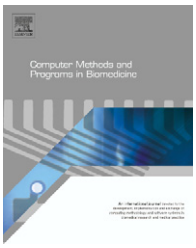
COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 106 (2012) 274–286

journal homepage: www.intl.elsevierhealth.com/journals/cmpb



A bootstrap approach for lower injury levels of the risk curves

*Yuan-Chiao Lu, Costin D. Untaroiu**

Virginia Tech and Wake Forest University School of Biomedical Engineering and Sciences Center for Injury Biomechanics, Blacksburg, VA, USA

ARTICLE INFO

Article history:
Received 30 September 2010
Received in revised form
10 February 2011
Accepted 23 March 2011

Keywords:
Bootstrapping
Injury risk criterion
Survival analysis
Statistical simulation

ABSTRACT

Survival analysis is widely applied to develop injury risk curves from biomechanical data. To obtain more accurate estimation of confidence intervals of parameters, bootstrap method was evaluated by a designed simulation process. Four censoring schemes and various sample sizes were considered to investigate failure time parameters corresponding to low-level injury probabilities. In the numerical simulations, the confidence interval ranges developed by bootstrapping were about two-third of the corresponding ranges calculated by asymptotical normal approximation and showed highest reduction for censored datasets with smaller sample size (≤40). In analysis of two experimental datasets with reduced sample sizes and mixed censored data, it was shown that the bootstrapping reduce significantly the confidence intervals as well. The results presented in this study recommend using bootstrapping in development of more accurate confidence intervals for risk curves in injury biomechanics, which consequently will lead to better regulations and safer vehicle designs.

1. Introduction

Current vehicle designs [1,2] and transportation regulations [3,4] depend heavily on injury criteria that define the risk of injury as a function of measured engineering parameters. However, injury data collected from biomechanical studies have limited applicability towards developing robust injury criteria models. First, the sample size of experimentally collected data is typically less than 50 which for the given variance in the subject population may not yield statistically significant inferences and potentially bias the prediction of the injury risk models [5]. Secondly, development of most standard injury risk models require assumptions about the underlying sample distribution which is difficult to determine due to the small sample size of the available injury data [6]. To account for the above limitations, "bootstrapping" procedure provides a statistical method capable of constructing more robust estimates along with confidence intervals (CIs) for data with small sample sizes. Essentially, bootstrapping is a resampling-based procedure to artificially create multiple random samples from a unique test sample to improve the statistical power for prediction of sample parameters and error analysis [7–12]. While a few studies investigated the use of bootstrapping in development of risk curves [13–15], the information about the effect of bootstrapping applied to datasets with various censoring types and sample sizes on lower-level injury risk curves was rarely seen in the literature [6]. The current study using a novel simulation approach tries to develop an understanding regarding the effectiveness of bootstrapping in analyzing biomechanical datasets.

The current study is organized as follows. The censoring injury data, point estimates of the lower-level failure time, and their asymptotical CIs are described in Section 2. In Section 3, two bootstrap CIs are proposed for the estimation of the lower-level failure time for different censoring schemes. The simulation results corresponding to each censoring scheme are presented in Section 4. In Section 5, the effectiveness of

* *Corresponding author*. Tel.: +1 540 231 1890; fax: +1 540 231 2953.
E-mail address: costin@vt.edu (C.D. Untaroiu).
0169-2607/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.cmpb.2011.03.015

bootstrapping versus typical asymptotical approach is illustrated on two experimental datasets.

2. Data censoring and simulation procedure

The development of accurate injury risk functions from biomechanical data is challenging due to the contents of the dataset points, which commonly are right-, left- or interval-censored. A stimulus value is considered to be right-censored if no injury is observed at the end of the test itself. For such case, the stimulus value is not large enough to cause injury of the specimens. In contrast, a stimulus value is said to be left-censored if injury is observed at the end of the test. In this case, the stimulus value is higher than the actual threshold value while the actual threshold value is unknown. Furthermore, when the actual threshold value is only known to occur within an interval, this case is called interval-censoring. A dataset that includes both right-censored and left-censored (or interval-censored) observations is called doubly censored data. It is also possible to obtain the exact threshold values in biomechanical injury data. These data are called uncensored data.

To simulate the process of biomechanical testing and construct injury risk curves, an underlying ("real") distribution of the biomechanical tolerance limit in the population has to be predefined. Since a certain validated statistical distribution of biomechanical injury threshold in the population is lacking in the literature, the normal risk function was assumed to represent the underlying distribution [6]. N samples (where $N = 20$, 30, 40, 50, 60, 70, and 80) were randomly selected from the underlying normal distribution to simulate actual injury values. To illustrate the bootstrapping methodology, the mean (350) and standard deviation (stdv) (70) for the normal distribution were chosen based on the mean fracture moment and stdv of the mid-shaft femur and thigh under bending loading [5,13]. To simulate censored data, "observed values" were randomly drawn from the uniform distribution in the interval (250, 450). This interval covers major observed experimental values, excluding outliers [5]. Then, each of the N injury values was compared with an observed value. If the sampled injury value was below the sampled observed value, the observation was considered to be an injury (left- or interval-censored); otherwise it was a non-injury (right-censored). The lower bound of tolerance limit for interval-censored data was set to be 100, the approximate lower bound of 99.9% coverage of the assumed normal distribution with mean 350 and stdv 70. To simulate the uncensored data, the censored data were replaced by their corresponding injury values from the underlying normal risk distribution. Each doubly censored sample set had half injury data and half non-injury data. For example, when $N = 50$, there were 25 injury values and 25 non-injury values. Four different censoring types were investigated: (1) all uncensored injury data; (2) injury: uncensored; non-injury: right-censored; (3) injury: interval-censored; non-injury: right-censored; (4) injury: left-censored; non-injury: right-censored.

Parametric survival models have been widely used in failure time survival analysis for prediction purpose because parametric models provide an analytical form for the risk function and are available in numerous statistical software packages [16]. Some parametric models such as Weibull, log-logistic, and lognormal distributions with a lower bound at zero were commonly applied to injury risk data [12,13,16,17]. Among all available parametric survival models, Weibull distribution has been demonstrated as central to the parametric analysis of failure time data. First introduced to represent the probability distribution of the breaking strength materials [18], the Weibull distribution is commonly used in reliability analysis and is therefore chosen for modeling biomechanical injury tolerance data [12,13,19].

Let X be a non-negative random variable. The probability density function (pdf) of the two-parameter Weibull distribution is

$$f(x) = \alpha\lambda x^{\alpha-1} e^{-\lambda x^{\alpha}} \quad x \geq 0; \alpha \text{ and } \lambda > 0 \tag{1}$$

The cumulative distribution function (cdf) of the Weibull distribution can be determined in a closed form:

$$F(x) = 1 - e^{(-\lambda x^{\alpha})} \quad x \geq 0; \alpha \text{ and } \lambda > 0 \tag{2}$$

Taking the log transformation of X such that $Y = \ln X$, and redefining the parameters as $\lambda = \exp(-\mu/\sigma)$ and $\alpha = 1/\sigma$ [19], the pdf and the cdf of Y can be written as

$$f(y) = \left(\frac{1}{\sigma}\right) e^{[(y-\mu)/\sigma - e^{[(y-\mu)/\sigma]}]} \tag{3}$$

and

$$F(y) = 1 - \exp(-e^{[(y-\mu)/\sigma]}) \tag{4}$$

The maximum likelihood estimators (MLEs) of μ and σ, denoted as $\hat{\mu}$ and $\hat{\sigma}$, are known as the intercept and scale parameters of the Weibull distribution, which can be found numerically in most statistical packages [19]. Their corresponding 95% asymptotical CIs, $(\hat{\mu}_{2.5\%}, \hat{\mu}_{97.5\%})$, and $(\hat{\sigma}_{2.5\%}, \hat{\sigma}_{97.5\%})$, are obtained using the delta method [20].

