Human and Ecological Risk Assessment, 11: 1127–1137, 2005
ISSN: 1080-7039 print / 1549-7680 online
DOI: 10.1080/10807030500278586

Insider Censoring: Distortion of Data with Nondetects

Dennis R. Helsel
U.S. Geological Survey, Lakewood, Colorado 80225, USA

ABSTRACT

Environmental data often include low-level concentrations below reporting limits. These data may be reported as "<RL," where RL is one of several types of reporting limits. Some values also may be reported as a single number, but flagged with a qualifier (J-values) to indicate a difference in precision as compared to values above the RL. A currently used method for reporting censored environmental data called "insider censoring" produces a strong upward bias, while also distorting the shape of the data distribution. This results in inaccurate estimates of summary statistics and regression coefficients, distorts evaluations of whether data follow a normal distribution, and introduces inaccuracies into risk assessments and models. Insider censoring occurs when data measured as below the detection limit (<DL) are reported as less than the higher quantitation limit (<QL), whereas values between the DL and QL are reported as individual numbers. Three unbiased alternatives to insider censoring are presented so that laboratories and their data users can recognize, and remedy, this problem.

Key Words: detection limit, statistics, censoring, less-than, nondetect.

INTRODUCTION

Environmental data often include low-level concentrations below reporting limits. These data may be reported as "<RL," where RL is one of several types of reporting limits. Some values may also be reported as a single number, but flagged with a qualifier to indicate a difference in precision as compared to values above the RL. Data that include values below (or above) thresholds are called "censored data" in the statistics literature. Methods for statistical analysis of censored environmental data, including computation of summary statistics, hypothesis tests, and regression, were recently presented by Helsel (2005).

A currently used method for reporting censored environmental data, here called "insider censoring," produces a strong upward bias while distorting the shape of the data distribution. Insider censoring uses internal information (the measured

Received 15 October 2004; accepted 14 December 2004.

Address correspondence to Dennis R. Helsel, U.S. Geological Survey, P.O. Box 25046, MS 964, Lakewood, CO 80225, USA. E-mail: dhelsel@usgs.gov

value from the laboratory instrument) to decide which reporting limit is applied to an observation. The resulting (inadvertent) bias produces inaccurate estimates of summary statistics and regression coefficients, distorts evaluations of whether data follow a normal or other distribution, and introduces inaccuracies into risk assessments and models. This is true whether those estimates come from simplistic methods such as substitution of one-half the reporting limit for less-thans, or from more valid and complex methods such as maximum likelihood or regression on order statistics. Three unbiased alternatives to insider censoring are presented so that laboratories and their data users can recognize, and remedy, this problem.

CENSORING AT REPORTING LIMITS

There are currently at least two competing conventions for calculating and naming reporting limits. The convention currently followed by many analytical laboratories in the United States uses methods and terminology developed at the U.S. Environmental Protection Agency (USEPA). The two reporting limits used in this convention are called detection limits and quantitation limits (USEPA 2003). Both limits are calculated by assuming the standard deviation of measurement remains constant from low concentrations down to a level of zero concentration. Data above a detection limit (DL) are considered to have nonzero concentrations; data below are not significantly different from zero. Quantitation limits (QL) are thresholds higher than a detection limit, representing the level at which individual values may first be reliably quantified. Measurements above the QL are assigned individual values. Measurements below the DL are considered nondetects indistinguishable from zero, and are reported with a less-than value. Practices differ among laboratories for what to do with measurements in the region between the two thresholds. Here the chemist generally believes that the analyte is present in the sample, but at concentrations that cannot be quantified with full precision. Some laboratories will report a value of <QL for these in-between measurements. Other labs report single numbers below the quantitation limit, but qualify them with a "remark" (Oblinger-Childress *et al.* 1999; USEPA 1989) indicating that the value reported is "estimated." The letters E or J are common qualifiers, and these data are sometimes called "J-values." Most data users incorporate J-values as if they were equivalent to values above the quantitation limit, ignoring the remark.

The second convention is based on terms presented by Currie (1968) for radiochemical data, and later applied to water or air measurements. The threshold above which observations are considered nonzero is called the "critical level" or "decision limit" rather than detection limit. A second, higher threshold that avoids false nondetections as well as false detections is labeled the detection limit. The concept of a quantitation limit is similar in both conventions. Proponents of this convention have also recently suggested that determination of reporting limits be performed without assuming that the standard deviation of measurement remains constant (Gibbons 1994; Gibbons and Coleman 2001).

