Chemosphere 80 (2010) 7–12

Contents lists available at ScienceDirect

Chemosphere

journal homepage: www.elsevier.com/locate/chemosphere

# A comparison of techniques for assessing central tendency in left-censored data using PCB and p,p′DDE contaminant concentrations from Michigan's Bald Eagle Biosentinel Program

Katherine F. Leith [a,*], William W. Bowerman [a], Michael R. Wierda [a], Dave A. Best [b], Teryl G. Grubb [c], James G. Sikarske [d]

[a] *Department of Forestry and Natural Resources, Clemson University, Clemson, SC 29634, USA*
[b] *United States Fish and Wildlife Service, East Lansing Field Office, East Lansing, MI 48823, USA*
[c] *United States Forest Service, Rocky Mountain Research Station, Flagstaff, AZ 86001, USA*
[d] *College of Veterinary Medicine, Small Animal Clinical Sciences, Michigan State University, East Lansing, MI 48825, USA*

ARTICLE INFO

*Article history:*
Received 23 March 2010
Accepted 30 March 2010
Available online 22 April 2010

*Keywords:*
Kaplan–Meier
Substitution
Outliers
Multiple imputation
Non-detect

## ABSTRACT

Monitoring of contaminants in the environment is an important part of understanding the fate of ecosystems after a chemical insult. Frequently, such monitoring efforts result in datasets with observations below the detection limit (DL) that are reported as 'non-detect' or '<DL' and no value is provided. This study explored the effects of non-detect data and their treatment on summary statistics. The data analyzed in this paper are real-world data. They consist of both large ($N = 234$) and moderate ($n = 12$–$64$) sample sizes with both good and marginal fit to an assumed distribution with a log-transformation. Summary statistics were calculated using (1) the '0.0001' near-zero method of substitution, (2) substitution with '1/2 $*$ DL', (3) multiple imputation, and (4) Kaplan–Meier estimation. Median was used for comparison. Several analytical options for datasets with non-detect observations are available. The general consensus is that substitution methods ((1) and (2)) can produce biased summary statistics, especially as levels of substitution increase. Substitution methods continue to be used in research, likely because they are easy to implement. The objectives were to (1) assess the fit of lognormal distribution to the data, (2) compare and contrast the performance of four analytical treatments of left-censored data in terms of estimated geometric mean and standard error, and (3) make recommendations based on those results.

## 1. Introduction

Monitoring contaminant levels in the environment is an important part of understanding the fate of ecosystems after a chemical insult. As levels decrease over time, limitations in analytical equipment create a lower bound below which contaminant levels cannot be accurately reported. This results in datasets with observations below the detection limit (DL) that are reported only as 'non-detect' or '<DL' and no value is provided. This type of distribution is called 'left-censored', as the low-end observations that are unknown generally occur near the origin of the *x*-axis in figures. Several options for the analysis of these datasets have been investigated leading to the conclusion that methods replacing all non-detects with a single value (substitution methods) are frequently inferior (Liu et al., 1997; Singh and Nocerino, 2002; Baccarelli et al., 2005; Helsel, 2005a,b, 2006; Eastoe et al., 2006; Needham et al., 2007; Antweiler and Taylor, 2008). Specifically, it has been shown that the bias caused by substitution increases dramatically as the percent of observations censored increases (Eastoe et al., 2006). In spite of this, various substitution methods continue to be used in research, frequently with little regard for the proportion of observations censored.

In addition to censored observations, another complication when analyzing environmental contaminant data is the possibility that datasets display a right skew. This occurs when a few samples show very high concentrations while the general tendency is for concentrations to be lower. This distribution is common in environmental data and can frequently be accommodated by log-transformation. The median has traditionally been an accepted measure of central tendency for data which do not fit a normal distribution well, and this approach has been used in almost every field of scientific inquiry. Focus has shifted to newer approaches as more complex methods have been developed and computing power has grown to make them feasible for the average researcher. While the median is still useful in that it is not based on indefensible

---

* Corresponding author. Address: Department of Forestry and Natural Resources, 261 Lehotsky Hall, Clemson University, Clemson, SC 29634, USA. Tel.: +1 864 656 5332 (O), +1 864 903 2728 (C); fax: +1 864 656 3304.
*E-mail addresses:* kleith@clemson.edu, leith.kate@gmail.com (K.F. Leith).

