HOW RELIABLE ARE INJURY RISK CURVES?

Norbert Praxl
PDB – Partnership for Dummy Technology and Biomechanics
Germany

Paper Number 11-0089

ABSTRACT

Injury risk curves are the basis for assessing automotive occupant safety. They are used in regulation, consumer test ratings, safety system design, and for the evaluation of the effectiveness of safety systems. Therefore, an injury risk curve should be accurate and credible. But how reliable is the risk prediction of an injury risk curve?

The objective of this study was to identify and illustrate factors influencing the reliability of injury risk curves. Thereby, highlight the need for a more thoughtful construction and use of injury risk curves as well as the need for additional statistical measures when publishing injury risk curves. The results of this study will lead to a better understanding of injury risk curves and can also be used for a better design of experiments in biomechanical testing.

Four factors affecting the reliability of injury risk predictions were evaluated exemplarily in this study:

- The sampling, i.e. the uncertainty due to statistical inference from a sample on the population.
- The censoring of data, i.e. the uncertainty introduced by the imprecision of the tolerance limit determination.
- The test severities, i.e. the uncertainty in injury risk prediction as a result of test severities used in the biomechanical tests.
- The statistical model, i.e. the bias introduced by the method used to calculate the injury risk curve.

Although all of the findings presented can be explained by statistical theory, this paper demonstrates the effects of different factors on the reliability of injury risk curves in a visual manner. Statistical simulation is used to replicate biomechanical testing and injury risk curve construction.

The statistical simulations comprise several steps including the definition of a distribution of the biomechanical tolerance limit in the population, the sampling and biomechanical testing of specimens as well as the construction of the injury risk curve.

The statistical simulations clearly illustrate the effect of the sample size and data censoring on the uncertainty of injury risk curves. It can be concluded that the interpretation of an injury risk curve without a proper measure of confidence is meaningless. Exact data of the biomechanical tolerance limit improve the reliability of the injury risk curve – however only with the use of an appropriate statistical method.

The range of criterion values used in the injury risk curve construction systematically affects the shape and reliability of the curve. Biomechanical tests should be done over a wide range of test severities in order to avoid bias in the risk estimation.

It is demonstrated that the use of an unsuitable - nevertheless widely used - statistical method for constructing the injury risk curve can lead to unrealistic injury risk predictions.

INTRODUCTION

Injury risk curves are used for the safety assessment of passenger cars regulated by law, for occupant safety ratings in consumer tests, for the design of safety systems, and for evaluating the effectiveness of safety systems. Injury risk curves are the basis for assessing, improving, and evaluating automotive occupant safety. The prediction of the injury risk by an injury risk curve plays an essential role in vehicle safety. Thus the risk prediction needs to be reliable.

But how reliable is the risk prediction of an injury risk curve?

Before looking at the reliability of injury risk curves, it should be clear what an injury risk curve is and how it relates to the biomechanical tolerance limit.

Biomechanical Tolerance Limit

The maximum mechanical load a person can withstand without getting injured is called the biomechanical tolerance limit of this person. Different persons – most likely – have different biomechanical tolerance limits. The biomechanical tolerance limit varies within the population.

The distribution of the varying biomechanical tolerance limit can be expressed by a probability density function (PDF), an example of a possible PDF is shown in figure 1. The real probability density function of the biomechanical tolerance limit of a population is unknown and generally can't be determined. The PDF or distribution can only be estimated statistically by using a random sample from the population of interest (if the sample isn't random, all conclusions will be highly biased!).

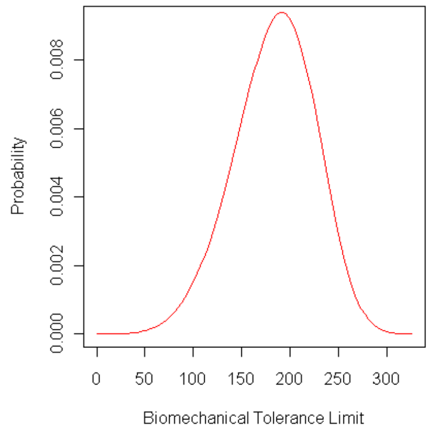

Figure 1. A possible probability density function (PDF) of the biomechanical tolerance limit of a population.