The invariance property of the MLE provides that the MLEs of λ and α are given by

$$\hat{\lambda} = e^{(-\hat{\mu}/\hat{\sigma})} \quad \text{and} \quad \hat{\alpha} = \frac{1}{\hat{\sigma}} \tag{5}$$

The asymptotical variances and covariance for $\hat{\lambda}$ and $\hat{\alpha}$ can be obtained by applying the delta method [19].

$$\begin{aligned} \mathrm{Var}(\hat{\lambda}) &= e^{(-2\hat{\mu}/\hat{\sigma})}\left[\frac{\mathrm{Var}(\hat{\mu})}{\hat{\sigma}^2} + \frac{\hat{\mu}^2\,\mathrm{Var}(\hat{\sigma})}{\hat{\sigma}^4} - \frac{2\hat{\mu}\,\mathrm{Cov}(\hat{\mu},\hat{\sigma})}{\hat{\sigma}^3}\right], \\ \mathrm{Var}(\hat{\alpha}) &= \frac{\mathrm{Var}(\hat{\sigma})}{\hat{\sigma}^4}, \text{ and} \\ \mathrm{Cov}(\hat{\lambda},\hat{\alpha}) &= \exp\left(\frac{-\hat{\mu}}{\hat{\sigma}}\right)\left[\frac{\mathrm{Cov}(\hat{\mu},\hat{\sigma})}{\hat{\sigma}^3} - \frac{\hat{\mu}\,\mathrm{Var}(\hat{\sigma})}{\hat{\sigma}^4}\right] \end{aligned} \tag{6}$$

While the injury probabilities less than 50% are commonly used as safety limits in vehicle or blast design, the lower-half region of injury risk curves was investigated in this study. Five target injury values (t_{10}, t_{20}, t_{30}, t_{40}, and t_{50}) corresponding to five injury probabilities (10%, 20% 30%, 40%, and 50%) were

obtained from the MLEs $\hat{\lambda}$ and $\hat{\alpha}$ using R statistical package [21]. For example, the MLE of t_{10} is calculated as

$$F(\hat{t}_{10}) = 1 - e^{(-\hat{\lambda}\hat{t}_{10}^{\hat{\alpha}})} = 10\%, \quad \text{and,} \quad \hat{t}_{10} = \sqrt[\hat{\alpha}]{\frac{\log(10/9)}{\hat{\lambda}}} \tag{7}$$

Similarly, we can derive the MLEs for t_{20}, t_{30}, t_{40}, and t_{50}.

$$\hat{t}_{20} = \sqrt[\hat{\alpha}]{\frac{\log(5/4)}{\hat{\lambda}}}, \quad \hat{t}_{30} = \sqrt[\hat{\alpha}]{\frac{\log(10/7)}{\hat{\lambda}}}, \quad \hat{t}_{40} = \sqrt[\hat{\alpha}]{\frac{\log(5/3)}{\hat{\lambda}}}$$
$$\text{and } \hat{t}_{50} = \sqrt[\hat{\alpha}]{\frac{\log(2)}{\hat{\lambda}}} \tag{8}$$

t_{50} is also known as the "median survival time."

To construct the asymptotical CIs for $t_i = g(\alpha,\lambda)$, where i = 10, 20, 30, 40, and 50, the distributions of the estimates of α and λ are assumed to be approximately normally distributed. Therefore, the asymptotical CIs for t_i can be obtained using the delta method, that is, $\hat{t}_i \overset{a}{\sim} N(t_i, \text{var}(\hat{t}_i))$ [20]. Under the delta method, the asymptotical variance of $\hat{t}_i = g(\hat{\alpha}, \hat{\lambda})$ is given by

$$\text{Var}[g(\hat{\alpha},\hat{\lambda})] = \left[\frac{\partial}{\partial\alpha}g(\hat{\alpha},\hat{\lambda})\right]^2 \text{Var}(\hat{\alpha}) + \left[\frac{\partial}{\partial\lambda}g(\hat{\alpha},\hat{\lambda})\right]^2 \text{Var}(\hat{\lambda}) + 2\left[\frac{\partial}{\partial\alpha}g(\hat{\alpha},\hat{\lambda})\right]\left[\frac{\partial}{\partial\lambda}g(\hat{\alpha},\hat{\lambda})\right]\text{Cov}(\hat{\alpha},\hat{\lambda}) \tag{9}$$

where

$$\frac{\partial}{\partial\alpha}g(\hat{\alpha},\hat{\lambda}) = -t_i\alpha^{-2}\log\left(\frac{\log(100/(100-i))}{\lambda}\right)\Bigg|_{\alpha=\hat{\alpha},\lambda=\hat{\lambda}} \tag{10}$$

$$\frac{\partial}{\partial\lambda}g(\hat{\alpha},\hat{\lambda}) = -\frac{\lambda^{-1}}{\alpha}t_i\Bigg|_{\alpha=\hat{\alpha},\lambda=\hat{\lambda}}, \quad \text{where } i = 10, 20, 30, 40 \text{ and } 50 \tag{11}$$

3. Bootstrap confidence intervals

In this section, the steps for the construction of the bootstrap CIs for the parameters, t_{10}, t_{20}, t_{30}, t_{40}, and t_{50} are introduced. The advantage of the bootstrap is that the joint distribution of the MLEs is not assumed to be normal, unlike in the delta method [10,15]. In the bootstrapping procedure, the non-parametric resampling bootstrap was first utilized to construct the point estimators of t [10]. The percentile method (bootstrap-p) [7,10,15] and bias corrected and accelerated method (BCa) [9,10,22] were then used to construct the 95% bootstrap CIs of t. The bootstrap-p method is computationally simpler than other existing methods (e.g. bias corrected method) and no invalid parameter values can be included in the interval. However, the coverage error of bootstrap-p method, defined as the probability that the bootstrap estimate of a parameter exceeds the true value of the parameter, is often substantial if the distribution of the estimated parameter is not nearly symmetric [10]. On the other hand, the BCa method is more accurate in terms of the coverage error than the percentile method, but more time consuming [10].

Let $U = (x,\delta)$ be the observed data where $x = (x_1,\ldots,x_n)$ is the vector of exact/censored data points and $\delta = (\delta_1,\ldots,\delta_n)$ is the vector of indicators of censored observations. The steps for the construction of bootstrap CIs are shown as follows.

3.1. Bootstrap-p

[a] Sample n observations randomly with replacement from U to obtain a bootstrap dataset $(x_1^*, \delta_1^*), \ldots, (x_n^*, \delta_n^*)$, denoted U^*.
[b] From the bootstrap sample in [a], find the MLEs of t_i, i = 10, 20, 30, 40, and 50, denoted by $\hat{t}_i^*$.
[c] Repeat steps [a] and [b], $B = 2000$ times.
[d] From $\hat{\underline{t}}_i^* = (\hat{t}_{i(1)}^* \le \hat{t}_{i(2)}^* \le \cdots \le \hat{t}_{i(B)}^*)$, find a 95% bootstrap CI given by $(\hat{t}_{i(q_1)}^*, \hat{t}_{i(q_2)}^*)$ where $q_1 = 2000 \times 0.05/2$, $q_2 = 2000 - q_1$, i = 10, 20, 30, 40, and 50.

3.2. Bias corrected and accelerated method (BCa)

[a'] Sample n observations randomly with replacement from U to obtain a bootstrap dataset $(x_1^*, \delta_1^*), \ldots, (x_n^*, \delta_n^*)$, denoted U^*.
[b'] From the bootstrap sample in [a'], find the MLEs of t_i, i = 10, 20, 30, 40, and 50, denoted by $\hat{t}_i^*$.
[c'] Repeat steps [a'] and [b'], $B = 2000$ times.
[d'] Count the number of members of [c'] that are less than $\hat{t}_i$ (i = 10, 20, 30, 40, and 50) calculated by the asymptotical method. Call this number p and set $b = [\Phi(p/B)]^{-1}$, where $\Phi(\bullet)$ is the standard normal cdf.
[e'] Calculate a^*, which is defined by

$$a^* = \frac{\sum_{j=1}^{n}\left(\tilde{t}_i - \hat{t}_i^j\right)^3}{6\left[\sum_{j=1}^{n}\left(\tilde{t}_i - \hat{t}_i^j\right)^2\right]^{3/2}}, \quad i = 10, 20, 30, 40, \text{ and } 50 \tag{12}$$

where $\hat{t}_i^j$ is the jackknife value of a statistic $\hat{t}_i$ based on the original sample with the jth point removed, and $\tilde{t}_i$ is the mean of $\hat{t}_i^j$.
[f'] The BCa CI endpoints are also given by percentiles of the bootstrap distribution, but not necessarily the same ones as by the percentile method. The BCa $100(1-\alpha)$% CI is given by α_1th and α_2th percentiles as lower and upper bounds, respectively, where

$$\alpha_1 = \Phi\left(\frac{(b + (b + z^{(\alpha/2)}))}{(1 - a^*(b + z^{(\alpha/2)}))}\right), \quad \text{and} \quad \alpha_2 = \Phi\frac{(b + (b + z^{(1-\alpha/2)}))}{(1 - a^*(b + z^{(1-\alpha/2)}))} \tag{13}$$

Here $Z^{(\alpha)}$ is the 100αth percentile point of a standard normal distribution.

