Statistical Considerations in the Development of Injury Risk Functions

TIMOTHY L. McMURRY and GERALD S. POPLIN

1 University of Virginia Department of Public Health Sciences, Charlottesville, Virginia
2 University of Virginia Center for Applied Biomechanics, Charlottesville, Virginia

Received 11 July 2014, Accepted 20 November 2014

Objective: We address 4 frequently misunderstood and important statistical ideas in the construction of injury risk functions. These include the similarities of survival analysis and logistic regression, the correct scale on which to construct pointwise confidence intervals for injury risk, the ability to discern which form of injury risk function is optimal, and the handling of repeated tests on the same subject.

Methods: The statistical models are explored through simulation and examination of the underlying mathematics.

Results: We provide recommendations for the statistically valid construction and correct interpretation of single-predictor injury risk functions.

Conclusions: This article aims to provide useful and understandable statistical guidance to improve the practice in constructing injury risk functions.

Keywords: injury risk function, survival analysis, cadaver test, crash dummies, modeling, statistics

Introduction

Injury risk data are of the form \((X_1, Y_1), \ldots, (X_n, Y_n)\), where \(X_i\) is a predictor—for example, an experienced force or deflection in a dummy test—and \(Y_i\) is a binary outcome such as the occurrence or nonoccurrence of injury in a matched cadaver test. For example,

\[ Y_i = \begin{cases} 1 & \text{if force } X_i \text{ resulted in injury} \\ 0 & \text{if force } X_i \text{ did not result in injury.} \end{cases} \]

The goal in the development of an injury risk curve is a model that accurately relates the probability of injury to the force experienced; in this article, we focus on single-predictor models without confounding variables (e.g., age or gender), but many of the ideas we present have extension to multiple variable problems. In the single-predictor literature, Petitjean and Trosseille (2011) provide a wide-ranging survey of available methods and their relative strengths and weaknesses. We focus on logistic regression and survival analysis, the 2 best performing approaches in their simulations.

Further, the International Organization for Standardization (ISO) has developed a stepwise process for the development of injury risk functions (ISO 2014). Kent and Funk (2004) demonstrated that when information about the exact force or deflection experienced at the moment of injury is available, it is important to incorporate this knowledge into the analysis. Our recommendations are broadly complementary to each of these works but offer important clarifications and refinements.

The remainder of the article is divided into four sections, each addressing a misunderstanding commonly seen in the injury risk function literature or in personal communications with others involved in their development; the four sections are outlined below. Where possible, we provide examples to justify and better articulate our recommendations, and each section provides practical recommendations. Our examples are based on a data set provided by J. Crandall (University of Virginia, email communication, April 5, 2013) showing the survival of eggs dropped from various heights onto a padded surface (Table A1, see online supplement). We chose this data set to avoid the appearance of criticizing particular analyses; our wish is only to improve practice. The remainder of the article is structured as follows.

The next section addresses the relationship between logistic regression and survival analysis, the 2 most common techniques. These approaches are technically very similar, which explains why, as previous research indicates (Petitjean and Trosseille 2011), they produce results of similar quality. We use this section to make the mathematical connections and
introduce notation that will be useful in subsequent sections, which focus on direct application.

The following section addresses the construction and interpretation of confidence intervals for injury risk functions. A practitioner must make several choices that on the surface seem immaterial but result in dramatically different interpretation and performance. These choices include whether or not the intervals should be horizontal or vertical, and the scale on which the intervals should be constructed. We provide recommendations on the best approach and justify our recommendations with simulations.

The next section discusses the difficulty in choosing a functional form for the regression based on model fit criteria. We demonstrate that with sample sizes typical in biomechanics and injury risk function development, the Akaike information criterion (AIC) does not reliably choose the optimal model form. As such, we are left to choose the functional form based on how realistic the resulting shape is likely to be and our best physical understanding of the mechanisms causing injury.

The final section focuses on the problem with using repeated measurements on the same test subjects. For example, it is not uncommon to test a subject first in a lower impact test that is not expected to be injurious and then retest the same subject with a more potentially injurious test. These repeated tests are a substantial violation of the assumptions underlying both survival analysis and logistic regression and need to be handled carefully. We explain the concerns associated with repeated testing, their theoretical basis, and some thoughts on the way this problem might be handled.

Logistic Regression and Survival Analysis

Injury risk data can be seen as either binary outcome data or survival data. When the binary outcome viewpoint is taken, a force $X$ either produced an injury, so $Y = 1$, or no injury, making $Y = 0$. When faced with binary data, most statisticians’ instincts are to attempt a logistic regression model.