Injury Risk Curve

For an arbitrary person of a population of interest the injury risk curve shows the probability of getting injured at a specific mechanical load level. The mechanical load is expressed by an injury criterion value. This can be a physical measure like force or acceleration or more complex quantities like HIC or NIC. Figure 2 shows a possible injury risk curve.

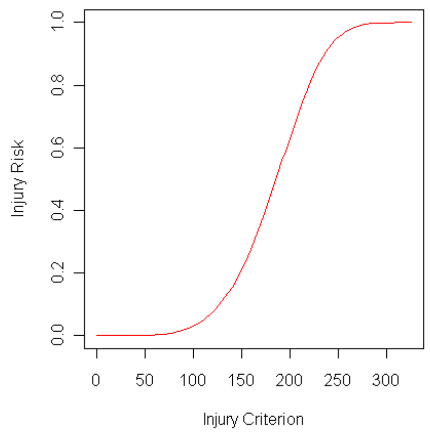

Figure 2. A possible injury risk curve or cumulative distribution function (CDF) of the biomechanical tolerance limit.

An injury risk curve displays the percentage of the population having a biomechanical tolerance limit lower than the associated criterion value. In the example used in this study (figure 2) 20% of the population have a biomechanical tolerance limit (expressed as an injury criterion value) lower than 150.

An injury risk curve is the cumulative distribution function (CDF) belonging to the distribution of the biomechanical tolerance limit. Consequently the distribution of the tolerance limit and the injury risk curve contains the same information. However, the presentation of this information is different.

Like the PDF of the biomechanical tolerance limit the injury risk curve of a population is not known and can only be estimated by statistical inference, i.e. inference from a sample on the population. The distribution of the biomechanical tolerance limit within the sample is the best estimate of the distribution of the biomechanical tolerance limit in the population and the best predictor for the injury risk of an arbitrary person of the population.

The injury risk curve is calculated on the basis of the biomechanical tolerance limits found in biomechanical tests mostly done on cadavers. The cadavers are the sample for the population of interest.

Since every injury risk curve is a statistical estimate, the question about the reliability of this estimate is obvious.

Objective

The development of an injury risk curve is a statistical procedure. Some of the problems and the complexity of this procedure are addressed in this paper.

The main intention of this study is to analyse and highlight factors adding to the uncertainty of an injury risk curve. Another objective is to provide the basis for an improved assessment of predicted injury risks and to show the need for a precautious interpretation as well as a more thoughtful construction and usage of injury risk curves. Furthermore, this study will reveal the benefit of additional statistical measures or supplementary information in the publication of injury risk curves.

The effect of different factors on the injury risk curve is demonstrated in a visual way using statistical simulation. All of the findings presented can be explained by theory. However, theoretical considerations are not the purpose of this study.

Only the baseline risk without any predictors (e.g. age) is used. The analysis of more complex statistical models is not within the scope of this study.

METHODS

Four factors affecting the uncertainty of an injury risk curve have been evaluated exemplarily in this study:

- The uncertainty due to the statistical inference from a sample on the population, named the effect of sampling.
- The uncertainty and bias introduced by the imprecision of the determination of the biomechanical tolerance limit in biomechanical tests, denoted as effect of censoring.
- The bias in the injury risk curve as a result of the test severities used in the biomechanical tests, herein after called effect of testing.
- The effect of statistics, i.e. the bias in the injury risk curve as a result of the statistical method used in calculating the injury risk curve.

The effects of these four factors were evaluated by statistical simulation.

Statistical simulation

To simulate the process of biomechanical testing and injury risk curve construction an underlying ("real") distribution of the biomechanical tolerance limit in the population was predefined. A Weibull distribution [1] with shape parameter 5 and scale parameter 200 was used as underlying distribution of the biomechanical tolerance limit in the simulations (see figure 1), except for the additional analysis of the logistic regression (see figure 19 and 20). Here the shape parameter was changed to a value of 2.5. The CDF of the distribution shown in figure 1 is presented in figure 2. This is the "real" injury risk curve of the population and will be used for the analysis of the effect of the factors listed above.

