

3

*INJURY BIOMECHANICS RESEARCH
Proceedings of the Thirty-First International Workshop*

Risk Curve Boundaries

L. DiDomenico and G. Nusholtz

This paper has not been screened for accuracy nor refereed by any body of scientific peers and should not be referenced in the open literature.

ABSTRACT

Many and multiform are the experiments that have been conducted to allow extraction of injury risk functions. In particular, human surrogate impact test results are commonly mapped into injury risk curves for the purpose of characterizing the stimulus injury response over a given range. However, it is not always clear that the quality and quantity of the data along with the experimental design are sufficient to reliably determine the desired outcome. Therefore it is desirable to obtain some form of guidance, or heuristic rules, as to the usability and appropriateness of injury risk curves with respect to sample size, stimulus distribution over the critical range, censoring, shape of the underlying risk function, and the inclusion of “actual” (uncensored) along with censored data. To accomplish this goal the Consistency threshold and Extended Consistency threshold methods along with Monte Carlo simulation are used to evaluate the experimental design of human surrogate testing. The results imply that the total amount of tests needed to generate a risk curve with a given confidence bound is dependent on the shape of the risk function along with the stimulus distribution over the critical range. This dependence can also be a function of the relative contribution of censored and actual data. However, the results from this analysis also indicate that for “large” biomechanical injury data sets there is no advantage to using actual data; censored data will yield the same injury risk curve as actual data. Therefore, for “small” biomechanical injury data sets the inclusion of actual data will significantly improve the quality of the resulting risk curve but not for large data sets. Confidence intervals are presented for the thoracic injury risk and the head injury risk to show the influence of data distribution on the goodness of the risk function estimation.

INTRODUCTION

The development of human tolerance levels is a difficult task, but one that is essential for the assessment of automotive restraint systems. The main reason for its difficulty is the necessity to obtain injury tolerance information through indirect methods such as Post Mortem Human Subjects (PMHS) or animal testing resulting in undefined system specification errors. In addition,

the limited amount of data, the high variability, and the inherent complexity of performing biomechanical (PMHS) tests presents additional challenges. Along with the quality and the quantity of injury data, the characteristics of the data and the experimental design are critical factors for the goodness of the risk curve estimation itself.

For example, an important characteristic of many biomechanical impact data sets is their censored nature (Mertz, 1996). Censored data are biased in one direction or another. The sign of the bias is known but not the magnitude. This complicates the applications of conventional techniques since these methods assume data to be free from bias. Lately, some biomechanical laboratories have started using devices, (such as an acoustic device) that can obtain the actual time and stimulus in which the injury occurs during a test. The availability of actual data (instead of only censored data) should theoretically improve the estimation of the underlying risk function.

A common approach to estimate the underlying (unknown) distribution from censored/actual biomechanical data is to assume a form for the injury risk distribution (Ran et al., 1984, Hertz, 1993). The parameters that control the shape of the assumed distribution are fit using some kind of optimization technique such as maximum likelihood or least squares. However, many of the advantages and desired properties of parametric approaches are lost if the form of the risk distribution is not appropriate.

Moreover, in a previous paper (Di Domenico et al., 2003), it has been shown that the injury risk function of biomechanical injury data could present drastic changes in slope. For example, it could resemble a step function in some part of the stimulus interval used. In the same paper, it has also been emphasized that usually there is no basis for the underlying distribution - all these characteristics should encourage the use of non-parametric models specific for censored/mixed data. A non-parametric estimate (appropriate for censored/mixed data and with the maximum likelihood property) has the flexibility to be closer than the parametric estimate to the actual distribution, in particular at the extreme values of the stimulus - since it is not constrained by a prior specified risk form.

Among all the non-parametric methods for censored/actual impact data, the only one that has the appealing statistical property of being a maximum likelihood estimate is the Consistent-Threshold (CT) for double censored and the Extended-Consistent-Threshold (ECT), for a mix of double-censored and actual data (Nusholtz et al., 1999). For all the reasons above, the CT/ECT estimation methods have been considered for this simulation study. In this paper, injury risk curves will be analyzed in terms of their usability and appropriateness with respect to: sample size, censoring, shape of the underlying risk function, the inclusion of actual along with censored data, and stimulus distribution over the *critical range* (i.e. the interval that spans from the highest stimulus value for which there is no injury to the lowest stimulus value for which there will be always an injury).

NOTATIONS AND TERMINOLOGY

A stimulus value associated with a specific test is considered to be *right censored* if neither serious-injury (i.e. MAIS 3+) nor death is observed at the end of the test itself. For such tests, the loading stimulus values used were not enough to cause the event of interest (injury/death), and for this reason these tests are also called losses. If at the end of the test an injury/death is observed, the test is considered to be *left censored*, since the actual (unknown) threshold value that caused the injury is at most equal to the loading stimulus value used, i.e., the actual value is to the left of the one used. In some cases, it is possible to infer from the time history of an injury test the exact (actual) threshold value that causes the injury, in such case, the test data is called *actual*. A data set that includes both *right censored* and *left censored* observations is called *doubly censored*. A data set that includes *right censored*, *left censored* and actual observations is called *mixed*.

The following notation will be used:

- T a random variable denoting the loading stimulus threshold value (e.g., Axial Femur Force, Thoracic Peak Acceleration, HIC value, etc.) at which the event (injury) occurred.
- t_1, t_2, \dots, t_n a sequence of loading stimulus values used by the experimenter.
- d_1, d_2, \dots, d_n the number of left-censored (serious injuries/deaths) observations at each loading stimulus value. There were d_k deaths (serious injuries) at the end of a test in which the experimenter used a maximum stimulus value t_k , with $k=1,2,\dots,n$. This implies that for d_k tests the actual threshold value is at most t_k .
- l_1, l_2, \dots, l_n the number of losses at each loading stimulus value. There were l_k losses at loading stimulus value t_k .
- a_1, a_2, \dots, a_n the number of death/serious injuries at each loading stimulus value. There were a_k deaths (serious-injuries) at loading stimulus value t_k .
- $S(t)$ is the survival function at $T=t$, that is, $S(t) = Prob\{T = t\}$.
- $R(t) = Prob\{T < t\}$ is the cumulative injury probability (risk function) at the loading stimulus value $T=t$.
- \tilde{R}_k will denote either CT or ECT Risk function estimate at t_k .
- \tilde{S}_k will denote either CT or ECT Survival function estimate at t_k .

DESIGN OF THE SIMULATION STUDY

The simulation study was designed so that it reflects practical applications of biomechanical data risk estimation. Synthetic (hypothetical, artificial) data was created that mimics the results from a series of biomechanical tests in terms of a peak stimulus result and if injury occurred. The synthetic data does not represent any single test series but resembles a generic (hypothetical) test series. For example, assume a three-point bending test in which a femur is supported at both ends and a weight is dropped in the middle. The peak of the contact force threshold of the weight would be the stimulus and if there is a fracture it is an injury. The synthetic data is created through a Monte Carlo method.

