a Percentage of the Estimate for Severely Injured Occupants in Rear Impacts

Wrong Methodology for Computing the Standard Error of the National Estimates

Viano testified that he did some reading and talking to individuals about how the NASS CDS data are collected and correctly stated that the NASS CDS is a stratified sample. The stratified sample design must be included in any computer code used to properly analyze the NASS CDS.

When questioned about the lines in his compute code that were used to generate the standard error, Viano identified coding under the heading of PROC SURVEYFREQ in Exhibit 6 (*Viano Deposition*, 20:1-25; 21:1-5, March 14, 2013). However, in order for the computer code to properly analyze NASS CDS as a stratified sample, it is necessary to specify the sample strata and clusters. Viano's computer code fails to do this. In fact, a well-trained statistician knows that the design of a stratified sample must be explicitly stated so that the standard errors are properly computed using stratified sampling procedures.[17]

As a result of his weak understanding of statistical methods, he selected the *wrong method* to compute all the standard error values in Exhibits 2, 3, 4 and 5.[18] Since I used the same statistical software package as Viano used in his analysis, I was able to reproduce his wrong standard error estimates using his computer code in Exhibit 6.

Next, I corrected Viano's computer code to specify the stratification procedure and produced the corrected standard error values. As an illustration, Table E lists some of the standard errors published in Exhibit 5 by Viano together with the corrected standard errors using the stratification procedure. For example, whereas he stated that the standard error on his estimate of severely injured occupants in front impacts was 4,535, the correct standard error is 18,870, a difference of 416%. The magnitude of this difference highlights the important impact of the stratification procedure on the standard error estimates, and therefore the reliability of Viano's estimates. In fact, throughout his Exhibit 5, his standard errors are consistently smaller than the actual, correct standard errors.

[17]See any sampling design text book for an explanation, e.g., *Sampling: Design and Analysis* by S. Lohr, Pacific Grove: Duxbury Press, 1999.

[18] See Appendix F for the correct computer code, as recommended by federal analysts at NHTSA.

Crash Type	Viano's Standard Error on His National Estimate Using the Wrong Methodology	The Correct Standard Error Using the Stratification Procedure	Percent Increase of Viano's Standard Error When Correct Procedure is Applied
Front	4,535	18,870	416%
Side	5,151	21,267	413%
Rear	983	1,469	149%

Table E. Comparison of Viano's Standard Error in the Last Column of Exhibit 5 and the Corrected Standard Error for the Estimates of Severely Injured Occupants

Contrary to Viano's testimony in his September 11, 2012 deposition, the standard error of an estimate is a measure of the reliability of the estimate. It provides information about both the amount of variability in the data used to construct the estimate and the sample size. When the standard error is large (due to a large amount of variability and/or small sample sizes), relative to the estimate, then the estimate is considered less reliable. When the standard error is small (due to a small amount of variability and/or large sample sizes), relative to the estimate, then the estimate is considered more reliable. By publishing incorrect, smaller standard errors he overstated the reliability of his estimates.

Wrong Methodology for Computing the Standard Error of the Risk Rates

In addition to providing the standard errors on his national estimates of occupants and severely injured occupants, Viano provided standard errors on his risk rates, as shown in the third block of data in Exhibit 5 and Table A. He computed the standard error on his risk by performing a division operation. This is consistent with his remark that he merely did "pure mathematics" (*Id*., 45:15-16). In fact, he did arithmetic and incorrectly computed these standard errors on his risk rates.

Since the risk rate is a statistic, it also has its own standard error. That standard error is not obtained by mere division but must be calculated using calculus and sophisticated statistical software, such as SAS. The mistake made here by Viano is compounded by both his lack of understanding of the risk rate as a statistic and the stratified sample issue discussed above.

A comparison of some of Viano's standard errors from Exhibit 5 and the corrected values is given in Table F. For example, whereas he stated that the standard error on his risk rate of severely injured occupants in front impacts was 0.029%, the correct standard error is 0.079%, a difference of 272%. Throughout his Exhibit 5, his standard errors are consistently smaller than the actual, correct standard errors on his risk rates.

Crash Type	Viano's Standard Error on His Risk Rates Using the Wrong Methodology	The Correct Standard Error Using the Stratification Procedure	Percent Increase of Viano's Standard Error When Correct Procedure is Applied
Front	0.029%	0.079%	272%
Side	0.075%	0.184%	245%
Rear	0.043%	0.058%	135%

Table F.Comparison of Viano's Standard Error of Risk Rates in the Last Column of Exhibit 5 and the Corrected Standard Error for the Estimates of Risk Rates of Severely Injured Occupants

Use of Non-Standard Confidence Intervals

A confidence interval provides a range on the estimate obtained from the NASS CDS. Instead of a single estimate such as the risk rate of 0.99% for severely injured occupants in all impacts, it is possible to construct an interval and say, "The severe injury risk in all impacts in between 0.71% and 1.28% with confidence of 95%." It is standard practice to use 90%, 95% or 99% confidence intervals when performing statistical analysis.[19]

When questioned about his use of a non-standard confidence interval of 68%, he responded that he "absolutely did not" use any confidence intervals on his materials. He said that anyone else can do that, but he did not (*Id*., 128: 1-6).