The underlying distribution is an assumption for the purpose of the simulation study. There is no justification that the distribution of the biomechanical tolerance limit in a population is following a certain or even common statistical distribution. Nevertheless, the effect of the factors evaluated in this study will in principle be the same on similar distributions.

In the next step of the simulation process a random sample of biomechanical tolerance limits was taken from the predefined distribution of tolerance limits [2]. This equals a random sampling of specimens (cadavers) from the population of interest.

To simulate the biomechanical testing of the sample, a range of test severities was defined, i.e. a minimum and maximum test severity. The test severities in the simulation study are expressed as criterion values, thus, the same quantity as for the biomechanical tolerance limit was used. The test severity in the simulation study is not equal to the physical test condition (e.g. pendulum mass and velocity) but equivalent to the response of the physical test condition. Test severities were randomly taken out of the test severity range and randomly assigned to the sampled specimens [2].

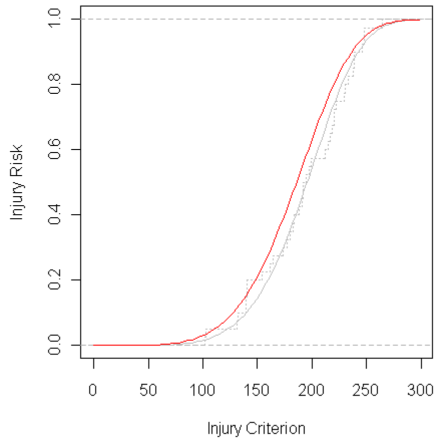

Figure 3. Sample injury risk curve (grey continuous line), "real" injury risk curve (red line), non-parametric CDF of the sample (grey dotted line).

By comparing the biomechanical tolerance limit of a specimen with the assigned test severity (expressed as resulting injury criterion value) a binary outcome is produced: injury or no injury.

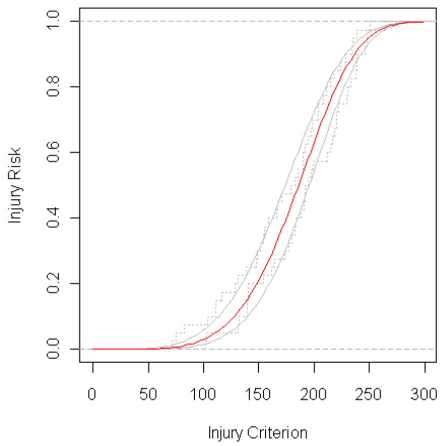

Figure 4. Injury risk curves of two samples (grey continuous lines), "real" injury risk curve (red line), non-parametric CDFs of the samples (grey dotted lines).

The last step of the statistical simulation was the calculation of the injury risk curve. Except for the analysis of the effect of statistics, a survival analysis with Weibull distribution assumption was used to calculate the injury risk curve of the samples [1, 3]. A baseline survival analysis, i.e. without predictors, equals a distribution fitting. The fitting of a Weibull distribution to data sampled from a Weibull distribution should give a reasonable estimation.

The statistical software R was used for the calculations and the plotting of the results [4].

Figure 3 shows a comparison of the result of one simulation run (sample size 40) with the "real" injury risk curve. Figure 4 shows a comparison of the results of two simulation runs (sample size 40) with the "real" injury risk curve.

The simulation process explained before was repeated hundred times and the resulting injury risk curves were compared to the "real" injury risk curve (see figure 5 as example). The scatter of the sample injury risk curves demonstrates the uncertainty of the injury risk curve estimation.

Effect of sampling

To show the effect of sample size on the reliability of the resulting injury risk prediction, hundred simulation runs with a sample size of 10, 40, and 160 have been performed.

Effect of censoring

Data censoring means that the exact biomechanical tolerance limit is not known for a specimen. If an injury is known to have occurred below a certain value of the injury criterion this data is called "left censored". If it is known that an injury will occur above a certain value of the injury criterion this value is called "right censored".

By comparing the biomechanical tolerance limit of a specimen from the sample with the assigned test severity, the binary result (injury or no injury) will lead to left and right censored data. If the assigned test severity is above the biomechanical tolerance limit, the outcome is an injury and the test severity is a left censored data and vice versa.