As a third alternative, some scientists advocate the numerical reporting of all readings, including those below detection and quantitation limits, along with a measure of their analytical error, that is, 0.3 ± 0.78 (Porter et al. 1988; Currie 1995). Specialized

procedures based on weighting the information content of each measurement by something like the reciprocal of its variance, so that less precise measurements receive lower weight, would need to be employed to interpret data reported in this way. Rocke and Lorenzato (1995) present a model for estimating the variance of low-level analytical measurements, so that the errors of each measurement can be reported and used in subsequent computations.

Insider censoring results from reporting observations as less-thans below an improper threshold value. This could occur if either of the first two conventions were used improperly. It is avoided completely by the third convention, where no measurements are censored. The errors of insider censoring occur most frequently with use of the first convention, so that convention is used as illustration in the remainder of this article.

INSIDER CENSORING

Insider censoring occurs when data measured as <DL are reported as <QL, whereas in-between J-values are reported as individual numbers. In essence, the reporting limit changes based on insider information, that is, the measured value of the observation. For J-values, the reporting limit is the DL. For values measured below the DL, however, the reporting limit is the higher QL. This shift in reporting limit produces a bias, distorting the distribution of data.

Insider censoring confuses the relative ordering of data—their ranks, and so their percentiles. When the relative ordering is distorted, so is all analysis based on that ordering. Essentially every interpretation method applied to censored data uses the relative ordering of data, either in the form of their percentiles (the cumulative distribution function, or cdf) or in the proportions of data falling below each censoring threshold. If laboratories employ insider censoring, and several do, all interpretations made by a data user will be distorted as a consequence. This is true whether the interpretations use a simplistic and incorrect method such as substituting one-half the reporting limit for less-thans, or use a more complex and valid method such as maximum likelihood estimation.

In the United States, the detection limit is most frequently set using a multiple of 3.14, the *t*-statistic having a 99% inclusion rate for a normal distribution with sample size of $n = 7$, so that the DL occurs at a concentration of $3.14s$ (USEPA 2003). Any measurement whose concentration is at the detection limit of $3.14s$ or higher is then considered to have no more than a 1% chance of originating from a true concentration of zero. In truth it is the mean of seven replicate analyses whose true concentration is zero that has no more than a 1% chance of exceeding the DL using this confidence interval approach; for probabilities concerning an individual measurement, a prediction interval should be used instead (Gibbons 1994). However, determining appropriate methods for setting detection limits are outside the scope of this article. What is of concern is the current practice in regards to reporting data below a higher limit, sometimes called the quantitation limit.

Limits higher than a detection limit arise from two perceived needs. The first perceived need has been for a threshold above which reliable single numbers can be reported (USEPA 2003), called the quantitation limit (QL). QLs based on this need

are generally computed as 10 times the standard deviation of a low standard such as the one used to define the detection limit. The factor of 10 is arbitrary, but has been around for a number of years, and a concentration 10 times the background variability is considered large enough by most chemists that a single number might be comfortably reported. The result is a threshold that is a little over 3 times the value of the detection limit $\frac{10s}{3.14s} = 3.18$.

The second perceived need is for a threshold that protects against false negatives. This is the idea behind the "detection limit" of Currie (1968). From a laboratory's point of view, a serious error is made if an individual sample whose true value is at or above the detection limit is reported as being below the detection limit. This is called a "false negative," or Type II error. The argument goes that a true concentration exactly at the detection limit (0.78 in Figure 1) has a 50% chance of being recorded as below the detection limit, and so a 50% chance of erroneously being reported as a <DL. To avoid this error, a higher threshold is instituted. This must be understood in the light that any measurement with a true concentration of X has a 50% chance of being reported as less than X, and a 50% chance of being reported as greater than X, assuming there is no analytical bias (100% recovery). True concentrations of samples are never known, so there is a 50% chance of the measurement being below the true value for every chemical measurement, due simply to random variation. But for the special case of the detection limit, some analysts consider the possibility of a false negative to be a problem. To remedy this problem, a probability distribution is set around an upper threshold (the bell curve at the right in Figure 1), sliding the threshold value upward until there is only a small probability that a measured value will fall below the detection limit (the shaded area). Assuming that the standard deviation of measurements at the higher limit is the same as the value computed using replicates at a low standard, the probability of a false negative is 1% when the

