

A REVIEW OF STATISTICAL METHODS FOR DATA SETS  
WITH MULTIPLE CENSORING POINTS\*

Richard O. Gilbert  
Staff Scientist  
Battelle Pacific Northwest Laboratories  
Richland, Washington

July 6, 1995

1.0 SUMMARY

RECEIVED  
DEC 17 1997  
OSTI

This report reviews and summarizes recent literature on statistical methods for analyzing data sets that are censored by multiple censoring points. The following references are particularly useful for obtaining an overview of these statistical methods: Akritas et al. (1994), Helsel and Hirsch (1992), Atwood et al. (1991), Helsel (1990), Hertzler et al. (1989), and Millard and Deverel (1988).

This review indicates that for the case of multiple censoring points:

- In general, substitution methods (wherein censored values are replaced by a value such as zero, one half the detection limit (DL), or the detection limit itself) are not recommended when the objective is to estimate the mean and variance of the underlying distribution.
  - Some authors indicate that using one half the DL may be appropriate when testing hypotheses, even though it is not appropriate for estimating the mean and variance
- Robust probability plotting methods are recommended for estimating the mean and variance.
- Maximum likelihood estimation conducted assuming an underlying lognormal distribution is favored for estimating percentiles of a variety of distributions realistic for environmental studies when sample sizes are large ( $n > 25$ ).

\*This report was prepared by Battelle Pacific Northwest Laboratories for the U.S. Environmental Protection Agency, Quality Assurance Division (QAD), Washington, D.C. under EPA Contract 68-D4-0091, Subcontractor Work Assignment BAT-6052-009.

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED

MASTER

19980504 084

## **DISCLAIMER**

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

- Maximum likelihood estimation is recommended for estimating the mean and variance if sample sizes are large and the assumed underlying distribution is correct.

Both robust probability plotting methods and maximum likelihood methods may product biased and imprecise estimates of the mean, variance and percentiles when the assumed underlying distribution does not match the data.

- Nonparametric two-sample scores tests (some of which are extensions of the Wilcoxon Rank Sum test) that were originally developed for right-censored survival analysis data, can be used if the censoring mechanism is the same for both populations.
- Tobit regression may be used for detecting differences between two populations if data are normally distributed.
- Contingency tables and methods based on the binomial distribution are useful if there is a large percentage of censored values, say >50%. However, these methods test for differences in the proportion of the data above the DL in each group. They do not test for differences in means or medians.

Published papers that discuss estimating means, variances, trends, percentiles and prediction limits, as well as papers that discuss conducting regression analyses and testing hypotheses for two independent data sets and paired data are cited here.

This report is organized as follows. Following the introductory comments in Section 2, a brief discussion of detection limits is given in Section 3. Sections 4 and 5 focus on data analysis methods for estimating parameters and testing hypotheses, respectively, when data sets are left censored with multiple censoring points. A list of publications that deal with a variety of other applications for censored data sets is provided in Section 6. Recommendations on future research for developing new or improved tools for statistically analyzing multiple left-censored data sets are provided in Section 7. The list of references is in Section 8.

## 2.0 INTRODUCTION

Analyzing data sets is less than straight forward when some or all of the data are "flagged" as being less than some censoring limit (reporting limit, RL, or detection limit, DL). The problem becomes more complex when the limit is not the same for all measurements, i.e., when there are multiple censoring points in the data set. For example, if 2 measurements in a data set of 5 measurements are reported as having concentrations less than their DL of 5 and 8, respective, then the reported data set might be <5, 6, <8, <8, 9. The publications discussed and referenced here deal with how to statistically analyze this type of censored data set when the goal is to estimate parameters of a distribution (such as the mean, variance, and percentiles) or to test hypotheses (e.g., that residual concentrations following remediation at a site are not greater than those in a background area).