Three indexes were calculated and used to evaluate the performance of the bootstrapping, and 1000 trials were generated in order to calculate the averaged indexes. First, the ranges of CIs calculated by asymptotical, bootstrap-p and BCa methods for parameters t_i (i = 10, 20, 30, 40, and 50) were computed with the range defined as the difference between the upper and lower limits of a CI [15,23]. Two-sample t-test was conducted to test the equality of the mean values of CI ranges between different cases. Second, the relative bias (RB) is computed as the difference between the

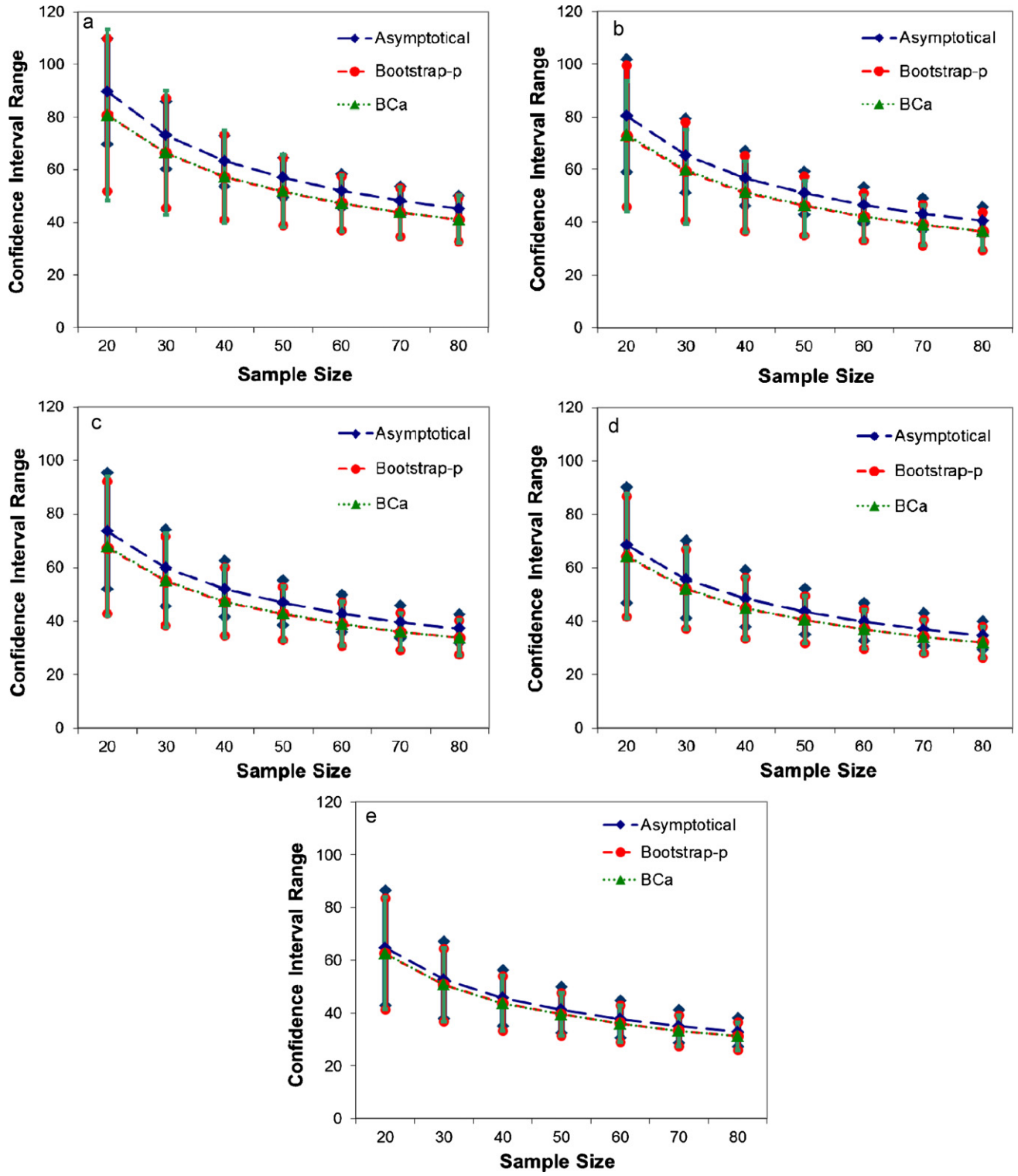


Fig. 1 – Confidence interval range for uncensored data: (a) 10% injury risk; (b) 20% injury risk; (c) 30% injury risk; (d) 40% injury risk; (e) 50% injury risk.

$\hat{t}_i$ (or the average of $\hat{t}_i^*$) and the true t_i divided by the true t_i [23]. Third, the observed coverage probability (CP) was recorded, i.e. the fraction of the 1000 computed CIs which contained the true t_i of the underlying normal distribution [24–26].

4. Results

The CI range, CP, and RB were calculated for all simulation combinations. Each point in Figs. 1–4, represents the averaged

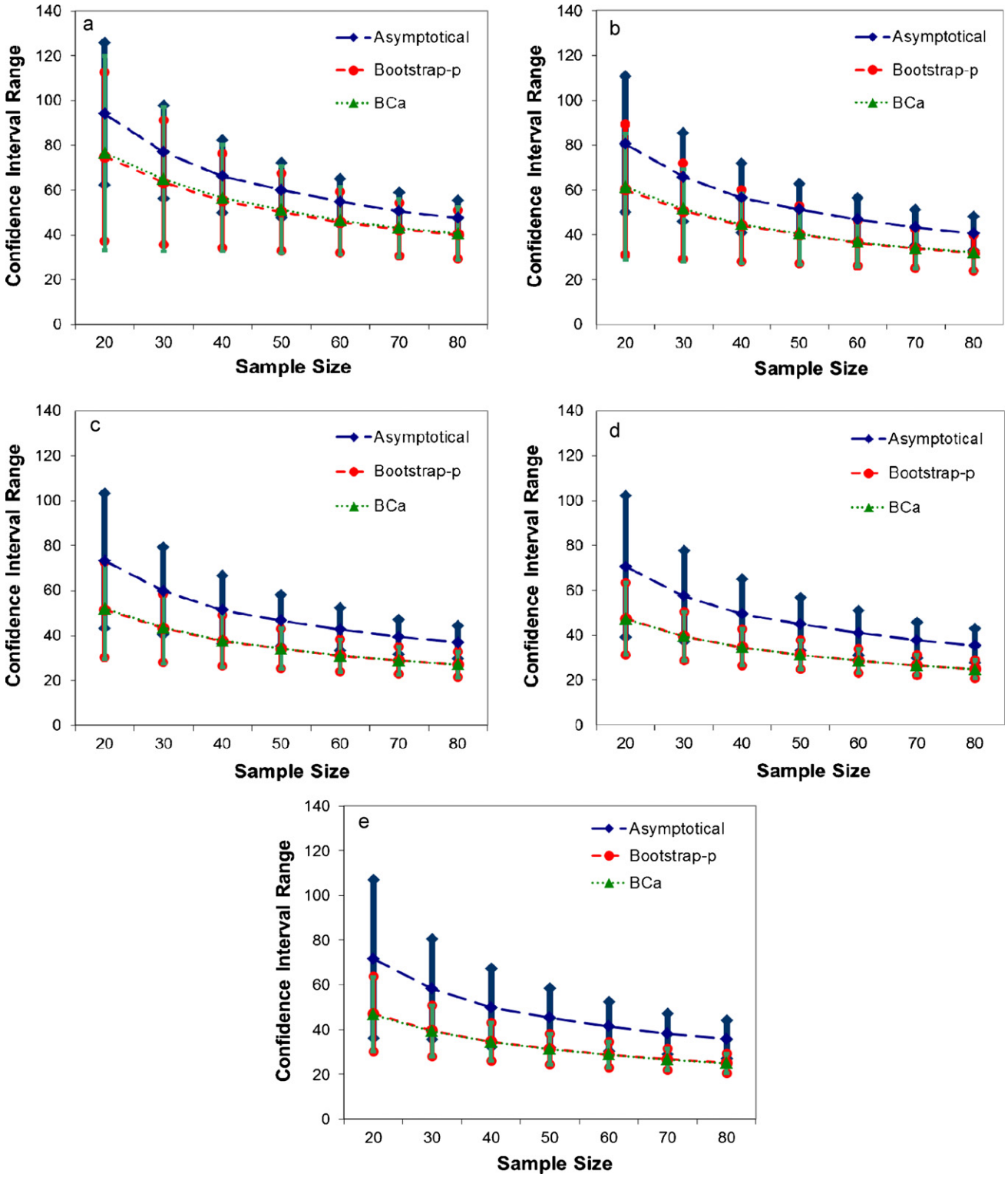


Fig. 2 – Confidence interval range for uncensored injury data and right-censored non-injury data: (a) 10% injury risk; (b) 20% injury risk; (c) 30% injury risk; (d) 40% injury risk; (e) 50% injury risk.

