Benefits of Monte Carlo Imputation of Non-detects in Environmental Data

Kirk Cameron, Ph.D.

MacStat Consulting, Ltd.

10330 Mill Creek Ct., Colorado Springs, CO 80908

Abstract

Common practice of either ignoring non-detects (NDs) or replacing them with 'simple substitution' values, e.g., half the reporting limit (RL), has led to calls for greater statistical sophistication via left-censored data models. Less clear is if the assumptions underlying a censored model are always satisfied, or whether such a model leads to more accurate results.

Simple substitution of half the detection/reporting limit (RL) can be viewed as replacing each non-detect with its expected value under a mixture model that draws non-detects from a uniform distribution on the interval [0, RL]. We propose an extension to this model whereby non-detects are drawn with Monte Carlo sampling from one of a class of bounded distributions on [0, RL], e.g., uniform, beta, triangle. We also propose repeated draws from the mixture model to generate a series of data realizations, from which the statistical properties of any desired estimator can be computed.

The benefits of this combined mixture model and computational strategy are explored, including algorithmic and computational feasibility, and better visualization and assessment of ND-associated uncertainty.

Key Words: Monte Carlo, non-detects, imputation, simple substitution, left-censored

1. Introduction and Background

Statisticians are popularly teased as 'anal-retentive.' As the old joke goes, how many statisticians does it take to change a light bulb? Answer: 3.67 statisticians, on average. Yet making sense of mathematics, statistics, and evidence is complex. To understand how things actually work, as opposed to *Just So* explanations, mundane, nitty-gritty assumptions and models must be tested.

So it goes with a ubiquitous issue in environmental data analysis: the little goblins known as non-detects. Ostensibly, non-detects are a statement of nature. A chemical or substance is tested and found lacking in a physical observation. In practice, laboratories can rarely make such a claim. Instead, due to analytical limitations and measurement imprecision, they assert that the measured value (e.g., concentration) of a non-detect is no greater than a threshold (e.g., detection/reporting limit [RL]). Thus, non-detects are akin to statistically left-censored observations, with 'true' value located in the interval [0, RL].

Statistically-minded observers may think the problem already solved under this conceptual framework. A quick reworking or extension of survival analysis methods for right- or interval-censored data ought to do the trick. Indeed, efforts have been made to apply more sophisticated censored-data models and survival techniques to non-detects. These have ranged from MLE estimates for left-censored data, to non-

parametric and parametric adaptations of Kaplan-Meier, an alternative to Kaplan-Meier known as regression on order statistics (ROS), to censored regression methods such as tobit regression, etc. (Cohen, 1959; USEPA, 2009; Helsel & Gilliom, 1986)

Nevertheless, the little goblins refuse to go away. One problem is very practical and ingrained. For a long time, environmental scientists — sometimes due to lack of statistical training or perhaps not wanting to bother — have taken the path of least resistance. Spreadsheets and Excel do not easily accommodate censored-data models without significant effort. Much easier is simple substitution: replacing each ND by a single guesstimate (e.g., 0 or some fraction of the detection/reporting limit, say RL/2, but sometimes the RL itself). Though statisticians have hollered in protest (Helsel, 2005) at the crudeness and probable bias associated with simple substitution — especially for smaller samples and larger proportions of non-detects — the practice has mostly gone on, unchanged.

A second problem is the difficulty in accurately modeling the modern analytical measurement process. Chemists routinely but inconsistently use text-based 'flags' (i.e., notations about the physical sample) to record additional information about measurements alongside their detection status and measurement threshold. An example is the contrast between a 'U' versus a 'J' flag. Both measurements might be reported with the same reporting limit (RL), yet 'U' often indicates an 'undetected' substance, while 'J' usually means the substance was probably detected yet still somewhere below the threshold. The reality can be fuzzier still. A 'U' flag may be interpreted as a value below the method detection limit (MDL), as opposed to a 'zero,' while 'J' might represent a value between the MDL and the larger RL.

Complicating matters, MDLs and RLs themselves are statistically-derived on the basis of repeated measurements of either 'blank' or 'spiked' physical samples. It is a probabilistic statement to say that a given measurement is assumed to be less than the MDL or RL, as the thresholds are not fixed or known physical properties of the lab apparatus. In fact, reporting limits can and do vary for the same substance measured by the same lab using the same analytical technique. And the variation may not be particularly or necessarily related to the distribution of the target substance. Measurements of a given chemical compound may suffer 'interference' by the presence of similar or 'nearby' compounds (e.g., on a spectrograph) or may be challenged by certain physical properties of the physical sample, such as high turbidity in groundwater or significant heterogeneity of grain size or composition in sediment and soil.

Indeed, a special sort of problem arises with non-detects associated with high dilution factors, where the chemist may repeatedly dilute the physical sample in order to 'block out' or 'wash out' interferences, only to determine that the target substance was not detected. Standard practice is to then 'reverse' or 'back down' the chain of dilutions before reporting the value, often resulting in unusually high reporting limit thresholds.

The bottom line is that the goblins are quite real. From lab to lab, and dataset to dataset, it may not be clear precisely what non-detects represent or how they relate to the subset of detected values. With this lack of clarity or consistency, what is the best or most accurate way to handle non-detects in statistical estimates? Further, what is the impact of different statistical treatments or assumptions about non-detects, and does it really matter? These are the questions explored in this paper.

2. Framing the Problem

While the 'obvious' statistical answer to non-detects is a left-censored data model, it is worth noting that censored models make strong assumptions, and usually require moderate to large samples for purposes of estimation. The strongest assumption is that non-detects occupy some portion of the lower tail of the overall measurement

distribution, so that, considered as random variates *sans* censoring, detects and non-detects are identically distributed. In other words, if one could 'peek behind the curtain,' the 'actual' values of non-detects would be consistent with those of the detected observations.

A reasonable (if 'anal-retentive') question might be: is this assumption typically valid? And, even if true, how would we check it? In extremely rare instances, labs have been convinced to report the 'raw,' unqualified instrument readings that are typically guarded prior to post-processing and flag assignment, via quality assurance and quality control (QA-QC) checks. More typically, non-detects are reported simply as 'less thans' (e.g., <5) or, somewhat better, as 'J'-flagged observations that may have an attached numerical estimate. The current best-case scenarios are probably one or both of the following:

1. the presence of numerical J-values combined with variation in the reporting limits for these same non-detects;

2. a degree of intermixing in nominal measurement levels between detected observations and the set of reporting limits.

The first scenario provides at least a crude look at the distribution of some portion of the non-detects. Comparing these values against projected estimates from a left-censored model where all the non-detects are assumed to be 'less thans' may offer a small test of the left-censored assumption. The second scenario extends this argument to include some portion of the detected values. Though this latter has the disadvantage of not providing alternate values for the non-detects (e.g., a J-value of 1.3 vs. <2), it does suggest that the 'less thans' may have actual values similar to those detected measurements with which the levels are intertwined.