0045-6535/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.chemosphere.2010.03.056

assumptions about the shape of the distribution, it does not make use of all the information contained in a dataset. A final complexity is added by the fact that lognormal data are frequently summarized using the geometric mean, which is particularly sensitive to the choice of substitution value.

In 1997, the Michigan Department of Environmental Quality (MDEQ) implemented a Bald Eagle Biosentinel Program (BEBP) to monitor trends of a suite of organic pollutants under the Clean Michigan Initiative (MDEQ, 1997). The data analyzed here are 234 observations of total polychlorinated biphenol (PCB), which is a sum of all observed PCB congeners, and p,p′-dichlorodiphenyldichloro-ethylene (p,p′DDE) concentrations found in nestling bald eagle plasma samples from throughout the State of Michigan. Current statistical methods include tests of significant differences between regions at several geographic scales, and calculating descriptive statistics in the form of geometric means. Substitution is currently used in cases of non-detect, or left-censored, observations.

This choice of methods for addressing the non-detects does not affect testing of significant differences between regions. The monitoring program uses nonparametric Kruskal–Wallis and Wilcoxon tests which are rank-based and make no assumptions about distribution and so, are not sensitive to the problems of substitution (Helsel, 2005b). However, summary statistics are reported as geometric means, which are affected by the choice of substituted value. The BEBP currently substitutes the near-zero value of ‘0.0001’ for concentrations at non-detectable levels when analyzing contaminants that are summed, such as total PCBs. This near-zero value might appear to have little influence on the resulting calculations to those accustomed to arithmetic mean calculation because the arithmetic mean is a function of addition, for which ‘0’ is the identity value. Geometric means on the other hand are a function of multiplication, for which the identity value is ‘1’, while near-zero values (like ‘0.0001’) have a drastic impact on the product. For contaminants that do not involve summation, a value of one-half of the quantification level is substituted for non-detects.

The desire to summarize data more accurately has fueled recent comparisons of proposed analytical alternatives to substitution or simple median reporting (Helsel, 2005b). One alternative, maximum likelihood estimation (MLE), forces the researcher to assume the shape of the underlying distribution, but is powerful if this assumption is correct. These MLE methods have been explored in a variety of environmental applications (Singh and Nocerino, 2002; Helsel, 2005b, 2006; Antweiler and Taylor, 2008; Jain et al., 2008). Kaplan–Meier (KM) estimators have also been proposed. This estimator began as a nonparametric method of estimating the central tendency in right-censored survival data, but is gaining popularity for left-censored datasets. Among those who have explored the application of KM calculations in left-censored environmental data are Antweiler and Taylor (2008), Eastoe et al. (2006) and Helsel (2005a,b). Multiple imputation (MI) has been proposed as a ‘fill-in’ technique that can be used to first estimate an appropriate distribution shape based on uncensored values, then samples from the values that would be found in the censored tail. This technique has been addressed in comparison studies by Antweiler and Taylor (2008), Baccarelli et al. (2005), Eastoe et al. (2006), Helsel (2005a,b), Krishnamoorthy et al. (2009) and Singh and Nocerino (2002). Right-skewed, left-censored data were the focus of Singh and Nocerino (2002), who applied many analysis techniques common for left-censored data and assessed their performance when the observations also displayed a right skew. They found that left-censored datasets were more difficult to accurately summarize in the presence of a right skew.