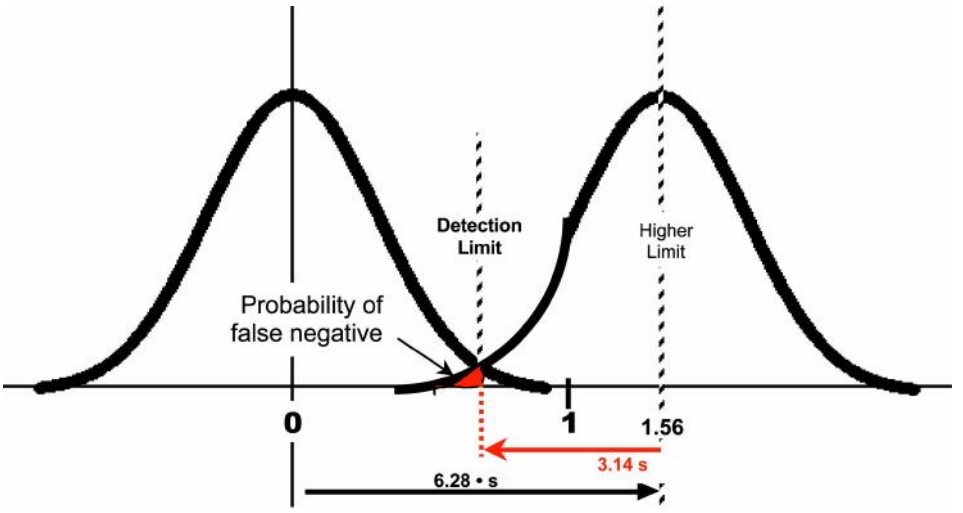


Figure 1. Establishing a higher limit at twice the detection limit, or 6.28 times the standard deviation, to protect against false negatives. The y axis is the probability of observing a value at a given concentration.

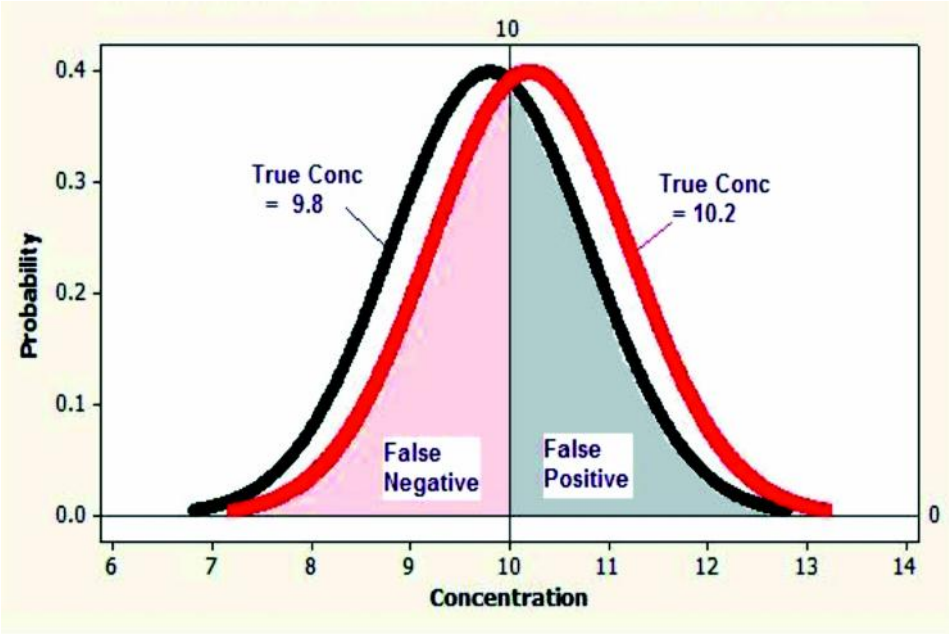


Figure 2. Balanced probabilities of false negatives and positives for true concentrations just below and above a detection limit of 10.

higher limit is double the detection limit, or 6.28·s (Figure 1). Based on avoiding false negatives, the upper limit is often twice the detection limit (USEPA 2003).