To simulate exact data, instead of using the test severity, the tolerance limit was used in the injury risk curve calculation when an injury was detected.

For the analysis of the influence of data censoring when using different statistical methods, injury risk curves with right censored and exact data were calculated with the use of logistic regression [1, 2] and survival analysis with logistic distribution assumption. The logistic distribution assumption in the survival analysis was used to avoid the influence of different distribution assumption in the compared methods.

Effect of testing

Usually, biomechanical tests are done at a certain mechanical load or at least within a certain range of loading severities. Very low and very high loadings are not used because it is assumed that such tests will not give valuable information since the outcome is known beforehand. To show the influence of the test severity range used in biomechanical testing on the injury risk curve, different ranges of test severities have been defined in the simulation: a low range, a small centred range, a high range and a large range. The lower and upper limit of the test severity range was defined with respect to the "real" distribution, i.e. quantiles of the "real" distribution were used for the definition of the test severity ranges (table 1).

Table 1.
Quantiles used for the definition of the test severity ranges

Range	Lower limit	Upper limit
low	0.001	0.25
high	0.75	0.999
small	0.35	0.65
large	0.01	0.99

Effect of statistics

To visualize an important effect of the statistical method on the injury risk curve, simulations using logistic regression as well as survival analysis with Weibull distribution assumption were performed. To show the dependency of this effect on the sample size, simulations with a sample size of 20 were conducted in addition to the simulations with a sample size of 40. In order to further analyse the influence of the statistical method on the injury risk curve, a different "real" distribution of biomechanical tolerance limits were used. The shape parameter of the Weibull distribution was changed from 5 to 2.5, i.e. the distribution was shifted to the left (compare figure 16 to figure 19).

RESULTS

Effect of sampling

In figure 3 the difference between the “real” underlying injury risk curve (red curve) and the calculated injury risk curve based on a random sample of size 40 is shown. The underlying risk curve is a Weibull distribution and the calculation method is a survival analysis with a Weibull distribution assumption (which basically is a fitting of a Weibull distribution).

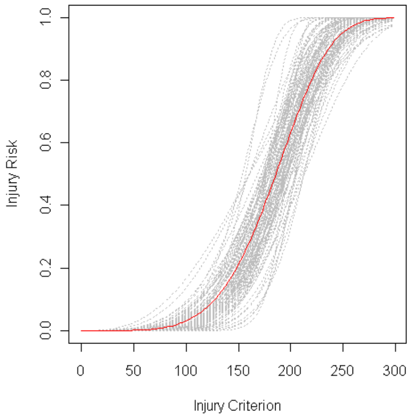

Figure 5. Scatter of injury risk curves in 100 simulation runs with a sample size of 10.

The difference between the underlying injury risk curve and the calculated risk curve is due to the sampling and not because of different distributions in the statistical calculation (the non parametric CDF additionally shown in figure 3 as dotted grey line supports this statement).

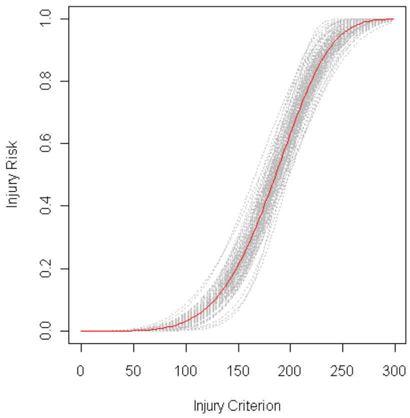

Figure 6. Scatter of injury risk curves in 100 simulation runs with a sample size of 40.

Figure 3 shows that a sample does not reproduce the “real” distribution. That is why the injury risk curve of a sample (most likely) will deviate from the “real” injury risk curve. The actual difference between the “real” injury risk curve and the injury risk curve based on a sample is up to the random sample. The randomness of sampling leads to uncertainty in the resulting injury risk curve, i.e. it is not known how good the calculated injury risk curve represents the real injury risk curve.