Underlying Risk Functions

Before choosing a specific distribution for the underlying risk function, it is necessary to study if trends and insights from the Monte Carlo simulation could be sensitive to this choice. To address this issue, both discontinuous and parametric risk functions were used; for example, the parametric models considered include the Gamma, the Weibull, the lognormal, etc. (Figure 1). It has been noticed that the general -CT estimation error- result did not depend on the risk function family chosen, but it depends on the overall rate of increase over the stimulus range. This is mainly due to the fact that CT is a non-parametric estimation method so that its performance is not influenced from the particular underlying (parametric) risk function used. As such, no loss of generality will occur if only one set of risk functions are used - the normal risk functions.

Therefore, the focus will be on the results for three normal risk functions: all three with mean 10, but with different standard deviation ($stdv$), i.e., $stdv=1,3,5$ (Figure 2). Thus, all three risk functions considered have the same mean survival stimulus value (10 units) but the risk (of injury) increases at different rates (as the stimulus value is increased). The slowest risk increase is represented by a standard deviation of five and the steepest risk increase is represented by a standard deviation of one. If any biomechanical stimulus risk function would be normalized to a

stimulus range of 20 units, then the range of standard deviation values considered encompasses the common/usual rate of risk increase observed for biomechanical injury data.

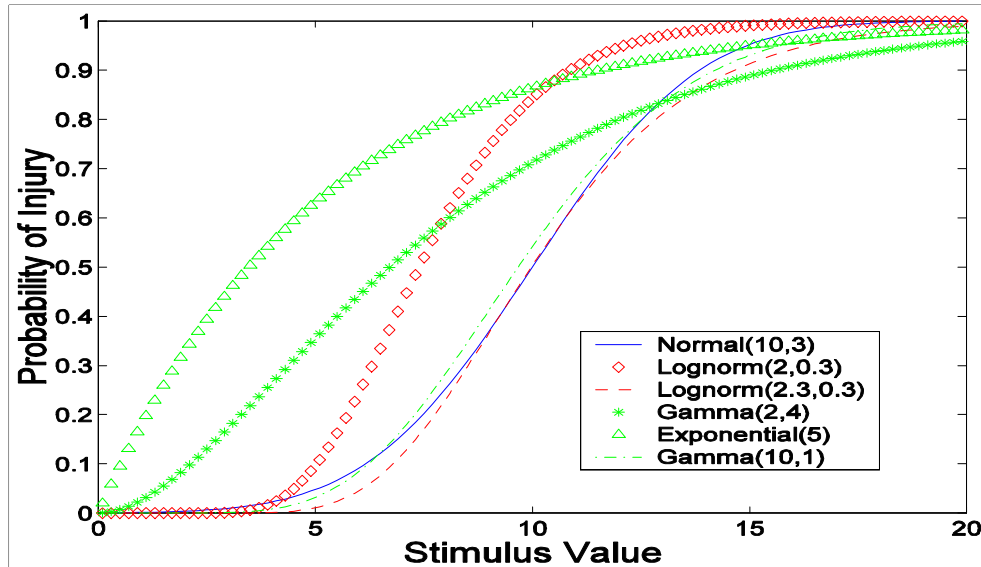


Figure 1: Common Risk function models.

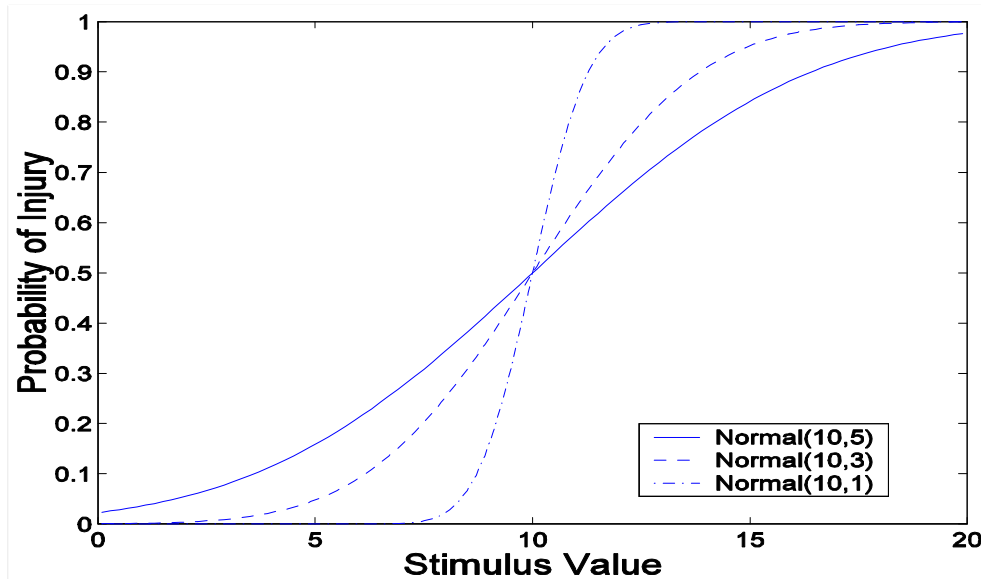


Figure 2: Risk functions used in the simulation study.

Sample Size (N) and Number of Actual Data (A)

Rarely biomechanical impact data sets get above 100 samples, and in many cases they are significantly smaller. To encompass the general range of sample size, a factorial arrangement with sample size $N=10,20,30,40,50,60,70,80$ and actual sample size $A=5*i$, for $i=0,1,2,\dots,16$ is used. In particular, a data set in which $N=20$ and $A=0$ (i.e. $i=0$), consists of 20 samples all censored (i.e. no

actual data). While, a data set in which $N=30$ and $A=5$ (i.e. $i=1$) consists of 30 samples of which 5 are actual and 25 are censored.

Simulation Procedure

N samples (pseudo-random numbers) of possible *injury values* were generated from the normal underlying risk function with mean 10 and *stdv* (where $stdv=1,3,5$). Of the N samples, only A will be used in the risk analysis as actual data; the remaining $(N-A)$ will be censored. This implies that $(N-A)$ “censored values”, instead of the actual values, will be used in the risk estimation. To generate the *censored test* value, pseudo-random numbers were uniformly drawn from the interval (2, 20). Then, each of the $(N-A)$ *pseudo-normal injury* values was compared with a *pseudo-uniform* number (test value). If the sampled *injury value* was below the sampled (*censored*) test value the observation was considered to be an injury (left-censored), otherwise it was a no-injury (right-censored).