However, censoring to avoid false negatives ignores the balancing of measurement errors inherent in a random process. Consider again the concern that a true concentration at the DL will be reported as below the DL. If the true concentration were just a hairs-breadth below the detection limit, there is an almost 50% chance that the measured value will exceed the detection limit, and so be erroneously reported as a detected value. The Type I and Type II errors surrounding each true concentration will tend to balance each other out (Figure 2). Without adjusting for false negatives, the proportion of values falling within each interval of concentration remains correct if each measurement has the same variability. The same percent of values near each boundary will by chance fall into a higher category as those from the higher falling into a lower category. So from a perspective of the overall data set, there is little need to censor data based on an avoidance of false negatives. They are balanced by errors in the opposite direction.

Regardless of whether the establishment of a limit higher than the detection limit is due to a concern for false negatives, or to a concern for precise quantitation, and regardless of the name assigned to that limit, insider censoring can result when the choice of whether to assign the detection or higher (quantitation) limit for any individual sample is based on the value of the analytical measurement inside the laboratory.

Insider censoring, advocated as "Type B" censoring by Ellis (1989), has been adopted by several laboratories including the U.S. Geological Survey

(Oblinger-Childress *et al.* 1999), and is *required* by the RAGS risk assessment guidelines for the Superfund Program of the USEPA (1989). An example from Ellis (1989) with 1 as the DL and 2 as the QL illustrates the process:

Analyst's result:	3.0	0.1	1.3	1.7	−0.4	0.9	2.3
Reported values:	3.0	<2	1.3**J**	1.7**J**	<2	<2	2.3

Measurements between the limits, between 1 and 2, are reported as single numbers. They may have associated qualifiers (J-values) assigned to them, but these are either officially or unofficially ignored when using the data for interpretation. For data between the limits, the censoring limit below which an observation will be called a "less-than" is therefore the detection limit. However, measurements below the detection limit of 1 are assigned a <2 rather than a <1. The limit at which these data are censored is the higher limit of 2. It is this change of censoring thresholds, based on the analyst's measurement information, that produces the bias of insider censoring.

CONSEQUENCES OF INSIDER CENSORING

A set of concentrations as measured by a lab instrument as a bar graph are shown in Figure 3a. Forty percent of observations are measured between 0 and the detection limit (the white bar in Figure 3a). Twenty-five percent are measured between the detection and quantitation limits (light gray bar), and the remaining higher measurements are reported as detected values (dark bars). In Figure 3b are shown the same data after insider censoring. The difference is that values measured below the detection limit are now reported as being below the quantitation limit or "<QL," as if they might belong anywhere from zero up to the quantitation limit. The probability (40%) that observations may fall below the detection limit is spread evenly along the entire range from zero to the quantitation limit. This is pictured in Figure 3b as white bars totaling 40%, evenly split between two categories, 20% of observations in each category. The result of insider censoring is that the probability that an observation might fall between the detection and quantitation limits is exaggerated, and the probability that it would fall below the detection limit is underestimated, in comparison to the proportions actually measured. The shape of the histogram has been changed, and so too will all interpretations that follow.

To see the distorted shape caused by insider censoring, data were generated from a lognormal distribution. These data should fall along the central straight line when plotted on a lognormal probability plot like those in Figure 4. After being generated, the data were censored in two ways. For Figure 4a, all values falling below 2 were censored as a <2, an unbiased censoring at the detection limit. All data above 2 are plotted on the "censored probability plot" (Helsel 2005) of Figure 4a as individual values, with their plotting positions accounting for the proportion of values falling below 2. Their overall shape is seen as lognormal because the data follow a straight-line pattern. In Figure 4b the same data are plotted after insider censoring. All the <2 values were reported as being below the QL, or <4. Individual values between 2 and 4 are reported and plotted along with all data above 4. The positions of plotted data are adjusted for the proportion of data reported as censored at <4. The

resulting censored probability plot in Figure 4b shows a pronounced curvature resulting from the censoring method. Evaluation of the shape of the data distribution is a common first step in deciding whether to use a parametric or nonparametric analysis procedure. Whether a graphical or numerical assessment of data shape is used, insider censoring distorts the shape and makes an accurate assessment impossible. An inaccurate presentation of shape will result in an inaccurate decision of