Again, contrary to his testimony, he used 68% confidence intervals on Figure 15 (page 17) and Figure 1 of Appendix I (page 50) of his July 8, 2012 report (Exhibit 2). Figure 15 is reproduced here as Figure A. Furthermore, when he expressed his opinion in Opinion 56 in Exhibit 3, he stated the risk rates comparing front, side and rear using only one standard error, which is a 68% confidence interval.

Viano was adamant about not including any confidence intervals in his analysis (*Id*., 128: 14-21). In my opinion, there are two fundamental problems with Viano's analysis in this regard. First, all the estimates that he created for his Exhibit 5 are, in fact, statistics and therefore statistical reasoning, such as confidence intervals, should be used for any supporting opinions. Furthermore, Viano used notation, e.g., (0.58%±0.031%), that typically represents a 68% confidence interval with one standard error value, in spite of his testimony that he did not create a confidence interval.

[19] National Institute of Standards and Technology, Engineering Statistics handbook (http://www.itl.nist.gov/div898/handbook/eda/section3/eda352.htm) and American Society for Testing and Materials (http://www.astm.org/SNEWS/JA 2011/datapoints ja11.htm)

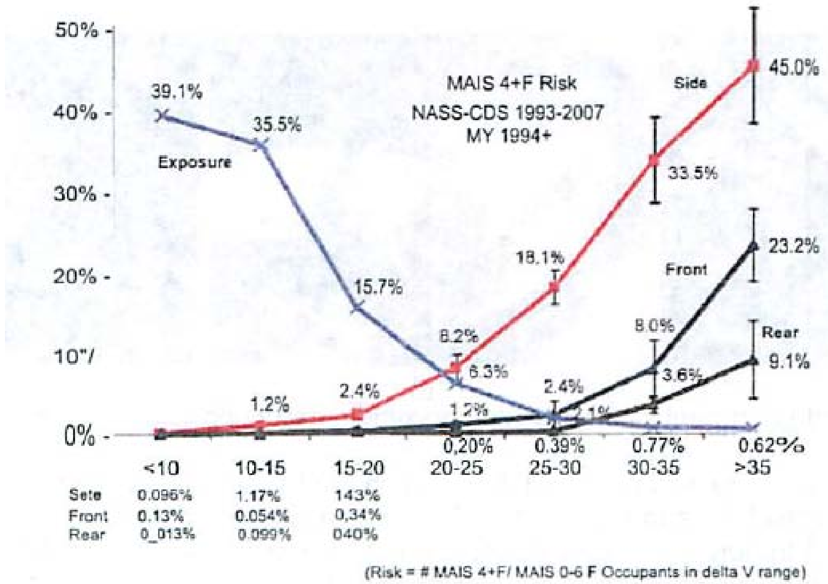

Figure A. Copy of D. Viano's from Exhibit 2 (Figure 15 of His July 8, 2012 Report) with His Caption: "Risk for severe-to-fatal injury (MAIS 4+F) with one standard error bars for NASS-CDS tow-away crashes (Viano, Parenteau 2010)."

When the correct methodology is used in computing the standard errors and when appropriate statistical reasoning is applied to the NASS CDS risk rate estimates in Exhibit 5, the results stand in stark contrast to Viano's conclusion.

- For the occupants in vehicles with computed delta V, when 95% confidence intervals are properly constructed on the severe injury risk rates, there is no statistically significant difference between frontal impacts (0.31%, 0.74%) and rear impacts (0.073%, 0.36%), i.e., the two confidence intervals overlap.
- For all occupants in frontal and rear impacts, the 99% confidence intervals on the severe injury risk rates are (0.36%, 0.83%) and (0.12%, 0.46%), respectively. Consequently, there is no statistically significance difference between the risk of severe injury in frontal and rear impacts, at the 99% confidence level.

In other words, using the appropriate statistical reasoning with the correct procedures, risk rate in rear impacts is not the lowest, as stated by Viano in his Opinion 56 of Exhibit 3.

Incorrect Standard Errors Associated with the Exposure Rates

In the last block of data in Exhibit 5, Viano listed the percentage of occupants exposed in a particular impact direction and delta V range. For example, according to his data 33.6% of the occupants in frontal impacts were occupant in vehicles with a delta V of less than 10 mph. He listed 1.4% as the standard error of this estimate. He obtained this number by another division operation.

There are two problems here. First, ignoring the stratification procedure omission, the standard errors in his Exhibit 5 do not match the standard errors in his own computer printouts in Exhibit 7. Second, standard errors on the exposure rates should be computed using the correct stratification procedure. Table G shows the contrast between some of Viano's estimates and the correct estimates.

Delta V Ranges	<10 mph	10-15 mph	15-20 mph	20-25 mph	25-30 mph	30-35 mph	35-40 mph	40-45 mph	>45 mph
Viano "se"	1.4%	1.2%	1.0%	0.56%	0.18%	0.072%	0.040%	0.024%	0.031%
Correct "se"	1.76%	1.2%	0.63%	0.67%	0.18%	0.096%	0.060%	0.030%	0.025%

Table G. Standard Errors of the Exposure Rates for Frontal Impacts Listed by Viano in Exhibit 5 Compared to the Correct Standard Errors Obtained Using the Stratification Procedure

Widespread Faulty Statistical Analysis

The lack of statistical expertise manifest in these problems of using extremely small samples, incorrect statistical analysis procedure resulting in the wrong standard errors, missing data bias and non-standard confidence intervals are repeatedly found in Viano's publications. Such problems make many of his published results wrong, unreliable, misleading and deceptive.