Logistic regression ignores an important feature of biomechanical data: zero impact corresponds to zero risk of injury. To solve this, much of the injury risk literature instead turns to survival analysis, treating all data as either left or right censored. For example if data $(X, Y) = (15, 0)$ is observed, then the subject experienced an impact force of 15 and no injury resulted. These data can be treated as right censored because the force required for injury is now known to have been greater than 15, but it is not known how much greater the required force would need to be. Conversely, if the data $(X, Y) = (15, 1)$ is observed, then a force of 15 was applied and an injury resulted. It is now known that the force required for injury was at most 15, and the threshold needed for injury could have been less. Such data can be viewed as left censored.

The remainder of this section demonstrates that from a model fitting perspective, the 2 approaches are strongly related; for this reason they tend to produce similar results.

Logistic Regression

In the broader statistical literature, logistic regression is the most common approach for modeling binary outcome data. It assumes a model of the form

$$P[Y = 1|X] = \frac{\exp[\beta_0 + \beta_1 X]}{1 + \exp[\beta_0 + \beta_1 X]}, \tag{1}$$

where we use $P[Y = 1|X]$ to denote the probability of injury in an impact of magnitude $X$. The regression coefficients $\beta_0$ and $\beta_1$ are typically unknown and estimated from the data. The logistic model is flexible enough to provide a reasonable, although likely never perfect, model for $P[Y = 1|X]$ in cases where this probability either strictly increases or strictly decreases with an increase in $X$. Furthermore, transformations of $X$ can be used to improve model fit, although in practice it is often difficult to know which transformation, if any, would be most appropriate.

The coefficients $\beta_0$ and $\beta_1$ allow the logistic curve to be shifted left/right and to capture different rates of risk increase, analogous to changing the intercept and slope in linear regression. These coefficients are typically estimated by a process known as maximum likelihood. The idea is that the best estimates of these coefficients are the values that make the observed data most probable. A detailed discussion can be found in any text on mathematical statistics; see, for example, Hogg et al. (2004). The result is that the coefficients are chosen to maximize

$$L_L(\beta_0, \beta_1) = \prod_{i \in I} \frac{\exp[\beta_0 + \beta_1 X_i]}{1 + \exp[\beta_0 + \beta_1 X_i]} \prod_{i \in N} \left(1 - \frac{\exp[\beta_0 + \beta_1 X_i]}{1 + \exp[\beta_0 + \beta_1 X_i]}\right), \tag{2}$$

where $I$ is the set of experiments in which injury occurred and $N$ is the set of experiments with no injury. Defining

$$F_L(X; \beta_0, \beta_1) = \frac{\exp[\beta_0 + \beta_1 x]}{1 + \exp[\beta_0 + \beta_1 x]},$$

Eq. (2) can be rewritten as

$$L_L(\beta_0, \beta_1) = \prod_{i=1}^n F_L(X_i; \beta_0, \beta_1)^{Y_i} [1 - F_L(X_i; \beta_0, \beta_1)]^{1-Y_i}, \tag{3}$$

where for each $i$, the binary outcome $Y_i$ acts as a switch to choose one of the two potential terms in the product.

The values $\hat{\beta}_0$ and $\hat{\beta}_1$ that maximize (3) do not have closed-form solutions, but reliable algorithms to estimate them are implemented in all statistical software packages.

Survival Analysis

An alternative to logistic regression is to use survival analysis, treating all data as either left or right censored. Impacts resulting in injury are left censored because the force at which injury occurred is now known to be less than or equal to the
applied force. Impacts not resulting in injury are right censored because the force required for injury is now known to have been greater than the applied force. Though survival analysis takes a different view of the data than logistic regression, the end result is again a model relating the experienced force to the probability of injury, similar to (1) but with a different functional form. In fact, the ISO recommends comparing survival models with 3 different functional forms: Weibull, log-normal, and log-logistic. For simplicity and clarity we focus on the Weibull; the others are similar.

For example, if one assumes a Weibull model, the resulting injury risk function is of the form
\[
P[Y = 1|X_i] = 1 - e^{-(X_i/\lambda)^k}, \tag{4}
\]
where the parameters \(\lambda > 0\) and \(k > 0\), also referred to as the scale and shape, are unknown and estimated from the data.

Model (4) has several attractive features. First, the model always associates zero force with zero risk of injury. Second, because the parameters are constrained to be positive, the fitted curve always increases as the force increases. Finally, different values of \(k\) allow the fitted curve to take on different shapes, which might better model reality. In contrast, all potential logistic curves differ only in location and scale but essentially have the same shape.