Few case studies have been published and substitution is still in wide use (Baccarelli et al., 2005; Eastoe et al., 2006). This study explored the effects of non-detect data and their treatment on summary statistics. The data analyzed in this paper represent both large ($N = 234$) and moderate ($n = 12–64$) sample sizes with both good and marginal fit with a log-transformation. Summary statistics were calculated using the current method of substitution with ‘0.0001’, the common method of substitution with ‘$1/2 * \mathrm{DL}$’, MI, and KM estimation. The median was also calculated for comparison with all four methods. The objectives were to (1) assess the fit of lognormal distribution to the data, (2) compare and contrast the performance of the four methods of non-detect handling in terms of estimated geometric mean, comparison to the median, and standard error, and (3) make recommendations based on those results.

## 2. Methods

Concentrations of total PCBs and p,p′DDE (μg/kg wet weight) in plasma collected from nestling bald eagles across Michigan from 1999 to 2003 were used in these analyses. In addition to analysis as a single dataset representing the whole State of Michigan, data were also classified geographically by subpopulation, based on the classifications used in the BEBP. Subpopulations were defined by first subdividing the state spatially into the categories of Great Lakes and Inland breeding areas. Great Lakes breeding areas are defined as being within 8.0 km of Great Lakes shorelines and/or along tributaries open to Great Lakes fish runs and inland breeding areas are defined as being greater than 8.0 km from the Great Lakes shorelines and not along tributaries open to Great Lakes fish runs. These categories are then further subdivided into four Great Lakes and two Inland groups. The Great Lakes subpopulations consisted of Lake Superior (LS), Lake Michigan (LM), Lake Huron (LH), and Lake Erie (LE). The Inland subpopulations consisted of Upper Peninsula (UP), and Lower Peninsula (LP) (Wierda, 2009). The data analysis for this paper was generated using SAS® software, Version 9.1.2 of the SAS system for Windows. Copyright 2000–2004 SAS Institute Inc. SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc., Cary, NC, USA.

### 2.1. Assessment of fit

Data were analyzed for significant departure from the lognormal distribution using the UNIVARIATE PROCEDURE with the options ‘normal’ and ‘plot’ activated (SAS Institute Inc., 2000–2004). This procedure produces several fit statistics. The Kolomogorov–Smirnov statistic was used to assess the assumption of fit to the lognormal distribution. The fit was classified as ‘good’ with $P$-values $\geqslant 0.05$, ‘marginal’ with $P$-values between 0.01 and 0.05, and fit was rejected for $P$-values $<0.01$.

### 2.2. Geometric means and standard error calculation

Geometric means and standard errors were calculated for all of the proposed methods. The median was used for comparison and obtained from the univariate analysis discussed above. For substitution methods, geometric means and standard errors were calculated by log-converting observations, calculating the mean and standard error of the transformed data using the MEANS PROCEDURE (SAS Institute Inc., 2000–2004), and then converting back to the original scale. Monte Carlo simulations of 10 000 iterations were run in order to provide a significance test for the divergence of the geometric mean (using each method of substitution) from the median. Each of these simulations resulted in a ‘$P$-value’ representing the probability of the observed divergence occurring due to sampling error alone. Simulations resulting in $P$-values of 0.05 or less were considered evidence of a significant substitution method effect.

Geometric means and errors were calculated using the multiple imputation methods based on those described in Krishnamoorthy et al. (2009) and the MI PROCEDURE (SAS Institute Inc., 2000–2004). Ten imputations were used on the recommendation of Jain et al. (2008). The option 'EM' was used to implement the maximum likelihood method of adjusting the approximated distribution from which imputed values were drawn. Bounds were set to ensure that no negative values were imputed.

Kaplan–Meier estimates of geometric mean and standard error were calculated using the LIFETEST PROCEDURE (SAS Institute Inc., 2000–2004) on log-transformed data. This procedure is designed to perform survival analysis for right-censored data, so data were transformed to reflect a right-censored distribution. The transformation was conducted by subtracting all log-transformed observations from a number larger than the largest observation. This was done for PCBs by subtracting all observations from 12, and for p,p′DDEs by subtracting all observations from 10. Results were then transformed back to reflect geometric means and standard errors in the original units.