METHODS

Underlying Risk Estimation

For each combination of N , A and *stdv*, the underlying risk normal function was estimated using a non-parametric method: The CT method when all the data were censored ($A=0$) and the ECT method when some of the data were actual ($A>0$). The CT method is a non-parametric maximum likelihood method that provides an estimate of the distribution function of *doubly censored* data. A description of the CT method and an easy algorithm to compute the CT estimate of the risk function of doubly censored data is presented in (Di Domenico et al., 2003). The ECT method is the *extended* form of the CT method that can be used to handle data sets that include both *doubly censored* and actual observations (see Appendix). The method to evaluate the confidence intervals for the CT/ECT estimates is also presented in the Appendix.

Evaluation of Risk Estimation Error

The risk estimation performance has been evaluated using two different metrics, L_1 error and L_2 error:

$$L_1 = \int |\tilde{S}(x) - S(x)| dx = \int |R(x) - \tilde{R}(x)| dx \quad (1)$$

$$L_2 = \int |\tilde{S}(x) - S(x)|^2 dx = \int |R(x) - \tilde{R}(x)|^2 dx \quad (2)$$

In particular, the results are presented in terms of percentage of L_2 (error) defined by:

$$PercentageL_2 = 100 \cdot \frac{\int \{\tilde{S}(x) - S(x)\}^2 dx}{\int S(x)^2 dx} \quad (3)$$

The L_2 error is the most used metric to evaluate the estimate performance, and by the Holder inequality (Chung , 1974), the following relationship will hold between L_1 and L_2 :

$$L_1 \leq \sqrt{\{20 \cdot L_2\}} \quad (4)$$

so that the upper bound of the L_1 error could be estimated from the L_2 error. However, the exact L_1 error is also reported since the mean stimulus value that will cause an injury (μ) is related to the survival function by the following (see Appendix):

$$\mathbf{m} = \int S(x)dx = \int |S(x)|dx \quad (5)$$

Thus the L_1 error will provide an upper bound estimate of the error that occurs when the mean stimulus value μ is evaluated using the risk estimate (instead of the real underlying risk function). For each combination of N , A , and $stdv$, 100 simulations were run to evaluate the average L_1 error and L_2 error incurred in estimating the underlying risk function.

RESULTS

In this section, both the results of the simulation analysis (first subsection) and the results of “real” biomechanical data risk analysis (second subsection) are reported.

Results of the Simulation Analysis

Figures 3 to 10 present the mean L_1 error and the (percentage of) mean L_2 error of the risk function estimation as function of the sample size. From these plots it is evident that the estimation error is a decreasing function of the total number of biomechanical data (censored and actual) used for the estimation. Moreover, for the same number of samples, the L_1 and L_2 errors decrease as the number of actual data available increases, or in other words, as some of the censored data are substituted with their actual values.

In each plot, the distance between the lines decreases as the sample size increases. This implies that the quantity of information conveyed by the actual data beyond the one provided by the censored data decreases as the total sample size increases. In particular, availability of actual data is very influential to improve the estimation risk based on limited (less than 20) censored injury samples. Increasing the sample size decreases the difference, in terms of error estimation, between actual and censored injury data. When the biomechanical data set is moderately large (more than 50 injury points), there is no substantial improvement in terms of estimation performance when some of the censored data is substituted with actual injury data (Figure 9). Also, when the available biomechanical data consists of about 30 samples of which about half are actual injury data there is no significant improvement as the percentage of actual data is increased.

Comparing Figure 3 with Figures 4 and 5 it is noticed that the diamond-solid line in Figure 3 is above the diamond-solid line in Figure 4 and it is consistently above the diamond-solid line in Figure 5. In general, for each type-line considered, the line plotted in Figure 3 is above the one (of the same type) plotted in Figure 4 and this is above the line (of the same type) plotted in Figure 5. Since the only difference among the three plots (Figures 3, 4 and 5) is the standard deviation value of the underlying risk function (i.e. $stdv$ ranging from 1 to 5 units), the estimation error is a function of the shape of the underlying risk function: the greater the average slope of the risk function the smaller the number of samples (censored, actual, or a mix of both) needed to obtain a good estimation. Also, the greater the average slope of the risk function the greater the difference in estimation performance between censored and actual injury data.

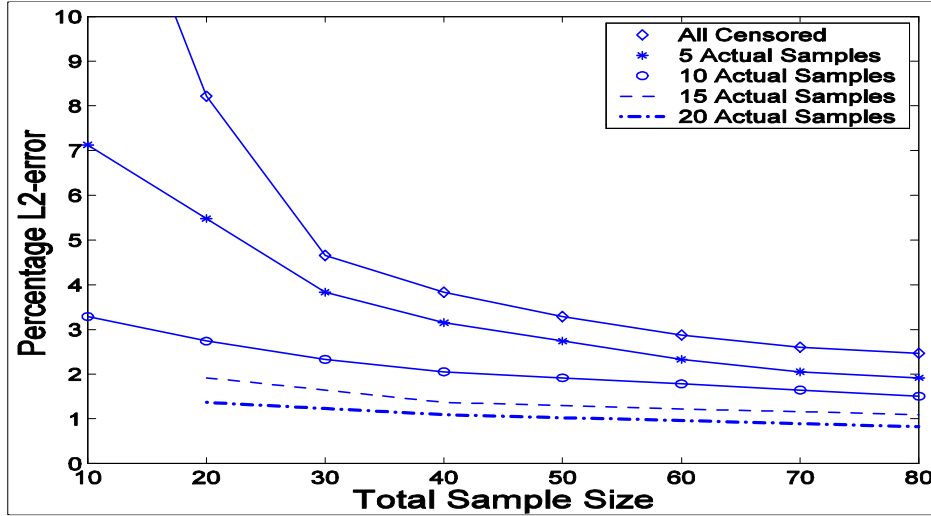


Figure 3: Results of the simulation study. Percentage of EQ L_2 error as function of the total number of samples for the case in which the standard deviation of the underlying risk function is five, i.e., $stdv=5$.

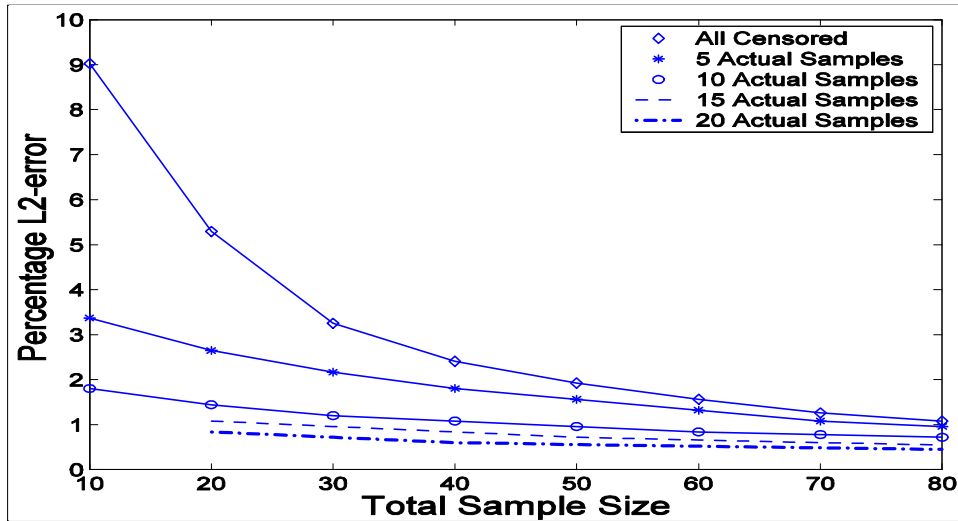


Figure 4: Results of the simulation study. Percentage of EQ L_2 error as function of the total number of samples for the case in which the standard deviation of the underlying risk function is three, i.e., $stdv=3$.