Many data analysis methods have been proposed for singly censored data sets, e.g., for data sets with only 1 censoring point. However, there is currently no EPA guidance on which of those methods should be used. To assist EPA's Quality Assurance Division (QAD) in preparing policy alternatives, this report summarizes the results of a review of the statistical literature conducted to identify state-of-the-art statistical methods for data

sets that have multiple censoring points. Based on this review, a list of research topics that are needed to develop improved methods of dealing with multiply censoring points is provided. Data analysis methods for data sets with a single censoring point will also be discussed as needed since they are the basis for most methods developed for the case of multiple censoring points.

We also take the opportunity here to reference publications that discuss the various types of RLs used. This is done because it is important for the statistician or data analyst to understand the process used to measure samples and decide whether or not a measurement will be censored. Davis (1994, p. 853) states that "one of the more pressing areas where research is needed involves the definitions of 'detection limits' and other limits."

It should be noted that censored environmental data sets almost always fall in the category of Type I left censoring. That is, the number of values less than one or more censoring points is a random variable. Type II censoring occurs when the number of censored values is an *a priori* specified percentage of the data set. Type II censoring is usually not encountered with environmental data sets and hence is not discussed here.

The key publications reviewed in this report that summarize the statistical literature for single or multiple censoring point problems are: Akritas et al. (1994), Helsel and Hirsch (1992), Atwood et al. (1991), Helsel (1990), Hertzler et al. (1989), and Millard and Deverel (1988). Also, Millard and Deverel (1988) and Atwood et al. (1991) discuss some of the causes of multiple detection limits. We note that the material in Helsel and Hirsch (1992) is essentially identical to that in Helsel (1990). Hence, we shall not reference Helsel (1990) further in this report.

### 3.0 DETECTION LIMITS

The definition of the detection limit (DL) is an important consideration in the data analysis of censored data sets. As pointed out by Gibbons (1994, Chapter 5) with regard to monitoring hazardous waste facilities, the detection of a single volatile organic compound in a groundwater sample may be taken as evidence of impact of the facility on environmental quality. This use of DLs makes it imperative that the definition of the DL and the methods used to determine the DL of each sample be carefully considered.

Berthouex and Brown (1994) note that the method detection limit is more a statistical than a chemical concept. They state (page 71):

The term has no scientific meaning until it is operationally defined in terms of a measurement process and a statistical method for analyzing the measurements that are produced. Without a precise statistical definition, one cannot determine a scientifically defensible value for the limit of detection, expect different laboratories to be consistent in how they determine the limit of detection, or be scientifically honest about declaring that a substance has (or has not) been detected. Beyond the statistical definition, there must be a clear set of operational rules for how this measurement error is to be determined in the laboratory. Most published definitions are weak with respect to these instructions, which must explain how to estimate the variances and what kind and number of blanks to be used.

Berthouex and Brown (1994) also state (page 79):

"Different chemists and different laboratories would estimate different MDLs working from the same specimens."

Prior to conducting data analysis of censored data sets, the data analyst should determine the statistical definition of the RL or DL (whichever has been used), the operational rules for estimating them, and the implications of these definition/rules on statistical data analyses and interpretations. This topic is discussed further by, e.g., Davis (1994), Gibbons (1994), Berthouex and Brown (1994), Clark and Whitfield (1994), Lambert et al. (1991), Atwood et al. (1991), Keith (1991, Chapter 10), and Clayton et al. (1987). Henceforth, in this report, DL will indicate the censoring value, however it is defined.

#### 4.0 ESTIMATION

In this section, statistical methods for estimating the mean, standard deviation, and percentiles of underlying population distributions using data from a multiple left-censored data set are provided. It should be noted that the statistical methods discussed are based on the assumption that the samples measured are collected using a probability sampling design and that all measurements are assumed to be independent (uncorrelated) and drawn from the same underlying distribution.

Helsel and Hirsch (1992) consider three classes of data analysis methods for estimating parameters of distributions:

1. Substitution methods
2. Distribution (parametric) methods
3. Robust methods

These three methods are now discussed.