## 3. Results

### 3.1. Fit of data to lognormal distribution

Kolmogorov–Smirnov (KS) tests did not force us to reject the assumption of lognormal distribution in all cases, but did suggest significant non-normality in others. Tests conducted for PCBs and p,p′DDEs at both the whole state and subpopulation level resulted in different conclusions.

At the state level, the PCB distribution was classified as marginal ($P = 0.0206$) and p,p′DDE distribution was classified as significantly differing from lognormal ($P < 0.01$). This was likely due to the presence of six moderate outliers in the upper tail of the p,p′DDE distribution. When the outliers were removed, the KS test resulted in no evidence for significant departure from lognormality ($P = 0.1338$), suggesting a good fit.

When broken down into the geographical units of subpopulation, KS analysis of PCB concentrations suggested that Lake Erie and Michigan coastal regions exhibited marginal evidence for departure from the lognormal distribution. Lake Huron and Superior coastal regions and both Upper and Lower Peninsula inland regions showed no significant departure from the lognormal distribution for PCB concentrations, suggesting a good fit of the data. For p,p′DDE concentrations, the Lake Erie coastal region showed significant departure from normality ($P < 0.01$), but all other regions showed good fit and the assumption of a lognormal distribution was considered sound.

### 3.2. Geometric means, medians, and standard errors

For comparisons made at the whole state level, measures of central tendency in PCBs ranged from 33 μg/kg for the current method, to 78 μg/kg using MI. The median PCB concentration was 77 μg/kg. For p,p′DDE, central tendency measures ranged from 6 μg/kg using the current method to 20 μg/kg using MI, with a median concentration of 17 μg/kg. In both cases the MI method produced the highest estimate of geometric mean, but was near the median and KM estimate, which was 69 μg/kg for PCBs and 18 μg/kg for p,p′DDE. Comparisons for both contaminants also resulted in the lowest estimate of geometric mean using the current method, as would be expected based on the mathematical underpinnings of geometric mean calculation. The method of substitution using half the detection limit was lower than the MI, K–M, and median, but much closer than the current substitution method. For PCBs, Monte Carlo simulation based tests of significance suggested that the current '0.0001' substitution and 'Half the DL' methods both produced estimates of the geometric mean that showed a significant substitution method effect ($P < 0.0001$ and $P = 0.0406$, respectively). Monte Carlo simulation based tests of significance suggested that for p,p′DDE the current '0.0001' substitution and MI methods both produced estimates of the geometric mean that showed a significant substitution method effect ($P < 0.0001$ and $P = 0.0149$, respectively). Geometric means for each of the methods discussed as well as the median are shown for PCBs and p,p′DDE in Fig. 1. In addition to summary statistics, the figure shows error bars representing one standard error above and below the geometric mean for each method and each contaminant.

The standard errors of the geometric mean at the state level for all of the methods discussed were similar. Table 1 displays the standard errors, number of observations ($n$) and the rate of censorship for each contaminant. For PCB concentrations, measures of standard error ranged from 1.1 using Kaplan–Meier, MI and half the DL methods to 1.3 using the current 0.0001 substitution value. For p,p′DDE concentrations, measures of standard error also ranged from 1.1 to 1.3. Again, the current method of substitution was higher, with the remaining three methods lower and in agreement.

Results at the subpopulation level follow the same trend as the whole state, with the current substitution method producing depressed geometric mean estimates and elevated standard errors. The results of the geometric mean analysis and the medians for PCB and p,p′DDE concentrations are summarized in Table 2. The standard errors for PCB and p,p′DDE concentrations are provided in Table 3. Tables 2 and 3 also display the number of observations ($n$) and rate of censorship for each subpopulation. Subpopulations that were omitted in these tables were those that had no censored values. In all cases with no censored values the KM method resulted in the same estimate as both substitution methods. Because these were instances in which no observations were missing and

