IEc

Treatment of Non-Detects in the Los Alamos National Laboratory Natural Resource Damage Assessment

DOE Contract DE-EM0003939
DOE Task Order DE-DT0011312

Final | June 20, 2017

prepared for:

Los Alamos National Laboratory Natural Resource Damage Assessment Trustee Council

prepared by:

Dr. Shahrokh Rouhani

NewFields
1349 West Peachtree Street
Suite 2000
Atlanta, GA 30309

and

Ms. Alix van Geel

Industrial Economics, Incorporated
2067 Massachusetts Avenue
Cambridge, MA 02140
617/354-0074

(This page intentionally left blank to facilitate double-sided printing)

TABLE OF CONTENTS

LIST OF EXHIBITS

ACRONYMS AND ABBREVIATIONS

ATS	Akritas-Theil-Sen line
CDF	Cumulative distribution function
DAP	Damage assessment plan
DARRP	Damage Assessment, Remediation, and Restoration Program
DL	Detection limit
DOE	Department of Energy
DOI	Department of the Interior
IDL	Instrument detection limit
KM	Kaplan-Meier method
LANL	Los Alamos National Laboratory
LLOQ	Lower limit of quantitation
LOD	Limit of detection
LOQ	Limit of quantitation
MDA	Minimum detectable activity
MDEQ	Michigan Department of Environmental Quality
MDL	Method detection limit
ML	Minimum level
MLE	Maximum likelihood estimation
NOAA	National Oceanic and Atmospheric Administration
ND	Non-detect
NRDA	Natural Resource Damage Assessment
NRDAR	Natural Resource Damage Assessment and Restoration
PAHs	Polycyclic aromatic hydrocarbons
PCBs	Polychlorinated biphenyls
PLE	Product limit estimation
PPW	Paired Prentice-Wilcoxon test
PQL	Practical quantitation limit
QL	Quantitation limit
ROS	Regression on order statistics
USFWS	U.S. Fish and Wildlife Service

EXECUTIVE SUMMARY

The Los Alamos National Laboratory (LANL) Natural Resource Trustee Council (Trustees) is conducting a natural resource damage assessment (NRDA) related to releases of hazardous substances from the LANL facility. All NRDAs require measuring the concentrations of one or more hazardous substances in environmental media and/or biota. NRDA decisions often rely on computations involving environmental concentration data that contain "non-detect" and/or "estimated" results. To help ensure the reliability of NRDA decisions, limitations in these measurements should be clearly understood, and computations should be performed in a manner that is appropriate, given a dataset's particular characteristics.

The major goals of this report are: 1) to review and summarize available information on the meaning, relevance, and significance of various reported limits and relevant approaches for treatment of non-detects in environmental data; and 2) to provide recommendations to the LANL Trustees for the treatment of non-detects and best practices for the LANL NRDA. More specifically, recommendations are presented for several general types of statistical tasks commonly undertaken in NRDAs. These include:

(a) **Summary statistics**, including measures of central tendency such as the mean and confidence intervals around the mean. Summary statistics are used to characterize contaminant concentrations at potentially exposed sites and at reference locations. Also included in this category is the calculation of certain total concentrations (e.g., tPAHs or ∑PCBs in a sample).[1]

(b) **Statistical comparisons.** A typical example would be comparing the concentrations of a hazardous substance in an affected area with the concentrations in a reference area.

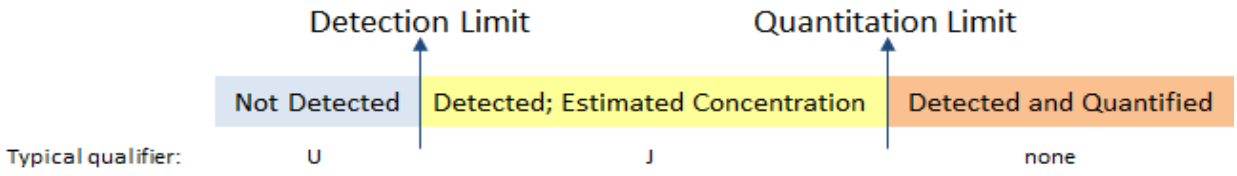
(c) **Correlation and regression analyses.** Examples of these analyses include determining correlations among various contaminants of concern, and evaluating the relationship between exposure of biota to contaminants with one or more measures of effect.

The above types of tasks can be complicated when the analytical chemistry dataset includes non-detect results and estimated values.

[1] PCBs are polychlorinated biphenyls. PAHs are polycyclic aromatic hydrocarbons. Both of these are classes of organic contaminants, which include individual congeners, the concentrations of which are often summed for purposes of consideration of their toxicity or fate and transport. (That said, evaluation of individual congeners or subsets of congeners, is also undertaken.)

Non-detects and estimated values arise in environmental datasets because methods used to measure contaminants are limited in their sensitivity. This sensitivity can be described using two general types of thresholds: the Detection Limit (DL) and the Quantitation Limit (QL). Results that fall below the DL, termed "non-detects" (NDs), are indistinguishable from blank results. Results that fall between the DL and QL are considered to be detected but are not reliably quantified, and the values are considered to be "estimated."[2] Data that fall below the relevant DL or QL are flagged (or "qualified") by the analytical laboratory with a code (e.g., U or J), which allows data users to identify such measurements (Exhibit ES-1).

EXHIBIT ES-1 RELATIONSHIP BETWEEN DETECTION AND QUANTITATION LIMITS, AND ASSOCIATED TYPICAL DATA QUALIFIERS



Trustees are faced with several challenges when analyzing datasets that include qualified measurements. First, it is important to understand the difference between DLs and QLs and the specific types of limits that are represented in a given dataset. Terminology can cause confusion: over the years different authorities have used and developed a variety of surrogate terms and associated definitions for values that can be thought of generally as a DL or QL. This proliferation of definitions has contributed to confusion in the field, and trustees must ensure that they correctly understand the types of limits used and reported in their specific datasets. Second, trustees often must perform calculations with datasets that include qualified data, and the statistical literature is not unanimous in its recommendations on how best to do this, even for relatively straightforward calculations such as determining the mean.

With these challenges in mind, the major components of this report include:

- Reviewing and summarizing available information on detection and quantitation limits, including those used in the LANL *Intellus* database;
- Reviewing how trustees have handled NDs in other NRDAs; and
- Recommending approaches for handling NDs in statistical tasks common in NRDAs.

[2] All analytic chemistry results are subject to a degree of uncertainty, including results that are not flagged by the analytical laboratory. As stated in ITRC (2003), "J-flagged results are those that do not meet the data quality requirements of the analytical laboratories in that they fall below the laboratory's quantitation limit (i.e., the smallest concentration at which analytical results will likely achieve specified or acceptable tolerances for precision and bias)." In other words, J-flags indicate a result's "lack of quantitative reliability" (ITRC 2003).

To begin, we consider the *Intellus* database, which houses environmental data for the LANL site. This database includes a number of fields that have information about detectability. Review of information in these fields suggests that those most important with respect to the issue of detectability are:

- The "Validation Qualifier" field, which includes qualifier codes (e.g., U, J) that identify results below the analytical method's relevant DL or QL;
- For non-radionuclides, the "Report MDL"[3] field, which identifies the sample-specific DL in standardized units;
- For non-radionuclides, the "Report Detection Limit" field, which identifies the sample-specific QL in standardized units;
- For radionuclides, the "Report MDA" field, which refers to the reported minimum detectable activity;[4] and
- The "Report Result" field, which presents the numeric analytical result for the analyte and sample in question.

KEY DEFINITIONS

Non-detect (ND): An analytical result that falls below the relevant detection limit (DL); the concentration of the analyte is said to be "<DL" and is qualified by the laboratory with the relevant flag, such as "U".

Distribution: a listing or function that shows all the possible values of a variable and the relative number of times (probability) that each possible value occurs. Examples of distributions include the normal, log-normal, and gamma distributions, among many others.

Censored data: A data condition in which the value of a measurement or observation is only partially known. If the value is known to fall below a certain level (although the exact value is uncertain), it is described as left-censored. Non-detect analytical results are an example of left-censored data.

Uncensored data: Data that are not subject to censoring; the value is reported.

Several broad categories of approaches are available for working with censored data (see "Key Definitions" text box): substitution methods, parametric methods, and non-parametric methods. Each category and each individual method has advantages and limitations, and the selected method will affect the computational results and may influence the ultimate NRDA injury quantification and damages determination.

Simple substitution methods (i.e., substituting a value such as ½ the detection limit for U-flagged data points) are the easiest to implement, and to-date, NRDA trustees as a group have favored this approach nearly universally. The statistical literature, however, has produced a wider variety of opinions as to the preferred method. Depending on the author, the preferred method can be a function of the distribution of the data (see "Key Definitions" text box) and other considerations.

Although the statistical literature is both complex and not unanimous, a critical review suggests that in many cases, simple substitution is not only the easiest approach to

[3] MDL refers to method detection limit.

[4] In *Intellus*, Report MDA is used to define (flag) radionuclide values as non-detect (i.e., Report MDA is considered a DL). This is in contrast with the Hanford Environmental Information System data dictionary (no such similar data dictionary is currently available for LANL *Intellus*), which defines MDA as equivalent to a QL (USDOE, 2014).

implement but also is frequently reasonable from a technical perspective. The following paragraphs provide more specific guidance based on the current state of the literature and practice with respect to analytically preferred methods for types of analyses commonly undertaken in NRDAs.

Calculating totals (e.g., total PCBs, total PAHs).[5] In calculating the total concentration in a sample of certain contaminants commonly considered jointly (e.g., all PCB congeners, or all PAHs), non-detect measurements of individual analytes should be substituted with zero to avoid the cumulative overestimation of effects.

Radionuclides. In *Intellus*, radionuclide results are uncensored (although some results are flagged as non-detect). The uncensored results may be used directly in calculations without treatment or estimation.

"Estimated" results. Use of uncensored "estimated" results (J-flagged) in NRDA-related statistical computations is recommended. Although these results do not meet the data reliability criteria of analytical laboratories, treating them as censored data is sub-optimal.

Summary statistics (e.g., means and confidence intervals). The recommended treatment method for non-detects in NRDA-related summary statistical computations depends on several factors, including the proportion of non-detects in the dataset, the distribution of detection limits within the dataset, and whether the distribution of the dataset as a whole can be reliably determined (see Exhibit ES-2). Notably, however, we expect that in the large majority of NRDA datasets, the pattern of detection limits will not be random but rather will be clustered towards the low end of the range.[6] In practice, therefore, ½ DL substitution, followed by calculation of the desired summary statistics, will be the preferred approach in most cases.

[5] PCBs are polychlorinated biphenyls. PAHs are polycyclic aromatic hydrocarbons. Both of these are classes of organic contaminants, which include individual congeners, the concentrations of which are often summed for purposes of consideration of their toxicity or fate and transport. (That said, evaluation of individual congeners or subsets of congeners, is also undertaken.)

[6] One exception to this general expectation may occur with certain older datasets with higher DLs, which may be more likely to meet the requirement of having a widespread distribution of detection limits.

EXHIBIT ES-2 RECOMMENDED NON-DETECT TREATMENTS FOR SUMMARY STATISTICS COMPUTATIONS

PROPORTION OF NON-DETECTS	PATTERNS OF DETECTION LIMITS			
	DLs MAINLY OCCUR WITHIN LOW CONCENTRATIONS		RANDOM OCCURRENCE OF DLs	
	Reliable Distribution Assumptions	Unreliable Distribution Assumptions	Reliable Distribution Assumptions	Unreliable Distribution Assumptions
<15%	½ DL substitution			
15-50%	½ DL substitution		ROS or MLE	Kaplan Meier
50-70%	½ DL substitution		Kaplan Meier	
>70%	Use alternative summary statistics, such as the proportion of detects or the proportion of exceedances			

ROS: regression on order statistics
MLE: maximum likelihood estimation

Statistical comparisons. For performing NRDA-related statistical comparisons (e.g., statistically comparing the mean concentration of a contaminant in samples from one area with the mean concentration of that same contaminant in samples from a reference area), the treatments listed in Exhibit ES-3 are recommended. If ½ DL substitution treatment is selected, appropriate parametric or non-parametric comparative tests (DON, 2002, Section 4.2) should be considered after non-detects are substituted, depending on the distributional characteristics of the resulting data. As noted above, because we expect that in the large majority of NRDA datasets, the pattern of detection limits will not be random, in practice, ½ DL substitution followed by an appropriate comparison test will be the preferred approach in most cases.

EXHIBIT ES-3 RECOMMENDED NON-DETECT TREATMENTS FOR STATISTICAL COMPARISONS

PROPORTION OF NON-DETECTS	PATTERNS OF DETECTION LIMITS	
	DLs MAINLY OCCUR WITHIN LOW CONCENTRATIONS	RANDOM OCCURRENCE OF DLs
<15%	½ DL substitution followed by appropriate comparison tests*	
15-50%	½ DL substitution followed by appropriate non-parametric comparison tests**	Generalized Wilcoxon for unpaired samples; Paired Prentice-Wilcoxon for paired samples
>50%	Test of proportions	

* For large, normally-distributed and/or low-variance datasets, parametric tests, such as Student's t test, are appropriate. Otherwise, non-parametric tests are recommended.
** Examples of non-parametric comparison tests include the Slippage, Quantile and Wilcoxon Rank Sum tests.

Correlation and regression analyses. For performing NRDA-related correlation or regression analyses, treatments listed in Exhibit ES-4 are recommended. If ½ DL substitution is selected, appropriate parametric or non-parametric methods should be considered after non-detects are substituted, depending on the distributional characteristics of the resulting datasets. Because we expect that for the large majority of NRDA datasets the pattern of detection limits will not be random, in practice, ½ DL substitution followed by an appropriate correlation and/or regression analysis will be the preferred approach in most cases.

EXHIBIT ES-4 RECOMMENDED NON-DETECT TREATMENTS FOR CORRELATION AND REGRESSION ANALYSES

PROPORTION OF NON-DETECTS	PATTERNS OF DETECTION LIMITS	
	DLs MAINLY OCCUR WITHIN LOW CONCENTRATIONS	RANDOM OCCURRENCE OF DLs
<15%	½ DL substitution followed by appropriate procedures*	
15-80%	½ DL substitution followed by appropriate non-parametric procedures**	Modified Kendall's tau for correlation; Akritas-Theil-Sen line for regression
>80%	Phi coefficient method for correlation; Regression analysis is not recommended	

* For large, normally-distributed and/or low-variance datasets, parametric procedures, such as Pearson's *r* for correlation and linear regression, are appropriate. Otherwise, non-parametric procedures are recommended.
** Examples of non-parametric procedures include Kendall's tau for correlation and Theil-Sen line[7] for regression.

Although the above tables provide guidance on how to handle datasets with non-detects, under certain circumstances, the Trustees may wish to undertake more than one analysis of the data to explore the effects of different methods on the ultimate result. However, before pursuing such sensitivity analyses, the applicability of alternative approaches should be demonstrated: in particular, the underlying assumptions associated with the selected alternative approach should be fully met.