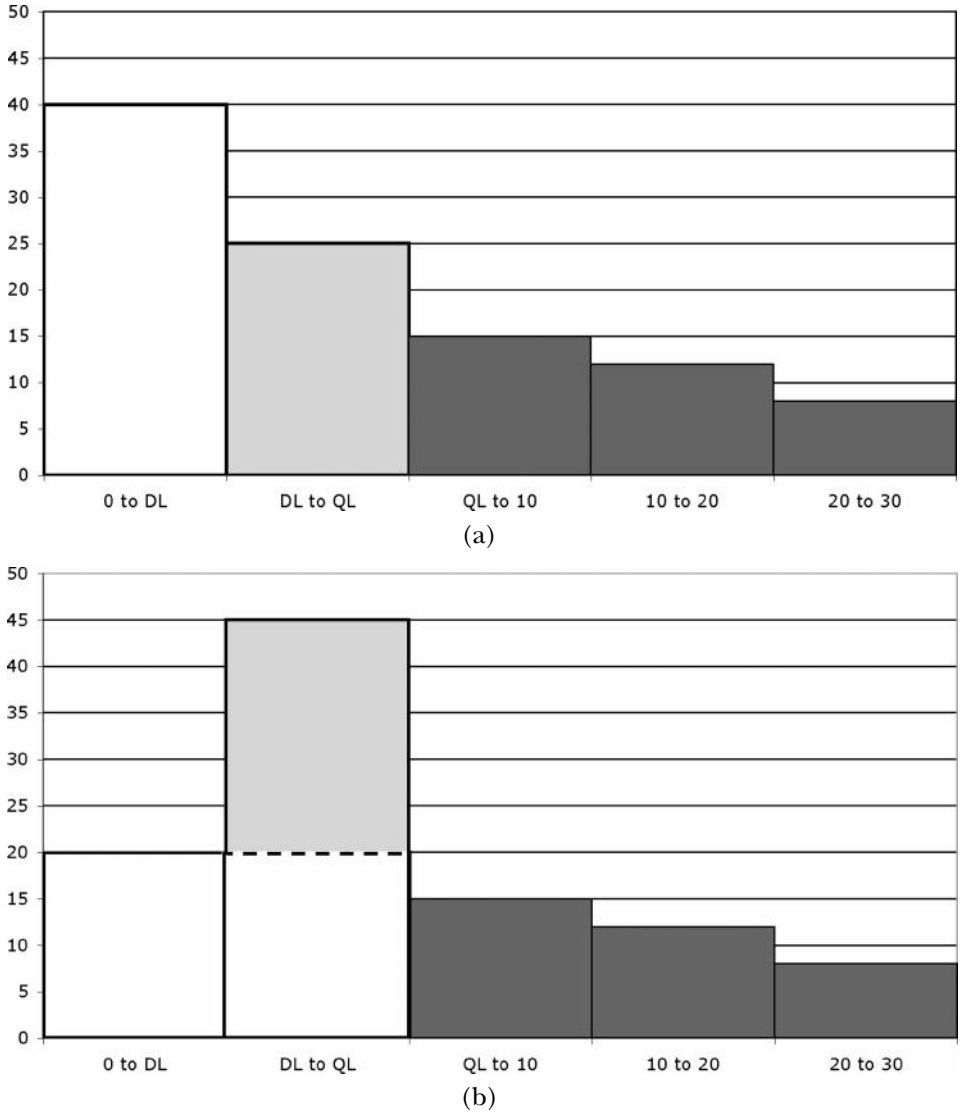
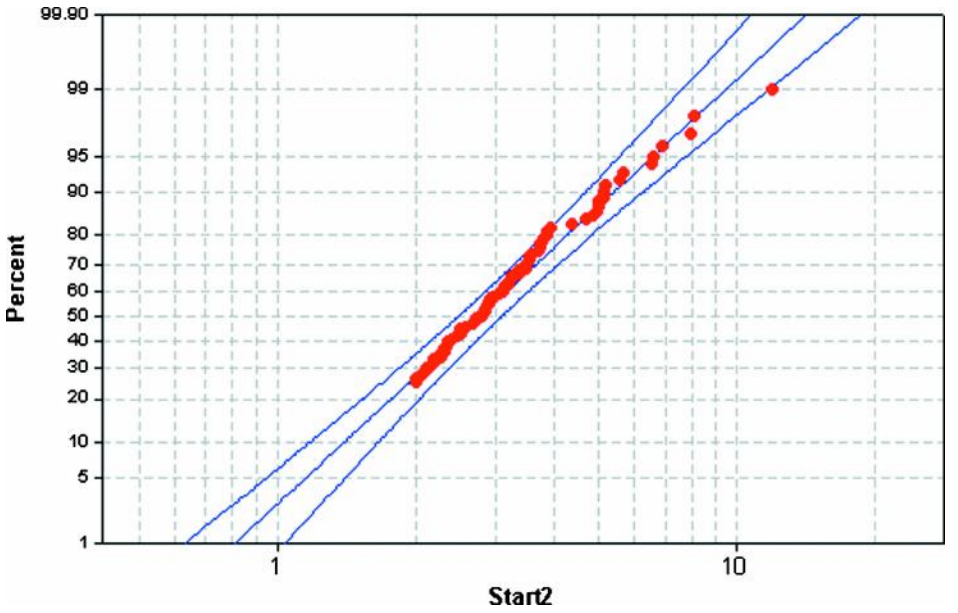
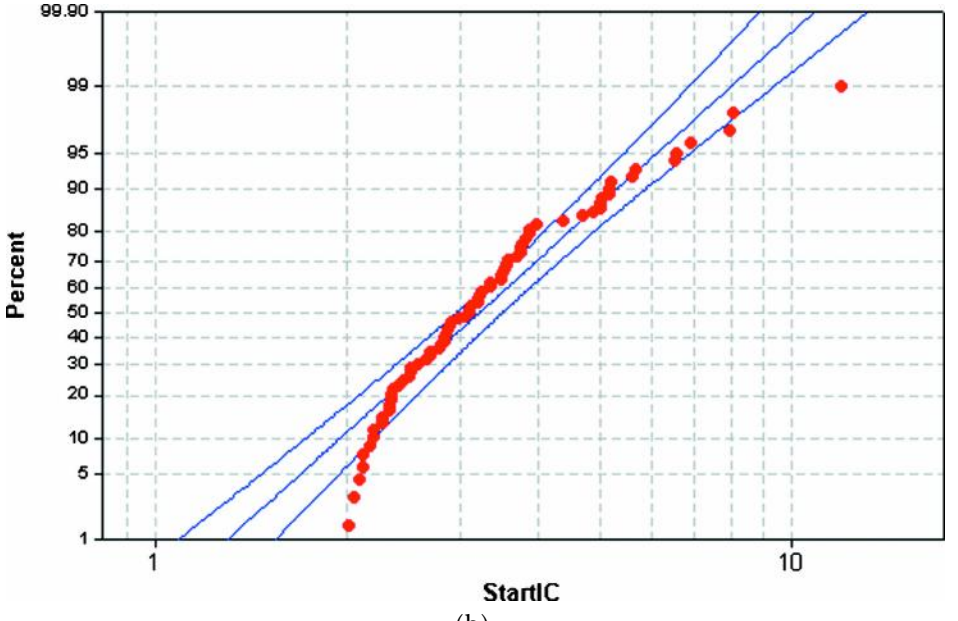