CI range from 1000 trials, accompanied by its 95% CI. The CI ranges for uncensored data (Fig. 1) calculated from the normal asymptotical method were found significantly higher than the ranges calculated from the bootstrap-*p* and BCa methods (p-values < 0.01). However, this difference became smaller when the risk probability increased from 10% to 50%. No significant CI range differences were found between the bootstrap-*p* and BCa methods for uncensored datasets

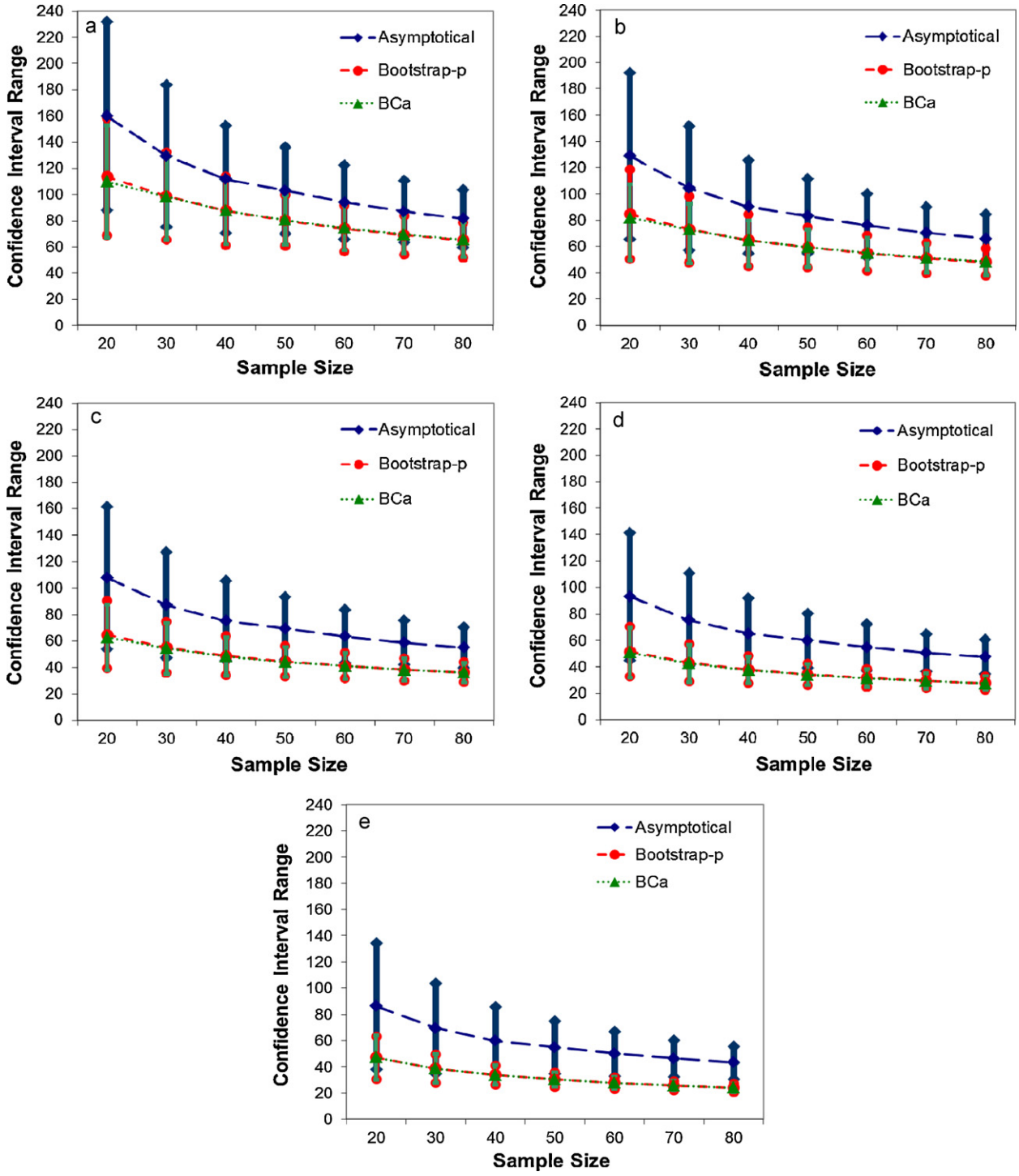


Fig. 3 – Confidence interval range for interval-censored injury data and right-censored non-injury data: (a) 10% injury risk; (b) 20% injury risk; (c) 30% injury risk; (d) 40% injury risk; (e) 50% injury risk.

(p-values > 0.1). Furthermore, regardless of what CI methods were used, higher CI range was found at lower risk probability, and when the sample size increased the CI range decreased.

In the cases where the injury data were uncensored and the non-injury data were right-censored (Fig. 2) the CI ranges calculated from the normal asymptotical method was again found significantly higher than the ranges calculated from the bootstrap-p and BCa methods (p-values < 0.01), while no differences of CI ranges between the bootstrap-p and BCa methods were observed (p-values > 0.01). The difference of the CI ranges between the normal asymptotical method and the bootstrapping methods became larger when the risk probability increased from 10% to 50%. Smaller CI range was found