Sensitivity analyses are most likely to be appropriate when multiple approaches are equally likely to be applicable. For example, when the proportion of non-detects is intermediate (i.e., neither very low nor very high), and when a dataset includes non-detects with both high and low detection limits throughout the range of results, the Trustees might consider implementing methods listed in the right-hand portions of Tables ES-2 through ES-4, as applicable, instead of ½ DL substitution. That being said, the Trustees should not feel obligated to conduct such sensitivity analyses, particularly when they have reason to believe that doing so is unlikely to materially affect the result.

[7] Note that the Theil-Sen line is not the same as the Akritas-Theil-Sen line; the latter is a modified version of the Theil-Sen line when some of the data are censored.

Finally, to correctly implement the recommended approaches, it is essential that the correct DL value be utilized, if known. For non-detect results, a common practice is to populate a "result" field with the DL. This practice has not been implemented consistently in *Intellus*: for non-radionuclide non-detect results, instead of populating the result field with a DL, the QL has often been used. To ensure the reliability of LANL NRDA decisions, the DL value, and not the QL value, should be used in calculations.[8] The DL values are available in the "Report MDL" field of the *Intellus* database.

[8] This recommendation is specific to non-radionuclide results in *Intellus*; as discussed above, for radionuclides in *Intellus*, the provided results are uncensored (although some results are flagged as ND). The uncensored results may be used directly in calculations without treatment or estimation. In addition, we recommend that investigators evaluate datasets prior to analysis to confirm, if possible, that U-flagging was conducted by the laboratory and does indeed correspond to the DL, and was not assigned by LANL after receipt of results from the laboratory based on samples being below the reporting limit or QL.

(This page intentionally left blank to facilitate double-sided printing)

CHAPTER 1 | INTRODUCTION

The Los Alamos National Laboratory (LANL) is a U.S. Department of Energy facility located in north-central New Mexico. Scientific research and other activities began at LANL in 1943 with U.S. government efforts to develop and test nuclear weapons. Over the years, its operations have been broadened to include a wide variety of activities. Historical site activities resulted in releases of radiological and other hazardous substances into the environment. Cleanup and decommissioning of contaminated areas were initiated in the 1970s and are expected to continue into the future.

The LANL Natural Resource Trustee Council (Trustees), consisting of representatives from the Department of Energy (DOE), U.S. Department of Agriculture acting through the Forest Service, Pueblo of Jemez, Pueblo de San Ildefonso, Santa Clara Pueblo, and State of New Mexico acting through the Office of the Natural Resources Trustee, are conducting a natural resource damage assessment (NRDA). The goal of the assessment is to replace, restore, rehabilitate, or acquire the equivalent of injured natural resources and resource services lost due to releases of hazardous substances. The LANL Trustees finalized a Damage Assessment Plan (DAP) in February, 2014. This DAP described the Trustees' current understanding of the assessment work necessary to complete the NRDA (LANLTC, 2014). Specifically, the DAP describes assessment activities to identify and quantify injuries to natural resources and the services they provide, and to identify, scale, and cost-out restoration actions necessary to compensate the public for these injuries and lost services.

One of the initial activities outlined in the DAP is to address the treatment of non-detects (NDs) in environmental data and to identify best practices in the context of NRDA. The final work plan, titled *Treatment of Non-Detects in Environmental Data*, was prepared by IEc under DOE Contract DE-EM0003939, DOE Task Order DE-DT0011312, dated September 2016, and describes the approach for implementing this assessment activity.

This report is prepared in accordance with tasks described in the work plan, which enumerates the goals of this work: 1) to review and summarize available information on the meaning, relevance, and significance of various reported limits and relevant approaches for treatment of NDs in environmental data; and 2) to provide recommendations to the LANL Trustees for the treatment of NDs and best practices for the LANL Natural Resource Damage Assessment (NRDA).

The objectives of this work are intended to allow the Trustees to achieve the goals described above, and include:

1. Review and summarize information on the meaning, relevance, and significance of various reported limits;
2. Review information on the treatment of NDs in environmental data;
3. Summarize approaches for the treatment of NDs used in other NRDAs;
4. Identify potentially applicable approaches to the treatment of NDs and any biases these approaches may introduce;
5. Identify representative sample datasets to demonstrate the implementation of applicable approaches to the treatment of NDs and any biases introduced by specific approaches;
6. Identify best practices and provide recommendations for the treatment of NDs in the LANL NRDA; and
7. Provide information necessary for the Trustees to discuss and come to agreement on best practices for the treatment of NDs within the assessment.

1.1 SCOPE OF STUDY

NRDA decisions often rely on computations involving environmental concentration data that contain ND results, i.e., results for which the exact or estimated magnitudes are not known, but which are reported as being below specified sensitivity limits. ND results are an example of what statisticians term "left-censored" data—that is, data for which the result is known to fall below a certain level although the exact value is uncertain.

Several broad categories of approaches are available for working with censored data: substitution methods, parametric methods, and non-parametric methods. The selected method will affect the computational results and may influence the ultimate NRDA decision. This report identifies and evaluates alternative methods to perform calculations with datasets that include NDs, focusing particularly on analytical methods that accomplish the following statistical computations:

- **Summary statistics:** including measures of central tendency such as the mean and confidence intervals around the mean. Summary statistics are used to characterize contaminant concentrations at potentially exposed sites and at reference locations. Also included in this category is the calculation of certain total concentrations (e.g., tPAHs or ∑PCBs in a sample).[9]
- **Statistical comparisons:** A typical example of this type of comparison in a NRDA would be comparing the concentrations of a hazardous substance in an affected area with the concentrations in a reference area.

[9] PCBs are polychlorinated biphenyls. PAHs are polycyclic aromatic hydrocarbons. Both of these are classes of organic contaminants, which include individual congeners, the concentrations of which are often summed for purposes of consideration of their toxicity or fate and transport. (That said, evaluation of individual congeners or subsets of congeners, is also undertaken.)

- **Correlation and regression analyses:** Examples of these analyses include determining correlations among various contaminants of concern, and evaluating the relationship between exposure of biota to contaminants with one or more measures of effect.

The above computations are relevant to assessing injury to natural resources or loss of natural resource services. Although this is not a comprehensive list of statistical computations likely to be performed as part of the LANL NRDA, these computations are among the most commonly used in NRDAs.

1.2 LITERATURE REVIEW

1.2.1 STATISTICAL TREATMENT OF NON-DETECTS

In accordance with the stated goals and objectives of this work, a comprehensive literature review was conducted to identify information about best practices with respect to the treatment of NDs when conducting statistical analyses of the types detailed above. Reviewed documents, as listed in the reference section, were identified as described below:

- **Recommended documents:** The LANL Trustees identified a number of documents to be considered during the review process (e.g., documents associated with EPA's ProUCL program).
- **ProQuest:** ProQuest's *Environmental Science Professional* literature database[10] was searched. For this purpose, the search was restricted to the words "detection limit," "non-detect" or "nondetect" in the title, subject, or descriptor fields of the records, and the word stem "statistic" anywhere in the records. The identified abstracts were then reviewed for relevance. Relevant publications were added to the list of documents to be considered during the review process.
- **EBSCO:** EBSCO's *Environment Complete* literature database[11] was searched. When searching in EBSCO, the term "left-censored" was added into the query described above for ProQuest based on initial review of results. Upon the review of identified abstracts, relevant publications were added to the list of documents to be considered during the review process.
- **Google Scholar:** A separate search was conducted in Google Scholar using the terms "non-detect" or "nondetect" and qualifiers including "environmental" and "statistics", which yielded too many results to be fully reviewed.[12] Top hits, sorted by relevance, were reviewed and relevant articles and publications not covered in the above database searches were added to the list of documents to be considered during the review process.

[10] http://www.proquest.com/products-services/agr_science.html

[11] https://www.ebscohost.com/corporate-research/environment-complete

[12] https://scholar.google.com

In total, more than 100 relevant documents were reviewed, which represented a comprehensive compendium of textbooks devoted to the topic of NDs, regulatory or professional guidance documents related to determination and treatment of NDs, review papers comparing various ND treatment methods, as well as publications devoted to discussions covering various ND issues. Appendix A provides a tabular summary of reviewed documents covering various ND treatment procedures and computational tasks.

1.2.2 NON-DETECTS IN NRDA

During the course of an NRDA, trustees commonly encounter ND results within the environmental datasets with which they are working. Consistent with the goals and objectives of this work, a comprehensive review was conducted to characterize the most common approaches that trustees have adopted in working with ND data.

We used two strategies to identify examples of NDs in NRDAs. First, we searched two publicly-available repositories of NRDA documents: the U.S. Department of the Interior's (DOI) Natural Resource Damage Assessment and Restoration Program's (NRDAR Program) online case map and document library, and the National Oceanic and Atmospheric Administration (NOAA) Damage Assessment, Remediation, and Restoration Program's (DARRP) online collection of case documents.

We employed different search strategies for each dataset, due to differences in the structure of these two repositories. The DOI repository allows for searching by incident type (chemical, mining, oil, or other), document type, and for a specific word within the documents' contents. Initial searching indicated that, if a keyword was not found in any document in the library, the search results would include all documents. At the time the search was conducted, the DOI NRDAR website included 1,154 documents. Searching only with the keyword "non-detect" produced 1,154 results, suggesting that the specific search term was not found. Instead, we limited our search to the word "detect" across all incident types and limited results to the following document types:

- Study Plan
- Study Report
- Preassessment Data Report
- Environmental Assessment
- Environmental Impact Statement
- Journal Links
- Agency Reports
- Restoration and Compensation Determination Plan (RCDP)
- Preassessment Screen (PAS)
- Preliminary Estimate of Damages (PED)
- Assessment Report
- HEA/REA

This search strategy resulted in identification of 153 documents, which we downloaded and reviewed for relevance. Appendix B summarizes key features of these, which include 46 documents and 20 NRDA sites. NOAA's case document repository does not

have a built-in capacity for searching. We therefore conducted a Google search on the domain as follows:

site:casedocuments.darrp.noaa.gov non-detect (21 hits)

and

site:casedocuments.darrp.noaa.gov nondetect (16 hits).

We reviewed all identified documents and summarize key features of the most relevant results in Appendix B.

Not all information generated during the course of an NRDA is, or becomes, public. As a supplemental source of information, we also conducted internal interviews of senior staff at IEc, who jointly have multiple decades' worth of experience in NRDA cases across the United States. These interviews served to confirm the reasonableness of conclusions drawn from public information.

1.3 LANL DATA

Throughout this report, LANL data are used for demonstration purposes. The source of LANL data is a compendium of publically available records made accessible on the *Intellus* New Mexico web portal (http://www.intellusnmdata.com/). Analytical data contained in the portal system are comprised of environmental data provided by LANL and the New Mexico Environment Department (NMED) DOE Oversight Bureau (DOE OB).

Data selected for use in this report include sediment, soil, water and groundwater results reported within the LANL property associated with radionuclide, dioxin/furan, metal as well as gasoline range, semi-volatile and volatile organic parameter groups. Rejected and duplicate records were excluded, resulting in a dataset for demonstration purposes of 352,707 records from 5,830 unique samples. Exhibit 1-1 provides a list of fields in this LANL dataset.

1.4 REPORT STRUCTURE

Consistent with the above goals and objectives, the chapters of the report are organized as follows:

- **Chapter 2:** This chapter reviews and summarizes information on the meaning, relevance, and significance of various reported detection limits. It also identifies and describes the types of ND information in *Intellus*, the environmental database used at LANL.
- **Chapter 3:** This chapter reviews information in the statistical literature on the treatment of NDs in environmental data. As part of this review, the chapter identifies potentially applicable approaches to the treatment of NDs and identifies biases that these approaches may introduce. It then illustrates some of the proposed best practices for handling NDs, using representative sample data queried from *Intellus*.

- **Chapter 4:** This chapter summarizes approaches for the treatment of NDs that have been used in other NRDAs.
- **Chapter 5:** This chapter presents a summary of the report's findings and identifies recommended approaches for handling and statistically analyzing LANL environmental data in the context of the NRDA.

EXHIBIT 1-1 LIST OF FIELDS IN THE SELECTED LANL DATASET

AirNET Background Exclusion Flag	East	Lab Uncertainty	Project No.	Sample Usage Code
Analysis Date	End Depth	Lab Units	QC Batch Sequence No.	Samplers
Analysis Deferred Flag	End Sample Date	Latitude	Quarter	Sampling Company
Analysis Lot ID	End Sample Time	Leachate Lot ID	Raw Lab Result	Sampling Event
Analysis Subcontracted	Error/Uncertainty	Leachate Volume	Released For Analysis Date	Sampling Method
Analysis Suite Code	Excavated Flag	Leachate Volume Units	**Report Detection Limit**	Sampling Plan ID
Analysis Type Code	Expected Value	Leached	**Report Instrument Detection Limit**	SDG
Analytical Group Name	Field Filtered	Leached Date	**Report MDA**	Start Depth
Analytical Method	Field Prep Code	Leaching Method	**Report MDL**	Std Ref Material
Approved By	Field Sample Comments	Location ID	Report Result	Subcontract Lab ID
Background Comparison Class Code	Field Sample ID	Longitude	Report Uncertainty	Task
Basis	Filename	Method Category	Report Units	Time Analyzed
Best Value Flag	Filtered	**Method Detection Limit**	**Reporting Limit Type**	Time Leached
Best Value Status Code	Final Leachate pH	Mod Par List	Result Comments	Time Sampled
Blank Correction Flag	Holding Time Flag	North	Result Record Source ID	**Unadjusted Lab Report Limit**
Calibration Reference ID	Holding Time Release Date	Original Lab Result	Result Type	**Unadjusted MDL**
Chain Of Custody No.	Init Prep Amt	Parameter Category	Retention Time	Uncertainty Type Code
Composite	**Instrument Detection Limit**	Parameter Code	Round No	Unknown Port Flag
Composite Description	Lab Blank ID	Parameter Group Name	Run Number	Use Flag
Composite Field Sample ID	**Lab Detection Limit**	Parameter Name	Rush TAT	Validated By
Composite Type Code	Lab ID	Percent Moisture	Sample Matrix	Validation Qualifier

Confidential Flag	Lab Matrix	Phase	Sample Name	Validation Reason Codes
Data Steward Code	**Lab MDA**	Port Sequence Number	Sample Prep Lot ID	Validation Set Type Code
Date Last Modified	Lab Parameter Name	Prep Amt Units	Sample Purpose	Validation Status Code
Date Sampled	Lab QC Reporting Option	Prep Basis	Sample QC Status Code	Vintage Code
Date Uploaded	Lab Qualifier	Prep Date	Sample Record Source ID	Visual Inspection Flag
Date Validated	Lab Recpt Date	Prep Method	Sample Requestor	Voided Sample Reason Code
Depth Units	Lab Report Date	Prep Time	Sample Retrieval Date	Web Publish Date
Detected	Lab Result	Primary or Confirmatory	Sample Retrieval Time	Work Order No.
Dilution Factor	Lab Sample ID	Program	Sample Type	WTR Source Flow Flag

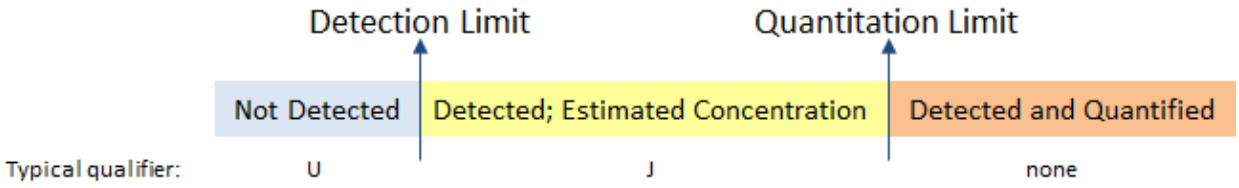
Notes:
Fields are listed in alphabetical order.
Bold and gray shading indicates fields potentially containing information about detection limits.