As with logistic regression, the parameters in Eq. (4) are typically estimated by maximum likelihood. In order to describe the appropriate likelihood function, we let
\[
F_W(x; \lambda, k) = 1 - e^{-(x/\lambda)^k},
\]
and we define
\[
f_W(x; \lambda, k) = \frac{d}{dx} F_W(x; \lambda, k) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}
\]
to be the probability density function (pdf) associated with \(F_W\).

With this notation, the appropriate likelihood function is (see Klein and Moeschberger 2003, p. 75)
\[
L_\lambda(\lambda, k) = \prod_{i \in E} f_W(X_i; \lambda, k) \prod_{i \in I} F_W(X_i; \lambda, k)
\times \prod_{i \in N} [1 - F_W(X_i; \lambda, k)] \prod_{i \in V} [F_W(R_i; \lambda, k) - F_W(L_i; \lambda, k)]. \tag{5}
\]
In (5), \(I\) is the set of experiments where injury occurred at an impact less than or equal to \(X_i\), \(N\) is the set of experiments where no injury occurred, \(E\) is the set of experiments where the exact force required to cause injury is known (as discussed in Kent and Funk 2004), and \(V\) is the set of interval-censored observations, where the force required for injury is known to be between a left endpoint, \(L_i\), and a right endpoint, \(R_i\). Note that in the case of a Weibull model \(F_W(0, \lambda, k) = 0\), so it is equivalent to treat an observation where injury occurred as being either left censored or interval censored with left endpoint 0.

Equation (5) highlights an additional feature of the survival approach: survival analysis is naturally designed to incorporate more detailed information about when the injury occurred, should that information be available. For example, it can correctly handle any cases where the exact force required for injury is known (set \(E\)) and cases where the force is known to be between two non-zero limits (set \(V\)). Nonetheless, such extra information is often unavailable; in these cases, sets \(E\) and \(V\) are empty, and (5) reduces to
\[
L_\lambda(\lambda, k) = \prod_{i \in I} F_W(X_i; \lambda, k) \prod_{i \in N} [1 - F_W(X_i; \lambda, k)]
= \prod_{i = 1}^n F_W(X_i; \lambda, k)^Y_i [1 - F_W(X_i; \lambda, k)]^{1-Y_i}. \tag{6}
\]

Comparison Between Logistic and Survival Approaches
A comparison of Eqs. (3) and (6) shows that when the outcomes are either either censored injury/no injury, logistic regression and survival analysis attempt to maximize very similar likelihood functions. The only difference is the shape of the curves used to interpolate risk probabilities between 0 and 1. In fact, one could do survival analysis with a logistic distribution instead of a Weibull distribution and get results identical to those from logistic regression. For these reasons, in most situations it is expected that the 2 techniques will produce similar results.

Which of the 2 models is preferable then comes down to which functional form (e.g., (1), (4), or the corresponding form for log-normal and log-logistic) can take on a shape closest to the unknown true risk function. Later we discuss the limitations of statistical model choice.

Recommendations
Our first recommendation is with regard to the choice between logistic and survival analysis. Both approaches have merit. The survival approach produces a model that is more appealing on physical grounds. Logistic regression has some added flexibility in that the slope and intercept are both unconstrained, which means that the model is likely to produce reliable fits in the range of the data.

Based on these considerations, we recommend the following:

1. If exact uncensored experimental data exist, this information must be accounted for via likelihood function (5), although both traditional survival distributions and the logistic distribution may be considered.
2. Fit both a logistic regression and survival analysis. If they are qualitatively very different—for example, the logistic curve is fairly flat while the survival curve increases—any further results should be viewed as unreliable.
3. If the logistic and survival curves are similar, then the practitioner is free to use a survival approach. The rationale is that the logistic model’s flexibility makes it a good benchmark in the middle of the data, but near \(x = 0\) the survival
model will be more accurate. As long as the 2 models agree in the middle of the data, the survival model is preferred.

4. Whichever model is used should be checked with a goodness of fit test, such as the Hosmer-Lemeshow test. This is particularly important with the survival approach because it may be more likely to produce a reasonable looking survival curve whether or not the curve reflects the data.

Confidence Intervals in Injury Risk Curves

Confidence intervals can be thought of as quantifying the uncertainty in an estimated injury risk function that could be expected if one were able to repeat the entire experiment. More precisely, a confidence interval should contain the unknown “truth” with a prespecified likelihood. For example, one could construct a 95% confidence interval for the probability of injury at a specified impact. In theory, this interval should contain the actual probability of injury 95% of the times the experiment is conducted and the interval constructed. In order to construct meaningful intervals, a decision must first be made about the appropriate direction for the interval. Then an appropriate procedure must be used.

