RECENT WORK WITH A METHOD FOR
THE FITTING OF INJURY VERSUS EXPOSURE DATA
INTO A RISK FUNCTION

Magnus Koch
Automotive Safety Centre
Department 93550-PV22
Volvo Car Corporation
S-40508 Göteborg, Sweden

ABSTRACT

In a report to the 1984 IRCOBI conference, the author and two colleagues demonstrated the application of the Maximum Likelihood method to biomechanical research data. The method was used to interpret field accident data or laboratory research data into a cumulative risk function for injuries. By using that method the problem with so called censored data - so common in biomechanical research - was easily overcome. The present report is a follow-up of the 1984 report, giving answers to received comments, and also including recent developments in connection with the method. A procedure to complement the risk function with a confidence band has been developed. It is shown that the method can comprise both censored as well as non-censored data in the same analysis. A heuristic goodness of fit measure is suggested. In conclusion, the described method seems to be an efficient tool for biomechanical research work.

1 BACKGROUND

Much work in biomechanics deal with the establishment of injury threshold functions. They are important since they can be interpreted as risk functions for larger populations, provided reliable inferences from sampled data are feasible. But violence is difficult to control, so injury data from accidents and laboratory experiments do seldom represent threshold injuries. In most cases the violence has been either too high or too low to inflict exactly a threshold injury. Therefore the translation of various injury data into threshold functions has been a problem.

This problem was brought into focus in Volvo's safety work in 1983, when the ISO-TC22-SC12-WG6 working group had to analyse certain biomechanical laboratory data [11,12]. The Volvo Car expert in that group, Mr. Hugo Mellander, initiated a study with the following results:

1) After a statistical analysis Mr. Arne Ran, at Volvo Data AB, showed that the problem could be solved with the Maximum Likelihood method, which is a known procedure in several scientific and technical areas [9].

2) Computer programming by Mr. Ran provided suitable algorithms for evaluation of such data. These were made available to Volvo's safety specialists.

3) A report on the study was presented to the 1984 IRCOBI conference [13].

The algorithms used at Volvo have since 1984 received some additions. The 1984 IRCOBI report has also been commented upon by other researchers. This report on these developments is appropriate for the 1988 IRCOBI conference with its emphasis on statistical methods. A review of the essentials from the 1984 report will be given before the account of the recent developments.

2. INTRODUCTORY TERMINOLOGY

The terminology in biomechanics is sometimes carefully discriminating in order to make clear whether an account refers to, for instance, test animals, crash test dummies, or human accident victims [1]. For the purpose of this paper presenting common statistical tools which can be applied to several areas of biomechanics, the words listed below will be used with the noted generalisations. They will be used as far as possible in the text, but deviations from this intention may occasionally be made in the interest of clarity.

Subject is used to signify an injury victim, a laboratory test animal, a sample of biological tissue, a human or a human substitute.

Damage is a term for injury, harm, impairment, fracture, rupture, death, that occurs to the subject. *Damages* are assessed after careful medical examinations, and they are usually grouped into a small number of classes like in the AIS scale, or often in a simple bi-level grouping: *significant* or *not-significant damage*.

Loading is a term for violence, dose, exposure, impact speed, deceleration magnitude, and other measurable physical influences that are supposed to have a strong correlation with, or even having caused, the damage to the subject.

Threshold is a term signifying that a loading has been just high enough to inflict a damage that will be assigned a certain significant damage classification. I.e. the structural tolerance of the tissues of the subject has been exceeded. If the loading has been slightly lower, the damage is lesser than the critical threshold and consequently assigned into a lower damage class. Hence, *threshold loading* and *threshold damage*.

2.1. Censoring - an important concept

In accident and laboratory data, the relation between loading and damage is not necessarily a straightforward cause and effect relation. A certain kind of bias is often present.

Recorded loadings have in many cases been more than sufficient to cause the classified and recorded damages. Thus, the recorded loading figures are probably too high - and never too low - to serve as a measure for the threshold damage levels. This tendency, that a recorded data point systematically overestimates another interesting variable, is by statisticians called *censoring*. The term is strictly technical and no improper tampering with data is implied. Implicit in the censoring concept is also the fact that you don't have any idea about the magnitude of the overestimation. On the other hand, if a loading has been too low to cause a threshold damage, then the loading number is a poor and too low estimate for the threshold loading. This is also called censoring, but the biasing uncertainty now has a reversed sign. Therefore censored data are data that are biased, the sign of the bias being known but not the magnitude.