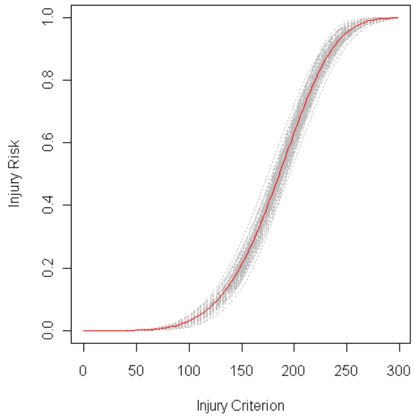

Figure 7. Scatter of injury risk curves in 100 simulation runs with a sample size of 160.

Figures 5 to 7 show the dependency of the difference between real and sample injury risk curve from the size of the sample. The smaller the sample size the bigger the uncertainty of the risk prediction (more scatter of the simulation results). In other words the reliability of an injury risk curve increases with increasing sample size.

Effect of censoring

Censored data substantially increases the uncertainty of an injury risk curve in comparison to the result of exact data (compare figure 8 and 9). Censored data contain less information, therefore, the risk prediction is less reliable. It is not shown here but can easily be imagined, that the more censored data are in the dataset the more scatter or uncertainty will be introduced. The example shown here (figure 8) illustrates the maximal effect because all data were left or right censored.

The censoring of data used in the calculation of the injury risk curve does not only introduce more uncertainty in the risk prediction (more scatter in the simulations) but can also lead to a bias in the risk prediction (shifting of the injury risk curve) when using an unsuitable calculation method. Logistic regression assumes left and right censored data. The use of right censored and exact data in logistic regression – as in all binary regression models – leads to a left shift of the injury risk curves (figure 10).

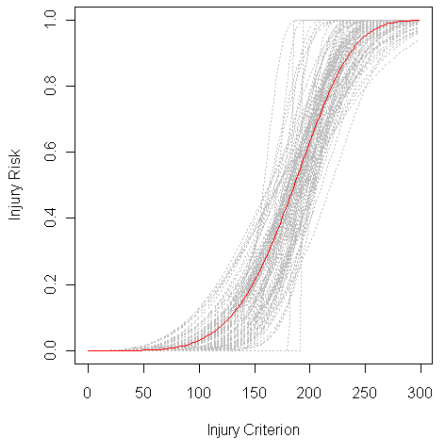

Figure 8. Simulated injury risk curves based on left and right censored data (sample size 40).

Using right censored and exact data with a survival analysis and logistic distribution assumption will have no bias (figure 11).

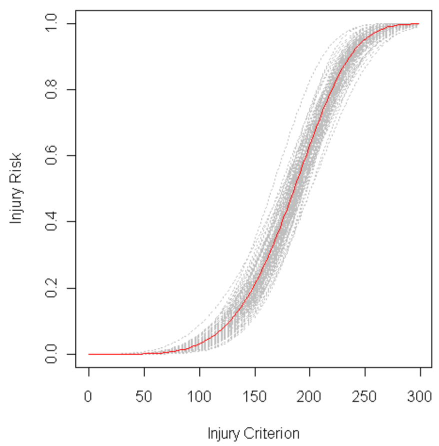

Figure 9. Simulated injury risk curves based on exact (uncensored) data (sample size 40).

The same distribution assumption (logistic) was used in survival analysis like in logistic regression. For this reason the result demonstrates that solely the calculation method (binary regression vs. survival analysis) is responsible for the bias in the injury risk curve and not the distribution assumption. The effects displayed in the figures 10 and 11 show the maximal effect since all injury data were treated as exact data. It is not shown but can easily be imagined that the strength of this effect depends on the percentage of exact data in the dataset: the more exact data the more bias in the injury risk curve calculated by logistic regression.

Effect of testing

The test severities used are influencing the injury risk curve. Outside the tested severity range the uncertainty is high. A low severity test range (figure 12) will probably lead to overestimated risks. In contrast a high severity test range (figure 13) can easily lead to underestimated risks. A small test severity range in the centre (figure 14) will likely lead to a steeper injury risk curve (overestimated risks above 50% risk and underestimated risks below 50% risk).