In Figures 3-10 are plotted the estimated mean L_1 and L_2 errors; the variance of these estimated errors can be substantial when the number of samples is limited. In general, the variance of the L_2 error and L_1 error decreases as the number of samples (or the number of actual data) increases. The estimated 95% upper bounds for the percentage of L_2 error are reported in Table 1 (Appendix). Table 1 shows that to insure (at 95% confidence level) an estimation error less than 5%, it is necessary to have either about 60-70 censored data or 20 actual data when the underlying risk function is slowly increasing, (i.e. $stdv = 5$); less data is needed for steeper risk functions (i.e. $stdv = 1, 3$).

Finally, in all the simulations the threshold injury values used by the experimenter are assumed to be evenly (uniformly) distributed in the stimulus range where the risk function

increased from about zero to almost 100% (critical range). Usually this does not happen for the real data, where the data could contain gaps (see, for example, the injury data analyses reported in the next sub-section). A preliminary analysis has shown that moderate changes of the threshold stimulus range do not affect the final results. More precisely, if the threshold region considered includes the region where the risk ranges from 10% to 90%, then there is no substantial difference in the estimation performance. However, if the stimulus range considered is restricted/shrunk or shifted substantially, the (overall) estimation error will increase; the further the considered range is decreased/shifted, the greater the (overall) estimation error will be. In particular, the error will be substantial in the region where there is a limited amount of data to perform the risk estimation. For example, if we consider the risk estimation for the maximum normalized thoracic deflection (Figure 11) or the peak head acceleration (Figure 12), the confidence intervals for the lower risk estimations are larger than the ones for the higher risk estimation (for which more data is available).

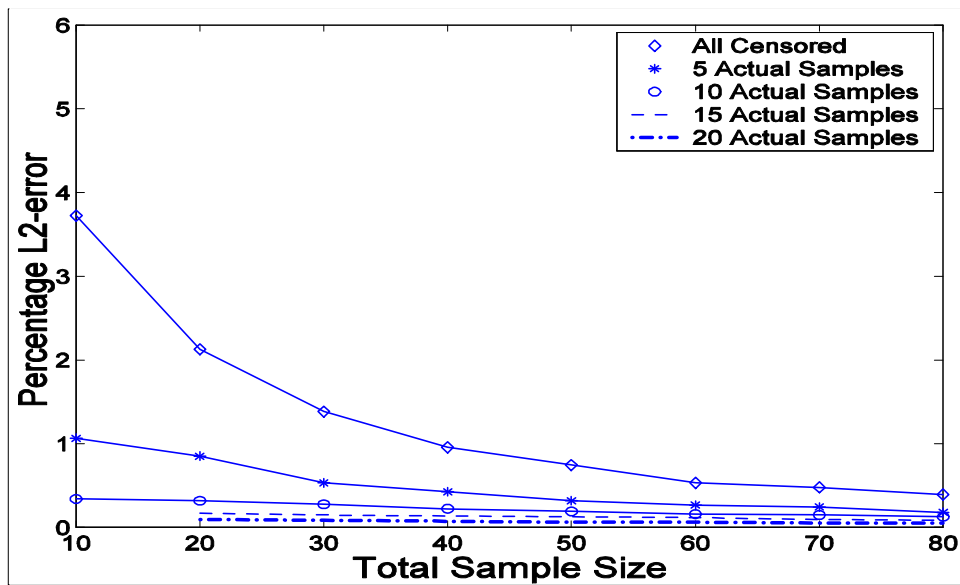


Figure 5: Results of the simulation study. Percentage of EQ L_2 error as function of the total number of samples for the case in which the standard deviation of the underlying risk function is one, i.e., $\text{stdv}=1$.

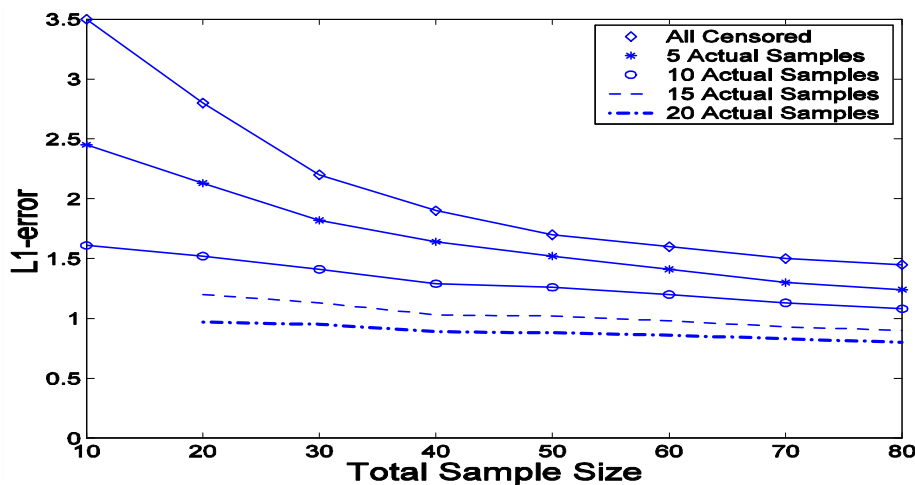


Figure 6: EQ L_1 error as function of the total number of samples when the standard deviation of the underlying risk function is five, i.e., $\text{stdv}=5$.