3. DETERMINATION OF THE DISTRIBUTION FUNCTION

One of the tasks in biomechanical research is to describe the threshold loading distribution for a certain population of subjects. The determination of this function, for a set of specimens, relies on a corresponding set of censored loading data. The method that we use to handle such data derives from Fisher's work with the Maximum Likelihood method in the 1920's [4].

The threshold loading for a certain type of damage in a certain population of subjects must be regarded as a random variable, the frequency density of which can be written as the function f(z), where z is the loading measure. The integral of f(z) is the cumulative frequency distribution F(z). It is our goal to determine the yet unknown function F(z) and its derivative f(z). See figure 1.

The function F(z) can be interpreted as the risk, or probability, that a loading of the magnitude z would cause a significant damage. Then, complementary, the probability for a non-significant damage at the loading level z must be 1-F(z). If, therefore, for a sample of certain subjects, there is a record of loading data together with associated damage classifications, the following addition can be made to each record: each observed significant damage has had an occurrence probability F(z), and each non-significant damage has had an occurrence probability 1-F(z).

Fisher introduced the "Likelihood" concept. It means that if you have a set of observations (numbered by q) which each had the probability P_q to occur, then the likelihood L, or joint probability, for this particular set of observations to occur is equal to the product of all P_q, i.e.:

$$L = \prod P_q \qquad (1)$$

Inserting our previous notations we get:

$$\begin{aligned} &L[\text{for the set of observations}] = \prod P[\text{each individual observation}] = \\ &= \prod P[\text{all significant damages}] \times \prod P[\text{all non-significant damages}] = \\ &= \prod \{F(z_i)\} \times \prod \{1-F(z_j)\} \end{aligned} \qquad (2)$$

Here the numbering variable i applies to significant damages, and j applies to non-significant damages.

If F(z) were to signify a number of different distributions, we would find that some of these gave a higher value to L than others. Therefore it is reasonable to imagine a smart manipulation of F, in order to yield a maximum for L, and then declare the most successfully manipulated F as the function which best models the recorded data! The candidate F distributions should be selected among the lot of sigmoid, or S-shaped, functions that are available. Common for these distributions are that they shall have a value close to 0% for low values of z, and a value close to 100% for high z values. They shall of course also be continuous and have a derivative dF/dz > 0. Let one class of such distributions be designated by F(z) = F(z;K), where K is a set, or a vector, of parameters that can be given certain values to trim the shape of the distribution F. Then the likelihood expression is:

$$L(\text{all } z_q;K) = \prod\{F(z_i;K)\} \times \prod\{1-F(z_j;K)\} \qquad (3)$$

Here we can see that L is a function of the known loadings z_q, and the yet undetermined parameter set K. The first product in the right member comprises the probabilities for the significant damages, the other product comprises those for the non-significant damages. To facilitate numerical work the logarithm is taken for both members:

$$\ln[L(\text{all } z_q;K)] = \sum \ln[F(z_i;K)] + \sum \ln[1-F(z_j;K)] \qquad (4)$$

An obvious way to proceed now is to study the partial derivatives of L. Putting them equal to zero yields an equations system (formally $0 = \partial L/\partial K$):

$$0 = \partial/\partial K \sum \ln[F(z_i;K)] + \partial/\partial K \sum \ln[1-F(z_j;K)] \qquad (5)$$

These equations (they are as many as there are elements in the set K) contain all the observed loadings, and the unknown parameter set K. Solving the equations for K yields as roots a set of constants denoted $\boldsymbol{K}$. Therefore the cumulative function $F(z;\boldsymbol{K})$ is the variant of the F function that has the highest likelihood, or gives the most credible fit, for the recorded loadings and damages. The "log-likelihood-number" $\ln[L(\text{all } z_q;\boldsymbol{K})]$ shall be recorded and used for further goodness of fit studies.

Note that in this method the censored loadings are accepted as the biased data they are, and no pretension is made to let them act as individual estimates of the threshold damages. Their joint result, the function $F(z;\boldsymbol{K})$, can, however, be used as the best estimate of the integral of the threshold damage function $f(z;\boldsymbol{K})$ for the set of damaged subjects.