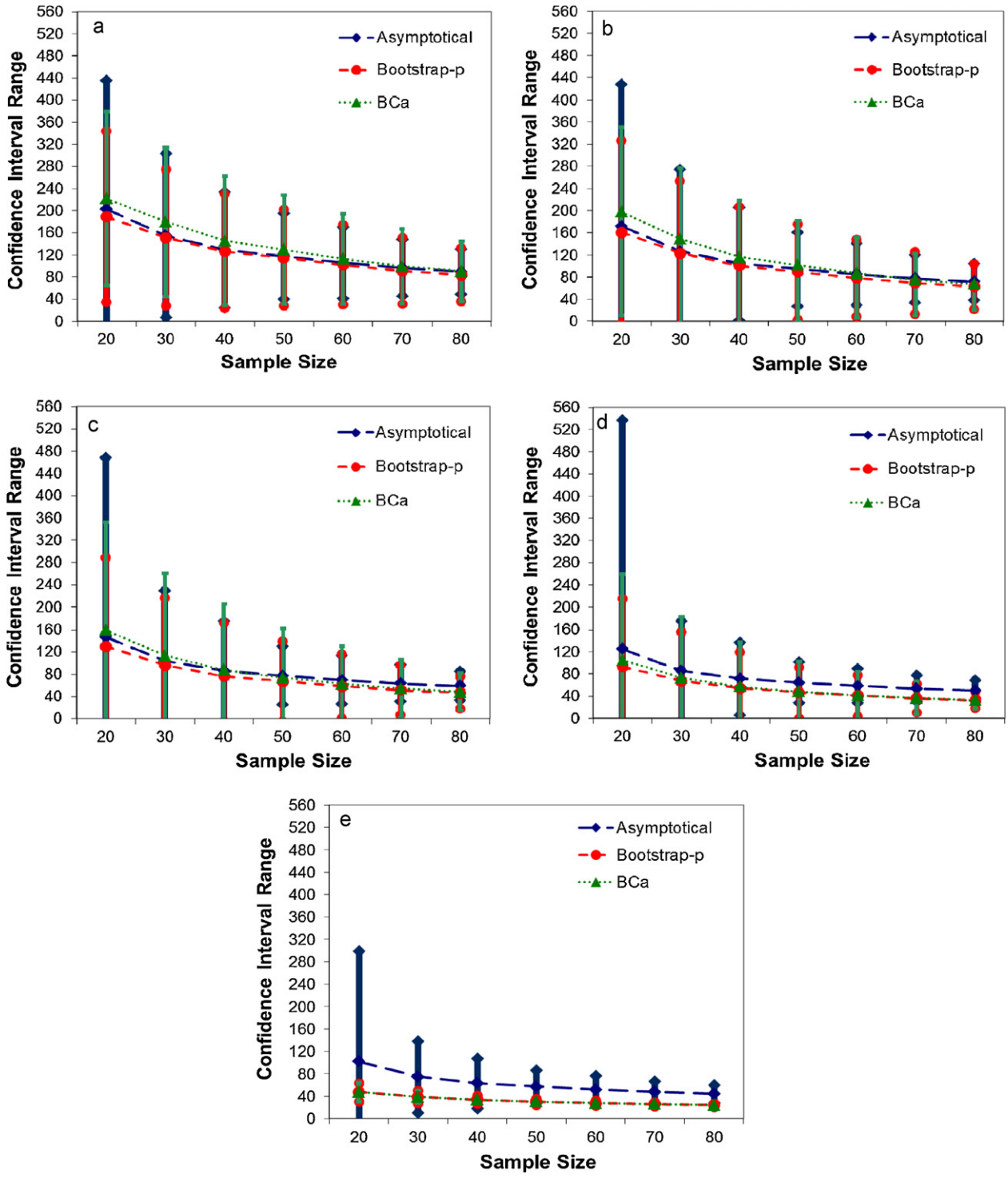


Fig. 4 – Confidence interval range for left-censored injury data and right-censored non-injury data: (a) 10% injury risk; (b) 20% injury risk; (c) 30% injury risk; (d) 40% injury risk; (e) 50% injury risk.

again when the sample size increased, and higher CI range was found at lower risk probability. When the injury data were interval-censored and non-injury data were right-censored, the effect of the bootstrap methods became more significant in terms of the CI range (p-values < 0.001) (Fig. 3). The CI range reduced approximately one third when comparing the asymptotical method to the bootstrap-p (or BCa) method. Moreover, the averaged CI ranges at risk probability of 10% was approximately twice of the averaged CI ranges at 50% risk, for both asymptotical and bootstrap CI ranges. No CI range difference was found between bootstrap-p and BCa methods (p-values > 0.01).

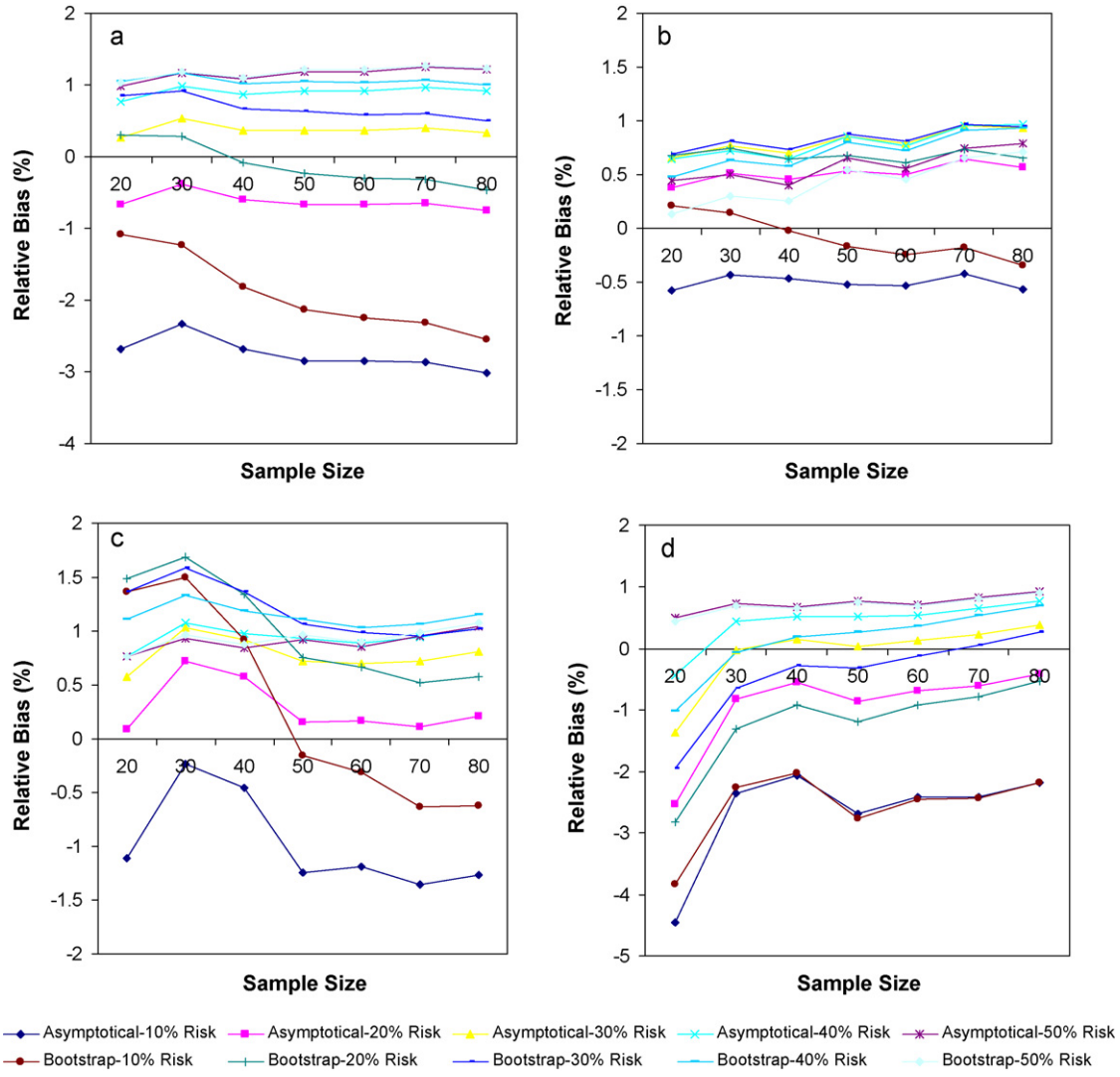


Fig. 5 – Relative bias: (a) uncensored data; (b) injury: exact; non-injury: right-censored; (c) injury: interval-censored; non-injury: right-censored; (d) injury: left-censored; non-injury: right-censored.

In the last censoring scenario when the injury data were left-censored and non-injury data were right-censored (Fig. 4) the CI ranges calculated by the bootstrap-*p* method were significantly less than the asymptotical CI ranges at 10% risk when $N \geq 70$, at 20% risk when $N \geq 50$, at 30% risk when $N \geq 40$ and at 40% and 50% risks for all sample sizes (*p*-values < 0.01). In contrast, the BCa CI ranges were higher than the asymptotical CI ranges at the risk probability of 10% and 20% when $N \leq 50$ (*p*-values < 0.005) but smaller than the asymptotical CI ranges at 20% risk when $N = 80$, at 30% risk when $N \geq 60$, at 40% risk when $N \geq 30$ and at 50% risk for all sample sizes (*p*-values < 0.001). No difference between bootstrap-*p* and BCa CI ranges was found at 40% and 50% risk (*p*-values > 0.01). Furthermore, higher CI range was found at the lower risk probability, and when the sample size increased, the CI range decreased. When comparing the interval-censored and left-censored injury data, higher CI ranges were observed for the left-censored injury data in all cases (Figs. 3 and 4).

The RB was computed using the difference between the $\hat{t}_i$ (or the average of $\hat{t}_i^*$) and the true t_i divided by the true t_i. For uncensored data, estimates were underestimated at 10% risk ($RB \leq -1\%$) and at 20% risk ($-1\% \leq RB \leq 0\%$), but overestimated from 30% to 50% risks ($0\% \leq RB \leq 1.5\%$) (Fig. 5a). Furthermore, bootstrap estimates tended to be higher than asymptotical estimates in lower injury probability region; however, the difference was insubstantial (less than 1.5%). When non-injury data were right-censored, point estimates were underestimated at 10% risk and overestimated from 20% to 50% risks; nevertheless, the absolute value of these relative biases were relatively small (less than 1%) (Fig. 5b). When injury data were interval-censored and non-injury data were right-censored, the RB tended to be smaller at lower risk probability again, but the absolute values of these RBs were less than 1.5% (Fig. 5c).

In the left-censored injury data and right-censored non-injury data, the RB for estimates at 10% risk was between −5% and −2%, and the RBs from 20% to 50% risks were higher

Table 1 – Coverage probability—uncensored data.