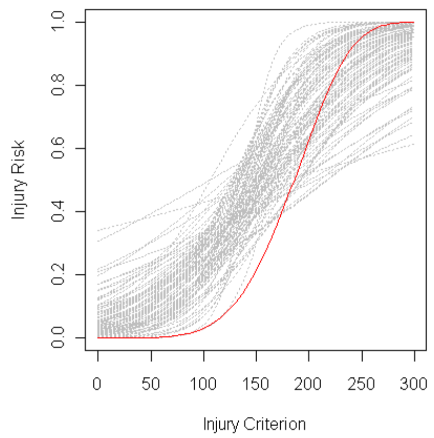

Figure 10. Logistic regression based on exact and right censored data (sample size 40).

These results are stochastic, i.e. a low test severity range does not necessarily lead to an overestimated risk. However, the probability of getting an overestimated risk is higher than getting a correct or underestimated risk.

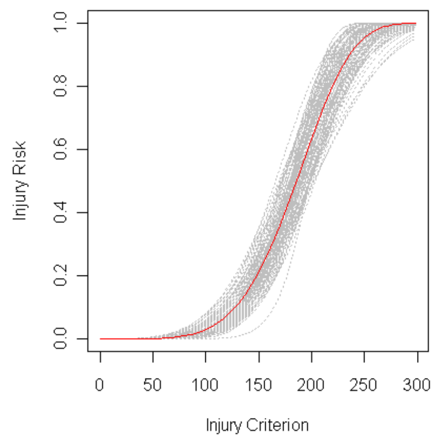

Figure 11. Survival analysis with logistic distribution assumption based on exact and right censored data (sample size 40).

The results for the other test severity ranges have to be seen analogous. A wide range of test severities gives the least bias in the injury risk prediction (figure 15). Due to the unknown “real” injury risk curve the width of the test severity range with respect to the "real" injury risk curve (or tolerance limit distribution) cannot be assessed, unfortunately.

Effect of statistics

An unsuitable statistical model can lead to a bias in the risk prediction. An injury risk curve calculated with the logistic regression does not pass through zero and will always predict a non-zero risk at zero load.

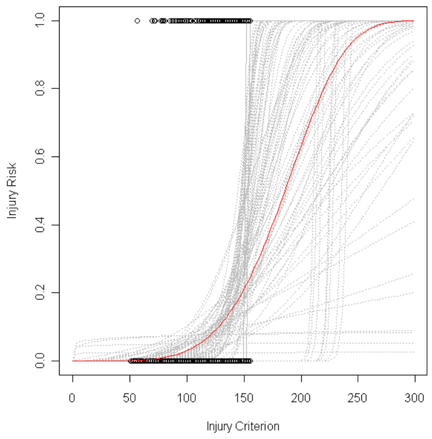

Figure 12. Low range of test severities (black dots showing the test results: 0.0=no injury, 1.0=injury).

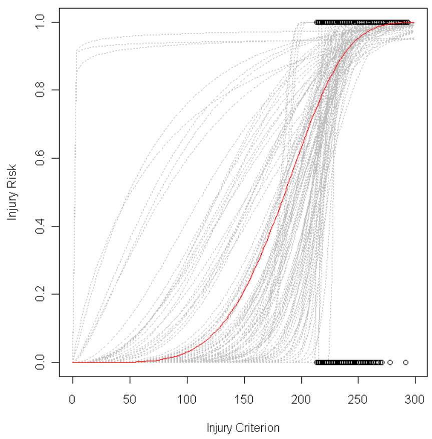

Figure 13. High range of test severities (black dots showing the test results: 0.0=no injury, 1.0=injury).

This is because a logistic distribution (like a normal distribution) is defined from minus eternity to plus eternity and is the basis of the injury risk curve calculation in logistic regression. This non-zero risk at zero load often is not obvious because the offset is very small (see figure 16). The probability of getting a substantial offset at zero risk depends on the sample size. The smaller the sample size the greater the probability of getting a substantial injury risk at zero load (compare figure 16 and 17).

If the underlying distribution has an early rise of the risk, the bias of a logistic regression injury risk curve will be more pronounced (figure 19).