##### Substitution Methods

Substitution methods are those where measurements reported as being less than the DL are replaced by values such as zero, DL/2, or the DL itself. Then standard computational formulas for the mean, standard deviation and percentiles are used on the (now) complete data set. It is possible to use simple substitution methods with either single or multiply censored data sets, but biased estimates of parameters will be obtained.

Several studies have evaluated the performance of simple substitution methods for the single DL case (Gleit 1985; Gilliom and Helsel 1986; Helsel and Gilliom 1986; Newman et al. 1989; El-Shaarawi and Esterby 1992). Helsel and Cohn (1988) consider the case of multiple DLs. Performance criteria typically used are bias, variance and mean squared error. Berthouex and Brown (1994, page 81) illustrate the bias of substitution methods. El-Shaarawi and Esterby (1992) compute explicit values for the expectation of the sample mean and variance (as well as the bias and mean square error) for substitution methods when the underlying distribution is normal or lognormal. They show that the direction and magnitude of the bias depends on the values of the parameters being estimated. They also show that

increasing the sample size does not reduce the bias.

Helsel and Hirsch (1992) conclude that the substitution methods should not be used for estimating distribution parameters. Also, Atwood et al. (1991) do not recommend using these methods because lognormal regression methods (discussed below) are often substantially better and never much worse. However, some authors (Davis 1994; Atwood et al. 1991) indicate that when testing hypothesis about differences between populations, as opposed to estimating parameters, using DL/2 for all less-than values may work well. See Section 5.0 for further discussion of this point.

### Distribution Methods

Distribution methods for handling single or multiple detection limits can be divided into two categories:

- Maximum Likelihood (ML)
- Probability plotting

*ML methods* are based on work by Cohen (1959, 1991). For these methods, the underlying distribution of the measurements must be assumed, which is difficult to do with certainty when the data set is small. ML estimators are known to be biased when the number of samples is small (even when the assumed distribution is correct), although methods for correcting for bias have been presented (El-Shaarawi and Naderi 1991).

For data sets with a single censoring value, El-Shaarawi (1989), El-Shaarawi and Dolan (1989), and Shumway et al (1989) evaluate the use of ML methods for estimating the mean and variance of normal data (or data that can be transformed to normality). Gleit (1985) showed that ML estimators produce estimates that are biased and have poor precision when data sets are small ( $n = 5, 10, 15$ ). El-Shaarawi (1989) shows how to correct the bias of means that are estimated using ML. He also provides methods for computing approximate confidence intervals for the mean and median of the lognormal distribution. Gilliom and Helsel (1986) found that ML methods performed better than probability plotting methods for estimating quantiles. Helsel and Hirsch (1992) indicate that Helsel and Gilliom (1986) and Gilliom and Helsel (1986) found that ML estimates obtained under the assumption of a lognormal distribution produced excellent estimates of percentiles for a variety of data distributions realistic for environmental studies when  $n = 25$  and  $50$ . However, the ML method did not work as well for estimating the mean and standard deviation.

ML methods for data sets with multiple censoring points (multiple DLs) are discussed by El-Shaarawi and Naderi (1991). They give the likelihood equations needed to estimate the mean and standard deviation from Type I multiply censored normal data sets or data sets that can be transformed to normality. Also, they note that ML estimators for the normal distribution can be obtained using many software packages (e.g., SAS and SYSTAT) and the likelihood equations in their article. El-Shaarawi and Naderi (1991) also show how to estimate confidence intervals for the lognormal mean for the multiple DL case when sample sizes are large.