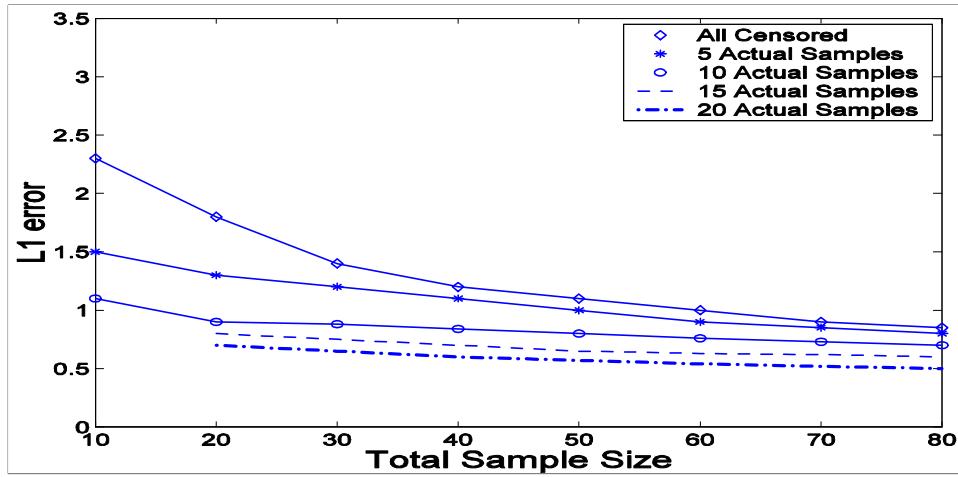


Figure 7: EQ L_1 error as function of the total number of samples when the standard deviation of the underlying risk function is three, i.e., $stdv=3$.

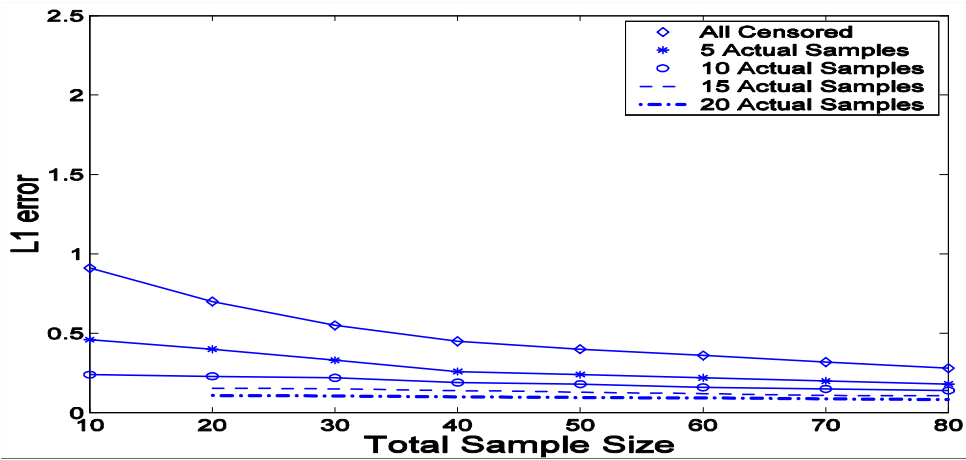


Figure 8: EQ L_1 error as function of the total number of samples when the standard deviation of the underlying risk function is one, i.e., $stdv=1$.

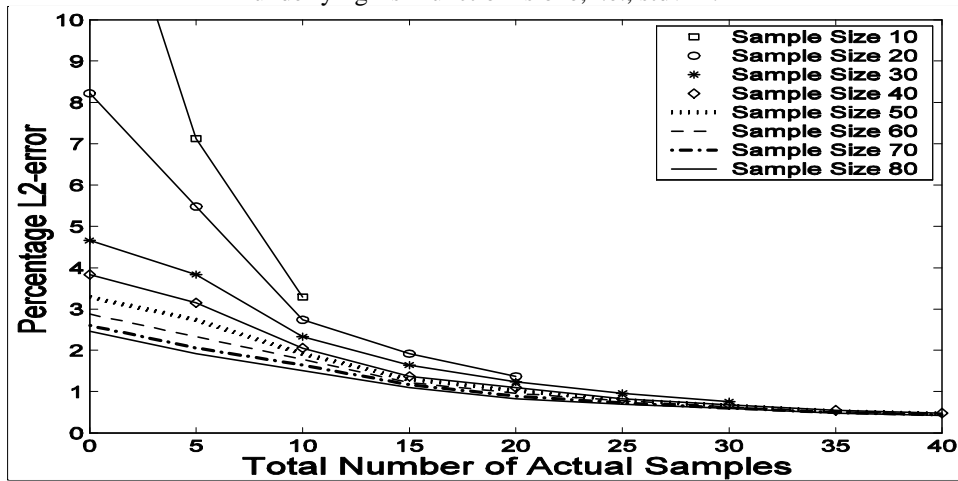


Figure 9: Percentage of EQ L_2 error as function of the total number of actual samples (A) when the standard deviation of the underlying risk function is five, i.e., $stdv=5$.

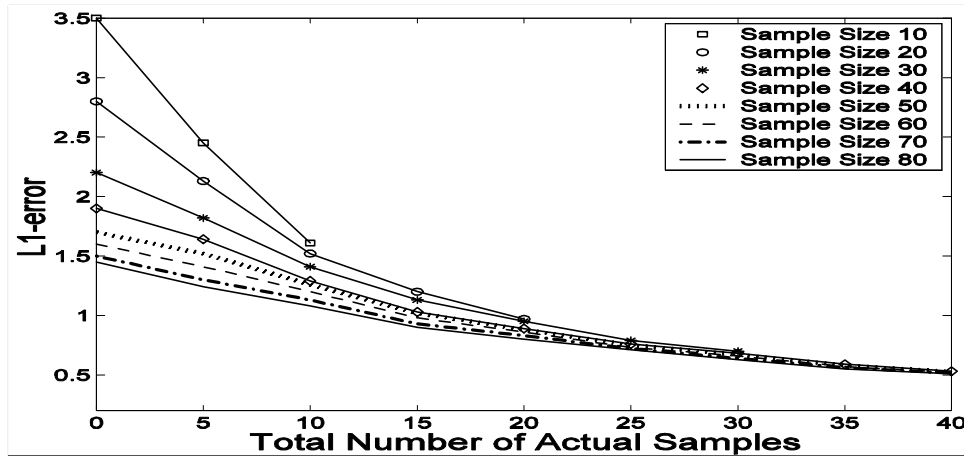


Figure 10: EQ L_1 error as function of the total number of actual samples (A) when the standard deviation of the underlying risk function is five, i.e., $stdv=5$.

Results of the Biomechanical Data Injury Risk Estimation

Thoracic Injury Risk Curve.

Thoracic risk functions and tolerance limits have been based on results from the analysis of seventy-one PMHS frontal impact tests. Details of the experimental procedure and test results can be found in Kuppa et al. (1998). As suggested by Prasad (1999) the following tests have been removed from the data set for analysis purposes: ASTS96, ASTS97, ASTS103, ASTS250, ASTS259 (performed at University of Virginia) and the test RC104 (performed at the Medical College of Wisconsin). Thus six tests have been removed and the remaining tests have been augmented with the tests from Kallieries et al. (1995). This final data set has been used to perform the CT estimated risk curve of serious-to-fatal PMHS thoracic injury as a function of maximum normalized deflection. The CT estimated risk of PMHS thoracic injury and the 90% confidence intervals are presented in Figure 11. The confidence intervals are large for the lower stimulus values where the amount of data is limited and it is not uniformly distributed.