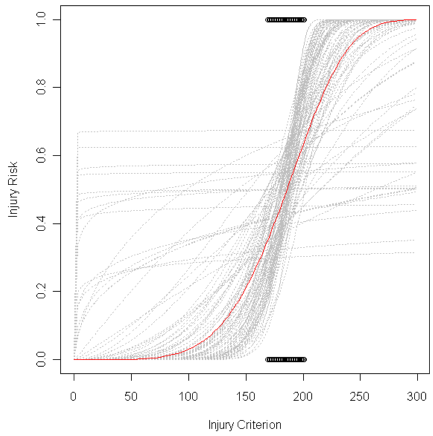

Figure 14. Small range of test severities (black dots showing the test results: 0.0=no injury, 1.0=injury).

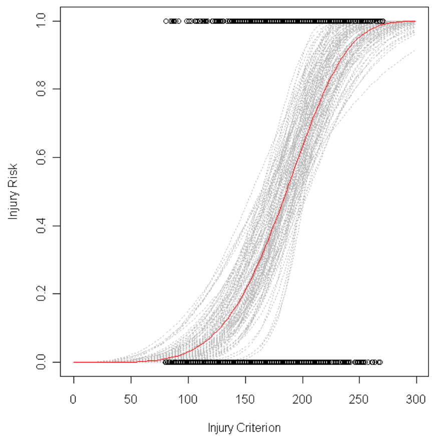

Figure 15. Large range of test severities (black dots showing the test results: 0.0=no injury, 1.0=injury).

A survival analysis with the assumption of a distribution which is defined only for positive values will always result in a injury risk curve passing through zero at zero load (as an example see figure 18 and 20 for a Weibull distribution).

CONCLUSION

This study shows factors influencing the injury risk curve, namely the data sampling, the data censoring, the test severities, and the statistical method. That means the reliability of the risk prediction of an injury risk curve depends on the size of the sample, the number of censored data, the range of test severities and the method used in the injury risk curve development. One or more of these factors may affect the injury risk curve in a way that leads to a significant difference of the predicted risk from the real risk.

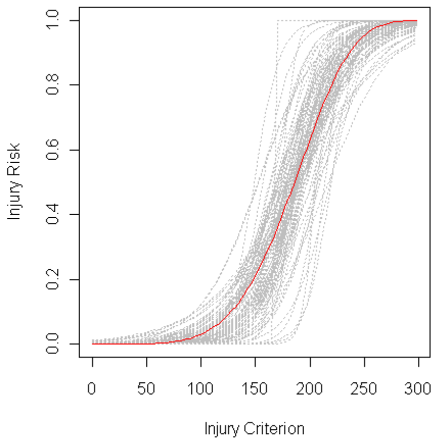

Figure 16. Logistic regression based on left and right censored data (sample size 40).

It is highly recommended to consider the reliability of an injury risk prediction in automotive safety. Before utilizing an injury risk curve the validity of the risk prediction should be assessed. Furthermore, the uncertainty of the injury risk prediction should be considered with respect to the required precision.

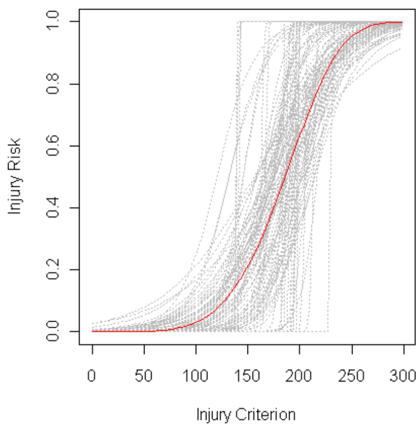

Figure 17. Logistic regression based on left and right censored data (sample size 20).

The calculation of a confidence interval (with description of the method used for its calculation) will give the user a measure of reliability. If there is no confidence interval given, the sample size and the number of censored data can provide an idea about the reliability of the risk prediction.

Logistic regression is not suitable for exact data in the dataset because it will lead to a bias in the injury risk curve and, therefore, in the risk prediction. Logistic regression should not be used with exact data. The same applies to all binary regression methods, e.g. probit regression.