For data sets with a single DL, *Probability plotting methods* for estimating the mean and variance of an assumed distribution are provided in Travis and Land (1990) and Gilbert (1987, page 181). As discussed in Helsel and Hirsch (1992), data both below and above the reporting limit are assumed to follow the same distribution, such as the lognormal. The

mean and standard deviation are estimated by a line fit to the data above the DL. As noted by Akritas et al. (1994, page 229), there are currently no analytical expressions available for computing the variance of estimators produced with probability plotting methods. Simulation-based methods for computing confidence intervals for parameter estimates were proposed by Helsel and Gilliom (1986)

Probability plotting methods for data sets with multiple censoring points fall under the category of "robust methods" and are discussed in the next section. These methods, which were studied in Helsel and Cohn (1988), are extensions of probability plotting methods used for data sets with single detection limits.

### Robust Methods

As stated in Helsel and Hirsch (1992) in their discussion of data sets with a single censoring point (DL), *robust methods* combine data above the DL with below-DL values to compute estimates of summary statistics. The below-DL values are obtained by first using ML or probability plotting to fit a distribution to the data above the DL. The fitted distribution is used to extrapolate a collection of values below the DL. These values are then combined with data above the DL to estimate summary statistics. Akritas et al. (1994, page 225) provide additional description and discussion of this procedure. The probability plotting approach to fitting the distribution is known as *normal probability regression (NR)* when the underlying distribution is assumed to be normal and as *lognormal probability regression (LR)* when the distribution is assumed to be lognormal. The LR method is explained in El-Sharrawi (1989). He also modifies the LR method to incorporate information about the DL. He recommends using ML methods rather than the LR method for making inferences about the mean of a lognormal distribution.

Helsel and Cohn (1988) and Hirsch and Stedinger (1987) modify the LR method for the case of multiple DLs. The method is illustrated in Akritas et al. (1994, page 227). Helsel and Cohn (1988) studied robust probability plotting, substitution, ML, and adjusted ML methods of analysis. They concluded that the robust probability plotting methods are more robust to misspecification of the underlying distribution than the ML and adjusted ML methods, and that the use of these robust methods should be encouraged when the underlying distribution is unknown. They also recommend avoiding substitution methods. They found that the bias-adjusted ML method of Cohn (1988) worked well when the underlying distribution was known.

Helsel and Hirsch (1992) indicate that the most crucial consideration when using distribution methods (ML or probability plotting) is how well the data are expected to fit the assumed distribution.

## 5.0 HYPOTHESIS TESTING

Helsel and Hirsch (1992) present advantages and disadvantages of methods for testing hypotheses when data sets are censored by either a single DL or multiple DLs. The methods can be classified into the same three groups used in Section 4 for estimation: *substitution methods, distribution methods and robust methods*.

Helsel and Hirsch (1992) do not recommend *substitution methods* because 1) no *a priori* arguments for fabrication of any particular value between 0 and the censoring point can be

made, and 2) conclusions of the test strongly depend on the value substituted. However, Davis (1994, page 848) suggests that, when comparing two populations using a statistical test, replacing less-than values by DL/2 in both data sets may be a reasonable way to proceed. McNichols and Davis (1988) provide additional discussion of this point. Also, Atwood et al (1991, pp. B-57 to B-112) concluded on the basis of a small simulation study for a two-factor fixed-effects analysis of variance model and lognormal data that using DL/2 was the method of choice (compared to LR and ML methods). This conclusion was based on the fact that the DL/2 approach was the only method that maintained the correct Type I error of 0.05 under censoring as high as 60%.

*Distribution methods* are also discussed by Helsel and Hirsch (1992). For the two-sample t-test situation they suggest using tobit regression (Judge et al. 1985), which can be used for censored data sets with a single DL or multiple DLs. This method requires using ML estimation assuming a normal distribution. Also, the variance of each group of data must be the same. These assumptions are difficult to verify with small data sets. Cohn (1988) has shown that tobit estimates are slightly biased. He devised methods to correct for the bias.

The *robust methods* discussed by Helsel and Hirsch (1992) are nonparametric tests. For the two-sample comparison situation with a single DL, the Wilcoxon Rank Sum (WRS) test may be used if it is assumed that the less-than values are tied at a value less than the smallest measured (detected) value in the combined data set (USEPA 1994). However, as pointed out by Lambert et al (1991), the assumption that all less-than values are less than the smallest measured value may not be true. Also, even if this assumption is true, the WRS test may not have very much power if less-than values constitute more than about 40% of the total samples. Helsel and Hirsch (1992, page 367) show an example where the WRS test makes efficient use of the information contained in the less-than values.