4.MIXING CENSORED AND NON-CENSORED VALUES

The above is, essentially, what was presented about the Maximum Likelihood principle in Volvo's 1984 IRCOBI paper. A recent addition is as follows.

Thanks to great care in planning, experimental set up and execution, and to careful damage inspection and analysis, it might sometimes be possible to determine that for some individual subjects the damages are "true" threshold damages, i.e. they need not be regarded as censored data. In many experimental situations such non-censored values are even the rule rather than the exception.

The analysis of such data is traditionally simple and well known - the calculation of a mean and possibly a standard deviation. In some cases when you are interested in determining the distribution of these threshold data, the Maximum Likelihood is as useful as any other assessment method. The statistical reasoning above is applicable with small adjustments. The probability for an experiment to yield a threshold damage at the loading z is:

$$P[\text{threshold damage, given } z] = f(z) \qquad (6)$$

Collectively the likelihood for a series of threshold recordings (numbered by k) is:

$$L[\text{all } z_k] = \prod P[\text{threshold damage at } z_k] \qquad (7)$$

$$L[\text{all } z_k;K] = \prod f(z_k;K) \qquad (8)$$

$$\ln[L(\text{all } z_k;K)] = \sum \ln[f(z_k;K)] \quad (9)$$

which should be analysed to determine the maximizing parameter set $\boldsymbol{K}$. The result is the probability mechanism that best fits the observations:

$$P[\text{threshold damage at } z] = f(z;\boldsymbol{K}) \quad (10)$$

Now it is obvious that we have the possibility to perform an analysis of a complete experimental series, containing censored as well as non-censored data, in one single maximization procedure. The total likelihood expression is:

$$L(\text{all } z_q;K) = \prod \{F(z_i;K)\} \times \prod \{1-F(z_j;K)\} \times \prod f(z_k;K) \quad (11)$$

and the equations system to solve for K is:

$$0 = \partial/\partial K \sum \ln[F(z_i;K)] + \partial/\partial K \sum \ln[1-F(z_j;K)] + \partial/\partial K \sum \ln[f(z_k;K)] \quad (12)$$

Here the partial sums in the right member contain: the significant damages, the non significant damages, and the threshold damages, in that order. This is an optimal utilization of experimental data, which probably is not much known. It is certainly worth more attention in planning and analysis of experiments in biomechanics as well as in other areas.

5. SELECTION OF A SUITABLE FUNCTION

In our first implementation of the Maximum Likelihood method the Weibull distribution was chosen for $F(z;K)$. The main reason was that this function, thanks to its three parameters has a large flexibility. It can model symmetrical distributions as well as skewed ones in either direction. The function is written:

$$W(z;\alpha,\beta,\gamma) = 1 - \exp[-\{(z-\gamma)/\alpha\}^{\beta}] \quad (13)$$

The parameters β and γ can be restricted to ≥ 1 and ≥ 0, respectively, in order to avoid solutions that would be physically unlikely. These restrictions will eliminate models with decreasing vulnerability as the loading increases, and injury measures less than zero, respectively. A detailed discussion on these aspects of the Weibull parameters can be found in the 1984 report [13]. In its present version of the algorithm there is included the possibility to handle mixed data sets containing censored as well as non-censored values.

Our Maximum Likelihood algorithm for the Weibull distribution is based on the N.A.G. software [10]. It has then been added as an extension to the statistical analysis programme package S.A.S. [14], which is available on Volvo Data's mainframe computers. Available as standard features in this package are Maximum Likelihood fitting programmes for the Probit [3] (easily transformable to Normal and Lognormal), and the Logistic distributions.

Recently it has been suggested that Extreme Value functions, that are used in fracture mechanics [8, chapter 19], might be an interesting addition. These functions are derived from theories of the weakest link in a construction, and they are skewed with a long tail towards high values. It appears as if such skew is present in several biomechanical data sets. You should, however, remember that accurate modeling of distribution tails requires a large number of samples, which rarely is available in biomechanical laboratory studies.

The computer programmes provide, as desired and feasible, confidence intervals to the fitted distributions. For the Weibull case the confidence algorithms are adapted from the work of Cheng and Iles [2]. Details about the algorithms for confidence bands in the S.A.S. programme are presently not analysed. However, when certain data sets are analysed with the Weibull and the Lognormal models the resulting output distribution graphs and their confidence bands appear quite similar by a visual inspection.