The difficulty in all this is that the observed measurement distribution depends not simply on the underlying presence or concentration of a substance, but also crucially on the sampling, collection, and handling procedures used to gather the physical samples and also the field or lab measuring process. As noted earlier, labs use various criteria to decide whether to censor a given observation. These criteria include and are influenced by:

- sample purity, homogeneity, and interferences from other chemicals or constituents;

- sample dilutions;

- contractual precision requirements;

- analytical and extraction methods, as well as instrument calibration.

On balance, a key question remains: how often do real data meet the assumptions of left-censored data models? If a left-censored model does not fit well, are there reasonable alternatives? And a related question: even if a left-censored model does apply, might an alternative model provide comparable performance?

Although left-censored models have gotten the most attention from environmental statisticians, lesser known models have been proposed and used. These generally fall under the umbrella of mixture models, including the delta and modified delta methods. Both propose treating non-detects distinctly from detects, with their own discrete distribution. The classic delta model attributed to Aitchison (1955) places a 'spike' at zero to represent NDs, with a lognormal distribution governing the detected data. More generally, the modified delta (see, for instance, Smith, Kahn, and Cameron, 1993) models the subset of non-detects as a discrete distribution with positive mass located at a (fixed) fraction of each distinct reporting limit, and with the

detected data governed by a separate continuous distribution (not necessarily lognormal).

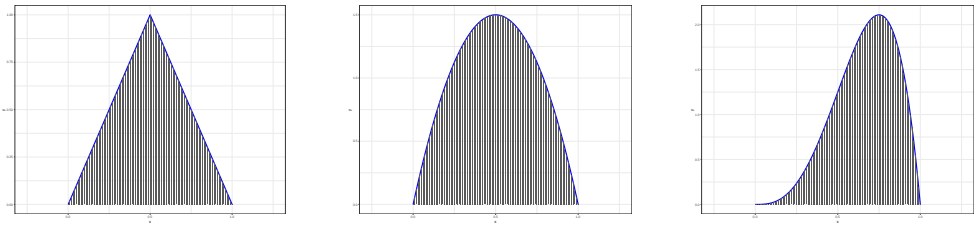
In practice, simple substitution falls under the modified delta umbrella, with non-detects replaced by their expected values under the discrete portion of the mixture model. Whether the modified delta is any more valid than a left-censored model is as difficult to verify as the latter, though there seem to be circumstances under which a mixture model is at least as plausible as a left-censored approach. Further, the modified delta is quite easy to implement; simple substitution is used all the time and can be computed even on tiny samples.

Nevertheless, when a left-censored model holds, estimating all non-detects at, say, half their reporting limits can lead to an artificial 'clumping' of the NDs, a potentially substantial degree of statistical bias, and mis-estimation of the sample variance. So the question remains: even when misspecified, do the outcomes from assuming a mixture model differ substantially from assuming a left-censored model? And, might the practical benefits of mixture models tend to outweigh a more technically precise censored model?

3. A Modest Proposal

Due to deficits in the approaches to date, we propose an extension of the mixture model. Instead of 'clumping' non-detects at a fixed fraction of their reporting or detection limits, consider each non-detect as distributed continuously throughout the range [0, RL]. A variety of bounded models might govern the non-detects, including, for example, the uniform, triangle, or beta distributions, and potentially others (see **Figure 1)**.

Figure 1. Some Bounded, Continuous Models: Triangle, Beta(2,2), Beta(4,2)



The choice of model (and the model parameters) could be informed by similar datasets, specific features of the measurement process, or perhaps its anticipated 'closeness' to the left-censored model, etc. In fact, a sensitivity analysis could be conducted on multiple mixture models to assess the impact of model choice.

But whichever mixture module is selected, we propose to apply Monte Carlo imputation to the non-detects. That is, we invoke repeated Monte Carlo sampling from the bounded model governing the non-detects to impute a series of realizations for the non-detect subsample. For each realization, we construct or compute the statistical estimate or test of interest, and then appropriately average the results across realizations to generate a final outcome.

The benefits associated with the Monte Carlo imputation strategy are several:

1. With modern computing power, ubiquity, and speed, it is very doable and practical;

2. It is relatively easy to program or automate, even for non-experts in statistical programming;

3. It is easy to explain and interpret how non-detects are being manipulated;

4. It can avoid the bias and criticisms of simple substitution, while remaining relatively easy to implement;

5. Results can be better than using a left-censored model, and may be more appropriate in many real-life datasets;

6. Results can be comparable to using a left-censored model in cases where the censored model *is* properly specified.

4. Method

To illustrate the utility of the Monte Carlo imputation strategy, we consider some comparisons and simulation study results for two common techniques: (1) visualizing exploratory box plots with non-detects; and (2) constructing Student-t confidence intervals around the mean. The comparisons include the following methods for treating non-detects:

- Simple substitution of either zero, RL/2, RL for each non-detect;

- Kaplan-Meier for left-censored data, in order to non-parametrically estimate the relevant statistical quantities (e.g, lower quartile for a box plot, mean for a confidence interval);

- Two variants of (robust) regression on order statistics (ROS). ROS imputes non-detects by fitting a linear regression to the detected values on a Q-Q plot and then extending this line to the censored portion of the overall distribution. Once imputed, relevant statistical estimates are calculated on the joint set of detected and imputed values;

- For confidence intervals, an order statistic-based, nonparametric interval around the median;

- Monte Carlo imputation of each non-detect, using each of five bounded models on the interval [0, RL], including the uniform distribution, the triangle distribution centered at RL/2, and three variants of the beta distribution, with shape parameters (2, 2), (2, 4), or (4, 2).

So that the comparisons would not naturally favor the mixture model, an R function was developed to generate underlying data according to either a left-censored normal or lognormal distribution with varying coefficients of variation (CV). The mechanism saves each uncensored realization, but then censors the data using a pre-specified number (1+) of increasing reporting limits (RL), along with a set of pre-specified mixing probabilities. Each mixing probability governs the chance that a generated measurement below a given RL will be observed (i.e., reported as is) or censored by the RL and reported as a ‘less than.’ This allows for more realistic modeling of intermixed samples of detects and non-detects.