Some nonparametric tests for the two-sample case, e.g., the *slippage test* (Rosenbaum 1954) and the *Quantile test* (Johnson et al, 1987; USEPA 1994) permit large proportions of the combined data sets to be censored. This advantage is due to these tests using only the largest few measurements in the data sets. The slippage test is based on the number of values in one data set (e.g., the site data set) that exceed the maximum value in the other data set (e.g., the reference site data set). The Quantile test uses only the largest  $r$  data from the two data sets, where  $r$  is typically from 5 to 10. These methods have great utility when there are multiple censoring points if enough data above the DLs can be collected so that the tests have adequate power.

Helsel and Hirsch (1992) indicate that contingency table analysis methods may be appropriate when severe censoring (near 50% or more) is present. Their suggestion springs from the fact that the WRS and other tests will have little power to detect differences in central values when such severe censoring is present. They state "The investigator will be inhibited in stating conclusions about the relative magnitudes of central values, and other characteristics must be compared." As an example, they indicate that contingency table analysis can test for differences in proportions of non-detects in different data sets. They also note that contingency table analyses can be used when 1) data are reported as only detected or not detected (i.e., no numerical less-than values or DLs are reported), and 2) when data can be categorized into three or more groups such as "below detection limits", "detected", below some health standard, or "exceeding standards". Another approach is described in Berthouex and Hau (1991), who give a simple rule based on the binomial distribution for judging compliance using highly censored samples for the case of a single censoring point.

*Nonparametric scores tests* can also be used for the case of multiple DLs. Millard and Deverel (1988) conducted computer simulations to evaluate the performance of several two-sample scores tests that had been originally developed for the analysis of right-censored data sets in survival analysis applications. Helsel and Hirsch (1992) discuss these tests and conclude from the work of Millard and Deverel (1988) that most of these tests were inappropriate for the case of unequal sample sizes. They also stress that these tests require the assumption that the censoring mechanism must be independent of the effect under investigation, i.e., that the probability of obtaining data below each censoring value is the same for all groups, assuming that the null hypothesis of "no difference" is true.

Helsel and Hirsch (1992) note that one robust method which can be used is to censor all data at the highest censoring value present in the data set and then perform an appropriate nonparametric test (e.g., the WRS test) on the new data set, which now has only one censoring point. They indicate that, although information is lost using this procedure, for some situations this is the best that can be done. Moreover, the power will be less than that of scores tests if the censoring mechanism is independent of the effect under investigation.

## 6.0 OTHER APPLICATIONS

Akritas et al. (1994) review the literature for a number of other applications: 1) estimation of the location difference in the 2-sample case, 2) multiple comparisons in the k-sample case, 3) linear regression (parametric and nonparametric methods), 4) time series regression, 5) goodness-of-fit tests, 6) capture-recapture studies, and 7) spatial modeling. The reader is referred to their paper for cited publications in these areas. Also, Verrill and Johnson (1988) extend the Shapiro-Wilk test for normality to the case of a data set with a single censoring value.

The literature review conducted for this report also uncovered the following additional papers that apply to the case of multiple DLs:

- Neerchal and Brunenmeister (1994) develop a new and better approach to trend estimation based on weighted regression.
- Stoline (1991) focuses on using the Box and Cox family of transformations when estimating percentiles of the underlying distribution of a data set with more than 1 censoring value.
- Lee et al (1993) develop a method for comparing two treatments in a matched-pairs design.
- Latta (1981) evaluated the power of several nonparametric scores test
- Liu et al. (1993) extend the class of all two-sample weighted logrank tests to test for ordered alternatives.
- Polissar et al. (1992) developed a method for obtaining plausible intervals for means when all values in the data set are below DLs.
- Helsel and Hirsch (1992) discuss contingency table analysis for measuring the strength of a linear relationship between two variables when both variables are censored.
- Stoline (1993) derived a test for comparing medians in a lognormal two-sample context. His work extends that of Harris (1991).