Every article authored by D. Viano in which the CDS data was analyzed, as reviewed by me, suffers from the same problems described above.[20] Consequently, any of his opinions that reference the following articles are not based on sound statistical analysis. The articles are:

Viano, D. and Parenteau, C. (2008). Fatalities of Children 0-7 Years Old in the Second Row, *Traffic Injury Prevention*, 9:231-237.

Viano, D. and Parenteau, C. (2008).Serious Injury in Very-Low and Very-High Speed Rear Impacts, SAE 2008-01-1485, Society of Automotive Engineers, Warrendale, PA, 2008.

Viano, D. and Parenteau, C. (2010).Crash Injury Risk for Obese Occupants, SAE 2008-01-0528, Society of Automotive Engineers, Warrendale, PA, 2008.

Viano, D. and Parenteau, C. (2010). Severe-to-Fatal Injury Risks in Crashes with Two Front Seat Occupants by Seat Belt Use, *Traffic Injury Prevention*, 11:294-299.

Viano, D. and Parenteau, C. (2010).Severe Injury to Near- and Far-Seated Occupants in Side Impacts by Crash Severity and Belt Use, *Traffic Injury Prevention*, 11:69-78.

Edwards, M. L., Parenteau, C. and Viano, D. (2009). Front-Seat Occupant Injuries in Rear Impacts: Analysis of the Seatback Incline Variable in NASS-CDS, SAE 2009-01-1200, Society of Automotive Engineers, Warrendale, PA, 2008.

[20] It is my intent to continue to review Viano's articles to assess their statistical correctness.

Failure to Review All Available Data

In addition to producing the sampled data in NASS CDS, the federal government compiles crash data on a census of fatal crashes on the nation's roadways. This Fatality Analysis Reporting System (FARS) is primarily based on the police accident report and other data collected by individuals in each of the states and US territories. (See Appendix G for a brief description of FARS.)

Out of the 58 rear-impacted occupants, Viano found only 5 occupants in Dodge Neons with serious or worse injuries in the NASS CDS, as shown in Table 8 of his Exhibit 3.[21]

While the FARS database contains information only on those crashes in which a fatality occurred, I found 233 occupants of Dodge Neons in the FARS database who were either killed or sustained incapacitating injuries in rear impacts in these fatal crashes, as shown in Table H below. Viano testified that he did not look at the injury and fatality data in FARS because the "amount of information available is very sparse" (*Id.*, 30:19-23). However, the FARS data provide some insight into the number of actual occurrences of severe injuries in Dodge Neons in rear impacts.

There is an additional reason not to solely examine the 5 severely injured Dodge Neon occupants in the NASS CDS. That problem centers on representation. Viano used the 5 occupants to determine if there were any serious spinal-skeletal injuries to these occupants. Not finding any such injuries, Viano cannot conclude that such injuries have not occurred. Nor can he support his statement that there are no such cases in the NASS CDS because such injuries occur at less than a rate of 20 per year (*Id*., 69:1-2).

Knowing that 233 occupants were either killed or severely injured in Dodge Neons but not knowing the nature of these injuries makes any conclusions about severe skeletal-spinal injuries in Dodge Neons based on the NASS CDS without foundation.

[21] The 58 occupants have an estimated total of 26,879. The 5 occupants have an estimated total of 249. Of these 5 injured occupants, 3 had severe or worse injuries, analogous to those in Exhibit 5, with a total weighted estimate of 187.

Occupants of Dodge Neon Vehicles with Model Year 1994+		
Calendar Year	Counts of Occupants in NASS CDS Counts from Table 8 of Viano's Exhibit 3 in Rear Impacts	Counts of Occupants in FARS with Incapacitating or Fatal Injuries in Rear Impacts
1994	0	0
1995	0	1
1996	0	8
1997	1	11
1998	0	11
1999	0	14
2000	2	12
2001	0	15
2002	0	21
2003	0	16
2004	1	25
2005	0	31
2006	0	16
2007	0	17
2008	1	13
2009	0	12
2010	0	10
Total	5	233

Table G. Comparison of Numbers Reported by Viano Using the NASS CDS Database in Exhibit 3 and the Numbers Found by Hubele Using the FARS Database

Concluding Comments

Viano repeatedly demonstrated that he lacks the statistical expertise necessary to provide sound statistical testimony using the NASS CDS. His statistical analysis also lacks specificity about the performance of the Dodge Neon and the JCI seatback in rear impacts. As a result, I believe that his attempt to use the NASS CDS to support his opinions about the safety of the Dodge Neon and the JCI seatback in rear impacts is extremely flawed.

I may be asked to supplement this report at a later date. My consulting rate is $375 per hour.

Norma Faris Hubele

Norma Faris Hubele, Ph. D.
Statistical Consultant

Appendix A.

NORMA FARIS HUBELE, PH. D.
REFRAC SYSTEMS
7201 W. OAKLAND STREET
CHANDLER, AZ 85226-2434
(800) 4-REFRAC OR (800) 473-3722
(480) 940-0068 LOCAL
(480) 940-9366 FAX
HUBELE@ASU.EDU

ACADEMIC EXPERIENCE

2007 - Emeritus Professor, Arizona State University (ASU)
2004 - 2006 Director of Strategic Initiatives, Ira A. Fulton School of Engineering, ASU
1998 - 2006 Full Professor in Statistics and Industrial Engineering
Department of Industrial Engineering (IE)
1995 - 1996 Interim Chair, IE, ASU
1992 - 1998 Associate Professor, IE, ASU
1984 - 1992 Assistant Professor, IE, ASU

INDUSTRIAL AND CONSULTING EXPERIENCE

1985 - Refrac Systems, Owner, Quality Control Advisor, Statistical Analyst. Statistical analysis work performed for more than 50 professional and manufacturing companies and extensive work with federal transportation accident databases.