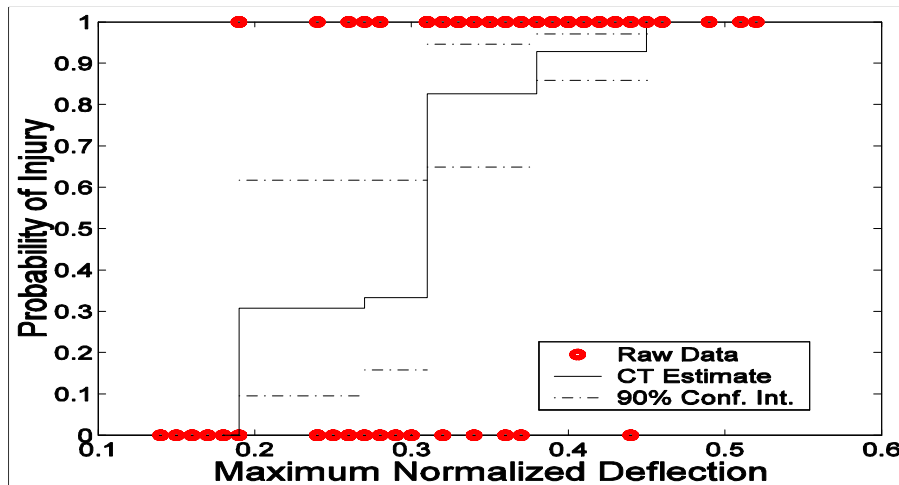


Figure 11: CT estimate and Confidence Intervals for the Serious-to-Fatal Thoracic Injury Risk.

Head Injury Risk Curve.

In Figure 12 is presented the CT estimated PMHS head injury risk and the 90% confidence intervals as a function of the peak resultant head acceleration. The risk estimation is based on the PMHS data reported in Mertz et al. (1996). Also, in this case the confidence intervals are large for the lower stimulus values where the amount of data is limited. For the higher stimulus values the estimation is mainly affected by the sparseness of the data.

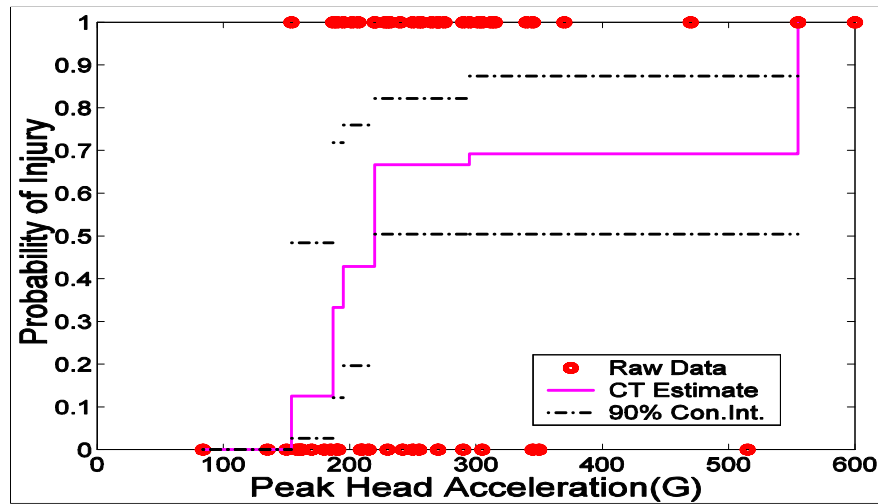


Figure 12: CT estimate and Confidence Intervals for the Serious-to-Fatal Head Injury Risk.

DISCUSSION

The risk estimation error appears to be a function of many parameters representing a response surface: the number of actual biomechanical data used in the estimation, the size of the data set (actual and censored), and the shape of the underlying risk function. This study shows that the estimation error will decrease as the sample size increases and this occurs regardless of the number of actual data. Also, the analysis indicates that the estimation error is related to the shape of the underlying risk function - the steeper the underlying risk function, the smaller the sample size needed to obtain a reliable estimation of the risk function itself, and the more informative will be the actual data with respect to the censored data. This result is intuitively easy to explain because the steeper the risk function, the “easier” it should be to discriminate the injury cases from the non-injury cases, up to the (unrealistic) situation in which it is possible to identify one single stimulus value that will discriminate the injury and non-injury tests (i.e. for which every loading force value above that value will cause an injury and every loading force value below will not cause an injury). An example of a sharply increasing risk function is provided by the Neck injury risk function as a function of the peak tension (Nusholtz et al., 2003).

In addition, it has been noticed that the error reduction effect of augmenting censored data with actual data depends substantially on the sample size and on the percentage of actual data already included in the data. The availability of actual data is very influential when the total sample size is relatively small. However, it should be emphasized that if the sample size is too small the overall error could be too high for practical use even if all the injury data is actual data. As a first order approximation, the error reduction for 20 censored samples can be approximated by

a factor of three if all samples are actual. But, when the available biomechanical data consists of more than 30 samples and approximately half of them are actual, there is not significant improvement in increasing further the percentage of actual data. When the sample size is large enough (more than 50 data points) it is almost irrelevant to have all censored data or a mix of censored and actual data or even all actual data (i.e. dashed, dash-dot and solid lines in Figure 9). Thus, augmenting an already large biomechanical injury data set with actual injury data should not lead to a substantial improvement in the risk estimation goodness. In other words, it is not crucial to acquire biomechanical actual data for the injury risk estimation of a body region for which a large censored data set is already available. Also, the efficiency and the complexity of the technical procedure may make obtaining actual data difficult or impossible.

In all the simulations it is assumed that the threshold values used by the experimenter are uniformly distributed in the critical range so that the available data (information) on which the risk estimation is based covers evenly the critical risk range. Although this is the best scenario, biomechanical injury data is usually not evenly distributed in the critical risk range. In some cases the available biomechanical injury data covers only a portion of the risk function (critical) domain so that it is possible to estimate, with high confidence, only a portion of the risk function. For example, the experimenter could decide to run the tests using mainly medium-high stimulus values so that the available injury data could be mainly spread in the region where the injury risk is medium/high. The risk estimation error for the low risk stimulus value (based on these tests) could be substantial but it will decrease as we approach the higher risk stimulus values. Therefore, the overall (L_2 and L_1) risk estimation error will be mainly driven by the risk estimation at the lower stimulus values, i.e., large confidence intervals at the lower stimulus values, see Figures 11 and 12.

Finally, in this study the risk estimation has been performed using non-parametric methods: CT for doubly-censored data and ECT for mixed data. For this simulation study, the logistic method (or a parametric risk analysis) would have given results that are too liberal/optimistic, since the underlying distribution was chosen to be normal. The two methods used (CT/ECT) are non-parametric methods so that they are more independent of the form considered for the underlying risk function. Thus, the results can have a much broader applicability.