## 7.0 RESEARCH RECOMMENDATIONS

There is a large literature on statistical methods for censored data sets, particularly for the case of a single DL. This report briefly summarizes the most recent publications. It is useful to consider what additional research and studies might be conducted to provide improved tools for analyzing censored environmental data sets. Some suggestions are offered below:

1. Conduct additional computer simulation studies of methods proposed or developed for estimating parameter and testing hypothesis for the case of multiple DLs. Conditions studied should duplicate real world situations applicable to hazardous waste remediation and cleanup sites (e.g., hot spots, multiple detection limits, subjective selection of sampling locations, etc.)
2. Expand the work of Neerchal and Brunenmeister (1994) by using computer simulation techniques to evaluate performance of their new trend estimation methods.
3. Expand the work of Polissar et al. (1992) on methods for obtaining plausibility intervals for means when all values are below detection limits. Expand their work to include plausibility intervals for percentiles. Computer simulations should be conducted to evaluate performance.
4. Conduct computer simulation studies for evaluating performance of nonparametric methods that incorporate retesting for estimating prediction limits for ground water (work discussed in Gibbons 1994, Davis 1994, and Davis and McNichols 1987, 1994a, 1994b).
5. Conduct research to develop nonparametric scores hypothesis tests that are appropriate for the case of unequal sample sizes. Include studies to develop methods that are appropriate when the assumption of equal probability of obtaining data below each reporting limit for all groups is relaxed.
6. Extend tobit regression methods to the case where data need not be normally distributed with equal variances in each group.
7. Conduct simulation studies to evaluate further if substituting  $DL/2$  for less-than values is an acceptable method when estimating parameters or testing hypotheses. This work would extend that of Atwood et al. (1991, Appendix B).
8. Develop a process for determining for a specified real hazardous waste site when it is acceptable to replace less-than values by  $DL/2$ .
9. Review software codes that handle multiple detection limits to determine what addition codes may be needed.

## 8.0 REFERENCES

- Akritas, M.G., T.F. Ruscitti and G.P. Patil. 1994. "Statistical Analysis of Censored Environmental Data," *Handbook of Statistics 12, Environmental Statistics* (G.P. Patil and C.R. Rao, editors), pp. 221-242, North-Holland, New York.
- Atwood, C.L., L.G. Blackwood, G.A. Harris, and C.A. Loehr. 1991. *Recommended Methods for Statistical Analysis of Data Containing Less-Than-Detectable Measurements*, EGG-SARE-9247, Idaho National Engineering Laboratory, EG&G Idaho, Idaho Falls, ID.
- Berthouex, P.M. and L.C. Brown. 1994. *Statistics for Environmental Engineers*, Lewis Publishers, Boca Raton, FL.
- Berthouex, P.M. and I. Hau. 1991. "A Simple Rule for Judging Compliance Using Highly Censored Samples," *Research Journal of the Water Pollution Control Federation* 63(4):880-886.
- Clark, M.J.R. and P.H. Whitfield. 1994. "Conflicting Perspectives About Detection Limits and About the Censoring of Environmental Data," *Water Resources Bulletin* 30(6):1063-1079.
- Clayton, C.A., J.W. Hines, and P.D. Elkins. 1987. "Detection Limits with Specified Assurance Probabilities," *Analytical Chemistry* 59:2506-2514.
- Cohen, A.C. 1959. "Simplified Estimators for the Normal distribution when Samples are Single Censored or Truncated," *Technometrics* 1(3):217-237.
- Cohen, A.C. 1991. *Truncated and Censored Samples, Theory and Applications*, Marcel Dekker, New York.
- Cohn, T.A. 1988. "Adjusted Maximum Likelihood Estimation of the Moments of Lognormal Populations from Type I Censored Samples," *U.S. Geological Survey Open File Report* 88-350.
- Davis, C.B. 1994. "Environmental Regulatory Statistics," *Handbook of Statistics 12, Environmental Statistics*, (G.P. Patil and C.R. Rao, editors), pp. 817-865, North-Holland, New York.
- Davis, C.B. and R.J. McNichols, 1987. One-sided Intervals for at least  $p$  of  $m$  observations from a normal population on each of  $r$  future occasions. *Technometrics* 29: 359-370.
- Davis, C.B. and R.J. McNichols. 1994a. "Ground Water Monitoring Statistics Update: Part I: Progress Since 1988," *Ground Water Monitoring and Remediation* 14:148-158.
- Davis, C.B. and R.J. McNichols. 1994b. "Ground Water Monitoring Statistics Update: Part II: Nonparametric Prediction Limits," *Ground Water Monitoring and Remediation* 14:159-175.
- El-Shaarawi, A.H. 1989. "Inferences About the Mean from Censored Water Quality Data," *Water Resources Research* 25(4):685-690.
- El-Shaarawi, A.H. and D.M. Dolan. 1989. "Maximum Likelihood Estimation of Water Quality