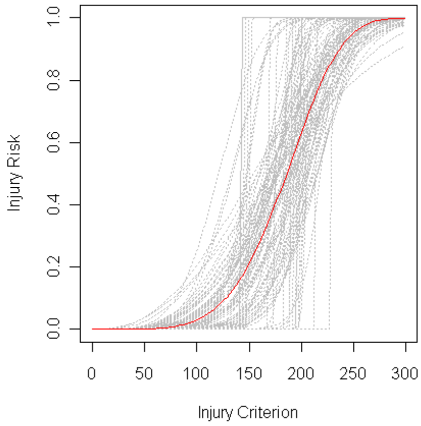

Figure 18. Survival analysis (Weibull) based on left and right censored data (sample size 20).

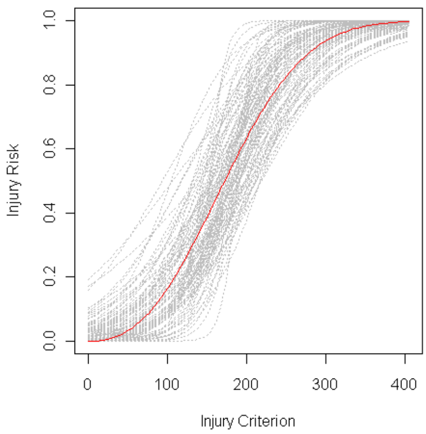

Figure 19. Logistic regression based on left and right censored data (sample size 40) and a early rising "real" distribution.

As have been demonstrated by this study the test severities used in the biomechanical tests have a significant influence on the result of the injury risk curve calculation. A wide range of test severities with respect to the underlying biomechanical tolerance limit distribution ensures a small bias in the risk prediction. In contrast to the simulation the underlying distribution of the biomechanical tolerance limits is not known in reality. Thus, the range of test severities is not known at which the specimens should be tested to be able to lead to optimal results. Nevertheless, it is clear that biomechanical tests should not be done in a small range of severities.

In general risk predictions outside the range of tested severities are less reliable than within the test severity range. Therefore, the range of test severities should be presented with the injury risk curve.

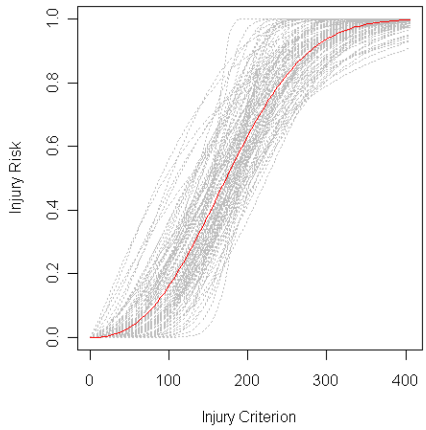

Figure 20. Survival analysis (Weibull) based on left and right censored data (sample size 40) and a early rising "real" distribution.

Exact data improves the reliability of injury risk curves, however, only with the use of an appropriate statistical method. The use of binary regression methods is critical because exact data will introduce bias in the injury risk prediction. In addition logistic regression shows an injury risk at zero load. This non-zero risk at zero load is more pronounced with lower sample size and with early rising "real" injury risk.

This study clearly shows, that an injury risk curve is affected by different factors and may lead to a unrealistic injury risk prediction. This can lead to misdirect the development of safety systems. Adapting safety systems to a misleading risk may have a negative impact on vehicle safety.

This study demonstrates that there is a need for a "quality control" for injury risk curves. Simply calculate a curve is not enough!

REFERENCES

[1] Sachs L. and Hedderich J. 2006. "Angewandte Statistik" Springer-Verlag, Berlin, Heidelberg.

[2] Dalgaard P. 2008. "Introductory Statistics with R" 2nd Edition, Springer Science +Business Media, LLC, New York.

[3] Hosmer D.W., Lemeshow S., May S. 2008. "Applied Survival Analysis" 2nd Edition, John Wiley & Sons, Inc., New Jersey.

[4] http://www.r-project.org

[5] Matloff N. 2008. "R for Programmers" http://heather.cs.ucdavis.edu/~matloff/R/RProg.pdf.