6. THE CREDIBILITY OF THE SOLUTION

The overall credibility of a derived statistical model is dependent not only on the Maximum Likelihood method, but also on a number of other circumstances, as listed below. For efficient handling of these problems, items 1 to 3 require an experimenter who knows the physics of the studied subjects and measurement procedures, and items 2 to 4 require a person with statistical training. Close cooperation between these two persons is important and rewarding.

1) The relevance of the measured variables must be assured. Experimental errors must be controlled. The traditional classification of errors in random, systematic and gross errors might be helpful during the assessment procedure.

2) Shortcomings in selection and representativeness of the studied subjects must also be controlled if you want to use sampled data for inferences about other subjects in larger populations.

3) The selection of a suitable model function for F, i.e. Normal, or Weibull, or something else, has no unique "best" solution. The magnitude of ln(L) - the log-likelihood-number - serves as an indication here; the best fitting function has the highest likelihood value. The choice of the "best" model is a task that requires experience with experimental data both from physical and mathematical points of view.

   The physicist knows the typical damage mechanisms and is familiar with models that are used for their characterization. This might sometimes lead to certain favoured distributions and parameter ranges. For instance, assuming the vulnerability of the specimens to increase by z, requires that the so called "Hazard" function:

$$H(z) = [dF(z)/dz]/[1-F(z)] \qquad (14)$$

   has a monotonic increase as z increases. This relation has not explicitely been much considered in biomechanical modeling, but is a standard concept in reliability and length of life studies.

   The mathematician knows the adaptability of different functions to various sets of scattered data. He/she also knows the operating characteristics of the maximizing computer programmes that are to be used.

   A discussion and cooperation between the two specialists will ensure a good selection of suitable functions.

4) The goodness of fit can be studied in a couple of ways.

A heuristic goodness of fit indicator can be defined as the quantity:

$$G = e^{[\ln(L)/n]} = \text{antilog}[\ln(L)/n] \quad (15)$$

where ln(L) is the log-likelihood-number, and n is the number of observations. Recalling the definition of L in eq. (1) and (2), G can be described as the geometrical average of all P_q values. In principle, this quantity would lie between 0.5 for meaningless fits, and 1.0 for perfect fits. See figure 2.

Many computer programme packages for model fitting automatically provide some kind of goodness of fit measure.

Confidence bands for the distribution found can be determined by traditional methods. Fisher's information matrix contains variances and covariances as partial derivatives $\partial L/\partial K$. From these the confidence band can be determined by algorithms such as those suggested by Cheng and Iles [2].

7.APPLICATIONS AND COMMENTS

Three application examples with accident and laboratory data were accounted in the 1984 report. Here will be described some additional findings and comments from ourselves and others.

The Maximum Likelihood method has for several years been used at Volvo Car Corp. to analyse accident data and relate these figures to laboratory test data [6], similar to what was proposed in a comprehensive study by Horsch in 1987 [5]. Studies of biomechanical data are also performed occasionally. Continuous evaluations are made of the user friendliness of the software, and modifications are introduced as needed.

The material in our 1984 IRCOBI report has been presented to various audiences on several occasions and has been well received. In one case it resulted in an (informal) readable tutorial on this application [7].

A researcher told us that his students are "intrigued and fiddle much with your Weibull distribution". A reflection to this is that the derived distribution primarily is a model summarising the experimental data. When beginning to use this method, one should be cautious, and learn from discussions with a competent statistician which interpretations that are permitted or discouraged when using models.

A comment that we received to the 1984 report was that the analysis of research HIC data is a "mixture of apples and bananas", since a skull that is fractured in an experimental situation has a somewhat lower HIC value due to the increased time of deceleration during the deformation and fracture of the skull. Therefore HIC values for fractured and non fractured skulls do not measure the same property. The available model permits an easy study of this. The list of HIC data in Prasad's and Mertz' paper [12] is used for a simple analysis. Suppose, without much actual data support for it, that a HIC value for a fractured skull should be 40% higher than recorded, if it is to be compared to the HIC values for non-fractured skulls. All HIC values for the fractured specimens in the Prasad data set are increased by 40% and a new distribution is fitted to the modified data set. The modification resulted in a better separation between the non-fractured and the fractured sets, which resulted in a more symmetric distribution. See figure 3.