- Concentrations from Censored Data," *Canadian Journal of Aquatic Science* 46:1033-1039.
- El-Shaarawi, A.H. and S.R. Esterby. 1992. "Replacement of Censored Observations by a Constant: An Evaluation," *Water Research* 26(6):835-844.
- El-Shaarawi, A.H. and A. Naderi. 1991. "Statistical Inference from Multiple Censored Environmental Data," *Environmental Monitoring and Assessment* 17:339-347.
- Gibbons, R.D. 1994. *Statistical Methods for Groundwater Monitoring*. John Wiley, New York.
- Gilliom, R.J. and D.R. Helsel. 1986. "Estimation of Distributional Parameters for Censored Trace Level Water Quality Data," *Water Resources Research* 22(2):135-146.
- Gleit, A. 1985. "Estimation for Small Normal Data Sets with Detection Limits," *Environmental Science and Technology* 19(12): 1201-1206.
- Harris, G.A. 1991. "Two-Sample Comparisons in the Presence of Less-Than-Detectable Data," *Proceedings of the Section on Statistics and the Environment*, American Statistical Association, Alexandria, VA, pp. 197-201.
- Helsel, D.R. 1990. "Less than Obvious," *Environmental Science and Technology* 24(12):1766-1774.
- Helsel, D.R. and T.A. Cohn. 1988. Estimation of Descriptive Statistics for Multiple Censored Water Quality Data. *Water Resources Research* 24:1997-2004.
- Helsel, D.R. and R.J. Gilliom. 1986. "Estimation of Distributional Parameters for Censored Trace Level Water Quality Data: 2. Verification and Applications," *Water Resources Research* 22:147-155
- Helsel, D.R. and R.M. Hirsch. 1992. *Statistical Methods in Water Resources*, Elsevier, New York.
- Hertzler, C.L., C.L. Atwook, and G.A. Harris. 1989. *Current Methods of Handling Less-Than-Detectable Measurements and Detection Limits in Statistical Analysis of Environmental Data*, EGG-SARE-8609, Idaho National Engineering Laboratory, EG&G Idaho, Idaho Falls, ID.
- Hirsch, R. and J. Stedinger. 1987. "Plotting Positions for Historical Floods and their Precision." *Water Resources Research* 23(4):715-727.
- Johnson, R.A., S. Verrill, and D.H. Moore II. 1987 "Two-Sample Rank Test for Detecting Changes that Occur in a Small Proportion of the Treated Population," *Biometrics* 43:641-655.
- Judge, G.G., W.E. Griffiths, R.C. Hill, H. Lutkepohl, and T.C. Lee. 1985. "Qualitative and Limited Dependent Variable Models, Chapter 18 in *The Theory and Practice of Econometrics*, John Wiley, New York.
- Keith, L.H. 1991. *Environmental Sampling and Analysis: A Practical Guide*, Lewis Publishers, Chelsea, MI.