DEGREES

Ph.D. Rensselaer Polytechnic Institute (Troy, NY), Computer and Systems Engineering, 1984
M. S. Rensselaer Polytechnic Institute, Operations Research and Statistics, 1977
B. S. University of Massachusetts (Amherst), Magna Cum Laude, Mathematics, 1975

AUTHORED BOOKS

Montgomery, D., Runger, G. and Hubele, N. F. (2011). *Engineering Statistics*, 5th Edition, Wiley & Sons (top selling text in its field since 2000). 4th Edition in 2007, 3rd Edition in 2003, 2nd Edition in 2000 and first printing in 1998. Translated into Italian, Portuguese and Chinese.

Keats, J.B. and Hubele, N. F., Editors, (1988). *Statistical Process Control In Automated Manufacturing*, Marcel Dekker.

EDITORIAL BOARD MEMBER

Journal of Quality Technology
Quality Technology and Quality Management

PROFESSIONAL SOCIETIES MEMBER

American Statistical Association
Sigma Xi
American Mathematical Society
Association for the Advancement of Automotive Medicine
American Society for Quality Control

PRINCIPAL PUBLICATIONS AND PRESENTATIONS

Hubele, N. And Arndt, M. (2012). "Vehicle Safety Standard Update: A Case Study in a Regulatory Debate Using Statistical Models" *Chance*, 25(4), 4-12.

Hubele, N. and Wood, P. (2010). "The Disappearance of Rank Correlation of Driving Accuracy and Money Earned on the PGA Tour: A Simple Explanation Unrelated to Grooves," *Joint Statistical Meetings Proceedings*, August 1-5, Vancouver, British Columbia, Canada.

Hubele, N. (2009). "A Critical Statistical Review of Some Transportation Safety Studies," *Joint Statistical Meeting*, August 2-6, Washington, DC.

Hubele, N. (2009). "A Critical Review of Rollover Field Accident Data Research and Meta-Analysis," Canadian Multidisciplinary Road Safety Conference XIX, June 7-10, Saskatoon, Saskatchewan, Canada.

Hubele, N. (2008). "A Fatality Risk Study of Rear-Seated Children in Rear Impact Collisions," *Protection of Children in Cars 6th International Conference Proceedings*, December 4-5, Munich, Germany.

Karady, G. G., Heydt, G. T., Gel, E. and Hubele, N. (2008). "Power Circuit Breaker Using Micromechanical Switches," in *Operation and Control of Electric Energy Processing Systems*, EPNES: Electric Power Networks Efficiency and Security, edited by J. Momoh and L. Mili, Hoboken: John Wiley & Sons/IEEE Press.

Arndt, M. and Hubele, N. (2007). "A Streamlined Initial Test Regime Utilizing the NHTSA NCAP Rollover Maneuver," 20th International Technical Conference on the Enhanced Safety of Vehicles (ESV), Lyon, France.

Haag, S., Hubele, N., Garcia, A. and McBeath, K. (2007). "Engineering Undergraduate Attrition and Contributing Factors" *International Journal of Engineering Education*, 23(5), 929-940.

Hubele, N., Abdelaziz, B. and Gel, E. (2005). "A Wald Test for Comparing Multiple Capability Indices," *Journal of Quality Technology*, 37(4), 304-307.

Daniels, L., Edgar, B., Burdick, R. and Hubele, N. F. (2005).“Using Confidence Intervals to Compare Process Capability Indices,” *Quality Engineering*, 7, 23-32.

Karady, G., Heydt, G. T., Gel, E. S. and Hubele, N. F. (2005). “The Utilization of Micromechanical Devices in a Power Circuit Breaker,” *Electric Power Components and Systems*, 33, 1159-1174.

Hubele, N. F., Gel, E. S., Berrado, A. (2005). “Probability Models, Control and Capability Comparison For Flatness Data,” *Proceedings of 2005 NSF DMII Grantees Conference*, Scottsdale, AZ.

Muccino, J. C., Hubele, N. F., and Bennett, A. (2004) “Significance Testing for Variational Assimilation,” *Quarterly Journal of the Royal Meterological Society*, 130, 1815-1838.

Hubele, N. and Vännman, K., (2004) “The Effect of Pooled and Un-pooled Variance Estimators on Cpm When Using Subsamples” *Journal of Quality Technology*, 36(2), 207-222.

Vännman, K., and Hubele, N. (2003) “Distributional Properties of Estimated Capability Indices Based on Subsamples,” *Quality and Reliability Engineering International*, 19, 111-128.

Wang, F-K., and Hubele, N., (2002) “Quality Evaluation of Geometric Tolerance Region in Form and Location,” *Quality Engineering*, 205-213.

Hubele, N. F. (2002) “Discussion” of “Process Capability Indices – A Review, 1992-2000,” *Journal of Quality Technology*, 34(1), pp. 22-22, Panel Organizer and Discussant.