The confidence bands used in industrial applications are usually selected at a 95% confidence level. Biomechanical experimental data do in many cases comprise so few samples that such a band becomes uncomfortably wide, or even impossible to determine. If 50% is chosen instead, the band becomes narrower and the picture appears more pleasant. See figure 4. However, one must recognize that the lower a statistical confidence level is, the more uncertain are any decisions based on such evaluations of data. Either one has to consciously accept this uncertainty, or try to reduce it by promoting basic biomechanical research for the increase of baseline data.

7.1. An Evaluation with a Surprising Result

One surprising outcome of a certain published study, by Prasad and Mertz [12], was the verdict that our "Three Parameter Weibull Maximum Likelihoood Method [is] clearly unable to provide a good approximation of the actual threshold curve *when data are highly scattered*" (italics added). This conclusion deserves some comments, since it has been founded on a very special case. The comments are included here for those who might want to look closer into the study.

The published study was devoted to the available biomechanical base data for the HIC criterion. Included in the study were several ways to model the head injury data into a risk function. In addition to testing risk function models directly to the HIC data, a study was also made with some hypothetical experimental data. A set of such data was selected from a Normal distribution, and an experimental strategy (load exposure to each hypothetical test specimen) was chosen so that a set of simulated observations was produced where no particular distribution could be discerned. The underlying Normal distribution had become invisible due to the experimental arrangement.

To these data the Weibull model, as used in the study, provided the correct interpretation that the risk for injury was 50% in the neighbourhood of the data. Since no other information was introduced to the Weibull model, the risk of 50% became extended infinitely to both ends along the z axis. This was not in agreement with the underlying supposed Normal distribution. Therefore the above declaration was made about the inability of the Weibull model. The conditional clause "*when data are highly scattered*" is an important one and should be applied to any other distribution tested in the same manner.

However, the Maximum Likelihood method seems not to have been used to its full potential in the study. Several reflections can be added.

First, with this set of data the goodness of fit indicator G will be 50%, which means that a meaningful fit is not possible to make. This shortcoming shall not be blamed at the Maximum Likelihood principle, but is the result of the scatter in data. It is one of the virtues of the Maximum Likelihood method to include such evaluation aids.

Next, a knowledge about the physics in the hypothetical experiment might have indicated that the specimens had an increasing vulnerability, in which case the Weibull β should have been restricted to values > 1; cf. eq. (14). If even a Normal distribution could have been assumed for the data, the Weibull β should have been fixed at $\approx$ 3.5; or still better - the Maximum Likelihood fit should have been performed with F(z) being a Normal distribution. These measures would have provided distributions with more appealing shapes. But the G value would have remained at 50%, because of the scatter of the assumed experimental data.

Finally, an analysis of the hypothetical arrangement of the experiment reveals that the selected strategy is a rare occurrence. A Monte Carlo simulation of permutations of the experimental procedure shows that scatter of this type occurs only in 0.6% of all possible cases. In other permutations the scatter is less, the G value is higher than 50%, and most Weibull fits approach the generating Normal distribution.

A visual indication of the discriminating power of the Maximum Likelihood Weibull fit can be found in the same study. Originally the data base contained 43 experiments with 23 significant brain damages (53%), but after a revision 5 experiments were removed, and the remaining set contained 38 experiments with 19 significant damages (50%). This difference between the two sets can be clearly seen when the Weibull curves are plotted together, see figure 5.

8.CONCLUSIONS

We have found that Fisher's Maximum Likelihood method is a suitable tool for the analysis of censored data in accidentiological and biomechanical investigations. Our first work with the model from 1984 has been complemented by several improvements both in the model and in the user interface. It has evolved as a useful tool in Volvo's safety work, and - it appears - also with several other researchers. The future will no doubt witness more improvements of this and similar powerful methods for traffic safety analyses.

The author is indebted to Arne Ran, Erik Elgeskog and many other friends and colleagues at Volvo for helpful discussions.

9.REFERENCES

1 B. Aldman, H. Mellander and M. Mackay: *The Structure of European Research into the Biomechanics of Impacts.* Proceedings Twenty-Seventh Stapp Car Crash Conference, paper no. 831610, pp 129 - 136, Society of Automotive Engineers Inc., 1983.