- Lambert, D., B. Peterson, and I. Terpenning. 1991. "Nondetects, Detection Limits, and the Probability of Detection," *Journal of American Statistical Association* 86:266-277.
- Latta, R. 1981. A Monte Carlo Study of Some Two-sample Rank Tests with Censored Data, *Journal of the American Statistical Association* 76:713-719.
- Lee, J.J., I. Ker-Chau, and R.M. Elashoff. 1993. "On Recensoring for Censored Paired Data," *Journal of the American Statistical Association* 88:104-118.
- Liu, P.Y., S. Green, M. Wolf, and J. Crowley. 1993. "Testing Against Ordered Alternatives for Censored Survival Data," *Journal of the American Statistical Association* 88:153-160.
- McNichols, R.J. and C.B. Davis. 1988. "Statistical Issues and Problems in Ground Water Detection Monitoring at Hazardous Waste Facilities," *Ground Water Monitoring Review*, Fall, 135-160.
- Millard, S.P. and S.J. Devereil. 1988. "Nonparametric Statistical Methods for Comparing Two Sites Based on Data with Multiple Nondetect Limits," *Water Resources Research* 24:2087-2098.
- Neerchal, N.K. and S.L. Brunenmeister. 1994. "A New Approach for Accommodation of Below Detection Limit Data in Trend Analysis of Water Quality," *Environmental Statistics, Assessment, and Forecasting* (C.R. Cothorn and N.P. Ross, editors), pp. 113-127, Lewis Publishers, Boca Raton, FL.
- Newman, M.C., P.M. Dixon, B.B. Looney, and J.E. Pinder, III. 1989. "Estimating Mean and Variance for Environmental Samples with Below Detection Limit Observations," *Water Resources Bulletin* 25:905-916.
- Polissar, N.L, T.E. Raghunathan, and S. Liao. 1992. *Plausibility Intervals for Means When All Samples Are Outside Detection Limits*, Report Number 1, The Mountain-Whisper-Light Statistical Consulting, Seattle, Washington.
- Rosenbaum, S. 1954. "Tables for a Nonparametric Test of Location," *Annals of Mathematical Statistics* 24:146-150.
- Shumway, R.H., A.S. Azari, and P. Johnson. 1989. "Estimating Mean Concentrations Under Transformation for Environmental Data With Detection Limits," *Technometrics* 31:347-356.
- Stoline, M.R. 1991. "An Examination of the Lognormal and Box and Cox Family of Transformations in Fitting Environmental Data," *Environmetrics* 2(1):85-106.
- Stoline, M.R. 1993. "Comparison of Two Medians Using A Two-Sample Lognormal Model in Environmental Contexts," *Environmetrics* 4(3):323-339.
- Travis, C. and M.L. Land. 1990. "Estimating the Mean of Data Sets with Nondetectable Values," *Environmental Science and Technology* 24:961-962.

USEPA. 1994. *Statistical Methods for Evaluating the Attainment of Cleanup Standards, Volume 3: Reference-Based Standards for Soils and Solid Media*, EPA 230-R-94-004, U.S. Environmental Protection Agency, Washington, D.C.

Verrill, S. and R.A. Johnson. 1988. "Tables and Large-Sample Distribution Theory for Censored-Data Correlation Statistics for Testing Normality," *Journal of the American Statistical Association* 83:1192-1197.

M98051240



Report Number (14) PNL--10652

\_\_\_\_\_  
\_\_\_\_\_  
\_\_\_\_\_

Publ. Date (11) 19950706

Sponsor Code (18) ~~MA~~ DOE/MA, XF

UC Category (19) UC-905, DOE/ER

DOE