The simulation study employed a range of sample sizes from 5 to 30, two confidence levels (0.9, 0.95), the two underlying distributional models discussed above with CV equal to either 0.5, 1, or 1.5, one to three distinct reporting limits ranging from 50% to 150% of the model mean, and a decreasing set of mixing probabilities (0.8, 0.6, and 0.4), designed to reflect the notion that larger-valued observations should be less likely to be censored than smaller observations. Simulations were repeated 100 times for each combination of parameters. The generating mechanism created samples with

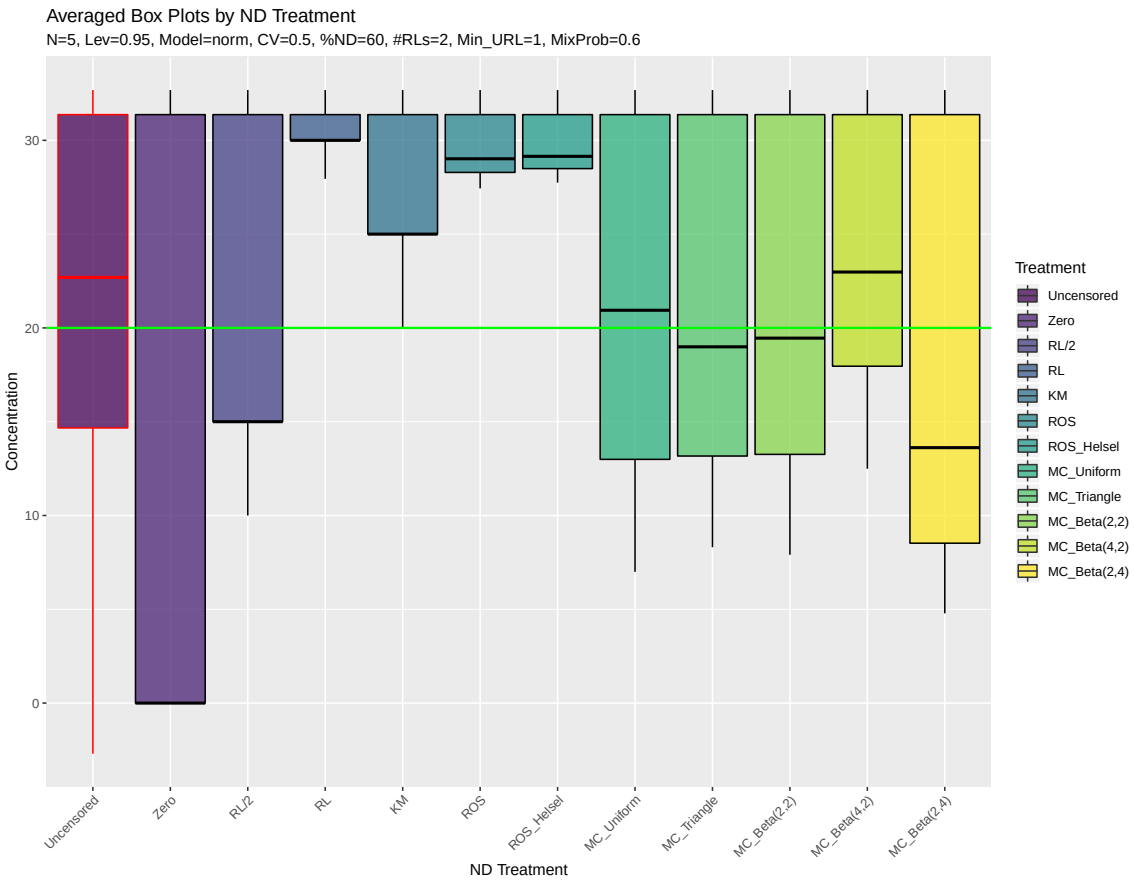
censoring levels ranging from 0 to 100%, with a median of 40% censoring across 1,260 simulation runs representing distinct permutations of the input parameters.

5. Box Plot Results

For exploratory box plots, simple substitution methods can lead to one of more of the following: visual artifacts, less apparent variation, and/or poor coverage of the uncensored median (see **Figure 2**). Very few left-censored model alternatives exist in standard software, though it would not be difficult to implement an approach using either Kaplan-Meier or ROS to estimate the needed summary statistics prior to plotting. With substantial fractions of non-detects, Kaplan-Meier and ROS estimates tend to break down. Helsel (2005) also recognized the large uncertainties associated with highly non-detect samples, suggesting 'erasing' or 'graying out' that portion of any box plot below the highest RL in the data. But this, of course, entails significant loss of information when there are non-detects with large dilution factors or a substantial fraction of detected measurements below the highest RL.

Monte Carlo imputation is an elegant but imperfect solution, avoiding many of the pitfalls of simple substitution, allowing construction of box plots regardless of non-detect percentage, while also providing a better visual sense of variation. Computationally, the number of Monte Carlo realizations can be varied to suit one's impatience. Further, for small to moderate sized samples, it can be helpful to overplot the data, including RLs for NDs, to aid in visual evaluation.

Figure 2. Example Box Plot Comparisons for N=5, Normal Model, Two RLs



In the simulation study, two measures were employed to compare box plots constructed under the 11 distinct treatments of non-detects: (1) coverage of the underlying model median (i.e., testing whether the interquartile range includes the true median), and (2) the root mean square discrepancy (RMSD) of the treatment-estimated lower quartile (LQ), median (Med), and upper quartile (UQ) versus the same statistics computed from the uncensored sample (e.g., LQ_0), as given by equation [1].

$$(1) \qquad RMSD = \sqrt{\frac{1}{3}[(LQ - LQ_0)^2 + (Med - Med_0)^2 + (UQ - UQ_0)^2]}$$

The methods were then ranked by average coverage and average RMSD across the set of simulations, both overall and broken down by specific factors (e.g., sample size, percentage censored, etc.). Lower (better) ranks corresponded to higher average coverage and lower average RMSD. **Tables 1** to **3** shows the rankings across sample sizes, models, censoring levels, and confidence levels for cases of one, two, or three distinct reporting limits. Despite all the data being generated according to a left-censored model, the lowest average RMSD values were associated with the Monte Carlo imputation methods almost uniformly, followed by the ROS methods in two cases. The Kaplan-Meier strategy ranked no higher than sixth of 11, but always fared worse than simple substitution of RL/2. Simple substitution of zero or the RL ranked last across the board.