2 R.C.H. Cheng and T.C. Iles: *Confidence Bands for Cumulative Distribution Functions of Continous Random Variables.* Technometrics, Vol.25, No.1, February 1983, pp 77-86.

3 D.J. Finney: *Probit Analysis,* 3rd edition, Cambridge University Press, 1971.

4 R.A. Fisher: *Statistical Methods for Research Workers.* First edition Edinburgh 1925, many later editions.

5 J.D. Horsch: *Evaluation of Occupant Protection from Responses Measured in Laboratory Tests.* Paper no. 870222, Society of Automotive Engineers Inc., 1987.

6 J. Korner (Volvo Car Corporation): *A Method for Evaluating Occupant Protection by Correlating Accident Data with Laboratory Test Data.* Tentative title, to be published.

7 R. Morgan: *Analysis of Injury vs. Exposure Data using Weibull Distribution Function.* Informal presentation at the the twelfth annual international workshop on Human Subjects for Biomechanical Research; Chicago Illinois, November 5, 1984.

8 M.G. Natrella: *Experimental Statistics,* National Bureau of Standards Handbook no. 91. Washington D.C. 1966.

9 W. Nelson: *Life Data Analysis for Units Inspected Once for Failure (Quantal Response Data).* IEEE Transactions on Reliability, Vol.R-27, No.4, October 1978, pp 274-279. This introduction contains 21 further references on analysis methods for cumulative risk functions.

10 Numerical Algorithms Group Ltd. NAG Central Office, 7 Banbury Road, Oxford, England.

11 *Position Paper on HIC Levels,* Document N-146 prepared by the U.S. delegation to ISO/TC22/SC12/WG6, June 1983. The same information, in an expanded and revised version, appears in reference [12].

12 P. Prasad and H.J. Mertz: *The Position of the United States Delegation to the ISO Working Group 6 on the Use of HIC in the Automotive Environment.* Government/Industry Meeting and Exposition, Washington D.C., May 20-23, 1985, paper no. 851246, Society of Automotive Engineers Inc., 1985.

13 A. Ran, M. Koch and H. Mellander: *Fitting Injury versus Exposure Data into a Risk Function.* 1984 International IRCOBI Conference, Delft.

14 SAS Institute, Cary, North Carolina, USA

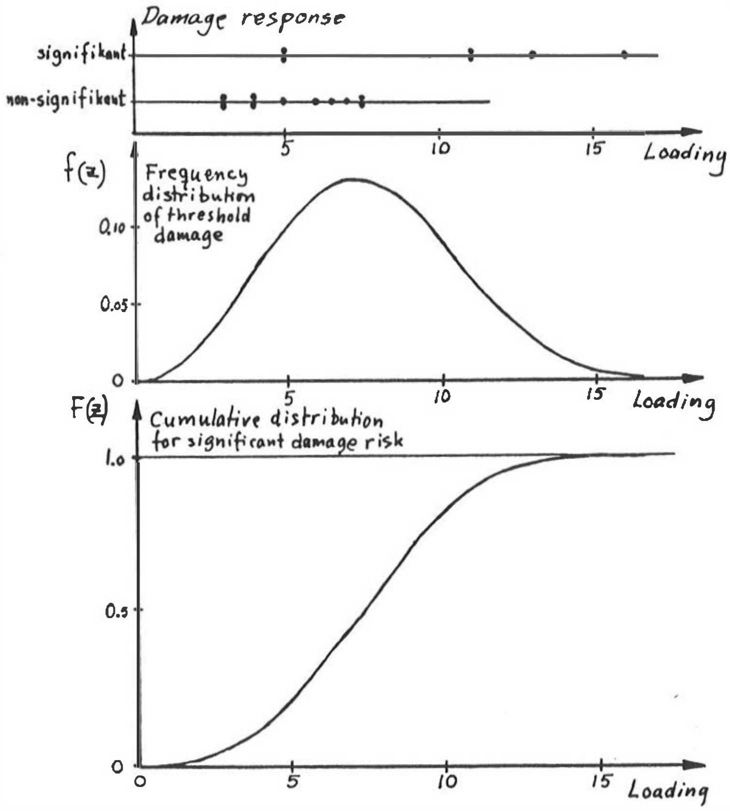

Figure 1. The graphs illustrate the relation between a set of measured censored data and its interpretation as statistical distributions of threshold loadings.