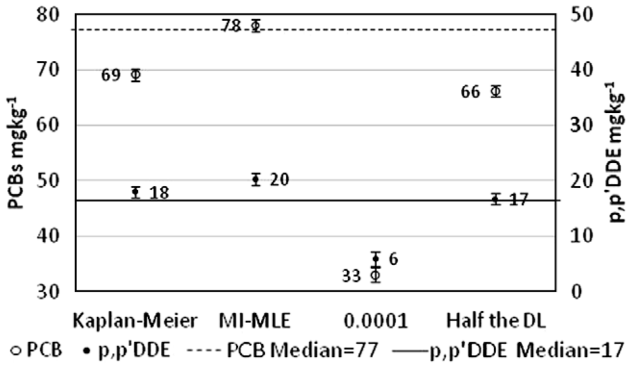

**Fig. 1.** Geometric means ± 1 SE resulting from four methods of calculation for PCB and p,p′DDE levels within the state of Michigan. Median is included for comparison. These 234 observations represent samples collected from 1999 to 2003.

**Table 1**
The standard error of the geometric mean resulting from four methods of calculation and the rate of censorship for the entire state of Michigan for PCBs and p,p′DDE. The 234 observations represent samples collected from 1999 to 2003.

| Error method | PCB (μg/kg) 6.41% Censored | DDE (μg/kg) 10.26% Censored |
|---|---|---|
| Kaplan–Meier | 1.1 | 1.1 |
| MI–MLE | 1.1 | 1.1 |
| 0.0001 | 1.3 | 1.3 |
| Half the DL | 1.1 | 1.1 |

**Table 2**
Medians and geometric means of total PCBs and p,p′DDE concentrations for each subpopulation and each method of calculation. Also included are the number of observations and rate of censorship for each subpopulation.

| Subpopulation | PCB (μg/kg) | | | | | | | p,p′DDE (μg/kg) | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Median | Kaplan–Meier | MI–MLE | Current 0.0001 | Half the DL | *n* | Censorship (%) | Median | Kaplan–Meier | MI–MLE | Current 0.0001 | Half the DL | *n* | Censorship (%) |
| Lake Huron | 135 | 121 | 124 | $43^{0.0224}$ | 105 | 12 | 8 | 36 | 29 | 29 | $11^{0.0001}$ | $25^{0.0182}$ | 12 | 8 |
| Lake Superior | – | – | – | – | – | – | – | 32 | 29 | 33 | $8^{*}$ | 25 | 45 | 11 |
| Lower Peninsula | 33 | 33 | 35 | $16^{*}$ | 31 | 49 | 6 | 12 | 11 | 12 | $2^{*}$ | $10^{0.0218}$ | 49 | 14 |
| Upper Peninsula | 36 | $29^{0.0077}$ | 36 | $4^{*}$ | $26^{0.0004}$ | 64 | 17 | 13 | 12 | 14 | $2^{*}$ | $11^{0.0306}$ | 64 | 17 |

Superscript *P*-values are provided for estimates that showed a significant substitution method effect based on Monte Carlo simulations.
There were no left-censored observations for PCB in the Lake Superior subpopulation.
[*] *P*-value was less than 0.0001.

**Table 3**
Standard errors of the geometric means of total PCBs and p,p′DDE concentrations for each subpopulation and each method of calculation. Also included are the number of observations and rate of censorship for each subpopulation.

| Subpopulation | PCB (μg/kg) | | | | | | p,p′DDE (μg/kg) | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Kaplan–Meier | MI–MLE | Current 0.0001 | Half the DL | *n* | Censorship (%) | Kaplan–Meier | MI–MLE | Current 0.0001 | Half the DL | *n* | Censorship (%) |
| Lake Huron | 1.4 | 1.4 | 3.4 | 1.5 | 12 | 8 | 1.2 | 1.2 | 2.9 | 1.3 | 12 | 8 |
| Lake Superior | – | – | – | – | – | – | 1.2 | 1.1 | 1.8 | 1.2 | 45 | 11 |
| Lower Peninsula | 1.1 | 1.1 | 1.6 | 1.1 | 49 | 6 | 1.1 | 1.1 | 1.8 | 1.1 | 49 | 14 |
| Upper Peninsula | 1.1 | 1.1 | 1.9 | 1.1 | 64 | 17 | 1.1 | 1.1 | 1.8 | 1.1 | 64 | 17 |

There were no left-censored observations for PCB in the Lake Superior subpopulation.

so, no substitutions were made, they were omitted to prevent them from being inappropriately interpreted as an instance of agreement between KM and substitution methods.