Figure 3. (a) Proportions of data within ranges of concentrations as originally measured. (b) Proportions of the 4a data within the same ranges after insider censoring. The lower end of the distribution has been shifted dramatically upward.



(a)



(b)

Figure 4. Lognormal probability plots of a set of lognormal data. (a) After censoring at DL = 2. Data shape is preserved—the data follow a straight line as expected for lognormal data. (b) After insider censoring. Data shape is distorted, as seen by the curvature on the plot.

whether or not to transform the data set, and an inaccurate judgment of whether a normal-theory procedure can be used.

SOLUTIONS FOR INSIDER CENSORING

The three reporting methods that follow avoid the distortions introduced by insider censoring. Laboratories could adopt any of the three methods in place of insider censoring when reporting data to users. Users with data affected by insider censoring can re-code their data with one of the three methods in order to avoid the hidden bias. The data user might select one of these three methods in consultation with the laboratory scientist.

1. Use the detection limit as the reporting limit. Values below the detection limit are censored as "<DL." J-values between the limits are reported as individual values. This provides the most information to the data user, recognizing J-values as higher than true nondetects.
2. Use the quantitation limit as the reporting limit. All values below the quantitation limit are considered "<QL." J-values between the limits are considered too unreliable to report as single numbers, and are reported as <QL, as are all values measured below the detection limit. This method might be chosen when caution regarding the reliability of data between the limits is warranted.
3. Report data as being interval-censored. This method is unfamiliar to many scientists, but is easily incorporated into methods for censored data analysis. Data below the detection limit are reported as "0 to DL," and data between the limits as "DL to QL." The ordering of data is preserved—the <DL group is considered lower than the in-between group—without assigning single values to observations in either group. See Helsel (2005) for more information about data analysis methods, both parametric and nonparametric, that incorporate interval-censored data.