Method	Risk	Sample size						
		20	30	40	50	60	70	80
Asymptotical	10%	0.973	0.983	0.977	0.972	0.963	0.954	0.949
	20%	0.978	0.976	0.986	0.991	0.988	0.987	0.991
	30%	0.981	0.972	0.989	0.987	0.988	0.982	0.987
	40%	0.981	0.970	0.988	0.983	0.985	0.981	0.982
	50%	0.981	0.965	0.978	0.980	0.971	0.969	0.974
Bootstrap-*p*	10%	0.959	0.963	0.964	0.963	0.958	0.938	0.934
	20%	0.962	0.955	0.970	0.972	0.978	0.969	0.969
	30%	0.965	0.950	0.974	0.974	0.975	0.970	0.975
	40%	0.970	0.959	0.968	0.969	0.964	0.963	0.973
	50%	0.978	0.965	0.971	0.969	0.968	0.955	0.963
BCa	10%	0.918	0.931	0.929	0.916	0.908	0.884	0.880
	20%	0.959	0.976	0.968	0.964	0.962	0.960	0.962
	30%	0.970	0.965	0.976	0.978	0.982	0.974	0.977
	40%	0.970	0.963	0.976	0.978	0.972	0.971	0.978
	50%	0.976	0.957	0.969	0.970	0.968	0.954	0.962

Table 2 – Coverage probability—injury: exact; non-injury: right censored.

Method	Risk	Sample size						
		20	30	40	50	60	70	80
Asymptotical	10%	0.957	0.965	0.967	0.985	0.976	0.980	0.975
	20%	0.954	0.963	0.962	0.983	0.978	0.977	0.972
	30%	0.957	0.970	0.972	0.985	0.983	0.975	0.978
	40%	0.962	0.974	0.979	0.986	0.988	0.984	0.986
	50%	0.967	0.965	0.985	0.984	0.986	0.990	0.990
Bootstrap-*p*	10%	0.894	0.922	0.915	0.944	0.931	0.938	0.932
	20%	0.891	0.898	0.907	0.934	0.925	0.924	0.929
	30%	0.899	0.903	0.916	0.922	0.924	0.910	0.914
	40%	0.905	0.898	0.920	0.916	0.918	0.908	0.915
	50%	0.913	0.892	0.922	0.930	0.939	0.926	0.928
BCa	10%	0.889	0.922	0.916	0.942	0.920	0.938	0.932
	20%	0.894	0.920	0.922	0.947	0.929	0.933	0.934
	30%	0.905	0.907	0.925	0.934	0.923	0.921	0.924
	40%	0.899	0.894	0.923	0.920	0.915	0.911	0.914
	50%	0.913	0.879	0.917	0.924	0.932	0.916	0.917

than −1.5% for $N \geq 30$ (Fig. 5d). In general, no differences were found between asymptotical and bootstrap estimates in terms of the RB, except the case at the 10% risk for uncensored data (p-values < 0.01) (Fig. 5a). Furthermore, at the 10% risk for uncensored data the absolute value of the RB for the bootstrap significantly increased as N increased (p-value = 0.0012 when comparing $N = 20$ and 80).

The CP was determined for asymptotical and bootstrap methods. In general, the asymptotical method provided approximate 1–8% higher CP than the bootstrap-p and BCa methods (Table 1). The results show that for uncensored datasets, the CPs were from 0.93 to 0.98 for all bootstrap-p and BCa cases except the CP at 10% risk for BCa method (0.88–0.93). Bootstrap-p worked better than BCa method at 10% risk for uncensored data. When the injury data were uncensored and the non-injury data were right-censored, the CP decreased to 0.89–0.94 for both bootstrap-t and BCa methods (Table 2). Higher BCa CPs were obtained at 30% and 40% risk for uncensored datasets (Table 1) and at 20% and 30% risk for right-censored non-injury datasets (Table 2) compared to bootstrap-p method.

When the injury data were interval-censored or left-censored and non-injury data were right-censored, the CPs were less than 0.96 for the bootstrap-p and BCa methods (Tables 3 and 4), and BCa method performed better at risks of 20%, 30%, 40%, and 50% for $N \geq 30$ in terms of the CP in most cases; however, this improvement is trivial. Overall, when N increased, the CP slightly increased.

5. Applications

To demonstrate the application of the methodology presented in this study, two injury biomechanical datasets reported in [5,27] were used.

The first dataset [5] used the results recorded in dynamic latero-medial bending tests on human cadaveric lower extremities. The scaled fracture moment for the bare femur

Table 3 – Coverage probability—injury: interval-censored; non-injury: right-censored.

Method	Risk	Sample size						
		20	30	40	50	60	70	80
Asymptotical	10%	0.967	0.955	0.950	0.972	0.963	0.978	0.970
	20%	0.970	0.946	0.961	0.975	0.965	0.983	0.970
	30%	0.981	0.957	0.965	0.980	0.969	0.984	0.974
	40%	0.992	0.972	0.979	0.985	0.977	0.991	0.984
	50%	0.992	0.976	0.989	0.995	0.992	0.998	0.995
Bootstrap-*p*	10%	0.927	0.918	0.900	0.929	0.927	0.940	0.936
	20%	0.929	0.907	0.891	0.923	0.920	0.933	0.935
	30%	0.929	0.885	0.894	0.905	0.919	0.922	0.926
	40%	0.938	0.872	0.895	0.907	0.910	0.914	0.912
	50%	0.918	0.866	0.915	0.908	0.912	0.917	0.911
BCa	10%	0.913	0.929	0.918	0.923	0.922	0.930	0.931
	20%	0.908	0.911	0.915	0.936	0.931	0.939	0.941
	30%	0.908	0.909	0.901	0.917	0.929	0.933	0.941
	40%	0.921	0.883	0.905	0.911	0.920	0.919	0.922
	50%	0.935	0.881	0.929	0.918	0.923	0.925	0.919

Table 4 – Coverage probability—injury: left-censored; non-injury: right-censored.

Method	Risk	Sample size						
		20	30	40	50	60	70	80
Asymptotical	10%	0.962	0.955	0.954	0.974	0.968	0.984	0.973
	20%	0.970	0.946	0.961	0.976	0.967	0.983	0.970
	30%	0.981	0.957	0.965	0.980	0.970	0.984	0.974
	40%	0.992	0.972	0.979	0.985	0.977	0.992	0.984
	50%	0.992	0.976	0.989	0.995	0.992	0.998	0.995
Bootstrap-*p*	10%	0.927	0.920	0.899	0.928	0.927	0.944	0.935
	20%	0.935	0.920	0.891	0.928	0.925	0.935	0.934
	30%	0.935	0.905	0.898	0.919	0.926	0.929	0.932
	40%	0.959	0.879	0.906	0.916	0.925	0.932	0.930
	50%	0.927	0.866	0.922	0.918	0.917	0.924	0.919
BCa	10%	0.913	0.931	0.920	0.924	0.921	0.930	0.927
	20%	0.910	0.922	0.924	0.934	0.934	0.940	0.944
	30%	0.913	0.916	0.906	0.928	0.930	0.939	0.946
	40%	0.940	0.890	0.918	0.924	0.936	0.926	0.942
	50%	0.932	0.881	0.931	0.920	0.924	0.927	0.926

and thigh specimen tests was used as failure time in the analysis. All 22 observations are uncensored data. Bootstrapping with $B = 2000$ resamples from the original data was conducted. The asymptotical and bootstrap CIs for the parameters t_{10}, t_{20}, t_{30}, t_{40}, and t_{50} were calculated (Fig. 6). As expected, in this uncensored case, bootstrap CIs had smaller CI ranges than asymptotical CIs. Also, larger CI ranges were observed at t with lower risk probability, and the range difference between asymptotical and bootstrap CIs became smaller for t with larger injury probability. No significant differences were found between the bootstrap-p and BCa CI ranges at five injury risks (p-values > 0.1).