Nahar, P.C., Hubele, N.F., and Zimmer, S.L. (2001) “Assessment of a Capability Index Sensitive to Skewness,” *Quality and Reliability Engineering International*, 17, 233-241.

Wang, F. K. and Hubele, N. F.(2001) Quality Evaluation of Geometric Tolerance Regions in Form and Location, *Quality Engineering*, 14(2), pp. 203-209.

Zimmer, L., Hubele, N. F., and Zimmer, W.J. (2001) “Confidence Intervals and Sample Size Determination for Cpm,” *Quality and Reliability Engineering International*, 17, 51-68.

Piplani, R., and Hubele, N. F. (2001). Enhancement and Evaluation of Pattern Recognition in Control Charts, *International Journal of Quality and Reliability Management*, 18(3), 237-253.

Montesinos, J., Gorur, R. S., Zimmer, L. and Hubele, N. (2000) “Statistical Models for Failure Modes of Polymeric Materials for High Voltage Outdoor Insulation” *IEEE Transactions on Dielectrics and Electrical Insulation*, 7(3), p. 408-415.

Wang, F. K., Hubele, N. F., Lawrence, F., Miskulin, J., and Shahriari, M. (2000) "Comparison of Three Multivariate Process Capability Indices," *Journal of Quality Technology*, 32(3), 263-275.

Dumitrescu, M. and Hubele, N. F. (1999). "An Entropic Framework for the Normal Distribution in Capability Analysis," *Communications in Statistics*. Theory & Methods, 28(6), 1361-1377.

Wang, F.K. and Hubele, N. F. (1999). "Quality Evaluation Using Geometric Distance Approach," *International Journal of Reliability, Quality and Safety*, 6(2), p. 139-153.

Zimmer, L.S., and Hubele, N. F. (1997). “Quantiles of the Sampling Distribution of Cpm,” *Quality Engineering*, Vol. 10, No. 2, pp. 309-329.

Roberts, C., Hubele, N., Henderson, M. and Stage, R. (1997). “Manufacturing Evaluation Using Resource-based Template-free Features,” *Journal of Intelligent Manufacturing*, Vol. 8, pp.323-331.

Ranaweera, D., Hubele, N. F. and Karady, G. (1996). “Fuzzy Logic for Short-Term Load Forecasting,” *International Journal of Electric Power and Energy System*, Vol. 18, No. 4, pp. 215-222.

Cheng, C. and Hubele, N., (1996). "A Pattern Recognition Algorithm for Control Charts," *IIE Transactions*, Vol. 28, pp. 215-224.

Hubele, N. F. and Arndt, M.W. (1996). "A Review of Crash Data Analysis in a Defect and Recall Investigation of the General Motors C/K Pickup Trucks," *Accident Analysis & Prevention*, Vol. 28, No. 1, pp. 33-42.

Rivera, L., Hubele, N. F., and Lawrence, F., (1995). "Cpk Index Estimation Using Data Transformation," *Proceedings of 17th International Conference on Computers and Industrial Engineering*, Phoenix, AZ., March 6, Published as Computers and Industrial Engineering, Vol. 29, No. 1-4, pp. 55-58.

DeKoster, D. R., Morror, K. P., Schaub, D. A. and Hubele, N. F. (1995). "Impact of Electric Vehicles on Select Air Pollutants: A Comprehensive Model," *IEEE Transactions on Power Systems*, (Aug.) Vol. 10, No. 3.

Hubele, N. F. (1995). "Discussion" of article by O. Mohammed, D. Park, R. Merchant, T. Dinh, C. Tong, A. Azeem, J. Farah, and C. Drake, "Practical Experiences with An Adaptive Neural Network Short-Term Load Forecasting System," *IEEE Transactions on Power Systems*, (Feb.) Vol. 10, No. 1, p. 263.

Shahriari, H., Hubele, N. F., and Lawrence, F., (1995). "A Multivariate Process Capability Vector," presented at the 4th Industrial Engineering Research Conference, Nashville, TN, May 24.

Ranaweera, D., Hubele, N. F. and Paplexopoulos, A., (1995). "Application of a Radial Basis Function Neural Network Model for Short-Term Load Forecasting," *IEE Proceeding C* (refereed). Vol. 142, No. 1, pp. 45-50.

Hubele, N. F. and Hwarng, H.B., (1994). "A Neural Network Model and Multiple Linear Regression: Another Point Of View," *1994 Artificial Neural Networks in Engineering, Intelligent Engineering Systems Through Artificial Neural Networks*, Vol. 4, Proceedings of the Artificial Neural Networks in Engineering (ANNIE '94), St. Louis, November 13-16, pp. 199-203.

Ranaweera, D., Karady, G. and Hubele, N. F., (1994). "Comparison of Artificial Intelligence-Based Models for Long-Term Load Forecasting," *Proceedings of International Conference on Power System Technology*, pp. 1158-1162.

Hubele, N. F., (1994). "Load Forecasting: A Contact Sport," Panel Discussion Chair, IEEE PES Summer Meeting, San Francisco, July 24-27.

Zimmer, L. and Hubele, N., (1994). "Example Average Run Lengths for a Cpm Control Chart," 48th ASQC Spring Conference, May, Las Vegas, pp. 577-582.

Hubele, N. F., Beaumariage, T. G., Baweja, G. S., Hong, S. -C and Chu, R., (1994). "Using Experimental Design to Assess a Vision Inspection System Capability: A Case Study," *Journal of Quality Technology*, Vol. 26, No. 1, pp. 1-11.