Horizontal vs. Vertical Confidence Intervals

In practice, 2 types of confidence intervals are commonly seen in the injury risk function literature. There are horizontal intervals that fix a probability of injury and ask how much uncertainty there is in the estimated impact resulting in that probability. The other type of interval is vertical in that it fixes uncertainty in the estimated impact resulting in the probability of injury. The other type of interval is vertical in that it fixes an impact force and asks how much uncertainty there is in the estimated risk of injury.

Both horizontal and vertical intervals have important application for injury risk, but they are not the same and the direction must be chosen based on context. If the goal is to set regulation based on a (for example) 30% chance of injury, a horizontal confidence interval describes the uncertainty in the estimated force resulting in this level of risk. If the goal is to describe the likelihood of injury resulting from a known impact, then a vertical interval would be appropriate.

Whichever confidence interval direction is chosen, many confidence intervals can be combined to produce confidence bands, such as those shown in Figure 1. These bands can only be interpreted pointwise, not as a confidence band for the entire injury risk function.

Vertical Confidence Intervals in Logistic Regression

We focus on vertical confidence intervals in logistic regression. Vertical intervals for survival models are less frequently implemented in software, so this section takes on additional importance when such intervals are required. For notational convenience, let \( p_x = P[Y = 1 | X] \) denote the probability of injury associated with an impact \( X \). With this notation, the logistic model (1) can be rewritten as

\[
\log \left( \frac{p_x}{1 - p_x} \right) = \beta_0 + \beta_1 X. \tag{7}
\]

The observed data \((X_1, Y_1), \ldots, (X_n, Y_n)\) are used to estimate the unknown coefficients \( \beta_0 \) and \( \beta_1 \) by maximum likelihood; we denote these estimates by \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \). Plugging these estimates into Eq. (7) and solving for \( p_x \) gives an estimated probability of injury

\[
\hat{p}_x = \frac{\exp[\hat{\beta}_0 + \hat{\beta}_1 X]}{1 + \exp[\hat{\beta}_0 + \hat{\beta}_1 X]}. \tag{8}
\]

A common and seemingly straightforward approach to constructing a \( 100 \times (1 - \alpha)/2 \)% confidence interval for \( \hat{p}_x \) in the injury risk literature is an interval of the form

\[
\hat{p}_x \pm z_{1 - \alpha/2} \text{SE}(\hat{p}_x), \tag{9}
\]

where \( z_{1 - \alpha/2} \) is the \( 100 \times (1 - \alpha/2) \)th percentile of the standard normal distribution, and \( \text{SE}(\hat{p}) \) is the standard error of \( \hat{p}_x \) as produced by most statistical packages.

The assumption underlying interval (9) is that the sampling distribution of \( \hat{p}_x \) is normal. For large samples with \( p_x \) not near 0 or 1, this assumption is justified by the delta method (e.g., Resnick 1999, p. 261). However, for typical biomechanical sample sizes and many values of \( \hat{p}_x \), the normal assumption is not adequate. This is easily seen because \( p_x \) and \( \hat{p}_x \) are constrained to be between 0 and 1, but the resulting confidence intervals are not. See Figure 1.

Fortunately, there is a straightforward solution to this difficulty. The simple solution is to construct the confidence interval on the log-odds scale and then transform the endpoints to the probability scale. In other words, the confidence interval starts with the confidence interval for \( \beta_0 + \beta_1 X \),

\[
(LL, UL) = \hat{\beta}_0 + \hat{\beta}_1 X \pm z_{1 - \alpha/2} \text{SE}(\hat{\beta}_0 + \hat{\beta}_1 X). \tag{10}
\]

where \( LL \) and \( UL \) denote the lower and upper limits of the interval. Interval (10) is then transformed to a confidence interval for \( p_x \) by applying the logistic function to \( LL \) and \( UL \),
Fig. 2. A fitted injury risk function from an egg drop experiment with 95% (vertical) transformed normal confidence intervals.

resulting in the interval

\[
\left( \frac{\exp[LL]}{1 + \exp[LL]} \exp[UL], \frac{\exp[UL]}{1 + \exp[UL]} \right). \tag{11}
\]

Interval (11) is symmetric on the log-odds scale, but asymmetric and constrained to be between 0 and 1 on the probability scale (see Figure 2) and comes closer to achieving the desired 95% confidence, as will be shown in our simulations.

The justification for interval (11) is based on the asymptotic normality of the maximum likelihood estimates \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \). Under fairly general conditions, applicable in this setting, maximum likelihood estimates have an asymptotic normal distribution (e.g., Hogg et al. 2004, p. 325). Unlike \( p_x \), \( \hat{\beta}_0 \) and \( \hat{\beta}_1 \) are unconstrained, which tends to make the asymptotic normal approximation for \( \hat{\beta}_0 + \hat{\beta}_1 X \) more accurate than that of \( p_x \). Because the inverse logit function is a monotonic one-to-one transformation, a 95% interval on the log-odds scale transforms into a 95% interval on the probability scale.