(This page intentionally left blank to facilitate double-sided printing)

CHAPTER 2 | DETECTION AND QUANTITATION LIMITS

In environmental laboratory analyses, low concentrations are often divided by two delimiters, hereinafter referred to as the detection limit (DL) and the quantitation limit (QL).[13] In a broad sense (and for purposes of this report), DL is defined as the lowest concentration that can be reliably detected and distinguished from the blank sample;[14] while QL is defined as the smallest detectable concentration that can be reliably quantified. A QL is typically much higher than the corresponding DL. Concentrations below DL are reported as ND values, while concentrations that fall between the DL and QL are detected but not reliably quantified, and are therefore considered estimated (Exhibit 2-1).[15] Data that fall below the relevant DL or QL data are flagged, or "qualified", by the analytical laboratory with a code (e.g., U or J),[16] which allows data users to identify such measurements. Only concentrations above the QL are reported as measurements meeting the reliability criteria of analytical laboratories. EPA (2007) provides similar general, "working definitions" of detection and quantitation limits.

EXHIBIT 2-1 RELATIONSHIP BETWEEN DETECTION AND QUANTITATION LIMITS, AND ASSOCIATED TYPICAL DATA QUALIFIERS



[13] DL and QL are analogous to the critical value or L_C and the minimum quantifiable value L_Q, respectively, as defined by the International Union of Pure and Applied Chemistry (IUPAC, 1997). IUPAC also defines a third, in-between delimiter as L_D or the minimum detectable value, which has not been adopted by EPA (2007, page 3). IUPAC (http://goldbook.iupac.org/L03540.html) provides new definitions which are worded differently from the original cited IUPAC (1997).

[14] A blank sample is a sample that is understood to contain no amount of the target analyte being measured and is used to calibrate an analytical method.

[15] All analytic chemistry results are subject to a degree of uncertainty, including results that are not flagged by the analytical laboratory. J-flagged results are additionally uncertain in that they do not meet the laboratory's data quality objectives.

[16] Intellus uses a variety of flags as data qualifiers, including U, UJ, J, J-, J+, and R, among others, and as partially described in online documentation (http://tinyurl.com/hwzvqgb). In the absence of a complete data dictionary, our review indicates that U and UJ represent non-detect results, while J, J-, and J+ are estimated values.

2.1 DETECTION LIMITS

Regulatory uncertainties[17] and professional considerations have led to the introduction of numerous types of limits that fall into the broader conceptual category of a DL (as defined above). Examples of types of DLs used within the environmental community include:

- **LOD (Limit of Detection)** is defined as "the lowest concentration level that can be determined to be statistically different from a blank" (Keith et al., 1983; Long and Winefordner, 1983). The IUPAC Gold Book provides the mathematical equivalent of this definition, defining LOD "as the concentration, c_L … derived from the smallest measure, x_L, that can be detected with reasonable certainty for a given analytical procedure. The value of x_L is given by the equation

  $$x_L = \overline{x_{bi}} + k s_{bi}$$

  where $\overline{x_{bi}}$ is the mean of the blank measures, s_{bi} is the standard deviation of the blank measures, and k is a numerical factor chosen according to the confidence level desired." In other words, the LOD is a more generalized version of an MDL (see below). Other guidance documents (e.g., see references in WDNR, 1996; API, 2002; Armbuster and Pry, 2008) have provided or compiled similar definitions for LOD.

- **MDL (Method Detection Limit)** was originally defined for Clean Water Act programs in 40 C.F.R. Part 136 Appendix B in 1984. In the current (2016) regulations, MDL is defined as "the minimum concentration of a substance that can be measured and reported with 99% confidence that the analyte concentration is greater than zero and is determined from analysis of a sample in a given matrix containing the analyte".[18] MDL is often calculated as 3.14 times the standard deviation of seven low-level spiked blank replicates.[19] Replicates are supposed to be prepared and used by taking into account instrument, matrix (e.g., water,

[17] The procedure for establishing the MDL for Clean Water Act programs was originally promulgated in 40 C.F.R. Part 136 Appendix B in 1984. The MDL procedure was adopted by many other EPA programs and written into many state and federal regulations (Sarver, 2015). As noted in EPA (2007, page 1), in 1999, several industry groups filed suit against EPA (Alliance of Automobile Manufacturers, *et al*. v. EPA, No. 99-1420, (D.C. Cir.)), and in October, 2000, the parties reached a settlement agreement that required EPA to assess procedures to determine Detection and Quantitation Limits under EPA's Clean Water Act (CWA) programs by November 1, 2004. Pursuant to this agreement, on March 12, 2003, EPA issued for public comment a draft report assessing various detection and quantitation procedures and a proposed rule amending EPA's Method Detection Limit (MDL) and Minimum Level (ML) definitions and procedures. Due to the critical nature of comments, EPA decided to withdraw the proposed rule and instead formed a Federal Advisory Committee (EPA, 2007; EPA, 2011). On February 19, 2015, EPA proposed changes to the MDL procedure (Clean Water Act Methods Update Rule for the Analysis of Effluent 2015).

[18] EPA has proposed to update the definition of MDL to "the minimum measured concentration of a substance that can be reported with 99% confidence that the measured concentration is distinguishable from method blank results" (Clean Water Act Methods Update Rule for the Analysis of Effluent 2015).

[19] See Clean Water Act Methods Update Rule for the Analysis of Effluent (2015) for EPA's proposed updates to MDL computations.

tissue, or sediment), and sample preparation variabilities.[20] Based on these definitions, an MDL is an example of a specific type of LOD.

- **IDL (Instrument Detection Limit)** is similar to MDL except it only reflects instrument variabilities. In other words, the low-level spiked replicates used to determine an IDL do not reflect matrix or sample preparation variabilities. In general, IDL values are lower than their corresponding MDL values. Keith et al. (1983) defines IDL as "The smallest signal above background noise that an instrument can detect reliably." For chromatographic methods, IDL may be defined as the product of three variables: (a) Student's t variate with alpha percent significance, (b) the relative standard deviation of the measured signal response areas of low-level spiked (fortified) blank replicates, and (c) mean spike concentration (Wells et al., 2011). EPA Method 1620 describes calculating an IDL "by multiplying by three, the average of the standard deviations obtained on three nonconsecutive days from the analysis of a standard solution… at a concentration 3-5x the instrument manufacturer's suggested IDL, with seven consecutive measurements per day" (EPA, 1989).

The above examples are provided to be illustrative; they do not represent a comprehensive list of alternative DL definitions; other authors have attempted more extensive compilations (e.g., API, 2002; EPA, 2010). Despite differences in specific definitions, one common feature of the above DL definitions is that DL values computed based on large numbers of replicates and multiple instruments under various matrix and sample preparations are bound to be higher than those representing specific conditions with fewer replicates and instruments.

2.2 QUANTITATION LIMITS

While DL surrogates have been defined and prescribed primarily by the regulatory community, determination of QL values or their surrogates are mainly left to data producers (EPA, 2006b). For example, many laboratories use the terms Reporting Limit (RL) or Reported Detection Limit (RDL) as QL surrogates, described as minimum concentrations above which results can be reliably reported.[21] Perhaps the most elaborate QL was proposed by the Federal Advisory Committee (EPA, 2007, Appendix D), according to which, QL was defined as the minimum concentration that meets specified false negative, accuracy (recovery), and precision rates. In practice, commercial environmental laboratories report a QL that represents the sample specific equivalent of the lowest calibration standard or a concentration that is three to five times higher than the DL. QL alternatives include:

[20] For example, early concerns about absence of typical instrument and sample variabilities in EPA proposed MDL computations led the United States Geological Survey (Oblinger Childress et al., 1999) to propose the LTMDL (Long Term Method Detection Limit). This alternative required larger numbers of low-level spiked blank replicates, collected over extended periods of time, while incorporating measurement variability that is typical for routine analyses in a production laboratory, such as multiple instruments, operators, calibrations, and sample preparation events.

[21] For an example, see https://alphalab.com/index.php/support-services/faq-frequently-asked-questions.

- **ML (Minimum Level)** is defined as the level at which the entire analytical system must give a recognizable signal and acceptable calibration point. It is equivalent to the concentration of the lowest calibration standard, assuming that all method-specified sample weights, volumes, and cleanup procedures have been employed (EPA CLP SOW, 2015a; 40 C.F.R. Part 136).
- **LLOQ (Lower Limit of Quantitation)** is defined as the lowest concentration at which the laboratory has demonstrated target analytes can be reliably measured and reported with a specified degree of confidence, which must be at or greater than the lowest point in the calibration curve (EPA CLP SOW, 2015b; EPA, 2003).[22]
- **Limit of Quantitation (LOQ) or PQL (Practical Quantitation Limit)** is the lowest concentration above which quantitative results may be obtained with an acceptable degree of confidence (Keith et al., 1983). As EPA (2006b)[23] states: "[w]hile the IDL is defined by the physics of the moment, and the MDL is defined by the statistical window, the PQL is essentially arbitrary. There are recommendations, PQL = IDL x 10 or MDL x 6 and others... It comes down to what the laboratory feels comfortable signing their name to, confidently, on a daily basis." Some state agencies have developed their own procedures to determine PQL (e.g., CDPHE, 2014).

As for DLs, the above examples of QL surrogates are provided to be illustrative; they do not represent a comprehensive list.

2.3 SOURCES OF CONFUSION

The above definitions are presented to illustrate the variety and multiplicity of DL and QL surrogates in use and are by no means exhaustive. For example, a review of EPA (2010) indicates that within EPA alone there are more than 14 different procedures to define DL and QL, ranging from statistical algorithms to subjective rules. Other examples of limits can be found in the Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) or the Multi-Agency Radiological Laboratory Analytical Protocols Manual (MARLAP).[24] API (2002) provides another compilation of definitions used in state and federal programs.

---

[22] A similar terminology is used for Contract Required Quantitation Limit or CRQL.

[23] https://www.epa.gov/sites/production/files/2015-06/documents/whatthel.pdf

[24] MARSSIM (2000) defines the lower limit of detection (L_D) as "the smallest amount of radiation or radioactivity that statistically yields a net result above the method background. The critical detection level, L_C, is the lower bound of the 95% detection interval defined for L_D and is the level at which there is a 5% chance of calling a background value 'greater than background.' Importantly, the L_D represents an "*a priori* calculation that represents the measurement capabilities of a system" (Consolvo and Sukosky, 2011). In contrast, the MDA (i.e., the smallest activity or concentration that yields a net count above sample background that can be detected with 95 percent probability), represents an *a posteriori* calculation, which "represents the limit for a particular sample count." The MDA changes from sample to sample and is affected by sample size, amount of time the sample is analyzed (counted), and other factors. The MDA is somewhat analogous to an MDL (Consolvo and Sukosky, 2011), and consistent with this, review of the LANL representative dataset indicates that any

The multiplicity of various limits in use has led to occasional confusion and misinterpretation of the reported limits. For example, International Union of Pure and Applied Chemistry (IUPAC, 1997a, Chapter 18, Section 437) states: "Unfortunately, a host of terms have been used within the chemical community to describe detection and quantification capabilities. Perhaps the most widely used is 'detection limit' (or 'limit of detection') as an indicator of the minimum detectable analyte net signal, amount, or concentration. However, because the distinction between the *minimum significant estimated concentration* and the *minimum detectable true concentration* has not been universally appreciated, the same term and numerical value has been applied by some, perhaps unwittingly, in both contexts."

Moreover, under ideal conditions, only concentrations below DL should be reported as ND, while those between DL and QL should be reported as "estimated" and flagged accordingly. Unfortunately, our review indicates that these rules have not been applied consistently in the literature, which can result in misrepresentation of NDs. Some sources of these misrepresentations include:

- **Definition ambiguities:** In some cases, DL and QL are defined clearly. For example, Federal Advisory Committee (EPA, 2011) provides meticulous descriptions for DL and QL based on well-defined measurement quality objectives (MQO) and data quality objectives (DQO). In most other instances, however, the definition of reported DL and QL surrogates are ambiguous or provided without justification. For example, Helsel (2005, page 21) states: "'Reporting Limit' is an intentionally generalized term that represents a variety of thresholds used to censor analytical results. It is a limit above which data values are reported without qualification by the analytical laboratory." Clearly the author is presenting the "reporting limit" as a QL surrogate. Despite this definition, the author proceeds to rely on a contradictory definition in the same page and paragraph by stating "…it is pointless to battle for changing what has become so common place. Therefore outside of this chapter [the author uses] the term 'detection limit' in its most generic sense – as a reporting limit." In other words, the author first defines the reporting limit as a QL but then replaces this QL with the term "detection limit".[25]
- **Semantic ambiguities:** In certain instances, the confusion about the wording of a limit leads to misapplications. For example, many laboratories use the term "Reported Detection Limit" (RDL) as a sample-specific concentration level above which results can be reliably quantified after adjustments made for dilutions or percent moisture.[26] Such a definition clearly describes RDL as a

reported radionuclide results less than the minimum detection activity or MDA is flagged as U. In other words, in the LANL representative dataset, the sole delimiter considered is MDA, which is analogous to a DL. However, unlike other parameters, radionuclide results are provided, even when U flagged.

[25] A subsequent publication, Helsel (2012), revises this approach and uses consistent definitions for DL and QL.

[26] For example, see: https://alphalab.com/index.php/support-services/faq-frequently-asked-questions

sample-specific QL surrogate. However, due to the presence of the term "detection limit" in RDL, some users erroneously treat RDL as a substitute for DL. As a result, the upper limits of NDs are incorrectly assigned to RDL instead of the correct DL value. Examples of such errors are presented later in this report.

- **Regulatory ambiguities:** Literal interpretation of some regulations causes further confusion about results that should be reported as ND. For example, 40 C.F.R. Part 136, Appendix A, 17.6.1.4.1 suggests that users should report results below the minimum level (ML) as "not detected." In other words, a type of limit that is clearly a QL surrogate is being recommended as the delimiter of NDs. Following such a recommendation leads to assigning inappropriately large upper limits to each ND value. Another example is EPA Method 301 (76 FR 28664) which describes "Limit of Detection" or LOD as the lowest level above which quantitative results with an acceptable degree of confidence can be obtained. This description is followed by the statement that "the LOD is defined as three times the standard deviation, S_o, at the blank level." In other words, although the LOD definition matches those of typical DL values, its description points to a QL-like surrogate.

In this work, to the extent possible, we refer to DL and QL to avoid any additional confusion. Recent NRDA studies have pursued procedures to minimize the confusion about DL and QL. For example, a review of publicly available *Deepwater Horizon* NRDA data[27] indicates that each record contains two fields, referred to as "Detection Limit" and "Reporting Limit," corresponding to DL and QL, respectively. NDs are censored, with "Max Result" values set equal to "Detection Limit." These definitions are consistent with their corresponding reported U and J flags.