Peng, T. M., Hubele, and N. F., Karady, G., (1993). "An Adaptive Neural Network Approach to One-Week Ahead Load Forecasting," *IEEE Transactions on Power Systems*, Vol. 8, No. 3, pp. 1195-1203.

Hwarng, H. and Hubele, N. F., (1993). "X-bar Control Chart Pattern Identification through Efficient Off-line Neural Network Training," *IIE Transactions*, Vol. 25, pp. 27-39.

Hwarng, H. and Hubele, N. F., (1993). "Back-Propagation Pattern Recognizers for X-Bar Control Charts," *Computers and Industrial Engineering*, Vol. 24, No. 2, pp. 219-235.

Hwarng, H. and Hubele, N. F., (1992). "Boltzmann Machines that Learn to Recognize Patterns on X-Bar Control Charts," *Statistics and Computing*, 2, pp. 191-202.

Roberts, C., Henderson, M. and Hubele, N. F. (1993). "An Architecture for Feature-based Manufacturing Evaluation Using Statistical Process Control Information," 2nd Industrial Engineering Research Conference, Los Angeles, May 26-27.

Hubele, N. F., Henderson, M. and Roberts, C. (1993). "A Framework for Feature-based Manufacturing Evaluation Using Statistical Process Control Information," Sixth National Symposium Statistics & Design in Automated Manufacturing, ASU, Tempe, AZ, February 17-19.

Hubele, N. F. and Si, J. (1993). "Discussion of Classical Forecasting Techniques and Neural Networks Applied to Short-Term Electric Load Forecasting," IEEE Power Engineering Systems, Winter Conference, Panel Discussion, Columbus, OH, February 2.

Cheng, C-S. and Hubele, N. F. (1992). "Design of a Knowledge-Based Expert System for Statistical Process Control," *Computers and Industrial Engineering*, Vol. 22, No. 4, pp. 501-517.

Peng, T. M., Hubele, N. F., and Karady, G. (1992). "Advancement in the Application of Neural Networks for Short-Term Load Forecasting," *IEEE Transactions on Power Systems*, Vol. 7, No. 1, pp. 250-257.

Evans, M., Hubele, N. F., Gochoel, A., and Cheng, C-S. (1992). "A Case Study of Family Formation for Statistical Process Control of Small Batch Manufacturing," Fall Technical Conference, ASQC.

Hubele, N. F., Hwarng, H.B. and Zimmer, L., (1992). "Capability Indices: Linking TQM with Concurrent Engineering," *1992 International Industrial Engineering Conference Proceedings*, May 17-20, Chicago, pp. 175-179.

Appendix B. National Automotive Sampling System Crashworthiness Data System Overview

National Automotive Sampling System Crashworthiness Data System (NASS CDS or CDS) has detailed data on a representative, random sample of thousands of minor, serious, and fatal crashes. Field research teams located at Primary Sampling Units (PSU's) across the country study about 5,000 crashes a year involving passenger cars, light trucks, vans, and utility vehicles. Trained crash investigators obtain data from crash sites, studying evidence such as skid marks, fluid spills, broken glass, and bent guard rails. They locate the vehicles involved, photograph them, measure the crash damage, and identify interior locations that were struck by the occupants. These researchers follow up on their on-site investigations by interviewing crash victims and reviewing medical records to determine the nature and severity of injuries.

Interviews with people in the crash are conducted with discretion and confidentiality. The research teams are interested only in information that will help them understand the nature and consequences of the crashes. Personal information about individuals - names, addresses, license and registration numbers, and even specific crash locations - are not included in any public NASS files.

The data collected by the PSU's are quality controlled by one of 2 NASS Zone Centers. Each Zone Center, staffed by the most experienced crash researchers, is responsible for half of the PSU field offices. Zone Centers have the responsibility for coordinating and supervising the activities of the field offices, keeping field offices informed regarding changes in functional and administrative procedures, sharing ideas and concepts throughout the system regarding new techniques, procedures, and components found on vehicles and updating field offices regarding changes in system hardware and software.

NASS case review is conducted at the Zone Center and may result in case data being sent back and forth between the Zone Center and the PSU several times until the case passes quality control standards built into the NASS data collection cycle. Once data is approved for inclusion into the NASS database, it will again be subjected to quality assurance checks before becoming publicly released as part of annual NASS data files.

The data collected by the CDS research teams become permanent NASS records. This information is used by NHTSA for a variety of purposes, including:

- Assessment of the overall state of traffic safety, and identification of existing and potential traffic safety problems.
- Obtaining detailed data on the crash performance of passenger cars, light trucks, vans, and utility vehicles.
- Evaluation of vehicle safety systems and designs.
- Increasing knowledge about the nature of crash injuries, as well as the relationship between the type and seriousness of a crash and the resultant injuries.
- Assessment of the effectiveness of motor vehicle and traffic safety program standards.
- Evaluation of the effect of societal changes, such as increased traffic flow and increased large truck traffic.

Content copied from NHTSA website, http://www.nhtsa.gov/Data/National+Automotive+Sampling+System+(NASS)/NASS+Crashworthiness+Data+System on January 25, 2012.

Appendix C. Some Statistical Terminology

A sample is a subset of a large group. In our discussions, we are interested in all types of crashes occurring across the U.S. Since it would be very, very expensive to study all crashes, the federal government has created a way to scientifically sample a representative subset of crashes. Here we are most interested in the NASS CDS described above. The purpose of the sample is to provide estimates about characteristics of crashes.