The second dataset was related to the biomechanical responses of the pediatric abdomen [27]. In the study, transverse, dynamic belt loading was conducted on the abdomen of 47 immediately post-mortem juvenile swine. The authors concluded that the best discriminator was the peak belt tension (N) due to its largest Goodman–Kruskal Gamma values. Therefore, the peak belt tension and the maximum

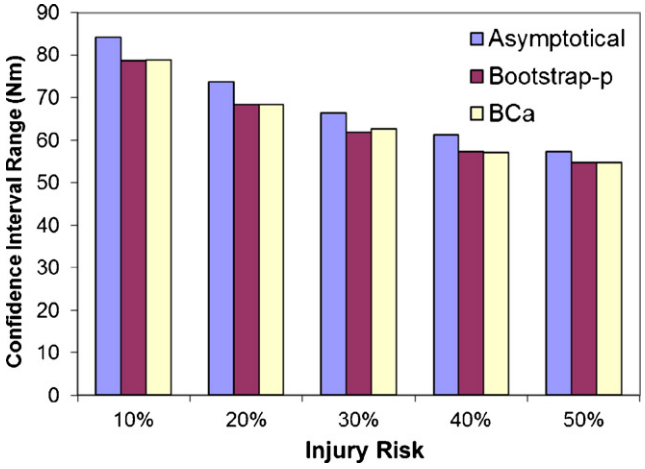


Fig. 6 – Confidence interval ranges for the scaled fracture moment data.

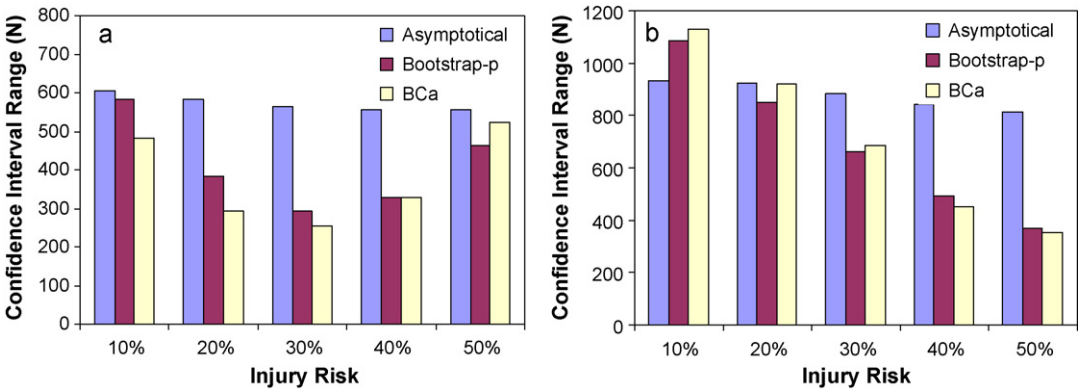


Fig. 7 – Confidence interval ranges for the abdominal peak belt tension data when assuming injury observations were: (a) interval-censored with 1100N as the lower bound; (b) left-censored.

Abbreviated Injury Scale (MAIS) 3+ were treated as injury predictor and injury observation, respectively. From this test series, 15 were non-injury data and the rest of 18 were injury data. Non-injury data were treated as right-censored and injury data were first considered as left-censored and then as interval-censored with lower bound at 1100 (*N*). In both cases, the asymptotical and bootstrap CI ranges for the parameters t_{10}, t_{20}, t_{30}, t_{40}, and t_{50} were calculated (Fig. 7). In the interval-censoring schemes, CI ranges obtained from bootstrapping were smaller than the asymptotical CIs ranges. Smallest bootstrapping CI ranges were found at 30% risk because this dataset tends to be right skewed. On the other hand, in the left-censoring schemes larger bootstrap CI ranges were observed for t_{10}, and BCa produced higher CI ranges than bootstrap-*p* at t_{10}, t_{20}, and t_{30}, as expected. Generally, larger ranges for the obtained CIs considering left-censored observations than considering interval-censored observations were obtained using both the asymptotical or bootstrap methods for small risks ($\leq$40%).

6. Discussion

The effectiveness of the bootstrapping method was evaluated in this study under different data censoring conditions and various sample sizes. Two bootstrapping methods to construct CIs were applied. The results show that the bootstrap could be quite useful in terms of the CI range. Better inference results were observed when comparing bootstrap-based methods to the asymptotical inference method based on the asymptotical normality of the MLEs. As discussed in [14], the difference between the asymptotic CI and the bootstrap CI may come from the fact that the variance obtained by the delta method depends on Taylor series approximation where the error terms are ignored; however, these terms are part of the variances computed by bootstrapping.

When comparing various censored type datasets for the asymptotical method, higher CI ranges were found for the censoring scheme when injury data were left censored and non-injury data were right-censored. Intuitively, uncensored data can improve the prediction of estimators due to gaining more information on the underlying distribution in comparison to right-, left-, or interval-censored data. In contrast, for the bootstrapping methods, the CI range was lower for censored datasets compared to uncensored datasets for higher injury risk (30–50%) except in the cases of 30% and 40% for left-censored injury data and right-censored non-injury data. This phenomenon is due to the resampling procedure on the same simulated dataset for each bootstrap calculation and the region of the sampled points from the underlying normal distribution. More specifically, in the censored datasets, a smaller range of observed data points (250–450) was used to represent the censored observations; however, the exact data points may spread on the wide range of the normal distribution. Therefore, more observations were obtained around the mean (350) in the censored datasets, leading to smaller CI ranges at 30–50% risk for censored datasets.

Furthermore, rarely do we see the comparison between left-censoring and interval-censoring for injury data. The numerical analysis in the current study shows that if injury data are considered to be censored and the lower limit can be pre-determined, interval-censored for injury data is recommended because the CI range significantly decreases when comparing to left-censored injury data, even though the CP remains very similar. One interesting observation from the study is the dramatic increase of the BCa CI range for left-censored injury data when the sample size decreases. This may be caused by the point underestimation for lower injury probabilities (10–30%) for left-censored injury data based on the information of RB and the a^* for constructing the BCa CI decreases as sample size decreases, leading the BCa CI to become wider. Therefore, BCa method is not suggested for constructing CI when injury data are left-censored and sample size is small, especially for predicting risk injury estimates at lower probabilities.

As expected, smaller CI ranges were observed [15] based on the bootstrap-*p* method in comparison with the CI ranges obtained from asymptotical method. In addition, the CI range decreases when the size of the biomechanical dataset increases [6,23]. However, larger CI range may be found for other bootstrap methods (e.g. bootstrap-*t* method), applied to different distribution models (e.g. proportional hazards model [10] and log-normal distribution [23]) with various censoring types. In those cases, criterions such as CP of the CIs and

normal quantile–quantile plots for target parameters should be considered to verify whether the normality assumption is appropriate.

In defining the injury threshold values for human protection during traffic crashes or blast, the lower-level injury probabilities (probability ≤ 50%) are usually considered in the design process. The current study shows that CI estimated at lower probability (10%) had higher CI range when comparing to CI estimated at higher probability (50%) for uncensored data. In the current sampling process, the "actual injury values" were selected from the underlying assumed normal distribution, and the "observed values" were drawn from a uniform distribution centered at the mean value (350) of the underlying normal distribution. This led to the fact that the range of the "observed data" was limited between 7.5% and 92.5% of risk and centered on the mean of the normal distribution. Therefore, more data points would be "created" around the mean value (350). This produced more accurate predictions around the mean value, which is corresponding to an injury probability of 50% for the underlying distribution. Therefore, within each censoring scenario, the CI range decreased as the injury risk increased for the asymptotical and investigated bootstrap methods. Thus, in future studies to better approximate the lower level injury domain, more samples are recommended in this lower risk region.

For an uncensored data set, the differences between the asymptotic and bootstrap CI ranges became smaller when injury probability increased from 10% to 50%. In contrast, for a dataset with censored points, the differences between the CI ranges became larger when injury probability increased from 10% to 50%, especially for the case of left-censored injury data. In the four censored types investigated in the analysis, bootstrapping provides better CI predictions for estimates at lower-level injury probabilities.

The choice of the lower bound of the tolerance limit for interval-censored data in this study depended on the approximate lower bound of 99.9% coverage of the assumed normal distribution. While a standard criterion for the lower bound of interval-censored data has not been established, the lower limit used in this study was chosen close to the lowest exact injury data point, as reported in other studies [28]. A sensitivity testing was further conducted to evaluate different lower bounds in this study. Two lower bounds, 150 and 170, corresponding to 99.5% and 99% coverage of the assumed normal distribution, were set for the lower limits of the interval-censored injury data, and the CI ranges were found to be within 0.5% change compared to lower limit of the 99.9% coverage in the simulations of interval-censored injury data and right-censored non-injury data. This demonstrates that the effect of the choice of a lower bound of the tolerance limit within a 1% range for interval-censored data would have no substantial influence on the CI range.