2.4 LANL CASE STUDY

As noted, above, we downloaded publicly available LANL data from the *Intellus* New Mexico web portal for demonstration purposes. For purposes of illustrating various treatments of NDs, we use more than 350,000 records of measured organic, inorganic and radionuclide concentrations in soil, sediment, and surface and groundwater samples. This large dataset provides a solid foundation to investigate the role of NDs and evaluate various treatment methods. Specific characteristics of the LANL data are explained below.

[27] The representative NRDA dataset was derived from publically available records made accessible on the NOAA *Deepwater Horizon* Natural Resource Damage Assessment Data web portal (rejected samples were excluded, resulting in 6,512 surface sediment benzo(a)pyrene concentration values measured and/or compiled during NRDA investigations that were initiated in 2010). See: https://www.diver.orr.noaa.gov/deepwater-horizon-nrda-data. This approach did not apply to the Response data, however, which are also available through the web portal, but were not collected by the Trustees.

2.4.1 DETECTION AND QUANTITATION LIMITS IN LANL DATA

LANL data reviewed for purposes of this work contain a number of fields that are specifically related to DL and QL, as listed in Exhibit 2-2.[28] For organic and inorganic analytes, the most prevalent limits are paired "Report Detect Limit" and "Lab Detection Limit" fields[29] and paired "Report MDL" and "Method Detection Limit" fields.[30] For radionuclides, the only reported paired limits are "Report MDA" and "Lab MDA."[31]

EXHIBIT 2-2 COUNT (PERCENT) OF RECORDS WITH AN ENTRY IN THE INDICATED FIELD WITHININ B THE LANL EXAMPLE DATASET

FIELDS RELATED TO DL OR QL	NOTES	COUNT AND PROPORTION (%) OF RECORDS WITH NON-NULL ENTRIES					
		SOIL AND SEDIMENTS			SURFACE AND GROUNDWATER		
		INORGANIC	ORGANIC	RADIONUCLIDE	INORGANIC	ORGANIC	RADIONUCLIDE
Report Detection Limit*	Values in these fields are identical after accounting for unit conversions.	2,103 (100%)	3,811 (97.5%)	0 (0%)	34,858 (62.9%)	277,593 (99.7%)	0 (0%)
Lab Detection Limit*		2,103 (100%)	3,811 (97.5%)	0 (0%)	34,858 (62.9%)	277,593 (99.7%)	0 (0%)
Method Detection Limit**	Values in these fields are identical after accounting for unit conversions.	2,103 (100%)	3,899 (99.8%)	0 (0%)	53,219 (96.1%)	278,433 (100%)	0 (0%)
Report MDL**		2,103 (100%)	3,899 (99.8%)	0 (0%)	53,219 (96.1%)	278,433 (100%)	0 (0%)
Report MDA+	Values in these fields are identical after accounting for unit conversions.	0 (0%)	0 (0%)	1,684 (100%)	0 (0%)	0 (0%)	11,192 (100%)
Lab MDA+		0 (0%)	0 (0%)	1,684 (100%)	0 (0%)	0 (0%)	11,192 (100%)
Reporting Limit Type	Most entries are "QL." There are also a lesser	524 (24.8%)	672 (17.2%)	207 (12.3%)	7,784 (14.1%)	28,162 (10.1%)	1,418 (12.7%)

[28] *Intellus* does not include a field indicating whether DLs or QLs have been adjusted for dilution factors. In our experience, standard practice is for reporting adjusted results and detection limits that have taken dilution into account. We cannot determine whether this has in fact been done for all samples in *Intellus*. However, comparison of the listed parameter-specific "Dilution Factors" to their corresponding "Report MDLs" indicates that in many cases these latter entries were increased to reflect dilution factors.

[29] Paired "Report Detection Limit" and "Lab Detection Limit" are either identical or adjusted for unit conversion.

[30] Paired "Report MDL" and "Method Detection Limit" are either identical or adjusted for unit conversion.

[31] "Report MDA," which is used for defining non-detects, is either identical or adjusted (probably to standardize units) based on its paired "Lab MDA."

FIELDS RELATED TO DL OR QL	NOTES	COUNT AND PROPORTION (%) OF RECORDS WITH NON-NULL ENTRIES					
		SOIL AND SEDIMENTS			SURFACE AND GROUNDWATER		
		INORGANIC	ORGANIC	RADIONUCLIDE	INORGANIC	ORGANIC	RADIONUCLIDE
	number of "NA."						
Instrument Detection Limit		0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Report Instrument Detection Limit		0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Unadjusted Lab Report Limit		0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Unadjusted MDL		0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Total Count of Records		2,103	3,907	1,684	55,380	278,441	11,192

Notes: Fields with counts (and percentages) of zero reflect fields present in the Intellus database but that were unpopulated in the example dataset.
* Used as QL.
** Used as DL.
+ Used in *Intellus* for identifying (flagging) NDs.

It appears that the "Report Detection Limit" and "Lab Detection Limit" fields are analogous to a QL, while the "Report MDL" and "Method Detection Limit" fields represent DLs.[32] This interpretation is supported by the following observations:

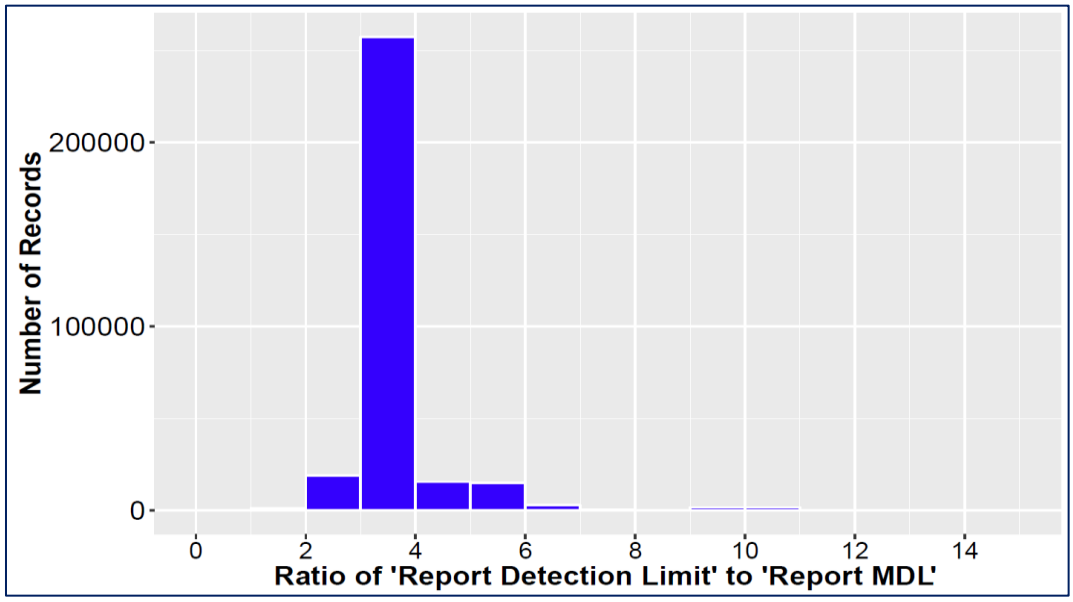
- As the histogram displayed in Exhibit 2-3 indicates, "Report Detection Limits" values for the vast majority of records, similar to common QL definitions, are two to six multiples of "Report MDL" values.
- The example LANL data contain 5,158 J-flagged (estimated) records for which the vast majority of "Report Results" fall between the "Report MDL" and "Report Detection Limit" values.
- Where provided, entries in the "Reporting Limit Type" field state "QL" for the most part (or "NA" otherwise).

Radionuclide records, on the other hand, contain only a single type of delimiter, referred to as the MDA or the minimum detectable activity. The Hanford Environmental Information System data dictionary (no such similar data dictionary is currently available for LANL *Intellus*) defines the MDA as "a sample-dependent estimate, typically dependent on the measured instrument background and sample yield… Generally, it

[32] This uncertainty may be resolved through ongoing consultation with DOE.

depends on the actual aliquot, count time, yield, efficiency, decay correction, and some measurement of background" (Rieger, 2013). Other U.S. DOE documents use a more quantitative definition for MDA, identifying the MDA as the smallest [average] amount (activity or mass) of an analyte in a sample that will be detected with a given probability of nondetection (β; i.e., Type II error) while accepting a given probability of erroneously deciding that a positive (non-zero) quantity of the analyte is present in an appropriate blank sample (α; i.e., Type I error) (USDOE 1998, 2013, both referencing ANSI Standard N13.30). The values for α and β are frequently set to 5% (USDOE 1996, 2014).

EXHIBIT 2-3 HISTOGRAM OF RATIO OF "REPORT DETECTION LIMIT" TO "REPORT MDL"



With respect to interpretation of a result that may fall below the MDA, ANSI Standard N13.30-2011 states that "The *decision level* shall be used to judge whether a net result has a statistically significant difference from the expected background result. The MDA shall *not* be used for this purpose" (emphasis added). Consistent with this standard, the Hanford Analytical Services Quality Assurance Requirements document states: "The question of whether the sample contains net activity is best answered by comparing the measurement result to the *decision level* [also referred to as decision level count rate or DLR] or considering the *confidence interval* for the measurement result, *not* by comparing the result to the estimated MDA or MDC [minimum detectable concentration]" (emphasis added, USDOE 2014).

The "decision level" referred to above is determined based on the selected α level (i.e., the Type 1 error rate) when the distribution represents background conditions. It follows from the mathematical definitions of the "decision level" and the MDA, that the "decision level" will always be lower than the corresponding MDA.

The "decision level" is, therefore, analogous to a detection limit. The MDA is not a detection limit; rather, it represents the average activity level that can be detected at

specified Type I and II error rates (USDOE, 2008b). Of note, and seemingly contrary to this guidance, in the LANL representative database, any radionuclide result less than its corresponding "Report MDA" is flagged as ND (but is reported at its uncensored measured value). We did not identify any field in *Intellus* that seems to correspond with a decision level.

2.4.2 CENSORED NON-DETECT RESULTS

ND results for organic and inorganic analytes are identifiable in the LANL data because they have a "U" flag in the "Validation Qualifier" field. Although these samples represent ND results, they nevertheless have a reported numeric value in the "Report Results" field. In nearly 97 percent of such cases, the value presented in the "Report Results" field corresponds with "Report Detection Limit" field; in the remaining three percent of cases, the "Report Results" value is that of the corresponding "Report MDL" field. As noted, in the LANL database "Report Detection Limit" is analogous to the QL, while "Report MDL" represents the DL. In short, it appears that in the vast majority of records in the LANL database, the upper limits of NDs have been set equal to the QL instead of the DL.[33] This approach is inconsistent with the generally accepted definition of NDs as those measurements that are below their corresponding DL, and thus, is not recommended.

The following example illustrates the effects of reliance on the QL value as the DL. In this example, which uses measurements of four analytes in LANL groundwater, typical NRDA statistics are computed using two common methods for handling NDs: substitution of ½ the detection limit, and Kaplan-Meier (KM).[34] The four analytes illustrated in Exhibit 2-4 were selected solely as examples for the purposes of demonstrating the impact of various methods for treating NDs on sample sets with different proportions of NDs; nothing is implied as to whether these substances are important injury drivers or not.

Although these treatment methods are explored in detail in Chapter 3, the results in Exhibit 2-4 indicates that incorrect assignment of DL can result in substantial over-estimations of both the mean concentrations and the upper confidence limit of the mean, especially in datasets containing a large proportion of NDs.

To ensure the reliability of LANL NRDA decisions, which may be based on datasets containing NDs, correct DL values should be used. Due to both the apparently inconsistent nature of the type of detection limit reflected in the "Reported Results" field, and the frequent use of a QL-like limit instead of a DL-like limit, we recommend that for non-radionuclides, values in the "Report Results" field should not be used for ND

[33] The reason for this is unclear. We recommend that investigators evaluate datasets prior to analysis to confirm, if possible, that U-flagging was conducted by the laboratory and does indeed correspond to the DL, and was not assigned by LANL after receipt of results from the laboratory based on samples being below the reporting limit or QL.

[34] Semi-parametric regression on order statistics (ROS) was also considered. However, the distribution of detected analytes could not be determined.

samples. Instead, the Trustees should base calculations on the value reported in the "Report MDL" field (within which units have been standardized).[35]

If the DL of a ND is missing, every effort must be made to determine this value either by contacting the laboratory or using detection limits for similar samples - ideally samples analyzed during the same period, by the same laboratory for the same analyte. A final determination about data usability should be made using guidance set forth in the Quality Management Plan (i.e., LANLTC, 2014, Appendix B).

EXHIBIT 2-4 EFFECTS OF INCORRECT DL ASSIGNMENTS

ANALYTE (% NON-DETECTS)	NON-DETECT TREATMENT METHOD	MEAN			95% UPPER CONFIDENCE LIMIT (UCL)		
		CURRENT LIMITS (PPB)	CORRECTED DL (PPB)	OVER-ESTIMATION (%)	CURRENT LIMITS (PPB)	CORRECTED DL (PPB)	OVER-ESTIMATION (%)
Vanadium (10% Non-detect)	½ Sub	5.40	5.19	**4%**	5.55	5.37	**3%**
	KM	5.47	5.25	**4%**	5.63	5.41	**4%**
Boron (48% Non-detect)	½ Sub	43.46	34.98	**24%**	48.98	40.57	**21%**
	KM	43.36	38.61	**12%**	49	44.34	**11%**
Arsenic (73% Non-detect)	½ Sub	2.6	1.34	**95%**	2.65	1.38	**92%**
	KM	2.55	1.93	**32%**	2.62	1.96	**33%**
TCE (96% Non-detect)	½ Sub	0.54	0.21	**164%**	0.57	0.23	**147%**
	KM	0.74	0.35	**111%**	0.8	0.38	**111%**

Notes:
Investigated data are groundwater concentrations measured within LANL Property.
The analytes were selected solely as examples for the purposes of demonstrating the impact of various methods for treating non-detects; nothing is implied as to whether these substances are important injury drivers or not.
EPA's ProUCL (2013) is used for computational purposes.
UCLs of 1/2Sub are calculated using nonparametric Percentile Bootstrap UCL.
UCLs of KM are calculated based on nonparametric KM Percentile Bootstrap UCL.