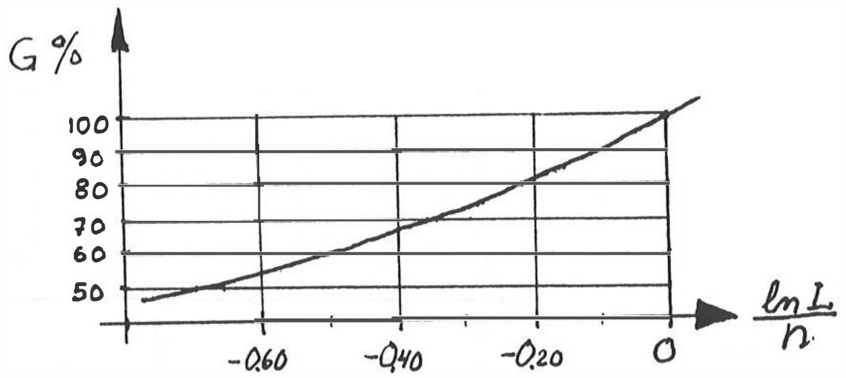

Figure 2. The relation between the log-likelihood-number and the goodness of fit indicator G, c.f. equation (12).

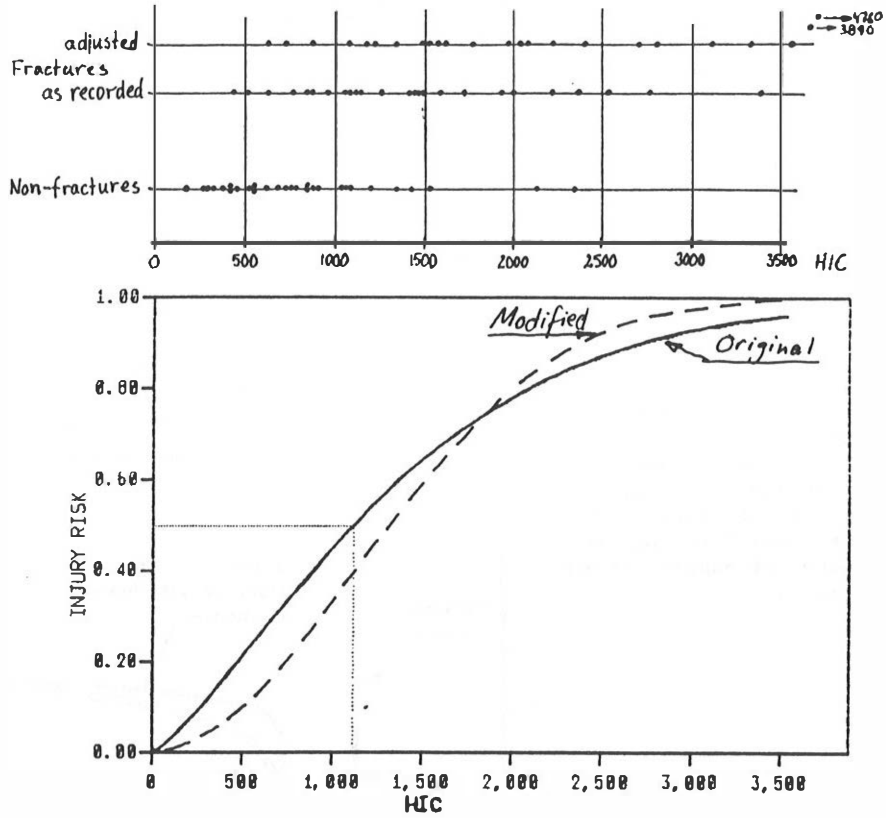

Figure 3. Comparison between two data sets. The original data are 54 samples from [12]. A two-parameter Weibull fit to the original data is drawn as ——. The data have been modified by multiplying the fracture HIC's by 1.4. A two-parameter Weibull fit for the modified set is drawn as — — — —. The modification resulted in a better separation between the non-fracture and fracture data sets, which gave the modified risk curve a better symmetry.