**Horizontal Confidence Intervals for Survival Models**

Survival models have an alternate representation to (4), which helps inform the construction of confidence intervals. The alternate form is (Klein and Moeschberger 2003, p. 46)

\[
\log(\text{Injury Force}) = \mu + \sigma W, \tag{12}
\]

where Injury Force is the exact force required to cause injury, and \( W \) is a random variable describing variation across the population. \( W \) can be assumed to have different probability distributions, with each distribution corresponding to a different class of survival model. For example, if \( W \) has an extreme value distribution, then model (12) produces a Weibull survival model. If \( W \) has a normal distribution, the resulting model is log-normal, and if \( W \) has a logistic distribution, then the resulting model is log-logistic.

From the point of view of (12), the most natural confidence intervals are for percentiles of Injury Force. In other words, in (12), a 10% chance of injury corresponds to the 10th percentile of \( W \), which for a given model (e.g., Weibull with \( W \) having an extreme value distribution) is known. The uncertainty in the required injury force then lies in the uncertainty of the maximum likelihood estimates of \( \mu \) and \( \sigma \). The result is that the most natural confidence interval is for the force corresponding to a fixed survival probability, which is a *horizontal* interval when viewed on a graph like Figure 1.

As a result of (12) there are again 2 plausible forms for the confidence interval. The first form, which we term data scale involves first estimating the required force and then constructing a confidence interval on the scale of the original data, resulting in the \( 100 \times (1 - \alpha) \)% confidence interval for the force that generates a 10% chance of injury of

\[
\exp[\hat{\mu} + \hat{\sigma} W_{0.10}] \pm z_{1 - \alpha/2} \text{SE}[\exp[\hat{\mu} + \hat{\sigma} W_{0.10}]], \tag{13}
\]

where \( W_{0.10} \) is the 10th percentile of the distribution of \( W \) (e.g., the 10th percentile of the extreme value distribution).

Alternatively, because the maximum likelihood estimates \( \hat{\mu} \) and \( \hat{\sigma} \) have an approximate normal distribution, confidence intervals could also be constructed on the log scale and then transformed to the injury force scale; the resulting interval will be termed the log scale interval. The force required for a 10% chance of injury starts with a \( 100 \times (1 - \alpha) \)% confidence interval for the log of the injury force:

\[
(\text{LL}, \text{UL}) = \hat{\mu} + \hat{\sigma} W_{0.10} \pm z_{1 - \alpha/2} \text{SE}[\hat{\mu} + \hat{\sigma} W_{0.10}],
\]

and the final confidence interval has the form

\[
(\exp[\text{LL}], \exp[\text{UL}]). \tag{14}
\]

Because injury force has much weaker constraints than the probability of injury modeled in logistic regression (i.e., probabilities are constrained to be between 0 and 1, whereas injury forces only need to be positive), the choice between (13) and (14) is less clear-cut. However, there is still some reason to think that (14) is preferable.

The reason we feel interval (14) is preferable is shown in the confidence intervals for the egg drop data in Figure 3. In this example, the intervals given by (13) curve to the left as the probabilities go up. This is counterintuitive because it seems reasonable to believe that the probability of injury should strictly increase as force increases. The interval (14) seems to
Injury Risk Functions

### Table 1. Logistic regression: coverage for the 2 types of nominal 95% vertical confidence intervals over 10,000 simulations. The top row shows the empirical coverage for the confidence intervals given by (9). The bottom row shows the coverage for the intervals given by (11)

| True risk of injury | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
|---------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Probably interval coverage | 0.87 | 0.90 | 0.92 | 0.93 | 0.93 | 0.92 | 0.91 | 0.88 | 0.84 |
| Log odds interval coverage | 0.96 | 0.97 | 0.97 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.97 |

better capture a physically realistic pattern. In our experience, the backwards curving shape is a frequent problem with interval (13) but less so with interval (14). In our simulations, the 2 intervals have similar coverage properties.

Remark: If model (12) is modified to Injury Force = μ + σW, where W is taken to have a logistic distribution, then the resulting survival model is identical to the logistic regression. As such, interval (13) can be used to produce horizontal confidence intervals for logistic regression; the necessary calculations are produced by many statistical packages. Because this model is directly fit on the scale of the data, there is no equivalent to interval (14).