2.4.3 UNCENSORED NON-DETECT RESULTS

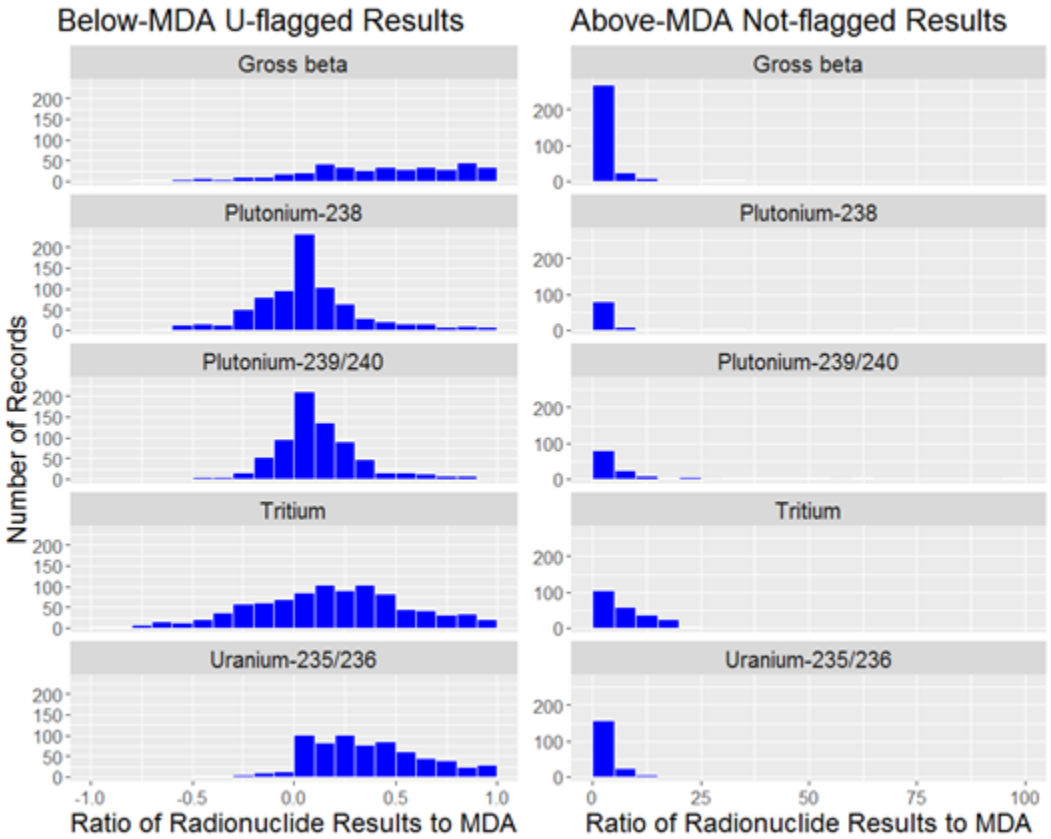
As with ND organic and inorganic results, ND radionuclides are also U-flagged; however, unlike non-radionuclides, the values in the "Reported Results" field are not censored. In other words, "Report Results" of radionuclides reflect the actual measured activities, which are U-flagged when less than their corresponding "Report MDA."[36] Reporting these actual measured activities, instead of censoring the results, is the typical practice for reporting radionuclides.

[35] This recommendation presumes that U-flagged samples are identified by the laboratory, and are truly below the DL, as opposed to being flagged by LANL in cases where laboratories only identify samples as being below the reporting limit or QL.

[36] Other uncensored low-reliability results in the LANL data include inorganic/organic "estimated" J-flagged results that exceed their corresponding "Report MDLs," but fall short of the "Report Detection Limits."

These uncensored U-flagged results offer a unique opportunity to concurrently explore the statistical distributions[37] of U-flagged results. As Exhibit 2-5 displays, radionuclide-specific[38] histograms of U-flagged (below-MDA) and not-flagged (above-MDA) radionuclide results, presented as fractions of their corresponding MDAs, demonstrate a number of distinguishable features:

EXHIBIT 2-5 HISTOGRAMS OF THE RATIO OF THE RESULT TO THE REPORT MDA VALUE, BY RADIONUCLIDE



- Histograms of U-flagged and not-flagged detected values indicate the presence of at least two or more distinct populations. Such histograms are typical of field data that often are mixtures of impacted and background values.
- While the not-flagged detected values display a highly right-skewed distribution, U-flagged (below-MDA) values have a nearly symmetrical distribution. Although these values may reflect measurement error (or "noise"), they also raise questions as to the appropriateness of statistical approaches that assume that the

[37] The distribution of a variable is a listing or function that shows all the possible values of the variable and the relative number of times each possible value occurs.

[38] Displayed radionuclides are those with at least 100 U-flagged and 100 above-MDA measurements.

NDs are simply the lower tail of a distribution characterized by detected values. As noted above, a dataset can contain multiple populations of data (e.g., a mix of background-level and exposed-area samples), each with a distinct distribution. As explained later, many statistical procedures rely on the assumption of NDs being part of a single distribution that also includes detects, and this assumption may not be valid if the dataset contains multiple populations.

- The symmetrical distribution of below-MDA values is far from being the lower tail of a single distribution represented by detected values. As noted by Consolvo and Sukosky (2011), "when a sample has little radioactivity, the analytical results should have a normal distribution of positive and negative results around zero." Such a distribution is a reflection of how radioactivity is measured, i.e., relative to the measured background.[39]

The above characteristics are contrary to treatment methods that favor assigning the full DL or QL to NDs.[40] They also illustrate a potential problem with statistical treatment methods that assume that the investigated data are always derived from a single population, the lower tail from which encompasses all the NDs. These statistical methods, which are examined in more details in the literature review section of this report, share a preconceived basis that NDs, belonging to the lower tail of a single distribution, are dominated by values close to DL. If NDs are in fact part of a separate distribution, this mischaracterization would lead to erroneous (biased-high) results.

[39] "When a sample result is subtracted from background and the sample value is less than that background, the result is a negative value. A negative result simply indicates that the radionuclide activity in the sample is low – so low that it approaches that of the analytical instrument's system background" (Consolvo and Sukosky, 2011).

[40] In this particular case, the use of uncensored non-detect values is superior to any form of substitution or statistical treatment of non-detects.

(This page intentionally left blank to facilitate double-sided printing)

CHAPTER 3 | APPROACHES FOR TREATMENT OF NON-DETECTS

The following sections summarize the findings of our literature review with regard to proposed or recommended treatments of NDs during various statistical computations. For each statistical task (e.g., generating summary statistics), we consider substitution methods, parametric methods, and non-parametric methods. For each, we identify relevant advantages and disadvantages such as:

- Ease of implementation;
- Degree and direction of bias with respect to the resulting parameter (i.e., over/underestimation);
- Need for *a priori* assumptions or knowledge (e.g., data distribution);
- Appropriateness for datasets with certain distributions;
- Flexibility in terms of being able to handle datasets that have different DLs; and
- Appropriateness for datasets with different proportions of ND values.

Over the years, many authors have considered the question of how to treat NDs in environmental and other datasets. The literature is not unanimous with respect to preferred approaches. Some of the differences in conclusions can be attributed to differences in the datasets with which the authors apply their statistical techniques: a technique that may work best in some circumstances may be sub-optimal in others. In addition, some studies are limited with respect to the methods examined, the type of distributions tested, the range of sample sizes tested, or other factors (Hewett and Ganser, 2007). In addition, studies' conclusions can be affected by their reliance on actual vs. simulated data. Studies that use simulated data have the advantage of being able to examine the performance of statistical methods across many hypothetical datasets (e.g., Hewett and Ganser, 2007); however, because the simulated datasets are drawn from one or more specific distribution(s) representing single populations, it is unsurprising if certain methods may perform better because they are a better match to the distribution(s) from which the random data are derived, and the results should be interpreted cautiously. Such analyses are especially inapplicable to typical field data that often represent mixtures of impacted and background populations. In short, care is necessary to understand the specific circumstances and context of each study, and how those may have influenced the study's results.

3.1 SUMMARY STATISTICS

Summary statistics are commonly calculated in NRDAs. Summary statistics are used to characterize contaminant concentrations at potentially exposed sites and at reference locations, and commonly involve the computation of mean concentrations and their upper confidence limits. One common reason for calculating summary statistics in an NRDA occurs during injury determination: for example, trustees may wish to compare the average concentration of a set of samples with a fixed injury threshold, such as a water quality standard.

When computing summary statistics, a variety of ND treatments are discussed in the statistical literature. Our review indicates that the most commonly discussed, proposed, or recommended methods are: (a) substitution or imputation (followed by appropriate computations to produce the desired statistics), (b) maximum likelihood estimation (MLE), (c) regression on order statistics (ROS), and (d) the Kaplan Meier method (KM). As discussed below, other related methods are also considered in this review.

3.1.1 SUBSTITUTION OR IMPUTATION

Substitution or imputation methods assign surrogate numerical values to NDs that are then treated as equivalent to detected values in statistical analyses. Typical surrogate values include 0, various fractions of the DL, the full DL or randomly assigned values between 0 and the DL. The most common form of imputation is ½ substitution, but there are also other proposed fractions such as $\frac{1}{\sqrt{2}}$ (Hornung and Reed, 1990; Antweiler and Taylor, 2015).

Substitution methods are often based on speculations or educated guesses about the shape of probability distribution functions of the censored data. For example:

- Zero substitution implies that all NDs are indistinguishable from blank concentrations and thus should be replaced with zero.
- ½ DL substitution assumes that the distribution of NDs are symmetric between 0 and DL. Under this assumption, the mean value of NDs is $\frac{DL}{2}$.
- In full substitution, all NDs are assumed to be at or close to the DL. This method is underlain by an assumption that NDs are dominated by values close to DL.
- $\frac{1}{\sqrt{2}}$ DL substitution is based on the general assumption that the distribution of NDs can be approximated by a right-triangular distribution with DL as its mode and zero as its lower limit. Under this assumption, the mean value of NDs is $\frac{2DL}{3}$, so a $2/3$ substitution should be used. However, Hornung and Reed (1990) recommend using the median value of NDs, which is $\frac{DL}{\sqrt{2}}$.. The assumed triangular distribution is primarily based on a preconceived notion that NDs are the lower tail of a single distribution encompassing both detected and non-detected values. This notion, however, may not be applicable to typical field data, as demonstrated by the LANL radionuclide results.

- Floit et al. (1996) note a substitution rule equal to $\frac{QL}{5}$ based on practical considerations.
- Randomly assigned values, also known as the bounding method (Sinha et al., 2006) is a flexible substitution method in that the form of the probability distribution function of ND values can be specified. For example, using the Excel random number generator function, RANDBETWEEN(), substituted values with a uniform distribution between 0 and DL can be assigned. The review of various methods by Sinha et al. (2006) for calculating the UCL, using simulated synthetic log-normal data, indicates that the bounding method produces reliable results for datasets with 51-81% NDs.
- β-substitution involves replacing NDs with DL x β, in which β is calculated by assuming a specific distribution for the detected values (Huynh et al., 2014). This approach specifically assumes that the investigated data, including NDs, derive from a single population. This assumption may not be applicable to field data that are mixtures of impacted and background values.

The advantages of substitution include:

- Most substitution methods are easy to understand and implement and in some cases produce results that are similar to certain more complex methods (e.g., Zoffoli et al., 2013).
- Simple substitution methods have the advantage of mathematical parsimony when compared to those that are more complicated or require specific distributional assumptions (Gauch, 2003).
- These methods can be applied to datasets with either one or multiple detection limits.
- Substitutions based on the central tendencies of ND values, such as ½ DL substitution, generally produce reasonably unbiased estimates of the mean. For example, Hornung and Reed (1990), She (1997), and Antweiler and Taylor (2008) found ½ DL substitution to perform acceptably for purposes of calculating the mean. Hewett and Ganser (2007) found that for datasets that follow a "contaminated lognormal" distribution, ½ DL substitution produced estimates for the mean with a bias ranging from -0.1% to -1.5% (underestimating).[41]

The disadvantages of substitution include:

- Some authors consider substitution procedures to be overly simplified. For example, Hewett and Ganser (2007) state: "past investigators have often rejected

[41] The authors created "contaminated lognormal" distributions by combining two lognormal distributions. In tests of datasets using single lognormal distributions, the bias of the mean ranged from -1.4% to -1.9% (underestimating), except in one set of circumstances in which the bias was 4.2% (overestimating). Those sets of simulated data represented a single lognormal distribution with a single laboratory and associated LOD, included between 1% and 50% non-detects, and had sample sizes between 20 and 100.

the LOD/2 substitution method, even though the method had similar or lower rMSE [root mean square error] values than their preferred method, on the basis that it had no theoretical basis." These authors specifically cite She (1997) who "… found that the LOD/2 substitution method often outperformed the KM method, but recommended the KM method over the substitution method because the LOD/2 method 'has no statistical theoretical basis.'" Such biased preferences are inconsistent with the principle of mathematical parsimony (Gauch, 2003).

- Zero and full substitutions produce biased-low and biased-high mean values, respectively. Other substitutions may generate biased mean values if actual ND distributions are inconsistent with their underlying assumptions. For example, ½ substitution results are biased if NDs are not distributed symmetrically between 0 and DL. In general, the extent of the bias is a function of the proportion of NDs in the dataset, and the degree to which detection limits are or are not elevated (Hewett and Ganser, 2007).
- In most cases, substitution methods produce biased-low estimates of population variance (Hewett and Ganser, 2007). Such results can affect the reliability of certain statistics, including the estimates of the upper and lower confidence limits that rely on variance. For example, in the Hewett and Ganser (2007) data simulation study, assuming that investigated data are derived from single populations, ½ DL substitution produced bias in values for the 95th percentile that ranged from -21.2% (underestimating) to 6.0% (overestimating).[42] ITRC (2013) also states that using ½ DL substitution produces estimates of UCL that cover less than 95 percent of the results.

In 2006, EPA guidance supported the use of 0, ½ DL, or DL substitution in datasets with less than 15% NDs (EPA, 2006a). EPA's 2015 ProUCL technical guide, however, provides different recommendations. While acknowledging the former recommendation for substitution methods under certain circumstances, this guide states that ½ DL substitution method has been retained only for historical and comparative purposes, and "its use is **not recommended** due to its poor performance" (emphasis original), even if the percentage of NDs is as low as five percent to 10 percent. Only if the proportion of censored data is less than five percent and if the data are mildly skewed, should substitution with ½ DL be utilized (EPA 2015c, page 128). The ProUCL statement stems from the fact that this software is designed for datasets that are derived from single populations (EPA, 2015c). Under such circumstances, NDs represent the lower tail of the distribution, yielding a highly asymmetric distribution between 0 and DL, and in these instances, simple substitution would perform poorly. However, typical field datasets often represent mixtures of impacted and non-impacted populations.[43] For such datasets, *a priori* assumption of distributional asymmetry of NDs is unwarranted. That this

[42] The discussed results are for various sets of "contaminated lognormal" distribution tests.

[43] This can also occur, for example, with datasets of concentrations of contaminants that cover spatial areas containing multiple hot-spots and less contaminated areas, as is common in NRDAs.

assumption is unwarranted is supported by the reasonable performance of ½ DL substitution as reported by many past investigators (Hornung and Reed, 1990; She, 1997; Antweiler and Taylor 2008).

In their extensive comparative evaluation of the performance of several methods for analyzing simulated censored datasets, Hewett and Ganser (2007) recognize the importance of ease of calculation/accessibility in dealing with datasets that include NDs, indicating that "when dealing with large datasets… [substitution methods] are certainly expedient and may be reasonably accurate, as was suggested by Hornung and Reed (1990), depending on the true (but unknown) underlying GSD [geometric standard deviation] and percent censored."