Table 1. Average Rankings of Box Plot ND Treatments, Single Reporting Limit

Treatment	RMSD Rank	Coverage Rank
Uncensored	1.00	1.00
MC_Uniform	4.15	1.03
MC_Beta(2,2)	4.28	1.11
MC_Triangle	4.79	1.13
MC_Beta(2,4)	5.77	1.16
MC_Beta(4,2)	5.80	2.00
ROS_Helsel	6.29	1.76
ROS	6.32	1.36
RL/2	6.61	1.89
KM	7.97	1.84
Zero	8.67	1.43
RL	8.94	5.44

Table 2. Average Rankings of Box Plot ND Treatments, Two Reporting Limits

Treatment	RMSD Rank	Coverage Rank
Uncensored	1.00	1.00
MC_Uniform	4.32	1.00
MC_Beta(2,2)	4.73	1.04
MC_Triangle	5.39	1.04
MC_Beta(4,2)	5.63	2.12
ROS_Helsel	6.42	1.48
ROS	6.73	1.48
RL/2	7.33	1.20
MC_Beta(2,4)	7.56	1.00
KM	7.93	2.69
RL	10.13	6.33
Zero	10.80	1.28

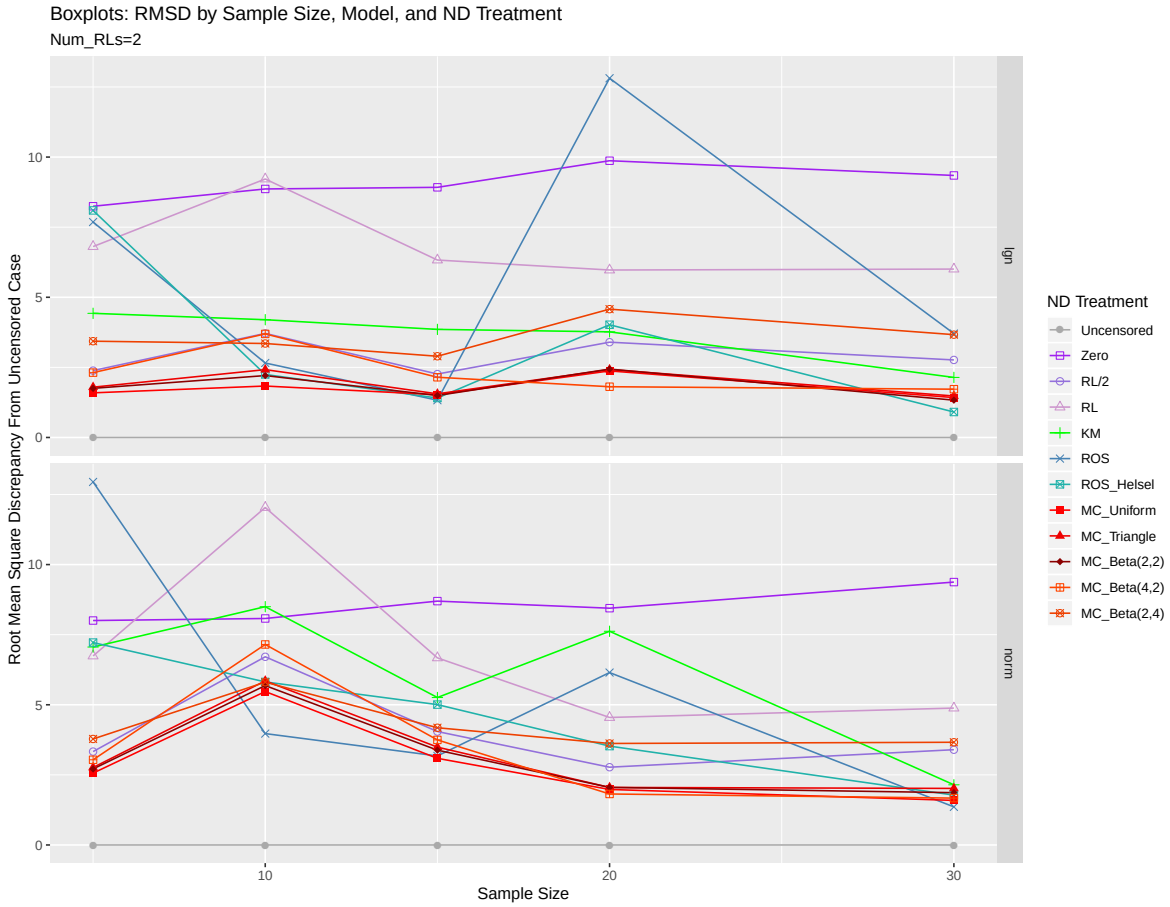
Table 3. Average Rankings of Box Plot ND Treatments, Three Reporting Limits

Treatment	RMSD Rank	Coverage Rank
Uncensored	1.00	1.00
MC_Beta(2,2)	4.43	1.10
MC_Uniform	4.67	1.10
MC_Beta(4,2)	5.00	2.17
MC_Triangle	5.07	1.27
RL/2	6.33	1.87
KM	7.00	2.83
ROS_Helsel	7.53	3.50
MC_Beta(2,4)	7.83	1.83
ROS	8.53	3.20
RL	9.90	7.03
Zero	10.70	3.07

Similar outcomes were seen for the average coverage of the model median. Four of the five Monte Carlo imputation methods ranked best, with some jostling among the simple substitution and censored methods. The Monte Carlo imputation methods also generally fared best when ranked separately by model (normal vs. lognormal) or sample size (see **Figures 3** and **4**), though results for the censored methods were rather similar to Monte Carlo imputation for the largest sample size (n = 30).

On balance, the simulation study supports the premise that Monte Carlo imputation offers advantages over simple substitution methods as well as over censored model techniques, especially for smaller sample sizes and heavier censoring. Note, however, that the most common form of simple substitution, RL/2, on occasion ranks better or similarly to the censored methods and some of the Monte Carlo imputation variants.