### Simulations

#### Logistic Regression

In order to demonstrate the statistical improvement achieved by using interval (11), we conducted a small simulation study based on the egg drop data shown in Figures 1 and 2. The simulation works by generating data in a setting where we know the true injury risk function and evaluating each confidence interval to see whether or not it contains the true risk of injury.

In the simulated data, drop heights were chosen by drawing samples of size 50 with replacement from the drop heights shown in Figures 1 and 2. In order to assess a wider range of designs, these drop heights were then randomly perturbed by adding normally distributed mean 0 standard deviation 4 random numbers. Finally, injury/no injury data were simulated so that the probability of injury is given by the fitted risk function shown in Figures 1 and 2. For the simulated data set, a new injury risk function was estimated and the confidence intervals (9) and (11) calculated at the drop heights that correspond to true injury risks of 10%, 20%, ..., 90%. Finally, we determined whether or not the confidence intervals contained the true probabilities of injury. An example simulation is shown in Figure 4. The sample size of 50 was chosen to minimize the impact of nonconverging model estimates.

The simulation experiment was repeated 10,000 times. The empirical proportions of times the confidence intervals contain the true injury risk are shown in Table 1; the target is intervals that contain the truth 95% of the time. The intervals given by (9) show substantial undercoverage for p close to 0 or 1. The intervals given by (11) are in this case mildly conservative but consistently close to the desired 95% coverage.

### Survival Model

In this experiment, we compared the accuracies of intervals (13) and (14). In each iteration, the probability of injury was taken to be the fitted risk function shown in Figure 3. The drop heights were generated as in the preceding simulation. For each simulated data set, a new injury risk function was estimated and both types of horizontal confidence intervals were calculated for the necessary drop heights to produce injury risks of 10%, 20%, ..., 90%. Finally, we looked to see whether or not the confidence intervals contained the true drop heights.

In terms of coverage, the two techniques performed similarly, as shown in Table 2. Interval (13) performed better in 3 cases, (14) performed better in 5, and there was one tie. However, as seen in Figure 3, the data scale intervals often curve backwards for high probabilities of injury. Though this did not manifest itself in decreased coverage rates in our simulations, it does seem physically unreasonable and potentially unreliable.

### Recommendations

We recommend that horizontal or vertical confidence intervals be chosen based on the desired interpretation. Once a direction has been chosen, construct symmetric confidence intervals on

![Confidence Interval Coverage](image-url)

**Fig. 4.** An example of a simulated data set and confidence intervals. The dashed risk function indicates the probability of injury used in the simulation. The solid risk function indicates the probability of injury as estimated from the data. At each of 10%, 20%, ..., 90% probabilities of injury, the 2 types of confidence intervals are shown; the intervals given by (9) are solid and those given by (11) are dash-dots.

### Table 2. Coverage for the 2 types of nominal 95% horizontal confidence intervals over 10,000 simulations. The top row shows the empirical coverage for the confidence intervals given by (13). The bottom row shows the coverage for the intervals given by (14)

|          | 10% | 20% | 30% | 40% | 50% | 60% | 70% | 80% | 90% |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Data scale | 0.92 | 0.92 | 0.93 | 0.94 | 0.96 | 0.93 | 0.90 | 0.88 | 0.86 |
| Log scale  | 0.91 | 0.91 | 0.92 | 0.95 | 0.96 | 0.95 | 0.92 | 0.89 | 0.88 |
the transformed scale and then transform them back to the scale of the problem as is done in intervals (11) and (14). Horizontal confidence intervals for logistic regression should be calculated via (13).

The Difficulty of Choosing the Functional Form

The ISO (2014) recommends using the AIC to choose between Weibull, log-normal, and log-logistic survival models. In our simulations, we demonstrate that with typical biomechanical sample sizes, the AIC makes a somewhat arbitrary choice. Before presenting our simulation results, we argue that the choice of survival model is strongly related to the choice of predictor variable, and we discuss some mathematical considerations that we feel, in the absence of additional information or a large sample size, favor the Weibull model over the other 2 survival models.

Biomechanical Considerations

In most biomechanical experiments, there is a choice of injury predictor. For example, when predicting egg breakage, one might consider any of drop height, momentum, or kinetic energy on impact as the x variable. These 3 quantities are strongly related, and to the extent that they are related, switching from one to another is simply a nonlinear rescaling of the predictor axis. From this point of view, the choice between these potential predictors is equivalent to a choice of the functional form for the statistical model. Because statistical methods offer little help with the choice of model, we also cannot expect them to adequately choose between strongly related predictors. We recommend choosing the predictor that best reflects available physical and mechanical understanding.