CONCLUSIONS

Development of injury criteria is often based on the injury risk estimation of PMHS specimens. Often these specimens are censored and their number could be limited. A simulation study has been performed to understand which factors are critical in the injury risk estimation and in which scenario the availability of actual biomechanical injury data could be relevant. This analysis shows that the injury risk estimation error depends on many factors: the total sample size, the number of actual data, the shape of the underlying risk function, and the way in which the censored data are distributed in the critical range. In general, the estimation error decreases as either the total sample size or the number of actual data increases. However, the quantity of information conveyed by the actual data beyond the censored data decreases as the sample size increases, and for a sample size large enough, there is no substantial difference between an estimation based on all censored, a mix of censored and actual, or all actual data. Also, the more sharply increasing is the underlying risk function, the smaller is the sample size needed to obtain good risk estimations and the less influential is the availability of actual injury data. Moreover, for censored data, the distribution of the threshold stimulus used to run the experiment is relevant; in the threshold region where there is not enough data available, the estimation performance could be very poor (i.e. large confidence intervals) so that the overall estimation error could be high.

Confidence intervals are presented for the thorax and the head (CT) estimated injury risk to show the influence of data distribution on the goodness of the risk function estimation.

APPENDIX

The Product Limit

The product limit (PL) is the standard method for non-parametric estimation of the survival function when the data contain losses but it is not doubly censored. First, the range of the data is partitioned by a sequence of stimulus threshold values, t_1, t_2, \dots, t_n where each t_k is the stimulus threshold value of at least one death or at least one loss, where, for now, we assume that deaths are not censored. Then, we have:

$$\Pr ob\{T > t_k | t_{k-1}\} = 1 - \frac{d_k}{Y_k} \quad (6)$$

where,

$$Y_k = \sum_{i=k}^n d_i + \sum_{i=k-1}^n l_i \quad (7)$$

Therefore, we have the estimate:

$$S_0^+ = 1 \quad (8)$$

$$S_k^+ = \prod_{i=1}^k \left(1 - \frac{d_i}{Y_i}\right) \quad (9)$$

As one can see from these formulae, the PL estimate at t_k is formed by computing the probability of a threshold value falling in the interval $(t_k, t_k]$ given that it is larger than t_{k-1} and taking the product with the PL estimate for S_{k-1} . For a more detailed discussion see (Kaplan et al., 1958, Turnbull, 1974).

The Extended Consistent Threshold Method

For the PL method, death stimulus values are assumed to be known exactly. Suppose now that only some deaths/serious-injury values are known exactly. The others are left censored. Denote by d_k^0 those deaths at stimulus values which are known exactly and denote by d_k those deaths that are censored.

Suppose also that we know the values of S_0, S_1, \dots, S_n . Consider a censored death at t_k . We know only that the threshold stimulus value lies at or below t_k , but since we know the distribution, we also know that

$$\Pr ob\{t_{j-1} < T < t_j | T \leq t_k\} = \frac{S_{j-1} - S_j}{1 - S_k} \quad (10)$$

Set

$$\mathbf{a}_{k,j} = \frac{S_{j-1} - S_j}{1 - S_k} \quad (11)$$

What we intend to do, instead of counting a censored death in any one interval, is to count it partially according to the above probability formula in every interval. Our new derived count will be considered not censored and we will thus be able to use the PL method. Define:

$$\mathbf{d}'_j = d_j^0 + \sum_{i=j}^n d_j \mathbf{a}_{ij} \quad (12)$$

$$Y'_j = \sum_{i=j}^n \mathbf{d}'_j + \sum_{i=j-1}^n l_j \quad (13)$$

$$\hat{S}_0 = 1 \quad (14)$$

$$\hat{S}_j = \left(1 - \frac{\mathbf{d}'_j}{Y'_j} \right) \hat{S}_{j-1} \quad (15)$$

These equations will provide us with a new estimate of the distribution, but we would like this new estimate to be the same as the one that we already have. This property of a distribution was defined by Efron (1967), and is called self-consistent. To find a self-consistent distribution we need to substitute \hat{S}_k for every S_k in the above equations and solve with the additional condition that $1 = \hat{S}_0 \geq \hat{S}_1 \geq \dots \geq \hat{S}_n \geq 0$. Turnbull (1974) suggests simple iteration, beginning with an initial guess of the PL estimate, ignoring left-censored deaths. It does seem to converge quickly. This procedure provides the unique maximum likelihood estimate of S_0, S_1, \dots, S_n (Ayer et al., 1955; Turnbull, 1974).

Confidence Intervals for CT

For the double censored data, Turnbull (1974) suggested as an estimator of the variance matrix of $\tilde{S}(t)$, the matrix $V=(V_{ij})$, where $V_{ij} = Cov(\tilde{S}(t_i)\tilde{S}(t_j))$, defined as follows. Let

$$A_{ij} = \frac{a_i}{(\tilde{S}(t_{i-1}) - \tilde{S}(t_i))^2} + \frac{a_{i+1}}{(\tilde{S}(t_i) - \tilde{S}(t_{i+1}))^2} + \frac{r_i}{(\tilde{S}(t_i))^2} + \frac{l_i}{(1 - \tilde{S}(t_i))^2} \quad (16)$$

where $i=1, 2, \dots, m-1$;

$$A_{mm} = \frac{a_m}{(\tilde{S}(t_{m-1}) - \tilde{S}(t_m))^2} + \frac{r_m}{(\tilde{S}(t_m))^2} + \frac{l_m}{(1 - \tilde{S}(t_m))^2} \quad (17)$$

$$A_{i+1,i} = A_{i,i+1} = -\frac{a_{i+1}}{(\tilde{S}(t_i) - \tilde{S}(t_{i+1}))^2} \quad (18)$$

where $i = 1, 2, \dots, m-1$; and

$$A_{i,j}=0 \quad (19)$$

for $|i - j| \geq 2$. The inverse of the matrix $A = (A_{ij})$ is the estimated covariance matrix V , $V = A^{-1}$. Thus, an approximate 95% confidence interval for $\tilde{S}(t_i)$ can be defined as (Borgan et al., 1990):

$$LowerBound(i) = 1 - (\tilde{S}(t_i))^{q_i} \quad (20)$$

$$UpperBound(i) = 1 - (\tilde{S}(t_i))^{1/q_i} \quad (21)$$

where $q_i = \exp\left(\frac{z_a \cdot V_{ii}}{\log(\tilde{S}(t_i))}\right)$ and z_a is the a - percentile of the standard normal distribution.

Upper Bound Table for the L_2 error

In Figures 3-5 are plotted the average over 100 simulations of L_2 errors for each combination of N, A and stdv. As we have noticed, to each of these errors will correspond a standard deviation that is not constant: it will decrease as either the total sample size or the total number of actual data will increase. In particular, Table 1 reports the estimated 95% upper bounds for the percentage of L_2 error when the standard deviation of the underlying risk function is five.