Figure 3. Box Plot RMSD Comparisons by Model and Sample Size, Two RLs



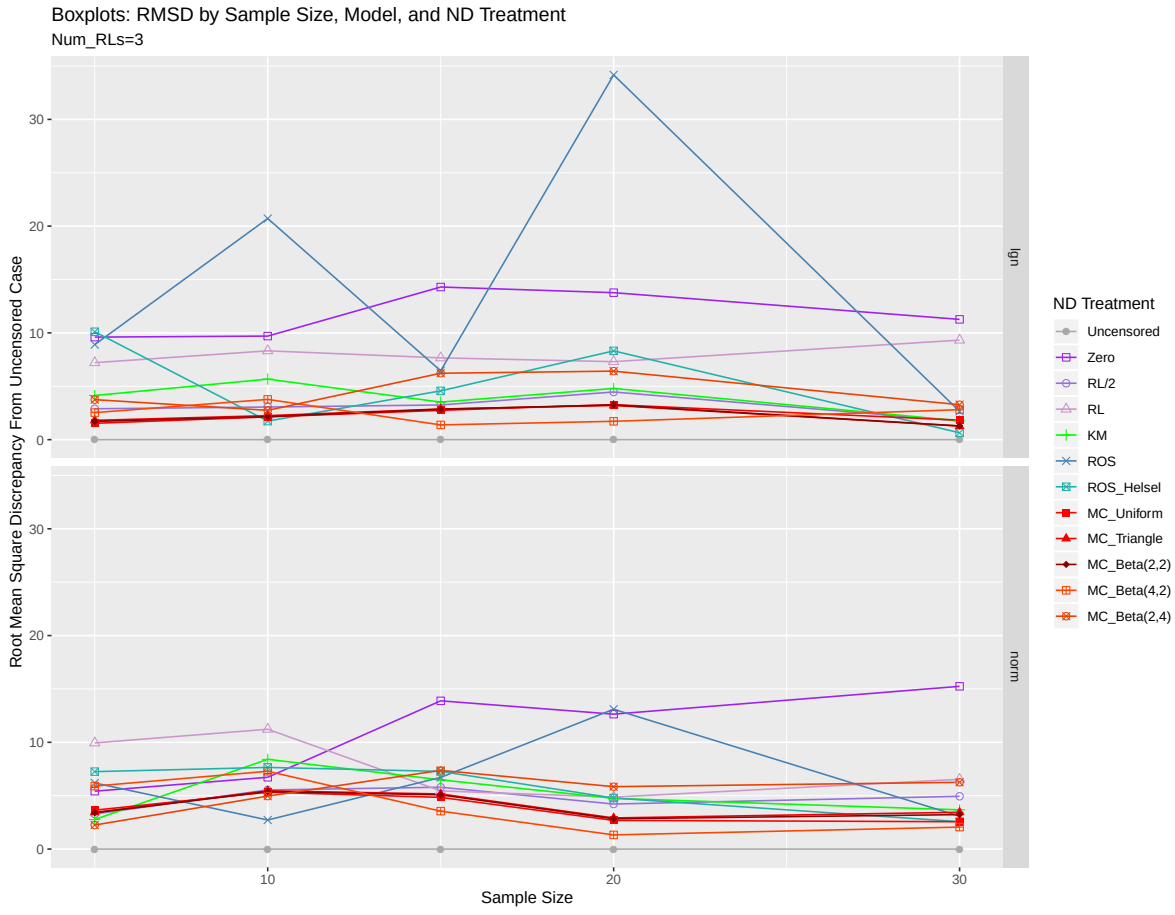
6. Confidence Interval Results

For samples with large fractions of non-detects, simple substitution has been dinged for underestimating the variance, and in turn, biasing interval width and coverage. These problems do not simply vanish when censored model techniques are used. For instance, nonparametric, rank-based methods must choose how to impute/estimate the lower confidence limit (LCL), as there is typically no way to explicitly rank the non-detects. This limitation can *also* impact or bias coverage and interval width (see **Figure 5**).

Resampling methods are not immune either. A percentile bootstrap or jackknife must rely on some combination of simple substitution, Kaplan-Meier, ROS, etc. to estimate the desired statistic (e.g., mean) within each bootstrap or jackknife replicate. Parametric bootstrapping of model parameters or sufficient statistics faces the same challenge.

Monte Carlo imputation entails a slightly greater computational burden, but the imputed non-detects can be easily combined with parametric (or nonparametric) estimation formulas for each Monte Carlo realization. Subsequently, confidence limits can be averaged across realizations. (Though not explored in this paper, one could alternatively calculate an extreme percentile across the Monte Carlo realizations to maximally account for non-detect uncertainty. This would be akin to the difference between beta-expectation vs. beta-content tolerance limits.)

Figure 4. Box Plot RMSD Comparisons by Model and Sample Size, Three RLs



We could also combine a bootstrap or jackknife approach with Monte Carlo imputation, where each Monte Carlo realization serves as the basis for a series of bootstrap or jackknife samples and estimated confidence interval. The resulting interval limits can again be averaged to get a final estimate. Note, in addition, that for heavily-censored samples, some bootstrap replicates may include too many non-detects to readily or accurately apply Kaplan-Meier or another censored-model technique. Monte Carlo imputation can always be computed, no matter what the composition of a bootstrap replicate or how many copies of specific non-detects it includes.

Although a hybrid resampling-Monte Carlo imputation approach was not tested in the present simulation study, Student-t intervals were compared across the same suite of ND treatments as applied to the box plot study, along with a rank-based nonparametric interval. Three measures were employed to compare confidence intervals constructed under the 12 distinct treatments of non-detects: (1) coverage of the underlying model mean, (2) coverage of the underlying model median, and (3) the root mean square discrepancy (RMSD) of the treatment-estimated lower confidence limit (LCL) and upper confidence limit (UCL) versus the same statistics computed from the uncensored sample (e.g., LCL_0), as given by equation [2].

(2) $$RMSD = \sqrt{\frac{1}{2}[(LCL - LCL_0)^2 + (UCL - UCL_0)^2]}$$

As with the box plots, the methods were then ranked by average coverage and average RMSD across the set of simulations, both overall and broken down by factors like sample size, degree of censoring, etc. Lower (better) ranks were assigned the higher the average coverage and the lower the average RMSD. **Tables 4** to **6** show the rankings across sample sizes, models, censoring levels, and confidence levels for cases of one, two, or three distinct reporting limits.