3.1.2 MAXIMUM LIKELIHOOD ESTIMATION (MLE)

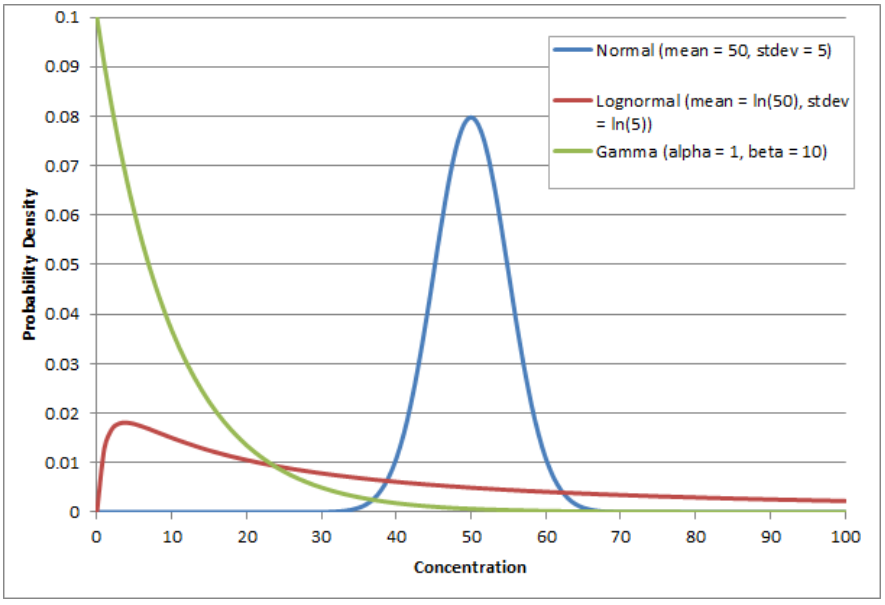
KEY CONCEPTS

Parametric statistics assume that the investigated variables follow a specific distribution.

Nonparametric statistics do not assume that the investigated variables follow a specific distribution.

Maximum likelihood estimation refers to a family of parametric methods that, in essence, estimate parameters of assumed distributions (see "Key Concepts" textbox and Exhibit 3-1) by maximizing the likelihood of the occurrence of the actual detected and ND values. As Helsel (2005, page 13) states, "MLE uses three pieces of information to perform computations: a) numerical values above detection limits, b) the proportion of data below each detection limit, and c) the mathematical formula of an assumed distribution…Parameters are computed that best match a fitted distribution to the observed values above each detection limit and to the percentage of data below each limit." The MLE-estimated distribution parameters are then used to calculate the summary statistics of the investigated data.

EXHIBIT 3-1 EXAMPLES OF DATA DISTRIBUTIONS: NORMAL, LOGNORMAL, AND GAMMA



The advantages of MLE include:

- MLE can be used under a variety of assumed symmetric and asymmetric distributions. The vast majority of MLE applications in the literature have been developed for normal or log-normal distributions, which in some instances may reasonably match observed distribution of NRDA datasets (Akritas et al., 1994; Nysen et al., 2015). Recent articles offer MLE solutions based on new classes of mixed-distributions,[44] which can be more representative of typical environmental datasets (Li et al., 2013).
- In their comparative evaluation of the performance of several methods for analyzing censored simulated datasets, Hewett and Ganser (2007) found "With the exception of the MLE_{mpv} method, the MLE-based methods performed well in the single distribution scenarios and were generally fairly robust in the multiple LOD and contaminated [*log-normal*] distribution scenarios".
- MLE can be applied to datasets with either one or multiple detection limits (Hewett and Ganser 2007; ITRC 2013).
- MLE does not require maximum or minimum values to be detected.[45]
- MLE is especially applicable to cases where the distribution of datasets can be reliably determined or assumed (Helsel, 2005; ITRC, 2013).

The disadvantages of MLE include:

- The primary cases where MLE is applicable are those in which the sample distribution can be reliably determined (Helsel, 2005; ITRC, 2013). These are often datasets, deriving from single populations, with larger sizes and/or a small proportion of NDs because a relatively large number of detects are required to reasonably ascertain the dataset's distribution. In datasets with multiple detection limits, sample sizes of 50 or above and a detection frequency of over 50 percent may be needed (ITRC, 2013). If the type of underlying distribution is incorrectly assumed or cannot be identified, the resulting MLE estimates could be misleading.
- Traditional MLE approaches are sensitive to outliers, the presence of which can result in erroneous parameter estimates (ITRC, 2013; EPA 2015c).[46]

EPA guidance supports the use of an MLE-based method when the proportion of NDs is no more than 50 percent, if the data without the ND values are normally distributed and if there is a single detection limit (EPA, 2006a). As noted above, ITRC (2013) similarly states "the sample size must be large enough to assess the best-fitting underlying

[44] Instead of modeling the data using a single distribution, the authors suggest modeling the data using a mixture of normal distributions to account for subpopulations with different distributions.

[45] As discussed later, Kaplan-Meier method requires the maximum and minimum values to be detected (uncensored) values.

[46] Khokan et al. (2013) propose an approach to mitigate the sensitivity of MLE approaches to outliers.

distribution." DON (2002) supports using the Cohen method (an MLE method) for calculating summary statistics if 15 to 50 percent are NDs; at higher frequencies "the loss of information is too great for descriptive statistics to provide much insight into the location and shape of the underlying distribution of measurements."

The most recent versions of EPA's ProUCL program have excluded parametric MLE methods, which EPA (2015c) describes as "poor performing", likely due to difficulties in verifying the distribution of left-censored datasets with multiple detection limits. In contrast, Hewett and Ganser (2007) found MLE methods in general to be strong performers when calculating the mean and 95th percentile values. Such a conclusion is not unexpected when considering that these authors rely on simulated datasets derived from single populations. Typical field data, however, are often associated with multiple populations of impacted and background values.

3.1.3 REGRESSION ON ORDER STATISTICS (ROS)

Regression on order statistics is a semi-parametric, imputation technique to estimate the summary statistics of censored data (Helsel and Cohn, 1988). To apply ROS, the results are ordered in accordance with their detected or DL values, as applicable. This step is followed by producing a plot of observed versus theoretical quantiles.[47] Among the plotted quantiles, those associated with detected values are then subjected to linear regression.[48] Each ND is then substituted with the predicted values based on the interpolated or extrapolated regression line using the order of their corresponding DL values (Helsel, 2005, page 68). In other words, ROS is an imputation method.

The advantages of ROS include:

- ROS can be applied to cases where the investigated data, including NDs, are derived from single populations. ROS is especially suitable for cases with 15 to 50 percent NDs.[49]:
- ROS can be applied to cases with either one or multiple detection limits (ITRC 2013; Sinha et al., 2006).
- ROS has no requirement regarding the maximum and minimum values in the dataset.[50]

The disadvantages of ROS include:

- ROS requires an *a priori* assumption about the distribution of the censored values: in particular, typical ROS applications assume that the distribution of the

[47] Also known as a Q-Q probability plot (DON, 2002).

[48] Sinha et al. (2006) present a similar approach for log-normally distributed data using a log-probit regression method.

[49] Using a simulated log-normal dataset, Sinha et al. (2006) found the ROS method based on log-probit regression to yield reliable results for purposes of calculating both the mean and UCL.

[50] As discussed later, Kaplan-Meier method requires the maximum and minimum values to be detected (uncensored) values.

investigated data is approximately normal or lognormal (ITRC, 2013).[51] If the type of distribution is incorrectly assumed, the resulting ROS estimates (for both mean and UCL values) could be erroneous.

- The reliability of ROS results depends on the accuracy of the underlying assumption that the extrapolated regression line is representative of quantile values of ND data. Such an assumption cannot be tested precisely when dealing with NDs. This is especially problematic when considering that ROS considers NDs as the lower tail of the single population. As noted above, typical field data are often derived from mixtures of impacted and background datasets, where ROS imputed values, computed based on elevated detected values, may not appropriately represent concentrations below low DL values.
- The presence of outliers can distort the regression estimates of slope and intercept that are used to impute values for the censored data (EPA, 2015c).

EPA's 2006 guidance on statistical methods does not address ROS (EPA, 2006a). However, in their extensive comparative evaluation of the performance of several methods for analyzing censored datasets, Hewett and Ganser (2007) found that overall, the log-probit-regression-based methods[52] (along with MLE methods) "performed well across all scenarios", as measured considering both bias and overall imprecision (as indicated by rMSE). Overall, ROS is most applicable to datasets that are derived from single populations, have limited skew, lack outliers, and have <50 percent NDs (EPA, 2015c). Datasets with 50 percent or more NDs should not be subjected to ROS calculations (ITRC, 2013; EPA, 2015c). ITRC (2013) further recommends that the approach should be applied with datasets representing at least eight to 10 measurements, while EPA (2015c) similarly recommends a minimum of four to six detected measurements.

3.1.4 KAPLAN MEIER METHOD (KM)

Verifying the distribution of datasets with NDs is not easy, particularly when multiple detection limits are present (Singh et al., 2006; EPA, 2015c); consequently, the use of nonparametric computation methods has potential advantages. The Kaplan-Meier (KM) method (Kaplan and Meier, 1958) is a nonparametric approach for construction of the cumulative distribution function (CDF) of a dataset that contains censored data. The constructed CDF in turn is used to estimate the summary statistics of interest. The KM method orders the dataset by detected and DL values and relies on the number of records at and below each detected value to compute its cumulative probability.

As stated in EPA (2015c, page 129) "[t]he KM estimation method…, also known as the product limit estimation (PLE) method, is based upon a distribution function estimate, like the sample distribution function, except that the KM method adjusts for censoring.

[51] The distributional assumption is not applied to the uncensored measurements; thus, ROS methods are described as being semi-parametric (ITRC 2013).

[52] LPR, or log-probit regression, is an ROS method.

The KM method is commonly used in survival analysis (e.g., dealing with right-censored data associated with terminally ill patients) and various other biomedical applications."

The application of the KM method to environmental studies is relatively recent, when algorithms were developed to reformulate the method for left-censored environmental data, i.e., NDs that are reported as less-than-DL. Helsel (2005) proposes to transform censored data from left to right by subtracting each detected and DL value by a "large" number (also referred to as "flipping the data"); whereas Popovic et al. (2007) adjust the algorithm formulae for left-censored data. This latter method has been adopted in EPA ProUCL (EPA, 2013, 2015c).

The advantages of KM include:

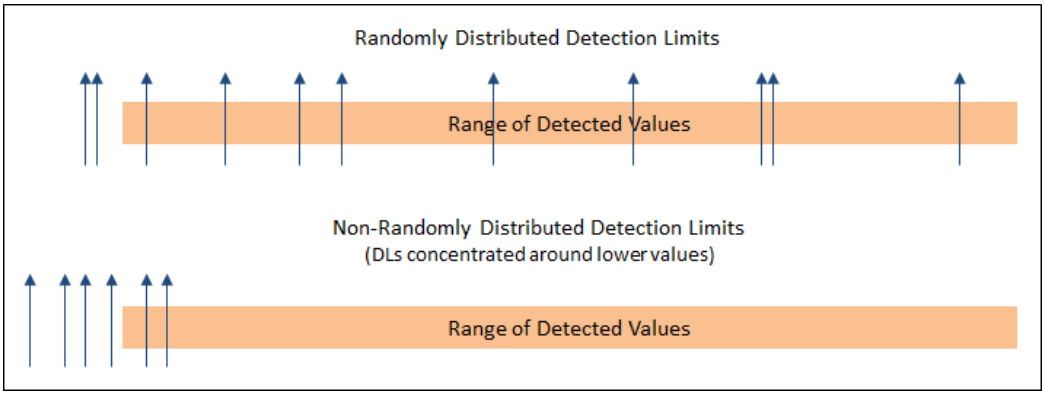
- KM does not require an assumption of data distribution or any substitution for NDs, and thus can be applied to cases where the distribution is not known or discernable (Hewett and Ganser, 2007; EPA, 2015c).
- KM, being a nonparametric method, tends to be insensitive to outliers (Antweiler and Taylor, 2008).

The disadvantages of KM include:

KM results are reliable only if the pattern of censoring is random[53] *and the probability of censoring is independent of DL values* (Schmoyeri et al., 1996 and She, 1997). This assumption means that the DL values associated with NDs in a dataset must occur at random without displaying any preference to any particular range of concentrations (see Exhibit 3-2). Unfortunately, DL values in typical environmental datasets do not display such random patterns. NDs are often associated with unique and/or low DL values which are lower than most, if not all, of the detected values. The lack of DL independence raises serious doubts about the appropriateness of KM applications in environmental studies. In fact, for most environmental datasets with DL values skewed towards the lower end of concentration ranges, KM mean and UCL results will be biased high. (Older datasets with higher DLs may be more likely to meet the requirement of having a widespread distribution of DLs.)

[53] In this context, the term random does not imply that the non-detects are randomly generated. Instead, it refers to the position of the non-detects when all observations are combined and sorted. In this process, non-detects are represented by their corresponding detection limits. Kaplan-Meier procedure assumes that the sorted non-detects are randomly scattered among the combined dataset. In other words, non-detects are not displaying any preference toward a specific range of concentrations. This condition can be tested by procedures such as the Wald-Wolfowitz runs test, in which the sorted dataset is converted to a binary sequence of 1's (detects) and 0's (non-detects). For more information about this test, readers are referred to NIST/SEMATECH (2013, Section 1.3.5.13).

EXHIBIT 3-2 HYPOTHETICAL EXAMPLES OF DETECTION LIMIT DISTRIBUTIONS



- The KM method requires more than one detection limit (ITRC, 2013, Section 5.7).[54] Analyte- and matrix-specific environmental datasets often have only one DL value less than the lowest detected value. In such cases, KM is equivalent to the full substitution approach, which always produces biased-high mean and UCL results.
- KM requires the maximum and minimum values to be detected (ITRC, 2013, Section 5.7); otherwise, the CDF cannot be accurately computed. For many environmental datasets, the lowest value is a ND. Under such conditions, KM artificially assumes that the lowest detection limit is a detected value (ITRC, 2013, Section 5.7). This requirement is one of the reasons for biased-high KM results when DL values are concentrated toward the lower range of concentrations.
- KM becomes less reliable when the proportion of NDs increases. PROPHET Stat Guide[55] warns against the use of KM in cases of heavy censoring or small sample sizes. Helsel (2005) recommends use of KM method on datasets with no more than 50 percent censored data, while ITRC (2013) states for Kaplan-Meier that "no more than 50-70% nondetects are recommended". Antweiler and Taylor (2008) recommend KM for summary statistics when datasets include less than 70 percent censored data.

KM has been promoted by many authors including Helsel (2005), Antweiler and Taylor (2008), and EPA (2015c), resulting in recommendations for its use within the environmental community. For example, Interstate Technology & Regulatory Council (ITRC, 2013, Section 5.7) recommends use of a "censored estimation technique to estimate sample statistics such as the Kaplan Meier method for calculating an upper confidence limit on the mean." However, this recommendation is not universal: Hewett and Ganser (2007) tested KM against substitution, MLE, and ROS and found that considering both bias and overall imprecision (as indicated by rMSE), KM "did not

[54] A dataset must include samples that together have more than one DL for a given analyte.

[55] See http://www.basic.northwestern.edu/statguidefiles/kaplan_ass_viol.html, viewed November 2016.

perform well for either the 95th percentile or mean and is not recommended," even when the dataset is suspected to contain multiple distributions. It may be that the poor performance is related to the non-random nature of censoring within the tested (simulated) datasets.