An estimate, sometimes called an educated guess, is a number. For example, an estimate would be the number of children severely injured in frontal impacts while riding in SUV's. We could use the CDS to obtain such an estimate.

Whether or not the estimate is a good guess depends upon many factors, including the amount of information used to arrive at the estimate. We like to have large, representative samples or large amounts of data to make our estimate. If we cannot have a large amount of data, then we are dependent on a scientifically-designed statistical sample to ensure that we have a good cross-section of data upon which to base our estimate. The NASS CDS is designed to have a good cross-section of data.

The problem that sometimes arises when using the NASS CDS to make a guess or estimate is that individuals 'drill down' or select a very small subset from the original NASS CDS subset of crashes. As we discuss in the body of this report, Viano drilled so far down that he had sample sizes of 1 or 6 or 7 injured occupants. His estimates were based on very little information.

If there is a lot of variety in the data (a good thing because it is representative) but the subsamples selected to make the estimates are very small samples (a bad thing) then the result will be an unreliable estimate. We can measure the degree of reliability of our estimate or guess by examining the standard error. If the standard error is too large, relative to the estimate, then our guess will be unreliable or poor.

The standard error is a reflection of the variability of the data and the sample size. We cannot change the variability of the data, just as we cannot change the variety of crashes. However, we can make sure that our sample size is big enough to make our estimate a good one. While it may be interesting to study the relationship between severe injuries in rear impacts and very fine intervals of delta V, it is not advisable because there is not enough data to obtain large enough samples to make our estimates good and reliable.

A confidence interval is also a measure of reliability. We like to have 90%, 95% or 99% confidence intervals on our estimates because the interval gives the reader an indication of how "sure" the analyst is about the guess or estimate. An interval, frequently shown as a vertical "error bar" on a point on a graph, is a visual aid for seeing the reliability of the estimate (the point). The percent confidence together with the length of the bars reflect the reliability of the estimate.

Appendix D. Detailed Explanation of Exhibit 5, Table A

Most of the data in Exhibit 5 are divided among the various delta V ranges: <10, 10-15, etc. The total of these row entries is contained in the second last column which Viano labeled “Total w/o unk” – meaning the total estimates in which the delta V is known.

Viano’s national estimates of occupants with his standard errors (se) are given in the first block of data labeled “MAIS 0+F all occupants” for all, front, side and rear impacts, as the rows are labeled. For example, in the first entry in line 1 of the table, Viano estimated 6,642,324 occupants were in all impacts with a delta V of less than 10 mph. He estimated his standard error “se” of his estimate as 200,338. For occupants in rear impacts with delta V of less than 10 mph, the corresponding numbers are 463,408 estimated occupants with his standard error of 61,744.

The second block of data labeled “MAIS 4+F all occupants” contains Viano’s estimates of occupants who suffered severe, critical or fatal injuries (for brevity I will call these severely injured occupants). The rows again refer to the national estimates and standard errors for all, front, side and rear impacts. For example, in the first entry, Viano estimated that 7782 occupants in all impacts with a delta V of less than 10 mph were severely injured with a “se” of 1,875. The corresponding numbers for rear impacted occupants are 50 severely injured with a “se” of 50.

The next block of data labeled “Risk MAIS 4+F all occupants” are Viano’s risk rates. He computed these by dividing the number of severely injured occupants by the total number of occupants for each of the impact directions and delta V ranges. For example, in the first entry, Viano estimated a risk rate of 0.12% (7,782/6,642,324) of an occupant being severely injured in all impacts with a delta V of less than 10 mph. The corresponding estimate for rear impacts is 0.011% (50/463,408). Furthermore, he computed a “se” of his rates by dividing his “se” of the severely injured occupants by the total number of occupants. The 0.12% risk rate, according to Viano, has a “se” of 0.028% (1,875/6,642,324). His “se” for the corresponding rear impact injury risk rate is 0.011% (50/463,408).

The last block of data in Exhibit 5 titled “Exposure MAIS 0+F all occupants” are percentages which Viano derived from the first block of data. The rows that are labeled all, front, side and rear contain entries that total to 100% and reflect the fraction of the occupants who were in vehicles with the corresponding delta V. For example, in line 1, Viano estimated that 38.8% (6,642,324/17,119,408) of occupants in all impacts with known delta V were in vehicles with delta V less than 10 mph. The “se” row is computed by Viano as the ratio of the standard error divided by the total number of occupants in that impact direction with known delta V. For example, 38.8% of occupants, according to Viano, have a “se” of 1.2% (200,338/270,784). The corresponding estimates for occupants in rear impacts with delta V less than 10 mph are 28.2% (463,408/1,642,977) with a “se” of 3.8% (61,744/1,642,977).

All the data described above are based on the occupants who had delta V recorded for their vehicle in their crash. As will be discussed below, there are a large number of occupants without this delta V information. Viano provided estimates and “se” values for all occupants, both with and without delta V, in the various impact directions in the last column of Exhibit 5. For example, his estimated 35,819,406 occupants were in the NASS CDS for all impact directions. He computed a corresponding “se” of 405,996. His estimate of severely injured occupants is 356,164 with a “se” of 9,174. The corresponding risk rate is 0.99% with a “se” of 0.026%.