Figure 5. Example Conf. Int. Comparisons, N=10, Normal Model, Single RL

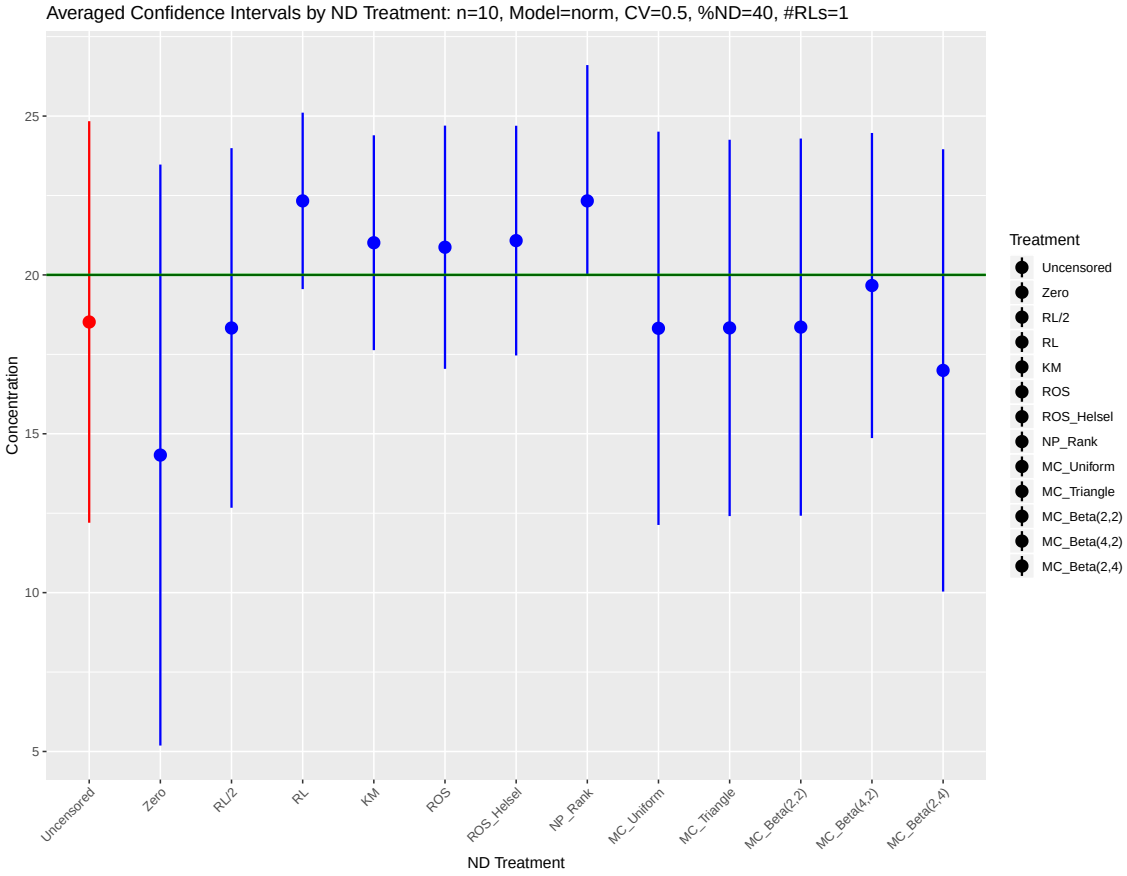


Table 4. Average Rankings of Conf. Int. ND Treatments, Single Reporting Limit

Treatment	RMSD Rank	Mean Coverage Rank	Median Coverage Rank
Uncensored	1.00	3.93	4.30
MC_Beta(2,2)	4.86	4.00	3.91
MC_Uniform	4.96	3.35	3.37
MC_Triangle	5.23	4.35	4.23
MC_Beta(2,4)	6.16	6.73	4.16
RL/2	6.20	5.74	5.65
ROS_Helsel	7.01	8.53	9.02
MC_Beta(4,2)	7.23	4.32	7.17
KM	7.94	9.29	9.98
ROS	7.95	6.87	6.39
Zero	8.96	8.46	5.33
RL	10.84	8.21	10.96
NP_Rank	12.65	5.52	6.70

Table 5. Average Rankings of Conf. Int. ND Treatments, Two Reporting Limits

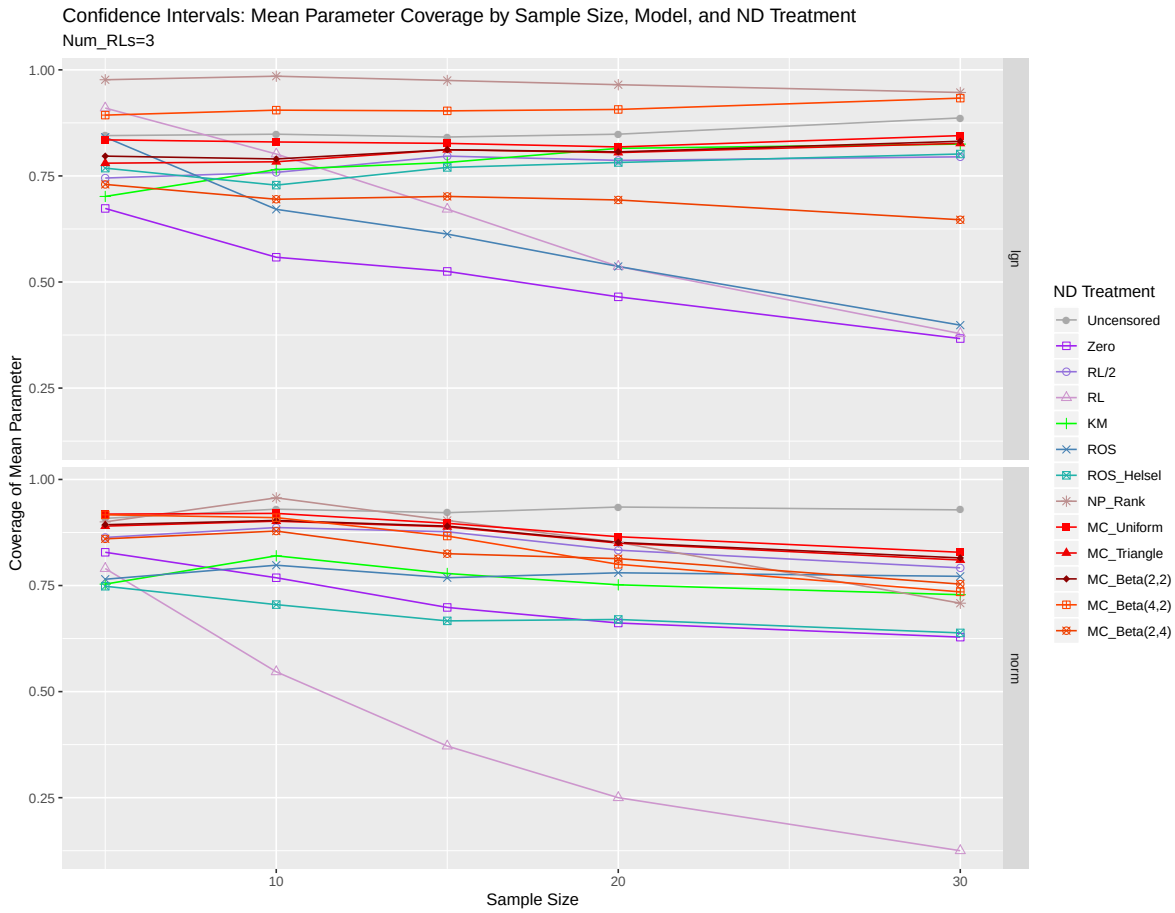
Treatment	RMSD Rank	Mean Coverage Rank	Median Coverage Rank
Uncensored	1.00	3.87	4.13
MC_Beta(2,2)	4.70	4.58	3.92
MC_Uniform	4.90	3.76	3.14
MC_Triangle	4.99	4.98	4.25
RL/2	5.76	6.73	5.93
MC_Beta(2,4)	6.35	7.65	4.62
MC_Beta(4,2)	6.38	4.43	6.87
ROS_Helsel	7.51	9.36	9.50
KM	8.01	9.78	10.39
ROS	8.88	8.22	6.99
Zero	9.05	9.55	6.36
RL	11.43	9.61	12.10
NP_Rank	12.04	2.98	7.69

Table 6. Average Rankings of Conf. Int. ND Treatments, Three Reporting Limits

Treatment	RMSD Rank	Mean Coverage Rank	Median Coverage Rank
Uncensored	1.00	3.33	3.73
MC_Beta(2,2)	4.22	4.78	3.83
MC_Triangle	4.60	5.17	4.12
MC_Uniform	4.62	3.60	2.63
RL/2	5.73	6.93	5.88
MC_Beta(4,2)	6.30	3.92	6.47
MC_Beta(2,4)	6.68	8.05	5.28
KM	6.83	8.78	9.33
ROS_Helsel	8.10	9.70	9.53
Zero	9.50	10.18	7.50
ROS	10.08	9.78	8.52
RL	11.43	10.12	12.43
NP_Rank	11.90	3.15	8.20

Similar to the box plot results, the lowest average RMSD values were associated mainly with the Monte Carlo imputation methods, although simple substitution of RL/2 also ranked fairly highly, sometimes better than the non-centered beta Monte Carlo imputation methods, and consistently better than the censored model methods or other simple substitution strategies.