3.1.5 SUMMARY RECOMMENDATION

For most environmental datasets, ½ DL substitution is recommended prior to performing appropriate computations.[56] Substitution with ½ DL is mathematically parsimonious while avoiding the known over- and under-estimation of the mean that would occur with full DL or zero-substitution, and the approach has performed reasonably well in past investigations. A disadvantage of ½ DL substitution is that it will usually produce biased-low estimates of variance; to avoid this, if data are suitable (i.e., DLs are widespread within the concentration range), then more complex approaches (ROS, MLE, or KM) may be employed. However, when the proportion of NDs is high (>70%), no statistical methods are available that can reliably calculate summary statistics. Therefore, the calculation of alternative statistics such as the proportion of detects or the proportion of exceedances, should be considered.

3.2 STATISTICAL COMPARISONS

Statistical comparisons of analytical chemistry data are commonly conducted in NRDAs. Much of the available statistical literature on handling NDs, however, is specific to the computation of summary statistics; the literature is less extensive when it comes to treatment of NDs in the context of other applications. That said, some authors have addressed issues related to the treatment of NDs when conducting comparisons, and this information is summarized below.[57]

3.2.1 APPROACHES

One common NRDA analysis involves comparing contaminant concentrations at impacted and reference sites. In addition, trustees often wish to determine whether, on average, concentrations in a dataset exceed a regulatory standard or other threshold. Analysts have several options when undertaking these comparisons with datasets that include NDs. One obvious alternative is to apply substitution to NDs followed by performing the appropriate comparison tests. The choice of the appropriate test is then determined by the statistical properties of the dataset without distinguishing any difference between the detected and substituted values. Such an approach carries the typical biases that are imbedded in any substitution method (see Section 3.1.1).[58]

[56] The choice of the appropriate procedure for computing summary statistics, such as UCL, depends on many factors. Readers are referred to EPA (2013, 2015c) for additional discussions about applicability of various parametric and non-parametric procedures for computing summary statistics.

[57] We note that some of the methods described in this section, including recommended methods, have not been widely applied in the environmental literature.

[58] Helsel (2005, page 162) warns against the use of substitution approaches in statistical tests but does not provide a quantitative justification for this warning.

Substitution has been supported as an appropriate method when conducting statistical comparisons: USACE (1995) summarizes results of a simulation study conducted to assess the performance of ten censored data methods, in the context of comparing concentrations of contaminants among groups of samples where sample sizes were small ("small" was not defined). Although the authors found that no censored data approach worked best in all situations, "In general, the simple substitution methods work best to maintain power and control type I error rate in statistical comparisons." The authors provide a complex table of recommended approaches as a function of the amount of censoring, variance, and distribution characteristics, and other factors, also concluding, however, that "If it is impossible to determine characteristics of the variances or statistical distribution for censored data samples, use DL for up to 40 percent censoring or DL/2 for 40 to 80 percent censoring", and furthermore, that "Beyond 60 to 80 percent censoring, it is unlikely that any technique will perform acceptably." EPA (2009) provides guidance on the statistical analysis of groundwater monitoring data, and states "at least 50% of the data should be detectable in order to compare either [groundwater] well means or medians." One disadvantage of substitution when conducting statistical comparisons results from the effect of substitution on variance: because substitution results in under-estimated variances, hypothesis testing of differences between groups is likely to produce lower p-values than would be the case absent censoring.

An alternative to simple substitution is to use a parametric comparison test that is capable of incorporating censored data. The reliability of such tests, however, depends on the appropriateness of the assumed distribution. Moreover, the reliability of the assumed distribution diminishes as the proportion of NDs in the dataset increases. For more detailed information about parametric alternatives, readers are referred to Helsel (2005, Chapters 9 and 10).

Based on the above facts, under certain conditions, pursuing nonparametric comparison tests capable of incorporating censored data for NRDA computations seems to be preferable. For example, if patterns of DL values are random (i.e., not tending to a specific range of concentrations),[59] then non-parametric methods based on KM-constructed cumulative distribution functions may be preferable. Non-parametric tests do not have any requirements about the underlying distribution of the data and avoid deficiencies associated with ½ DL substitution, such as under-estimation of the variance.

Non-parametric tests include the generalized Wilcoxon test for unpaired comparisons and paired Prentice-Wilcoxon or PPW test.

- Generalized Wilcoxon test (Peto and Peto, 1972; Prentice, 1978; Prentice and Marek, 1979) can be used to compare whether the medians of two groups differ statistically. In this test, datasets containing NDs are pooled together and subjected to a procedure identical to the KM CDF computation. Using the

[59] As stated previously, we expect that for the large majority of NRDA datasets, the pattern of detection limits will not be random but rather will be clustered towards the low end of the concentration range (i.e., the requirements for using the KM method would not be met).

computed CDF results, the score of each value is computed. In the final step, the test statistics are calculated based on the sum of the scores of the constituents of each dataset. This procedure, similar to other nonparametric methods, is insensitive to outliers and does not require any specific distributional assumption. However, the reliability of its results diminishes as the proportion of NDs increases (Helsel, 2005).

- Paired Prentice-Wilcoxon test (PPW; Helsel, 2005, page 158) is a variation of the Generalized Wilcoxon test, especially designed for paired datasets. As described by Helsel (2012, page 188): "To compute the PPW test, the data are stacked into one column, and a Kaplan–Meier estimate of the survival function for the combined data is computed. Scores, the estimated percentiles of the survival function minus 0.5, are computed for each observation, both censored and uncensored. The scores are then split back up into their respective groups. If the null hypothesis is true and the two distributions are the same, differences between pairs of scores should be small, hovering around zero. In other words, the two paired observations should be located at similar places in the combined distribution; therefore their score values should be similar. If the distributions of the two groups differ, the paired observations will be located at different points of the combined survival distribution, with the scores from one dataset consistently higher than the paired score from the other. The PPW test computes the differences between the paired scores, and determines whether the sum of these differences is significantly different from zero, using a normal approximation for the test statistic."

The above two methods have not seen widespread application in the environmental literature (as noted previously, most statistical literature addressing NDs has focused on summary statistics rather than other statistical tasks). Because applications have been limited, it is possible that available literature may not fully characterize these methods' advantages and disadvantages.

Datasets with high proportions of NDs are most suitable for non-parametric techniques involving comparisons of proportions of detected values, as opposed to statistical comparisons of central tendency measures. More specifically, in instances where the datasets contain large proportions of NDs (usually considered to be 50 percent or more NDs), statistical comparisons can be conducted in accordance with binary-based procedures such as the test of proportions (DON, 2002, page 97). In this test, the proportion of detects or the proportion of values in excess of a given threshold is calculated for each dataset. These proportions are then used to calculate the test statistics to determine whether statistically significant differences exist among the investigated datasets: for example, the test could determine whether the proportion of samples exceeding a threshold is higher in an affected area than in a reference area. As noted previously, however, to implement this test, the selected threshold must exceed the highest DL value in the dataset (DON, 2002, page 95). Unfortunately, this is not always the case in NRDA analyses, and it represents a limitation in the utility of such statistical approaches (and datasets with such high proportions of NDs) for NRDA purposes.

The advantage of the test of proportions is that it is non-parametric and therefore not dependent on the underlying distribution of the dataset; the disadvantage of the test is that it assumes that the investigated measurements are independent and devoid of any spatial and/or temporal correlations (DON, 2002, page 95).

3.2.2 SUMMARY RECOMMENDATION

For most environmental datasets, ½ DL substitution followed by appropriate comparison tests is recommended. When datasets are large with small proportions of NDs and low-variance, parametric tests, such as Student's t tests, are appropriate. Otherwise, non-parametric tests, including Slippage, Quantile, and Wilcoxon Rank Sum comparative tests (DON, 2002, Section 4.2) are recommended. Substitution with ½ DL is mathematically parsimonious, and the approach has performed reasonably well in past investigations of statistical comparisons. It also avoids the need to determine the distribution of the dataset, which can be difficult (and which is a requirement for parametric comparison tests).

A disadvantage of ½ DL substitution in the context of statistical comparisons is under-estimation of the true variance; to avoid this, if data are suitable (i.e., DLs are widespread within the concentration range), then more complex approaches (Generalized Wilcoxon for unpaired samples; paired Prentice-Wilcoxon for paired samples) may be employed. However, when the proportion of NDs is high (>50%), a test of proportions is recommended (DON, 2002, page 94).[60]

3.3 CORRELATION AND REGRESSION ANALYSES

Correlation and regression analyses can be part of an NRDA. For example, trustees may wish to determine correlations amongst various contaminants of concern, or they may be interested in evaluating the relationship between exposures of biota to contaminants with one or more measures of effect. As noted previously, the reviewed literature is mainly concerned with computing summary statistics based on datasets containing NDs. However, some authors have addressed issues related to the treatment of NDs in cases where correlation and regression analyses are performed using the dataset, as summarized in the following section.

3.3.1 APPROACHES

Typical NRDA correlation and regression analyses involve comparisons of concentrations of a contaminant of concern versus a measurement of biological status or function, collocated concentrations of another contaminant, or time. A number of treatment options are available when datasets contain censored data. One alternative is to apply ½ DL substitution to NDs followed by performing the appropriate parametric or

[60] When more than 50% of the data are NDs, then the median is also ND; such a condition raises questions about the reliability of comparison tests that treat measures of central tendency as non-censored values.

non-parametric correlation or regression analyses.[61] Some have argued against substitution in the context of correlation and regression. For instance, Helsel (2012) argued that substitution of a constant fraction of the reporting limit will obscure correlations (e.g., estimates of correlation coefficients and regression slopes).

An alternative to substitution involves using parametric correlation and/or regression analyses that are capable of incorporating censored data. The reliability of such tests, however, depends on the appropriateness of the assumed distributions. The reliability of the assumed distributions diminishes as proportions of NDs increase in the dataset. For more detailed information about parametric alternatives, readers are referred to Helsel (2005, Chapters 11 and 12).

Based on the above facts, under certain conditions, pursuing nonparametric correlation and regression analyses capable of incorporating censored data for NRDA computations seems to be preferable. For example, if the proportion of NDs exceeds 15 percent and the pattern of DL values is random (i.e., not tending to occur within a specific range of concentrations),[62] then non-parametric correlation/regression methods based on KM-constructed cumulative distribution functions could be employed. These alternatives include the modified Kendall's tau for correlation,[63] and the Akritas-Theil-Sen line (ATS) for regression.[64] When the proportion of NDs is high (>80 percent), the Phi coefficient for correlation can be used.[65] Applying regression approaches to datasets with very large proportions of NDs is not recommended.

In general, the advantage of the modified Kendall's tau, ATS, and Phi coefficient methods is that they are non-parametric and therefore not dependent on the underlying distribution of the dataset; the disadvantage of these approaches is that they require larger datasets to yield reliable results than those required by parametric procedures. Also of note, the above three methods have not seen widespread application in the environmental literature. Because applications have been limited, it is possible that available literature may not fully characterize the methods' advantages and disadvantages.

3.3.2 SUMMARY RECOMMENDATION

For most environmental datasets, ½ DL substitution followed by appropriate parametric and non-parametric correlation or regression method is recommended. Substitution with ½ DL is mathematically parsimonious, and it also avoids the need to determine the

[61] Parametric approaches include Pearson's r for correlation and simple linear regression. Non-parametric alternative include Kendall's tau for correlation and Theil-Sen line for regression.

[62] As stated previously, we expect that for the large majority of NRDA datasets, the pattern of detection limits will not be random but rather will be clustered towards the low end of the concentration range.

[63] Note that the *modified* Kendall's tau test referenced here is not the same as the previously-mentioned Kendall's tau test.

[64] Note that Akritas-Theil-Sen line method is a modified version of the Theil-Sen line method. See Appendix C for more information about these methods.

[65] The Phi coefficient for correlation involves converting datasets into binary values before assessing their correlation (Helsel, 2005). Methods for binomial data are most useful when values are severely censored, with more than about 80% non-detects (Helsel, 2005).

distribution of the dataset, which can be difficult (and which is a requirement for parametric comparison tests).

A disadvantage of ½ DL substitution is under-estimation of the true variance. To avoid this, when the proportion of NDs is elevated and if data are suitable (i.e., DLs are widespread within the concentration range), then more complex approaches (modified Kendall's tau for correlation, or ATS for regression) may be employed. When the proportion of NDs exceeds 80 percent, regression analyses are not recommended, but correlation analysis can be pursued by procedures such as the Phi coefficient. This requires converting the investigated data into binary values prior to performing the correlation analysis.

3.4 LANL CASE STUDIES

To demonstrate the effects of various ND treatments, five representative datasets were extracted from the example LANL data and used to generate summary statistics using EPA ProUCL (2013). The analytes illustrated in the examples in this section were selected solely as examples for the purposes of demonstrating the impact of various methods for treating NDs; nothing is implied as to whether these substances are important injury drivers or not. ProUCL is selected for these computations because: (a) the software has been developed and promoted by EPA, (b) the software contains a broad range of alternatives for calculation of summary statistics, and (c) the software has been updated numerous times to broaden its applicability to environmental applications. These datasets include measured concentrations of groundwater arsenic, boron, vanadium, tetrachloroethene (TCE), and sediment selenium concentrations within LANL property. All reported concentrations are based on EPA SW-846 methods.

Each of the above datasets has unique ND and sample-size characteristics. For example, the sediment selenium dataset only has 36 measurements of which 44 percent are NDs. This small dataset is used to facilitate the presentation process. The groundwater vanadium dataset has 1,270 measures with a small percent of NDs at 10 percent. The groundwater boron dataset has 1,270 measurements with 48 percent NDs. The groundwater arsenic dataset is also large with 1,271 measurements, but has a higher ND rate of 73 percent. The groundwater TCE dataset has the largest sample size with 2,219 measurements of which 96 percent are NDs.

LANL case studies are initiated by exploring the selenium dataset as listed in Exhibit 3-3. As this table indicates, the "Report Result" of U-flagged NDs have the same values as their corresponding "Report Detection Limit," which are much higher than their corresponding "Report MDL." As discussed, such assignments to NDs are inappropriate. In all the subsequent computations, the upper limit of NDs is set equal to their corresponding "Report MDL."