The results from the CP (Tables 1–4) show that the asymptotical method provided higher CP than the bootstrap methods. The results in Tables 1–3 were expected because the 95% CIs of the CI ranges in Figs. 1–3 for the asymptotical method were up to even 2 times higher than the CI ranges for the bootstrap methods. Accordingly, a trade-off between the choice of methods with acceptable CI range and CP should be considered. In the current study, while losing an approximate 1–8% CP, up to a one-half reduction of the CI range could be obtained, especially in the interval-censored injury datasets. Nevertheless, a special case was found for the left-censored injury data where the BCa method provided smaller CP than the asymptotical method (Table 4) while it produced higher averaged CIs than the asymptotical method (Fig. 4). In this case, the bootstrapping tended to slightly underestimate the CIs due to the property of the left-censored points, leading to the decrease of the CPs, even though the 95% CIs of the CI ranges at 10–40% risks were similar between the asymptotical and bootstrap methods (Fig. 4) except when the sample size was 20. This suggests that the bootstrapping may not well estimate CIs compared to asymptotical method when $N \geq 30$ at lower risks (10–40%) for the left-censored injury datasets.

The analysis of CP also demonstrates the discrepancies between the bootstrap-*p* and BCa methods. In most cases, BCa method gave higher CPs for injury risk greater than 20% while bootstrap-*p* method provided higher CP for injury risk at 10%. However, the differences of the CPs between bootstrap-*p* and BCa methods were small (<1.5%).

At the 10% risk for uncensored data, the RB for the bootstrap significantly increased as the sample size increased (p-value = 0.0012) between sample size 20 and 80 (Fig. 5a). It could be concluded that for the bootstrap approach the estimation is deteriorating (i.e. bias increases) as the sample size increases for uncensored datasets at lower risk (10%). This suggests that when dealing with lower risk (10%) where few injury points are collected in this region, bootstrap may not be suitable for determining the point estimate of t for uncensored datasets.

In the four censored types investigated in the current study, the simulated points contained half injury data and half non-injury data. While in real biomechanical experimental data the number of censored points may not be known before tests, this study focused on the comparison between the exact, interval- and left-censored injury data. Further investigations on different number of injury and non-injury observations for each censoring type are suggested, as discussed in [6].

The underlying injury data was assumed to be normally distributed in the current study; however, the actual underlying distribution remains unknown. In addition, distributions with skewness were observed in some biomechanical data. Further investigation on the influence of bootstrapping applied to various underlying distributions would be suggested. On the other hand, several distributions which can accommodate skewed data such as log-normal and log-logistic distributions in survival analysis were extensively attempted to fit the biomedical data [12,15,23]. Therefore, the application of the bootstrap with different parametric distributions in constructing the injury risk curves is suggested to be investigated in future studies.

7. Conclusions

This study investigated the applicability of bootstrapping methods in improving the estimation of the CIs of lower-level survival times. It was observed that bootstrapping performs better than asymptotical method especially in terms of the

CI range. When sample size is small (≤40), the injury observations are recommended to be used as interval-censored rather than left-censored, if their lower bound can be found. As expected, it was observed that the larger number of observations and uncensored data improve significantly the estimation of the CI.

Conflict of interest statement

All authors declare that they have no potential conflicts of interest associated with any financial, personal or other relationships with other people or organizations which could inappropriately influence or bias what is reported in the manuscript.

REFERENCES

[1] C.D. Untaroiu, J. Shin, J.R. Crandall, A design optimization approach of vehicle hood for pedestrian protection, Int. J. Carshworthiness 12 (6) (2007) 581–589.

[2] D. Bose, J.R. Crandall, C.D. Untaroiu, E.H. Maslen, Influence of pre-collision occupant state parameters on injury outcome in a frontal collision, Accid. Anal. Prev. 42 (4) (2010) 1398–1407.

[3] R. Eppinger, E. Sun, F. Bandak, M. Haffner, N. Khaewpong, M. Maltese, S. Kuppa, T. Nguyen, E. Takhounts, R. Tannous, A. Zhang, R. Saul, Development of Improved Injury Criteria for the Assessment of Advanced Automotive Restraint Systems II, National Highway Traffic Safety Administration, Department of Transportation, Washington, DC, 1999.

[4] R. Eppinger, E. Sun, S. Kuppa, R. Saul, Supplement: Development of Improved Injury Criteria for the Assessment of Advanced Automotive Restraint Systems II, National Highway Traffic Safety Administration, Department of Transportation, Washington, DC, 2000.

[5] J.R. Kerrigan, D.C. Drinkwater, C.Y. Kam, D.B. Murphy, B.J. Ivarsson, J.R. Crandall, J. Patrie, Tolerance of the human leg and thigh in dynamic latero-medial bending, Int. J. Carshworthiness 9 (6) (2004) 607–623.

[6] L. Di Domenico, G. Nusholtz, Risk curve boundaries, Traffic Injury Prevention 6 (1) (2005) 86–94.

[7] B. Efron, Censored data and the bootstrap, J. Am. Stat. Assoc. 76 (374) (1981) 312–319.

[8] B. Efron, R. Tibshirani, Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, Stat. Sci. 1 (1) (1986) 54–75.

[9] T. DiCiccio, R. Tibshirani, Bootstrap confidence intervals and bootstrap approximations, J. Am. Stat. Assoc. 82 (397) (1987) 163–170.

[10] J. Carpenter, J. Bithell, Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians, Stat. Med. 19 (9) (2000) 1141–1164.

[11] S.-L Jeng, W.Q. Meeker, Parametric simultaneous confidence bands for cumulative distributions from censored data, Technometrics 43 (4) (2001) 450–461.

[12] R.W. Kent, J.R. Funk, Data censoring and parametric distribution assignment in the development of injury risk functions from biomechanical data, in: SAE 2004 World Congress & Exhibition, 2004, 2004-01-0317.

[13] E.A. Kennedy, W.J. Hurst, J.D. Stitzel, J.M. Cormier, G.A. Hansen, E.P. Smith, S.M. Duma, Lateral and posterior dynamic bending of the mid-shaft femur: fracture risk curves for the adult population, Stapp Car Crash J. 48 (2004) 27–51.

[14] R.C. Gupta, O. Akman, S. Lvin, A study of log-logistic model in survival analysis, Biometr. J. 41 (4) (1999) 431–443.

[15] J. Mazucheli, E.A. Barros, J.A. Achcar, Bootstrap confidence intervals for the mode of the hazard function, Comput. Methods Programs Biomed. 79 (1) (2005) 39–47.

[16] L. Di Domenico, G. Nusholtz, Comparison of parametric and non-parametric methods for determining injury risk, in: SAE 2003 World Congress & Exhibition, 2003, 2003-01-1362.

[17] A. Petitjean, X. Trosseille, P. Petit, A. Irwin, J. Hassan, N. Praxl, Injury risk curves for the WorldSID 50th male dummy, Stapp Car Crash J. 53 (2009) 443–476.

[18] W. Weibull, A statistical theory of the strength of material, Proc. R. Swedish Inst. Eng. Res. 151 (1) (1939).

[19] J.P. Klein, M.L. Moeschberger, Survival Analysis: Techniques for Censored and Truncated Data, 2nd ed., Springer, 2003.

[20] G. Casella, R.L. Berger, Statistical Inference, 2nd ed., Duxbury, 2002.

[21] M.J. Crawley, The R Book, 1st ed., Wiley, 2007.

[22] B. Efron, R.J. Tibshirani, An Introduction to the Bootstrap, 1st ed., Chapman and Hall/CRC, 1994.

[23] H. Jiang, X.-H. Zhou, Bootstrap confidence intervals for medical costs with censored observations, Stat. Med. 23 (21) (2004) 3365–3376.

[24] N. Reid, Estimating the median survival time, Biometrika 68 (3) (1981) 601–608.

[25] R. Brookmeyer, J. Crowley, A confidence interval for the median survival time, Biometrics 38 (1) (1982) 29–41.

[26] J. Landon, N.D. Singpurwalla, Choosing a coverage probability for prediction intervals, Am. Stat. 62 (2) (2008) 120–124.

[27] R. Kent, S. Stacey, M. Kindig, W. Woods, J. Evans, S.W. Rouhana, K. Higuchi, H. Tanji, St.S. Lawrence, K.B. Arbogast, Biomechanical response of the pediatric abdomen. Part 2. Injuries and their correlation with engineering parameters, Stapp Car Crash J. 52 (2008) 135–166.

[28] S.W. Rouhana, R.E. El-Jawahri, T.R. Laituri, Biomechanical considerations for abdominal loading by seat belt pretensioners, Stapp Car Crash J. 54 (2010) 381–400.