Figure 6. Conf. Int. Mean Coverage by Model, Sample Size; Three RLs

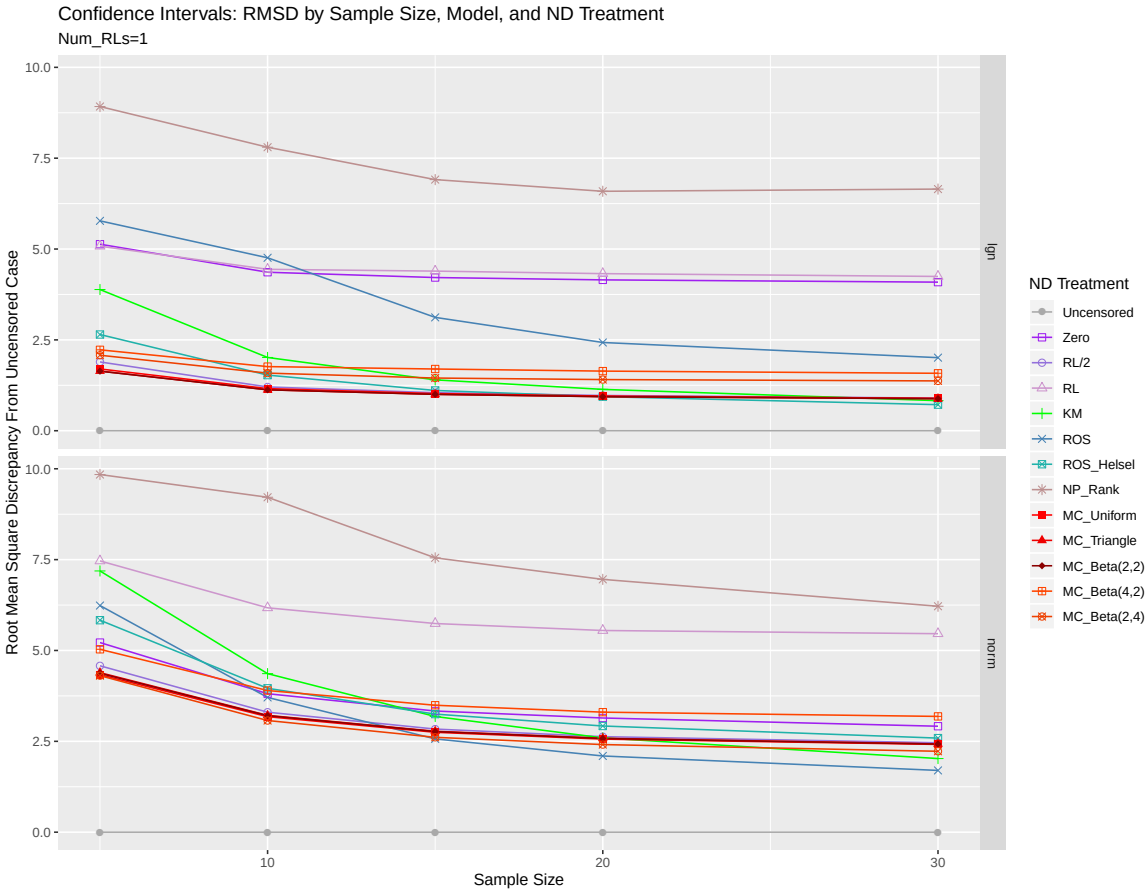


Average rankings of coverage of the mean and median again tended to favor Monte Carlo imputation, with one notable exception. The nonparametric rank-based confidence interval ranked better than all the other methods with respect to covering the mean, when there were two or three reporting limits, but towards the rear of the pack when covering the median. It seems possible that this result is partially an artifact of the method used to compute the interval. To enable its computation when the LCL (and occasionally the UCL) would be chosen from among a subset of non-detects, the censored data were treated ‘as is,’ i.e., reporting limits were used as nominal measurements, akin to simple substitution of the reporting limit. This tends to bias the interval upward, negatively impacting coverage of the median, but perhaps positively impacting coverage of the mean in the lognormal case.

Overall, the results of the simulation study again validate the robustness of the extended mixture model approach, even when data are generated according to a left-censored model. The Monte Carlo imputation methods generally fared best when ranked separately by model (normal vs. lognormal) or sample size (see **Figures 6** and

7), though again it must be noted that simple substitution of RL/2 ranked surprisingly well given the criticisms leveled against it.