Aesthetic Considerations

The mathematical considerations favoring the Weibull model over the log-normal and log-logistic models revolve around the hazard function. Let the random variable X denote the exact force needed to cause injury to a randomly selected subject, and let x denote a fixed force. Then the hazard function is defined as

$$ h(x) = \lim_{\Delta x \to 0} \frac{P[x \leq X < x + \Delta x | X \geq x]}{\Delta x}. $$

In other words, h(x)Δx is the probability that a subject who is able to experience all forces up to x uninjured would be injured by a force between x and x + Δx for small Δx. It can be shown that $h(x) = f(x)/(1 - F(x))$, where $F(x) = P[X \leq x]$ is the injury risk function and $f(x) = F(x)$ is the corresponding probability density. In the context of an injury risk function, it may be reasonable to expect that h(x) be increasing.

The Weibull distribution with shape parameter k and scale \(\lambda\) has hazard function $h_w(x) = kx^{k-1}/\lambda^k$. $h_w(x)$ is easily seen to be increasing for $k > 1$, constant for $k = 1$, and decreasing for $k < 1$. An injury risk model with nondecreasing hazard function seems appealing, but of the approaches we consider herein, only the Weibull model can have this property.

The log-logistic distribution has $h_l(x) = (\beta/\alpha)(\alpha x^{\beta-1}/(1 + (\alpha x)^\beta))$, where the values of \(\alpha\) and \(\beta\) can be derived from \(\mu\) and \(\sigma\) in (12). When $\beta \leq 1$, $h_l(x)$ is strictly decreasing. When $\beta > 1$, $h_l(x)$ has a single peak at $x = (\alpha/(\beta - 1))^{1/\beta}$, which corresponds to an injury probability of $1 - 1/\beta$. After this point, the hazard rate drops, which may be unnatural in our context. The log-normal hazard function is qualitatively similar to the log-logistic hazard function but does not have a simple closed form.

In practice, the hazard function is a derivative and is likely of secondary importance to the value of the fitted injury risk function itself. Therefore, we do not feel that these aesthetic considerations should be the deciding factor in choosing the functional form. However, in the absence of better understanding of the detailed mechanisms of biomechanical injury, we feel, this reasoning lends some favor to the Weibull model over the log-normal and log-logistic models.

Simulations

In order to demonstrate the difficulty in choosing the appropriate functional form of the survival curve, we conducted a small simulation experiment. We start off with the egg drop data to ensure a survival curve that represents an actual risk of injury in a realistic way. We then fit four injury risk functions to these data using logistic regression and survival analyses using the log-logistic, log-normal, and Weibull distributions.

We used the 4 fitted injury risk functions to simulate new data sets of size n = 50, large for a biomechanical injury risk function, where we know the true functional form. Finally, we use the AIC to choose the best functional form based on the simulated data. We simulated 10,000 data sets from each of the 4 fitted injury risk functions. The frequencies with which the AIC chose the different distributions are shown in Table 3.

| True risk function | Log-Logistic(%) | Log-Normal(%) | Logistic(%) | Weibull(%) |
|-------------------|----------------|--------------|-------------|------------|
| Log-logistic      | 6.76           | 58.35        | 24.08       | 10.81      |
| Log-normal        | 6.47           | 61.76        | 21.70       | 10.07      |
| Logistic          | 5.98           | 38.03        | 40.94       | 15.05      |
| Weibull           | 6.26           | 50.17        | 29.53       | 14.04      |

Table 3. Results from a simulation demonstrating the difficulty of choosing the correct functional form with AIC

The simulation shows that the AIC tended to choose the log-normal and log-logistic models no matter what the true risk function; this demonstrates that with small sample sizes the AIC cannot reliably select the best model. In order to ensure that this was a sample size problem, we reran the simulation instead using simulated data sets of size n = 5,000. With the larger sample size, the AIC was able to identify the correct functional form the majority of the time, with success rates ranging from 67% for the log-logistic model to 97% for the logistic model.
We should not view these small sample troubles as a shortcoming of the AIC. The fundamental cause is that the 4 families of models are each flexible enough to provide good (and usually similar) fits to the data, and other model fit metrics can be expected to have the same trouble.

**Recommendations**

1. Start with a logistic regression. Logistic regression provides a best linear fit in the log-odds space. This fit is typically reasonable near the middle of the data in the x direction for the same reason that one term Taylor approximations often work well over small domains.

2. If extrapolation to smaller risks is desired, fit a Weibull distribution in addition to the logistic regression. If the Weibull fit is similar to the logistic regression over the range of data, then the Weibull can reasonably be thought of as an improvement because it passes through (0, 0) and matches the logistic regression where the majority of data were collected.