A Monte Carlo analysis of insider censoring in comparison to the three unbiased methods is presented in Figure 5. One thousand sets of 50 observations were generated from a lognormal distribution with a mean of logarithms = 1 and standard deviation = 0.5. This moderately skewed distribution is similar to the shapes of many water-quality and air-quality variables. The mean of the uncensored data in units of concentration is 3.08. Data were then censored using each of the three unbiased methods, and with insider censoring. The mean of 50 observations was computed using maximum likelihood estimation (MLE) for each of the 1,000 sets for each of the censoring methods. MLE is a parametric method that can incorporate data below multiple detection limits without substituting a number for less-than values (Helsel 2005). The histogram bars show the 1,000 estimated means for each censoring method. The difference between the panels is only in how the uncensored "measurements" were censored and reported. Figure 5a uses unbiased method 1, censoring at a detection limit of 2. All data generated at values below 2 were assigned a <2 prior to MLE, which uses the information in the proportion of values occurring as a <2 when computing estimates. Figure 5b uses unbiased method 2, censoring at a quantitation limit of 4. All data generated below 4 were assigned a <4, and the mean estimated using MLE. Note that there is a greater variability here in estimates

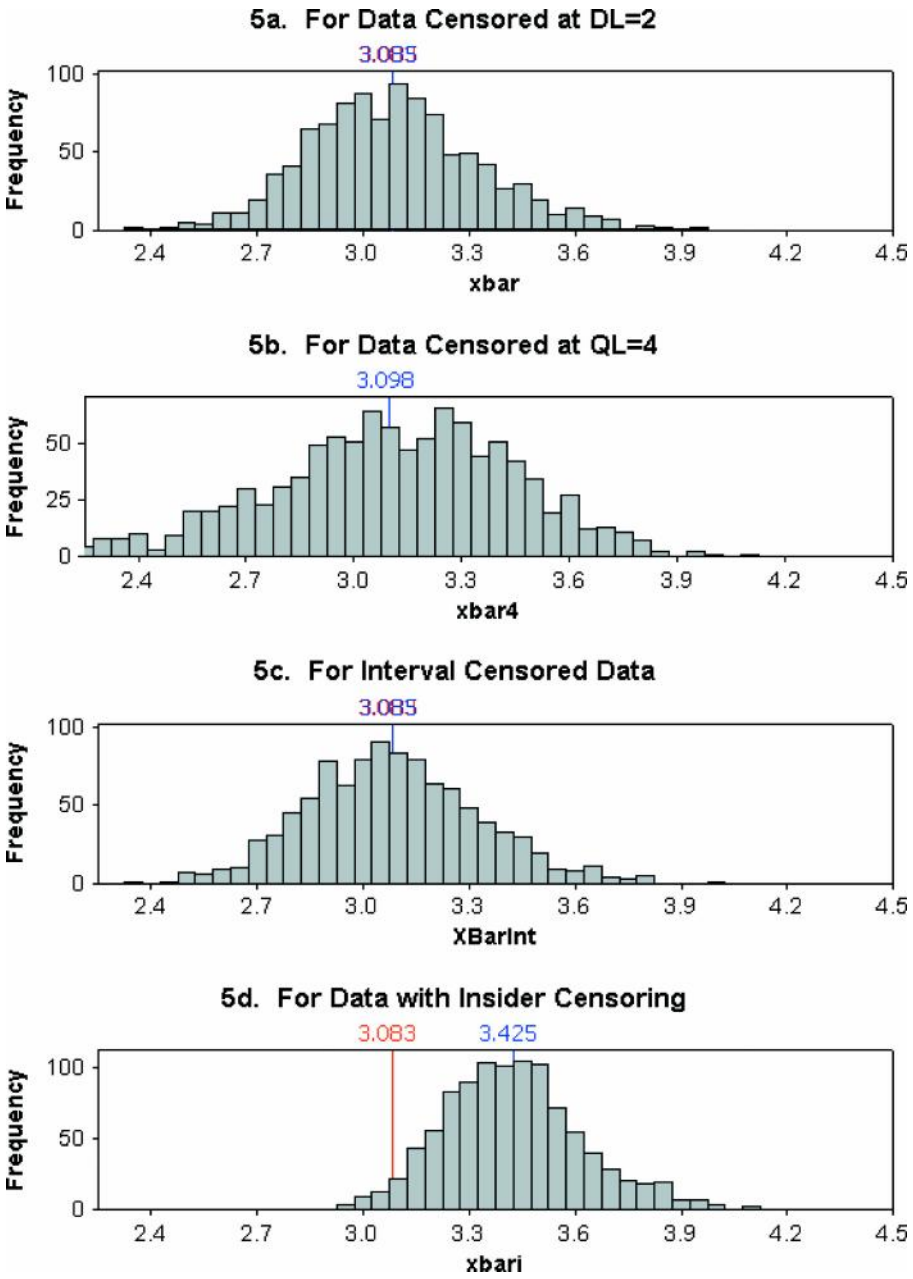


Figure 5. Histograms of 1,000 means for identical data censored by four methods. (a) Censoring at the detection limit of 2. Estimated means cluster around the true value of 3.08. Unbiased. (b) Censoring at the quantitation limit of 4. Unbiased. (c) Interval censoring: Censored data reported as "0 to 2" or "2 to 4." Unbiased. (d) Insider censoring. Data measured below 2 reported as <4. The histogram is offset almost 0.4 units above the true uncensored mean of 3.08. Biased.

of the mean than for method 1, due to the higher censoring threshold. Figure 5c uses unbiased method 3: data measured between 0 and 2 are reported as within that interval, whereas data measured between 2 and 4 are reported to be within that higher interval. MLE incorporates data specified only as within a range to estimate the mean. Note the variability of estimates for this method is similar to method 1, and smaller than method 2.