Figure 7. Conf. Int. RMSD Comparisons by Model, Sample Size; Single RL



7. Other Applications

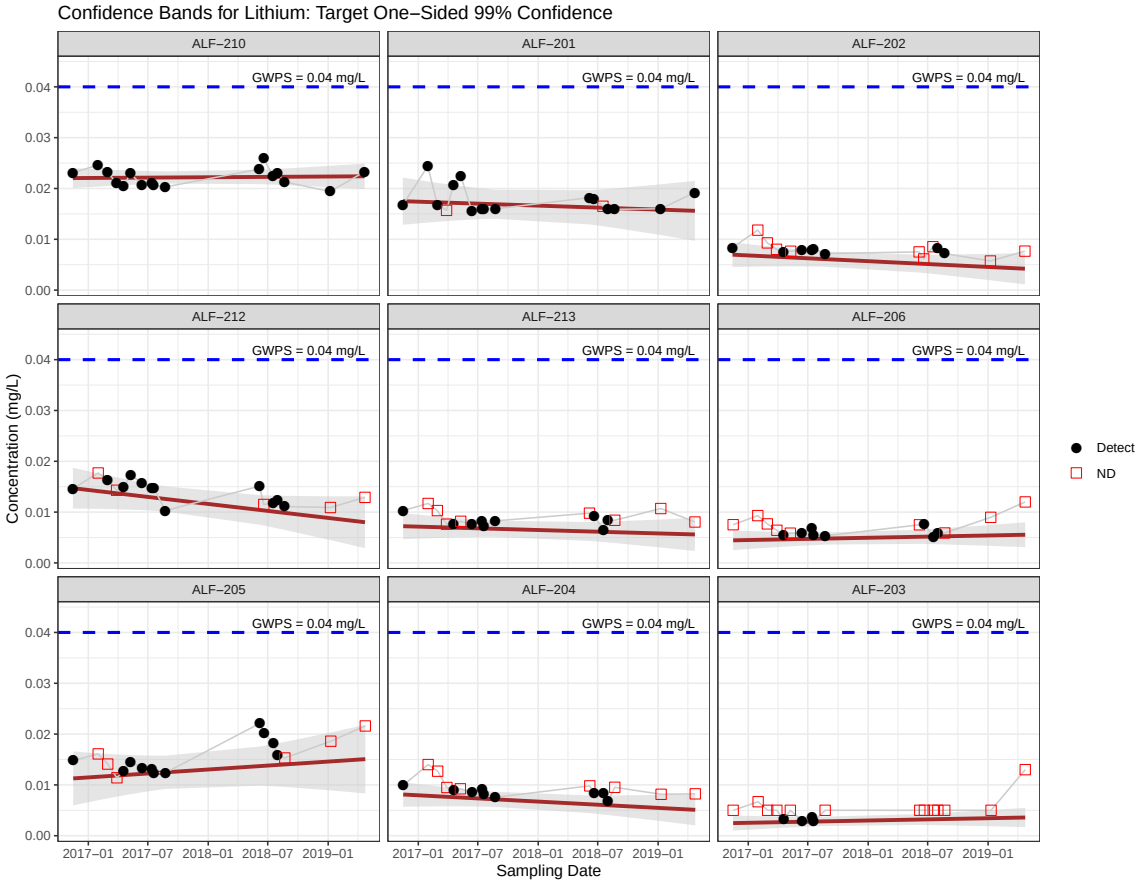
The two examples studied in this paper are merely the tip of iceberg. Almost any statistical approach involving data with non-detects could be massaged with Monte Carlo imputation. Furthermore, some estimates are quite difficult if almost impossible under a left-censored approach. Consider for instance regression or trend estimates and their associated confidence bands. The distributional assumptions of standard linear regression relate to the residuals around the regression or trend line — since the response variable (Y) is not stationary — but the distribution of censored residuals is harder to fit within the usual left-censored model framework (e.g., the interval of censoring is distinct and often non-positive for each residual depending on the difference between the fitted estimate at a given X or point-in-time and the reporting limit of a non-detect at that point).

Tobit (censored) regression can be applied to left-censored data, or perhaps the Akritas-Theil-Sen (ATS) nonparametric method (Akritas, Murphy, and Lavalley, 1995). The first of these assumes a normal linear model of the uncensored measurements, and tends to be less robust to model mis-specification. ATS is more computationally expensive, requiring careful tracking of potentially censored residuals and an iterative search for a slope estimate which makes Kendall's tau-b correlation as close to zero as possible.

Simple substitution is commonly used in these cases for all the reasons outlined in the **Introduction**, but can easily lead to artificially 'clumped' trend or regression estimates, with underestimated local variance and biased confidence bands.

Monte Carlo imputation, on the other hand, is straightforward. It can be combined with standard parametric or nonparametric trend or regression methods. For each Monte Carlo realization, a regression line or trend and associated confidence band can be computed, and then the results averaged across realizations. Some examples are shown in **Figure 8**.

Figure 8. Linear Conf. Bands Using Monte Carlo Imputation, Beta(2,2) Model



Another easy application is two or multi-sample comparisons (e.g., t-tests, ANOVA). Rank-based tests (e.g., Wilcoxon-Mann-Whitney, Kruskal-Wallis) treat non-detects as 'tied values,' even though the justification for such an assumption can get a bit messy. Under a left-censored framework, not only is the pair (<5, <5) considered a 'tie,' but in some sense so must the pair (<5, 3). With varying reporting limits and intermixing of the detects and non-detects, this can lead to overlapping subsets of differently tied values where portions of each subset may be tied with part but not all of other subsets (e.g., (<10, 7) and (<10, <5) are tied pairs, but (7, <5) is not).

One could alternatively use Gehan's test or similar, involving a partial ranking of the data, but again Monte Carlo imputation is probably more straightforward and practical, and possibly more powerful. The relevant test statistic can be computed on each Monte Carlo realization and then averaged prior to assessing significance.

8. Conclusion

Non-detects continue to be the little goblins of environmental data analysis. Simple substitution of a fixed fraction of the reporting or detection limit has been the common solution, but this has been roundly criticized. Ingenious, but sometimes complicated, methods have been developed to accommodate non-detects under a model of left-censoring. Unfortunately, the censored model may not be realistic or practical in many situations, and standard software may or may not provide appropriate censored model techniques.

We propose instead an extended mixture model in which non-detects are modeled as bounded, continuous variates on the interval [0, RL]. Practical computation with this model entails repeated Monte Carlo imputation of the non-detects. This strategy offers several practical and empirical benefits:

- Ease of algorithmic description and implementation;
- Doable with modern computing power/speed;
- Ease of interpretation;
- Useful even with very large censoring fractions and/or very small data sets;
- At least sometimes more realistic than a left-censored model;
- Provides similar or better results than left-censored techniques, even when data are generated according to a censored model;
- Better accounts for uncertainty due to non-detects than common simple substitution methods;
- Flexible enough to adapt to a variety of measurement process models and censoring mechanisms.

We encourage environmental scientists, statisticians, and statistical software developers to implement and offer this approach. Let the little goblins be vanquished!

9. References

Aitchison, J. 1955. On the distribution of a positive random variable having a discrete probability mass at the origin. Journal of the American Statistical Association, 50(272), 901-908.

Akritas, M.G., S.A. Murphy, and M.P. Lavalley. 1995. The Theil-Sen estimator with doubly censored data and applications to astronomy. Journal of the American Statistical Association, 90(429), 170-177.

Cohen, A.C. 1959. Simplified estimators for the normal distribution when samples are single censored or truncated. Technometrics, 1, 217-237.

Helsel, D.R. & J.R. Gilliom. 1986. Estimation of distributional parameters for censored trace level water quality data: II. verification and applications, Water Resources Research, 22, 147–155.

Helsel, D.R. 2005. Nondetects and Data Analysis. New York: John Wiley & Sons.

Smith, M., H. Kahn, & K. Cameron. 1993. Estimation of effluent limitations from datasets with small proportions of detected measurements. Contributed paper, 9th Annual EPA Conference on Statistics, Richmond.

USEPA. 2009. Statistical analysis of groundwater monitoring data at RCRA facilities: unified guidance, Office of Solid Waste, EPA 530-R-09-007.