3. If the Weibull and logistic fits differ substantially, the logistic regression should be taken as more reliable in the middle of the data, and the Weibull should not be used. The reasoning underlying this point is the same as the reasoning underlying Recommendation 1: logistic regression has enough flexibility to consistently produce a good fit in the middle of the data in the x direction. Because survival models pass through (0, 0), their flexibility is diminished and they should be viewed as less reliable in the middle of the data, even though they are typically more accurate for very small impacts.

**Repeated Measurements on the Same Subject**

It is not unusual to see repeated measurements on the same subject used as independent data points in an injury risk function development. For example, many injury risk functions are built on data where a cadaver is impacted once with a low and ultimately noninjurious impact and then impacted again at a higher, and potentially injurious, force. Doing this makes 2 fundamental mistakes:

1. It implicitly assumes that the first impact did not weaken the test subject in any way.

2. It ignores the difference between repeating tests on one subject versus conducting tests across the population.

Point 2 is the more subtle. Any particular subject has its own injury risk curve that is likely very steep. For example, each egg has a drop height at which it starts to crack. Below that drop height it survives, above that height it breaks. So, its injury risk curve is almost vertical at one height. Another egg has its own injury risk curve that is vertical at a different height, accounting for needing a different impact to crack. Population injury risk curves, like the ones we seek to fit, model the proportion of eggs that would have cracked when dropped from a given height. If we take multiple measurements on a single subject, we find out more about that subject’s injury risk function, but we do not collect another data point on the population as a whole.

There are numerous statistical techniques for dealing with repeated measurements on the same subject. Logistic regression has been extended to correlated data through random effects models and generalized estimating equations (GEE; Diggle et al. 2002), although only GEE produces a model with the desired population, level interpretation (termed a *marginal model* in the statistics literature). Survival analysis has been extended to handle correlation through frailty models (Klein and Moeschberger 2003, ch. 13), but these models again measure risk at the individual rather than population level. Lipsitz and Ibrahim (2000) discussed potential GEE-type extensions to parametric survival models.

In the context of repeated impacts on the same cadaver, we have found useful a third approach that better meshes with the survival interpretation of injury risk data. We illustrate our approach with an example. Suppose an egg survives a drop of 50 cm but is broken by a drop from 100 cm. We treat this as a single interval censored-observation, where the injury occurred at an unknown point between 50 and 100 cm. If the egg had also survived the 100-cm drop, we treat the egg as providing a single data point right censored at 100 cm. Mathematically, estimation is handled by the maximization of (5) with interval-censored terms. Interval censoring correctly handles the concern of testing the same subject at multiple impact levels. Like all other techniques for handling repeated measures, it does not do a good job accounting for potential damage that may accrue during noninjurious rounds of testing.

In other contexts, we would consider other models. For example, in a lower extremity test where both legs of a cadaver are impacted (either together or separately), the interval censoring approach is no longer reasonable. In this case, we would lean toward either a GEE model or a frailty model, although the latter would require some care in order to achieve a population-level interpretation of injury risk.

**Recommendations**

When modeling requires the use of repeated measurements on the same subject, a statistical method that correctly accounts for the correlation between these measurements must be used. The methods we describe are commonly used in other applications and widely implemented in statistical software.

**Acknowledgments**

The authors thank Jim Funk and Jeff Crandall for helpful discussions and 3 anonymous referees whose insightful recommendations greatly improved this article.

**Supplementary Materials**

Supplemental data for this article can be accessed on the publisher’s website.
References

Diggle P, Heagerty P, Liang KY, Zeger S. Analysis of Longitudinal Data. New York: Oxford University Press; 2002.
Hogg R, McKean J, Craig A. Introduction to Mathematical Statistics. 6th ed. Upper Saddle River, NJ: Prentice Hall; 2004.
International Organization for Standardization. Road vehicles—procedure to construct injury Risk Curves for the Evaluation of Occupant Protection in Crash Tests. Geneva, Switzerland: Tech. Rep. ISO/TS 18506. 2014.
Kent R, Funk J. Data Censoring and Parametric Distribution Assignment in the Development of Injury Risk Functions From Biochemical Data. Detroit, MI: Society of Automotive Engineers; 2004. Technical Paper 2004-01-0317.
Klein J, Moeschberger M. Survival Analysis: Techniques for Censored and Truncated Data. 2nd ed. New York, NY: Springer; 2003.
Lipsitz SR, Ibrahim JG. Estimation with correlated censored survival data with missing covariates. Biostatistics. 2000;1:315–327.
Petitjean A, Trosseille X. Statistical simulations to evaluate the methods of the construction of injury risk curves. Stapp Car Crash J. 2011;55:411–440.
Resnick S. A Probability Path. Boston, MA: Birkhäuser; 1999.