## 4. Discussion

### 4.1. Fit of data to lognormal distribution

While not all of the datasets analyzed conformed well to the lognormal distribution, most were either a marginal or good fit. The lognormal distribution is common when handling environmental data, so it is not surprising that many of these contaminant distributions are well approximated by it. Singh and Nocerino (2002) have cautioned that parametric methods of calculating central tendency measures in left-censored datasets can be less reliable when observations do not fit the assumed distribution. In this analysis, the MI method would have been vulnerable to such problems. In most cases, however, the MI geometric mean estimate was very close to the KM estimate, which is not vulnerable to such parametric assumptions. Likewise, in the calculation of standard errors, MI performed almost identically to the KM method. This suggests that MI is robust to at least minor deviations from an assumed distribution.

### 4.2. Geometric means, medians, and standard errors

This study shows that the current method of near-zero substitution for calculating geometric means with left-censored data and a right skew performs poorly relative to the methods used for comparison here. It resulted in the lowest estimate of central tendency, which was farthest from the median, and resulted in the highest estimated standard error when compared to other methods. Several investigations have provided evidence that methods replacing all non-detects with a single value (substitution methods) can introduce bias (Liu et al., 1997; Singh and Nocerino, 2002; Baccarelli et al., 2005; Helsel, 2005a,b, 2006; Eastoe et al., 2006; Needham et al., 2007; Antweiler and Taylor, 2008). This is especially true for datasets with a large proportion of censored observations because bias caused by substitution increases dramatically as the percent of observations censored increases (Eastoe et al., 2006). While several of the alternatives are analytically intensive, many statistical packages now recognize the need and are designed to conduct such analyses. As more programs accommodate this need the programming skill required will no longer be a prohibitive factor. The problems of substitution in general, are compounded by the use of geometric means and especially the choice of substituted value (Helsel, 2005b, 2006). While it may seem appropriate to choose a near-zero value such as ‘0.0001’, this inference is based on the mathematical underpinnings of the arithmetic mean, which differ from those of geometric mean calculation. Arithmetic mean calculation is governed by the properties of addition, for which ‘0’ is the identity value. This means that ‘0’ is the number which can be added to a series without changing the sum. In calculating the arithmetic mean, a ‘0’ allows the sum to remain unchanged while increasing *N*, which is the divisor. In this regard, the purpose of substitution is to serve as a place holder that lets *N* increase without changing the numerator, thereby allowing the non-detects to affect the quotient only by inflating *N*. Geometric mean calculation, however, is governed by the properties of multiplication, for which ‘1’ is the identity value. In multiplication, in contrast to addition, near-zero values have a dramatic effect, while values near one are of the lowest impact. Imagine the difference in this simple example, between the effects of:

$$10\,000 + 0.0001 = 10\,000.0001 \tag{1}$$

$$10\,000 * 0.0001 = 1 \tag{2}$$

When adding in (1), the result is very near the number with which we began. When multiplying in (2) though, the result is 1, which is a drastic decrease from 10 000.

Estimates of geometric mean and standard errors for all methods except the current substitution method were largely in agreement for datasets with rates of censorship near or below 10%. For these low rates of censorship, estimates were in close agreement with one another and with the median, which suggests that they are capturing the central tendency of contaminant concentrations and not overly sensitive to the censorship or skew in the dataset. This includes the other substitution method tested here of '$1/2 * \mathrm{DL}$', which is common practice for contaminant monitoring programs. Based on Monte Carlo simulation results, the current substitution method always resulted in a significant substitution method effect regardless of rate of censorship. The '$1/2 * \mathrm{DL}$' method showed a significant substitution effect for all datasets with greater than 11% and one dataset with only 8% substitution, suggesting that it is not a good method of handling left-censored data as the rate of substitution increases.