The overall mean of the 1,000 estimated means for all three methods is 3.08 (3.09 for method 2), showing that the three methods are unbiased. In Figure 5d insider censoring was used—measurements from 0 to 2 were assigned a <4. Data between 2 and 4 retained their individual values (J-values). The bias is visible as an approximate 10% offset of the 1,000 estimated MLE means from the uncensored mean of 3.08.

CONCLUSIONS

The occurrence and consequences of insider censoring are going unrecognized. Until laboratories use an unbiased method of reporting censored values to data users, or report uncensored values for all measurements along with estimates of standard deviation, users should be advised to re-censor their data with one of the three unbiased methods cited here. Otherwise the results of data summaries, determinations of whether data follow a (log) normal distribution, regression equations and hypothesis tests, comparisons of data to standards, and performing risk assessments, all will be distorted by insider censoring.

REFERENCES

Currie LA. 1968. Limits for qualitative detection and quantitative determination. Anal Chem 48:586–93

Currie LA. 1995. Nomenclature in evaluation of analytical methods including detection and quantification capabilities. Pure Appl Chem 67:1699–723

Ellis JC. 1989. Handbook on the Design and Interpretation of Monitoring Programmes. Report NS 29. Water Research Centre, Medmenham, UK

Gibbons RD. 1994. Statistical Methods for Groundwater Monitoring, p 98. Wiley, New York, NY, USA

Gibbons RD and Coleman DE. 2001. Statistical Methods for Detection and Quantification of Environmental Contamination. Wiley, New York, NY, USA

Helsel DR. 2005. Nondetects and Data Analysis: Statistics for Censored Environmental Data. Wiley, New York, NY, USA

Oblinger-Childress CJ, Foreman WT, Connor BF, *et al.* 1999. US Geological Survey Open-File Report 99–193. US Geological Survey, Reston VA, USA

Porter PS, Ward RC, and Bell HF. 1988. The detection limit. Environ Sci Technol 22:856–61

Rocke DM and Lorenzato S. 1995. A two-component model for measurement error in analytical chemistry. Technometrics 37:176–84

USEPA (US Environmental Protection Agency). 1989. Risk Assessment Guidance for Superfund (RAGS), Volume I. Human Health Evaluation Manual (Part A). EPA/540/1-89/002. Office of Solid Waste, Washington, DC, USA

USEPA (US Environmental Protection Agency). 2003. Technical Support Document for the Assessment of Detection and Quantitation Approaches. EPA-821-R-03-005. Washington, DC, USA