Table 1. Estimated 95% Upper Bound For The Percentage Of L2 Error, When The Underlying Risk Function Has Stdv=5. In The First Column, For Each Row (In Bold) The Total Sample Size (N) Is Reported.

N	Total Number of Actual data (A)								
	0	5	10	15	20	25	30	35	40
10	29.2	17.7	7.9						
20	16.4	13.3	6.4	5.5	4.7				
30	8.7	8.4	5.3	4.4	3.1	2.6	2.1		
40	7.4	6.7	4.4	3.5	2.7	1.9	1.5	1.4	1.3
50	6.1	5.2	4.0	2.9	2.4	1.9	1.5	1.3	1.3
60	5.0	4.5	3.6	2.8	2.3	1.8	1.4	1.2	1.1
70	4.5	4.1	3.3	2.5	2.0	1.7	1.4	1.1	1.0
80	4.0	3.8	2.9	2.2	1.8	1.5	1.3	1.1	1.0

Mean Survival Time

Theorem (Chung, 1974): If X is a positive random variable with probability density $f(x)$, then for its mean $E(X)$ holds:

$$E(x) = \int_0^{\infty} \left(\int_x^{\infty} f(y) dy \right) dx = \int_0^{\infty} P(X \geq x) dx \quad (22)$$

Since $P(X \geq x) = S(x)$, where $S(x)$ is the survival function, it follows that:

$$E(x) = \int_0^{\infty} S(x) dx \quad (23)$$

REFERENCES

- AYER, M., BRUNK, H. D., EWING, G. M., REID, W. T., and SILVERMAN, E. (Dec. 1955). An Empirical Distribution Function For Sampling With Incomplete Information. *Annals of Mathematical Statistics*.
- BORGAN, O. and LIESTOL, K. (1990). A Note on Confidence Intervals and Bands for the Survival Curve Based on Transformations. *Scandinavian Journal of Statistics* 17 : 35-41.
- CHUNG, K. L. (1974). *A Course in Probability Theory*. Probability and Mathematical Statistics, Second Edition, Academic Press.
- DI DOMENICO, L. and NUSHOLTZ, G. (2003). Comparison of Parametric and Non-Parametric Methods for Determining Injury Risk. SAE 2003-01-1362.
- EFRON, B. (1967). The two-Sample Problem with Censored Data. *Proceedings of fifth Berkley Symposium on Mathematical Statistics and Probability*, Univ. of California Press, 4.
- HERTZ, E. (1993). A Note on the Head Injury Criteria (HIC) as a predictor of the Risk of Skull Fracture. 37-th Annual Proceedings of the Association for the advancement of Automotive Medicine.
- KALLIERIS, D., RIZZETTI, A., MATTERN, R., MORGAN, R., EPPINGER, R., and KEENAN, L. (1995). On the Synergism of the Driver Air Bag and the 3-Point Belt in Frontal Collisions. SAE952700.
- KAPLAN, E. L. and MEIER, P. (June 1958). Non Parametric Estimation from Incomplete Observations. *Journal of the American Statistic association*, Vol. 53.
- KUPPA, S., EPPINGER, R. (1998). Development of an improved thoracic injury criterion. Proc. 42 Stapp Car Crash Conference, Paper 983153.
- MERTZ, J. H., PRASAD, P., and NUSHOLTZ, G. (1996). Head Injury Risk Assessment for Forehead Impacts. SAE 960099.
- NUSHOLTZ, G. S., DI DOMENICO, L., SHI, Y., and EAGLE, P. (Sept. 2003). Studies on Neck Injury Criteria Based on Existing Biomechanical Test Data. *Accident Analysis and Prevention*, Vol. 35/5, pp. 777-786.
- NUSHOLTZ, G. and MOSIER, R. (1999). Consistent Threshold for Doubly Censored Biomechanical Data. SAE 1999-01-0714.
- PRASAD, P. (1999). Biomechanical Basis for Injury Criteria used in Crashworthiness Regulations. International research Conference on the Biomechanics of Impact (IRCOBI), Spain, 1999. [9]
- Ran, A., Koch, M., Mellander, H. (1984). Fitting Injury Versus Exposure Data Into Risk Functions. IRCOBI.
- TURNBULL, B. W. (March 1974). Non Parametric Estimation of a Survivorship Function with Doubly Censored Data. *JASA, Theory and Method Section*, Vol.69, No.345.

DISCUSSION

PAPER: **Risk Curve Boundaries**

PRESENTER: *Guy Nusholtz, DaimlerChrysler*

QUESTION: Richard Kent, University of Virginia

One question and one comment: I've used some of these non-parametric methods and I'm having a difficult time finding out if you can do a multiple, kind of a multi-variant analysis with a non-parametric model in terms of using multiple predictors instead of just a single stimulus. Do you know about that? Is that possible? Has that been defined?

ANSWER: The reason I made that last comment was to answer that question. Let me try it again. We've done it. We did it with only two variables. We did it on the neck numbers where we did--and that's published and I can get you that paper--where we did a moment and force. The problem with that is it isn't a clean system where you have a program and you just put the variables in. You have to go through a series of pre-conditioning of the data and test statistical validity of the data before you can go ahead and do it. So, I think the problem you're really running into is there are no software package out there to do it. It requires a lot of additional pre-work. So, the answer to your question is both yes and no. It can be done, but it can't be done easily.

Q: My comment is--[fog horn sound]

A: Well, I don't know if I can answer that comment in the way it was stated. [laughter]

Q: Okay. What was my comment? So, you showed that whether data is uncensored or censored plays a big role in the error--in the error that you have between the actual and the predicted risk?

A: That's right.

Q: But, there's an even bigger point in that if you assume your data's uncensored when it's actually censored, you get the wrong answer. And in fact, we've probably all seen in the literature this backwards risk function where you have decreasing risk with increasing force, or something like that, and a lot of times that results from an assumption of having uncensored data when in reality you do have--an assumption of censored data when in reality you have accurate data. And so, I just want to make that point because I do see it quite a bit in the literature.

A: In other words, you're getting that problem from...assuming that it's censored when--

Q: Right. Well, I think--I think logistic regression is to blame because--I think--Yeah. It's become kind of standard practice if you're going to come up with an injury function, you hit some things; you break some; you don't break others, and then you do a logistic regression. Well if those points where you're causing injury are actually uncensored, if they're accurate data and you do a logistic regression with them, then you'll underestimate the injury threshold because the injury's actually occurring at that point, not somewhere after that point or before that point.

A: [nearly inaudible] ...will be much larger...almost a factor of say 3 or 4 with a small sample size. With a large sample size, it won't matter what...

Q: Right. Yeah, I'm thinking specifically for biomechanical data where you typically only have--

A: If you have 10 or 15 samples, it can make a big difference.