Though the maximum likelihood based MI estimates of geometric mean did not show a significant substitution effect based on our Monte Carlo simulations, they were consistently highest, which suggests that they are the most vulnerable to right skew of the methods considered here. Indeed, in Singh and Nocerino's (2002) discussion of handling censored data in the presence of a right skew, they warned that such distribution based methods were "particularly susceptible to problems caused by outliers". They concluded that for large sample sizes and only when distributions could be satisfactorily fit, MLE-based analyses were good alternatives. As stated above, some of our data were shown to be a poor fit to the assumed distribution. However, MI provides an advantage over strict maximum likelihood estimators in that when it is used to 'fill-in' missing observations, sampling is done *multiple* times. This provides the distinct advantage of estimating the variance resulting from the procedure itself versus the variability in the actual contaminant concentrations. Multiple imputation estimators have also been previously found to produce unbiased estimates when the proportion of uncensored values was less than 50% (Jain et al., 2008). Multiple imputation estimates of standard error were similar to estimates produced by all but the current substitution method.

Kaplan–Meier estimates were first derived as a way of determining mean survival in datasets in which not all members of the sample died at the end of the experiment. This resulted in right-censored distributions, which are common in engineering and medical trials (Helsel, 2005b). Increasingly, the common problems in analyzing right and left-censored data have drawn researchers in the environmental field to apply these techniques. Originally, left-censored data were simply transformed to make a right-censored distribution by subtracting them from an arbitrary value larger than the largest observation. As this method grows in acceptance, programs have begun to accommodate left-censored data without such transformations.

In this study, KM estimates did not seem as sensitive to the effects of right-skewed data, which is a major benefit of using a nonparametric analysis technique. The KM estimates of both geometric mean and standard error were overall quite similar to those produced by all but the current substitution method, though the geometric means estimates were consistently lower than the MI estimates, where differences occurred. In one case, the Upper Peninsula geometric mean PCB concentration, the KM estimate was determined to be exhibiting a significant substitution method effect based on simulation results. In other published accounts of data handling methods for left-censored datasets, KM was determined to perform best in the determination of summary statistics (Antweiler and Taylor, 2008). As a testament to its robustness, it has been used as the standard of comparison in other studies of left-censored data (Eastoe et al., 2006).

### *4.3. Recommendations*

Based on the findings here, KM statistics provide the best estimates of geometric means in data with both left hand censorship and a right skew, like those generated by the Michigan BEBP. Differences between KM estimates and MI estimates were minor, which may tempt the conclusion that they are equally valid for these data. However, based on the subtle trend here of MI to be pulled upward, and published evidence of the tendency of parametric analyses like MI to be biased by skewed data (Eastoe et al., 2006), KM seems the best option. Use of the nonparametric KM also provides theoretical consistency in that significant difference testing is already performed using nonparametric techniques.

For the dataset in this study, the common practice of substitution with '$1/2 * \mathrm{DL}$' resulted in estimates of both geometric means and standard errors that did not differ greatly from other methods compared. This must be interpreted with caution, however, since these data had low rates of censorship (6.41% for PCBs and 10.26% for p,p′DDE, at the state level). It has been shown that the bias caused by substitution increases dramatically as the percent of observations censored increases (Eastoe et al., 2006). The agreement between this substitution method and more complex methods is likely a reflection of the low levels of substitution in these data; it should not be interpreted as evidence of equivalence between the '$1/2 * \mathrm{DL}$' substitution method and MI or KM methods. It may be concluded that substitution of '$1/2 * \mathrm{DL}$' would be an acceptable treatment of censored values *only* for studies with low levels of censorship. Substitution is still common practice in toxicological studies, which makes it tempting to employ for the purpose of consistency. However, substitution will become an increasingly problematic solution as monitored contaminants become less prevalent and a larger proportion of samples contain contaminant levels in the non-detectable range. This is evident when comparing the data presented here to historical data. For example, in the years from 1987 to 1993 Lake Huron nestlings provided no samples with non-detectable PCB concentrations compared with 8% of samples with non-detectable PCB concentrations in these data. Only 4% of samples from Lower Peninsula nestlings and only 9% of samples from the Upper Peninsula had non-detectable PCB concentrations in the 1987–1993 dataset, compared with 6% and 17%, respectively here (Bowerman, 1993). We believe that as more studies of this nature are published and software increasingly accommodates left-censored data, substitution methods will become less prevalent.