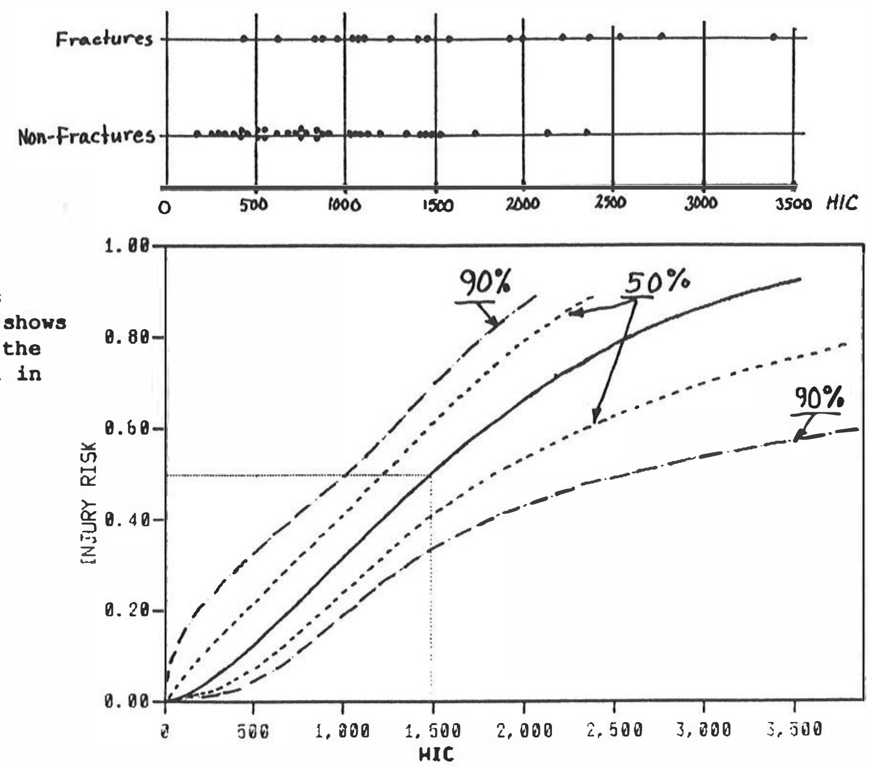

Figure 4. Confidence bands (two-sided) at 50 % and 90 % levels. The central risk curve shows a two-parameter Weibull fit to the 54 samples of HIC fracture data in [11].

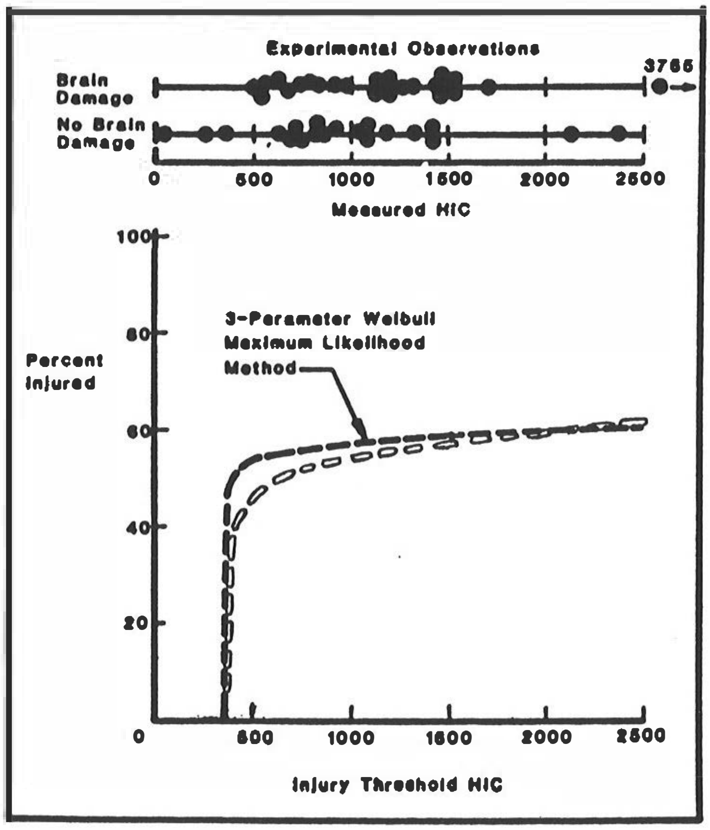

Figure 5. Combined distribution curves from figures 4 and 5 in [12]. The omission of four brain damages at 540, 1150, 1200 and 1500 HIC, and of one no-damage at 1185 HIC, lowers the curve from the solid to the open dashed line. Number of samples is 43 and 38, respectively.