Appendix E. Vehicle Body Types Included in Viano's Exhibits 2, 3, 4, 5, 7
(As listed in the NASS CDS Statistical Analysis Software format files)

01='Convertible(excludes sun-roof,t-bar)'
02='2-door sedan,hardtop,coupe'
03='3-door/2-door hatchback'
04='4-door sedan, hardtop'
05='5-door/4-door hatchback'
06='Station Wagon (excluding van and truck based)'
07='Hatchback, number of doors unknown'
08='Sedan/Hardtop, number of doors unknown'
09='Other or Unknown automobile type'
10='Auto-based pickup (includes E1 Camino, Caballero, Ranchero, SSR, G8-ST, Subaru Brat, Rabbit Pickup)'
11='Auto-based panel (cargo station wagon, auto-based ambulance or hearse)'
12='Large Limousine-more than four side doors or stretched chassis'
13='Three-wheel automobile or automobile derivative'
14='Compact utility (Jeep CJ-2-CJ-7, Scrambler, Golden Eagle, Renegade, Laredo, Wrangler, ...)'
15='Large utility (includes Jeep Cherokee [83 and before], Ramcharger, Trailduster, Bronco-fullsize ..)'
16='Utility station wagon (includes suburban limousines, Suburban, Travellall, Grand Wagoneer)'
17='3-door coupe'
19='Utility Vehicle, Unknown body type'
20='Minivan (Chrysler Town and Country, Caravan, Grand Caravan, Voyager, Grand Voyager, Mini-Ram, ...)'
21='Large Van (B150-B350, Sportsman, Royal Maxiwagon, Ram, Tradesman, Voyager [83 and before], ...)'
22='Step-van or walk-in van (<= 10,000 lbs. GVWR)'
28='Other van type (Hi-Cube Van, Kary)'
29='Unknown van type'
30='Compact pickup (GVWR <4,500 lbs.) (D50,Colt P/U, Ram 50, Dakota, Arrow Pickup [foreign], Ranger, ..)'
31='Standard pickup (GVWR 4,500 to 10,00 lbs.)(Jeep Pickup, Comanche, Ram Pickup, D100-D350, ...)'
32='Pickup with slide-in camper'
33='Convertible pickup'
39='Unknown (pickup style) light conventional truck type'
40='Cab Chassis Based (includes Rescue Vehicle, Light Stake, Dump, and Tow Truck)'
41='Truck based panel'
42='Light Truck Based Motorhome (Chassis Mounted)'
45='Other light conventional truck type'
48='Unknown light truck type (not a pickup)'
49='Unknown light vehicle type (automobile,utility vehicle, van, or light truck)'

Appendix F. Computer Code for Analyzing the NASS CDS Using the Stratification Procedure

From: Austin, Rory <NHTSA>
Sent: Thursday, January 17, 2008 6:55 AM
To: 'hubele@asu.edu'
Cc: Kahane, Chuck <NHTSA>; Kindelberger, John <NHTSA>
Subject: RE: Question about obtaining error bounds on estimates using SAS

Dr. Hubele,

Using SAS 9.1 to get the confidence intervals for PROC SURVEYFREQ is relatively easy. All of the proc survey commands require three elements – the sample weight, the (first stage) clusters, and the (first stage) strata.

So here is the basic proc surveyfreq for GES for one variable. I also added two options to the table statement to produce (95%) confidence limits [the SAS default] for the estimated proportion (cl) and the estimated total (clwt). (There are many other options for the table statement, but these are the basics for a univariate frequency.)

```
Proc surveyfreq data=xxxx;
Table variable /cl clwt;
Weight weight;
Cluster psu;
Strata psustrat;
Run;
```

Here is the basic proc surveyfreq for CDC [sic CDS].

```
Proc surveyfreq data=xxxx;
Table variable/cl clwt;
Weight ratwgt;
Cluster psu;
Strata psustrat;
Run;
```

I hope that this helps.
Rory
Rory A. Austin, Ph. D.
State Data System Technical Manager
US DOT/NHTSA/NVS-412

Appendix G: Fatality Analysis Reporting System Overview

The Fatality Analysis Reporting System (FARS) contains data derived from a census of fatal traffic crashes within the 50 States, the District of Columbia, and Puerto Rico. To be included in FARS, a crash must involve a motor vehicle traveling on a trafficway customarily open to the public and result in the death of a person (occupant of a vehicle or a non-motorist) within 30 days of the crash. FARS was conceived, designed, and developed by the National Center for Statistics and Analysis (NCSA) of the National Highway Traffic Safety Administration (NHTSA) in 1975 to provide an overall measure of highway safety, to help identify traffic safety problems, to suggest solutions, and to help provide an objective basis to evaluate the effectiveness of motor vehicle safety standards and highway safety programs.

Examples of specific FARS data uses include the evaluation of:

- Alcohol-Related Legislation
- Motorcycle Helmet Legislation
- Repeat Offenders
- Restraint Usage Legislation
- Speed Limit Laws
- Vehicle Safety Designs
- Large-Truck Safety
- Air Bag Effectiveness

FARS data can be used to answer a multitude of questions concerning the safety of vehicles, drivers, traffic situations, roadways, and environmental conditions. FARS data is also used at the State level by the FARS analyst to respond to State safety issues.

Content copied from NHTSA website,
http://www-nrd.nhtsa.dot.gov/Pubs/FARSBrochure.pdf on January 25, 2012.