## Acknowledgments

Special thanks to Dr. William C. Bridges of Clemson University and the reviewers who critiqued this work. This material is based upon work supported by NIFA/USDA, under project number SC-1700281, Technical Contribution No. 5792 of the Clemson University Experiment Station.

## References

Antweiler, R.C., Taylor, H.E., 2008. Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: I. Summary statistics. Environmental Science and Technology 42, 3732–3738.

Baccarelli, A., Pfeiffer, R., Consonni, D., Pesatori, A.C., Bonzini, M., Patterson, D.G., Bertazzi, P.A., Landi, M.T., 2005. Handling of dioxin measurement data in the presence of non-detectable values: overview of available methods and their application in the Seveso chloracne study. Chemosphere 60, 898–906.

Bowerman, W.W., 1993. Regulation of Bald Eagle (*Haliaeetus leucocephalus*) Productivity in the Great Lakes Basin: An Ecological and Toxicological Approach. Department of Fisheries and Wildlife, Institute for Environmental Toxicology, and Ecology and Evolutionary Biology Program, Michigan State University, East Lansing, MI, p. 291.

Eastoe, E.F., Halsall, C.J., Heffernan, J.E., Hung, H., 2006. A statistical comparison of survival and replacement analyses for the use of censored data in a contaminant air database: a case study from the Canadian Arctic. Atmospheric Environment 40, 6528–6540.

Helsel, D.R., 2005a. More than obvious: better methods for interpreting nondetect data. Environmental Science and Technology 39, 419A–423A.

Helsel, D.R., 2005b. Nondetects and Data Analysis: Statistics for Censored Environmental Data. John Wiley and Sons, Inc., Hoboken.

Helsel, D.R., 2006. Fabricating data: how substituting values for nondetects can ruin results, and what can be done about it. Chemosphere 65, 2434–2439.

Jain, R.B., Caudill, S.P., Wang, R.Y., Monsell, E., 2008. Evaluation of maximum likelihood procedures to estimate left censored observations. Analytical Chemistry 80, 1124–1132.

Krishnamoorthy, K., Mallick, A., Mathew, T., 2009. Model-based imputation approach for data analysis in the presence of non-detects. Annals of Occupational Hygiene 53, 249–263.

Liu, S., Lu, J.-C., Kolpin, D.W., Meeker, W.Q., 1997. Analysis of environmental data with censored observations. Environmental Science and Technology 31, 3358–3362.

MDEQ, 1997. A Strategic Environmental Quality Monitoring Program for Michigan's Surface Waters. MI/DEQ/SWQ-96/152. Michigan Department of Environmental Quality, Lansing, Michigan.

Needham, L.L., Naiman, D.Q., Patterson, D.G., LaKind, J.S., 2007. Assigning concentration values for dioxin and furan congeners in human serum when measurements are below limits of detection: an observational approach. Chemosphere 67, 439–447.

SAS Institute Inc., 2000–2004. SAS 9.1.2 Help and Documentation. SAS Institute Inc., Cary, NC.

Singh, A., Nocerino, J., 2002. Robust estimation of mean and variance using environmental data sets with below detection limit observations. Chemometrics and Intelligent Laboratory Systems 60, 69–86.

Wierda, M.R., 2009. Using Bald Eagles to Track Spatial and Temporal Trends of Contaminants in Michigan's Aquatic Systems. Department of Forestry and Natural Resources, Clemson University, Clemson, SC, p. 91.