EXHIBIT 3-3 INVESTIGATED SEDIMENT SELENIUM CONCENTRATIONS WITHIN LANL PROPERTY

FIELD SAMPLE ID	LOCATION ID	DATE SAMPLED	REPORT RESULT	REPORT UNITS	LAB QUALIFIER	DETECTED	REPORT DETECTION LIMIT	REPORT MDL
SFBDARHT-12-12442	15-600879	2012-04-24	0.39	mg/kg		Y	0.1	0.031
SFBDAR-13-29824	15-600879	2013-04-25	0.71	mg/kg		Y	0.1	0.052
DARHT-14-56699	15-600879	2014-04-25	0.78	mg/kg		Y	0.088	0.029
SFB-15-95573	15-600879	2015-04-21	0.67	mg/kg	B	Y	0.1	0.039
SFB-16-115030	15-600879	2016-05-03	0.48	mg/kg		Y	0.11	0.076
SFBDARHT-12-12444	15-600880	2012-04-24	0.4	mg/kg		Y	0.1	0.031
SFBDAR-13-29826	15-600880	2013-04-25	0.77	mg/kg		Y	0.1	0.051
DARHT-14-56701	15-600880	2014-04-25	1	mg/kg		Y	0.097	0.032
SFB-15-95575	15-600880	2015-04-22	0.76	mg/kg	B	Y	0.11	0.039
SFB-16-115031	15-600880	2016-05-03	0.8	mg/kg		Y	0.12	0.079
SFBDARHT-12-12443	15-600881	2012-04-24	0.34	mg/kg		Y	0.1	0.031
SFBDAR-13-29825	15-600881	2013-04-25	0.71	mg/kg		Y	0.11	0.053
DARHT-14-56700	15-600881	2014-04-25	0.89	mg/kg		Y	0.097	0.032
SFB-15-95574	15-600881	2015-04-21	0.71	mg/kg	B	Y	0.11	0.042
SFB-16-115032	15-600881	2016-05-03	0.67	mg/kg		Y	0.12	0.082
SFBDARHT-12-12445	15-600882	2012-04-24	0.39	mg/kg		Y	0.1	0.031
SFBDAR-13-29827	15-600882	2013-04-25	0.63	mg/kg		Y	0.1	0.052
DARHT-14-56702	15-600882	2014-04-29	0.89	mg/kg		Y	0.11	0.034
SFB-15-95576	15-600882	2015-04-22	0.48	mg/kg	B	Y	0.11	0.039
SFB-16-115033	15-600882	2016-05-03	0.49	mg/kg		Y	0.1	0.069
RE01-13-37758	LA-610960	2013-08-08	1.35	mg/kg	U	N	1.35	0.447
RE01-13-37759	LA-610960	2013-08-08	1.14	mg/kg	U	N	1.14	0.377
RE01-13-37764	LA-610964	2013-08-09	1.03	mg/kg	U	N	1.03	0.339
RE01-13-37765	LA-610964	2013-08-09	1.19	mg/kg	U	N	1.19	0.393
RE01-13-37770	LA-610966	2013-08-08	0.961	mg/kg	U	N	0.961	0.317
RE01-13-37771	LA-610966	2013-08-08	1.18	mg/kg	U	N	1.18	0.389
RE01-13-37772	LA-610966	2013-08-08	1.17	mg/kg	U	N	1.17	0.387
RE01-13-37791	LA-610966	2013-08-08	1.16	mg/kg	U	N	1.16	0.382
CALA-12-1679	Los Alamos above DP Canyon	2011-11-17	0.975	mg/kg	U	N	0.975	0.32
CALA-12-1680	Los Alamos above SR-4	2011-11-17	0.964	mg/kg	U	N	0.964	0.32
CAPA-13-24787	MDA G-6 Retention Pond Lower	2012-11-15	0.967	mg/kg	U	N	0.967	0.319
CAPA-13-24788	MDA G-7	2012-11-15	0.868	mg/kg	U	N	0.868	0.287

FIELD SAMPLE ID	LOCATION ID	DATE SAMPLED	REPORT RESULT	REPORT UNITS	LAB QUALIFIER	DETECTED	REPORT DETECTION LIMIT	REPORT MDL
CAPA-12-1694	PA-22890	2011-11-22	1.25	mg/kg	U	N	1.25	0.41
CAPA-12-1695	PA-603937	2011-11-22	0.94	mg/kg	U	N	0.94	0.31
CAWA-12-1713	Water at Beta	2011-11-23	1.2	mg/kg	U	N	1.2	0.4
CAWA-12-1715	Water at SR-4	2011-11-23	0.972	mg/kg	U	N	0.972	0.32

As listed in Exhibit 3-3, the detected sediment selenium concentrations range from 0.34 to 1 ppm. In contrast, NDs have "Report MDL" values ranging from 0.287 to 0.447 ppm. This means that the listed DL values mainly occur within the lower end of the concentration range. Such a pattern is inconsistent with the assumption of random occurrences of DL values. Note that under such conditions, KM has a tendency to yield biased-high results.

Using ProUCL, Version 5.0.00, two typical NRDA summary statistics are calculated: the mean and its 95% upper confidence limit (UCL). Each of these statistics is calculated by treating NDs in accordance with three methods, including ½ DL substitution, ROS and KM; MLE is not included as ProUCL does not include this method. Exhibit 3-4 presents the results.

EXHIBIT 3-4 COMPUTED SUMMARY STATISTICS OF SEDIMENT SELENIUM CONCENTRATIONS (ppm)

ANALYTE	COUNT	% ND	NON-DETECT TREATMENT METHOD	RESULT	
				MEAN	95% UCL
Selenium	36	44%	0 substitution	0.36	0.46
			DL substitution	0.52	0.58
			½ DL substitution	0.44	0.51
			ROS	0.47	0.55
			KM	0.49	0.56

Notes:
Selenium was selected solely as an example for the purposes of demonstrating the impact of various methods for treating non-detects; nothing is implied as to whether the substance is an important injury driver or not.
The upper confidence limit (UCL) of ½ DL substitution is based on nonparametric Percentile Bootstrap UCL.
ROS results are calculated based on gamma distribution.
UCL of KM is calculated based on t distribution.
Calculations performed in ProUCL, Version 5.0.00. MLE is not included as ProUCL does not include this method.

Exhibit 3-4 indicates that, as expected, the zero- and full-DL substitutions produce lower and higher mean and UCL values, respectively. The mean and UCL results based on KM are higher than those based on ½ DL substitution and ROS. This is mainly due to the fact that DL values associated with ND selenium results are concentrated towards the lower range of concentrations (as is common in typical environmental chemistry datasets). Under such circumstances, the KM method produces biased-high mean and UCL estimates.

With regards to ROS results, ProUCL warns that any dataset with 50 percent or more NDs should not be subjected to ROS calculations. The small size of the selenium dataset along with a 44 percent ND rate, and DL values being concentrated towards the lower range reduce the reliability of the ROS results. Under such conditions, the mean and UCL results produced by ½ DL substitution are preferred.

Similar computations are applied to the larger groundwater vanadium, boron, arsenic and TCE datasets. For these datasets, ProUCL could not identify any discernable distribution; therefore the semi-parametric ROS was not implemented. Exhibit 3-5 therefore presents the results associated with the KM and ½ DL substitution methods. As listed, in all datasets, the DL values mainly occur within the lower range of concentrations. Such patterns, as expected, resulted in biased-high KM mean and UCL values.

EXHIBIT 3-5 COMPUTED SUMMARY STATISTICS OF GROUNDWATER VANADIUM, BORON, ARSENIC AND TCE CONCENTRATIONS

ANALYTE	COUNT	% ND	MIN DL	MAX DL	MIN DETECTED	MAX DETECTED	NON-DETECT TREATMENT METHOD	MEAN	95% UPPER CONFIDENCE LIMIT (UCL)
Vanadium	1,270	10%	1	10	1.01	23.7	½ DL sub	5.19	5.37
							KM	5.25	5.41
Boron	1,270	48%	15	15	15	1310	½ DL sub	35.0	40.6
							KM	38.6	44.3
Arsenic	1,271	73%	1.7	8.5	1.7	5.3	½ DL sub	1.34	1.38
							KM	1.93	1.96
TCE	2,219	96%	0.3	1.5	0.31	28.5	½ DL sub	0.21	0.23
							KM	0.35	0.38

Notes:
The analytes illustrated in this table were selected solely as examples for the purposes of demonstrating the impact of various methods for treating non-detects; nothing is implied as to whether these substances are important injury drivers or not.
11 arsenic samples out of 1,271 have reported DL >1.7 ppb.
Three TCE samples out of 2,219 have reported DL >0.3 ppb.
Two vanadium samples out of 1,270 have reported DL >1 ppb.
UCLs of ½ DL substitution are calculated using nonparametric Percentile Bootstrap UCL.
UCLs of KM are calculated based on KM Percentile Bootstrap UCL.

(This page intentionally left blank to facilitate double-sided printing)

CHAPTER 4 | PAST TREATMENTS OF NON-DETECTS IN NRDA

During the course of a natural resource damage assessment, trustees commonly encounter ND results within the environmental datasets with which they are working. This chapter identifies some of the most common NRDA contexts in which trustees must decide how to handle NDs; it also characterizes the most common approaches that trustees have adopted.

4.1 DATA ACQUISITION

We used two strategies to identify examples of treatments of NDs in NRDAs. First, we searched two publicly-available repositories of NRDA documents: the U.S. Department of the Interior's (DOI) Natural Resource Damage Assessment and Restoration Program's (NRDAR Program) online case map and document library, and the NOAA Damage Assessment, Remediation, and Restoration Program's (DARRP) online collection of case documents.

Because the structure of these two repositories differs, we employed different search strategies for each. The DOI repository allows for searching by incident type (chemical, mining, oil, or other), document type, and for a specific word within the documents' contents. Initial searching indicated that, if a keyword was not found in any document in the library, the search results would include all documents. At the time the search was conducted, the DOI NRDAR website included 1,154 documents. Searching only with the keyword "non-detect" produced 1,154 results, suggesting that the specific search term was not found. Instead, we limited our search to the word "detect" across all incident types and limited results to the following document types:

- Study plan
- Study Report
- Preassessment Data Report
- Environmental Assessment
- Environmental Impact Statement
- Journal Links
- Agency Reports
- RCDP
- PAS
- PED
- Assessment Report
- HEA/REA

This search strategy resulted in 153 documents, which we downloaded and reviewed for relevance. Appendix A summarizes key features of the selected documents.

NOAA's case document repository does not have a built-in capacity for searching. We therefore conducted a Google search on the domain as follows:

site:casedocuments.darrp.noaa.gov non-detect (21 hits)

and

site:casedocuments.darrp.noaa.gov nondetect (16 hits).

We reviewed all identified documents and summarize key features of the most relevant results in Appendix A.

Not all information generated during the course of an NRDA is, or becomes, public; indeed, a substantial body of information is often generated during a NRDA prior to the filing of a legal case and/or during work that leads up to, or that is undertaken as part of, settlement discussions. This type of information is confidential and cannot be released; however, it represents the cutting edge of NRDA practice.

Therefore, as a supplemental source of information, we also conducted internal interviews of senior staff at Industrial Economics, Inc., who jointly have multiple decades' worth of experience in NRDA cases across the United States. The purpose of these interviews was to ascertain whether a significant body of non-public information about general trustee practices in handling NDs indicated any trends or changes in typical practice, which were not yet evident in public materials. The type of information gathered in these interviews was qualitative in nature but, as discussed below, nevertheless served to confirm the reasonableness of conclusions drawn from public information.

4.2 RESULTS

The types of documents gathered span the range of phases in NRDA from initial scoping investigations through planning, assessment, restoration planning, and restoration implementation. Identified documents include those specifically mentioned in DOI's NRDA regulations (e.g., the preassessment screen, damage assessment plan, preliminary estimate of damages, restoration compensation and determination plan, etc.) as well as other types of documents. The documents span a wide range of natural resources (sediments, surface water, soils, groundwater, biota) and years (1994 through 2016, although the bulk of the documents are from 2007 or later). Across all of these documents, three types of analyses with ND data are by far the most common:

- Calculating a total contaminant concentration in a given environmental sample by summing the concentrations of component contaminants—for instance, calculating the sum ($\sum$PCBs) of PCB congeners or Aroclors, or total PAHs (tPAHs).
- Calculating summary statistics across a group of samples—e.g., means, standard deviations, ranges, and counts.

- Comparing a concentration (typically, either an individual value or a mean) to a fixed threshold value. This threshold may be a promulgated value such as a water quality standard, or it may be a literature-based effects threshold.

It is notable that, in all documents identified, the documents' authors used the following general approaches to handling ND values:

- Excluding the data (typically, only in selected circumstances);
- Proportion (e.g., 4 out of 10 values were NDs);
- Presenting a range (e.g., as "non-detect" (or "<DL") to X µg/g), or
- Simple substitution (i.e., substituting NDs with either zero, half the DL, or the DL).

More sophisticated statistical approaches, such as those discussed in Chapter 3, were not utilized in any of the identified documents.

4.2.1 SIMPLE SUBSTITUTION

In applying the simple substitution method, some general patterns were evident (Exhibit 4-1). First, when contaminants are summed within a sample to produce a total (e.g., ∑PCBs or tPAHs), the most common approach was to substitute with zero (about 70 percent of examples); in the remaining instances, the authors used half the DL.

EXHIBIT 4-1 APPROXIMATE FREQUENCY (COUNT) OF DIFFERENT SIMPLE SUBSTITUTION APPROACHES

SUBSTITUTION VALUE	WHEN CALCULATING TOTALS	OTHER APPLICATIONS
Zero	69% (9)	19% (6)
½ the DL	31% (4)	61% (19)
Full DL	0% (0)	19% (6)

In other applications when simple substitution was used, half the DL was used approximately 60 percent of the time, while the remaining examples were split approximately equally between the other two approaches. Of note, the *Deepwater Horizon* NRDA—a major case of recent vintage—is not represented in the above collections of documents. In this NRDA, one of the primary measures was total polycyclic aromatic hydrocarbon (tPAH) surrogate-corrected concentrations, calculated based on the summation of 50 PAH analytes including parent PAHs and selected alkylated homologs (Forth et al., 2015). In these calculations, if the concentration of a given compound in a sample was not detected, it was treated as a 0 value in the summation (Rouhani et al., 2016).

In most cases, the documents' authors identify the substitution approach but do not provide much, if any, explanation for its selection. Based on our experience in NRDA, however, we would expect that the choice of substitution value would typically be influenced by the document's objectives, and these in turn, are in part a function of the

document's role in the NRDA. More specifically, we would expect that documents from earlier stages in the NRDA (e.g., screening analyses, PEDs) to be particularly concerned about missing potential evidence of harm and as such, might expect the authors to select substitution approaches more likely to over-estimate than under-estimate exposure and associated injuries—i.e., they would be more likely to select the full DL to represent NDs. Although sample sizes are small, this expectation appears to be borne out: the proportion of documents utilizing the full DL is highest among early-stage documents and decreases in mid- and late-stage documents (Exhibit 4-2).

EXHIBIT 4-2 APPROXIMATE FREQUENCY (AND COUNT) OF DIFFERENT SIMPLE SUBSTITUTION METHODS BY BROAD NRDA STAGE

SUBSTITUTION VALUE	EARLY	MID	LATE
Zero or half the DL	64% (7)	86% (6)	92% (12)
Full DL	36% (4)	14% (1)	8% (1)

Notes:
Examples of early-stage documents include the PAS, PED, assessment plans, and scoping/screening level work. Mid-stage documents include most injury and other technical reports, while late-stage documents include those developed for settlement or site restoration. The documents reflected in this table exclude those that used substitution solely to calculate totals (e.g., ΣPCBs or tPAHs).

The explanatory information provided by authors is generally consistent with this observation. For instance, in a data report and screening-level risk assessment of crabs on Vieques Island, Puerto Rico (NOAA and Ridolfi, 2006), the authors elected to substitute ND values with the MDL. The selection of the MDL was acknowledged to result in an overestimate of concentrations to avoid missing the potential for effects: the report "applied conservative assumptions about exposure that are not necessarily realistic or appropriate for characterizing actual risk". This may have been an issue of particular concern as the results of the investigation "will assist the USFWS in determining whether selected refuge areas can be opened to [subsistence] crab harvesting." That said, the document points to the potential need for additional evaluation to "fully identify and characterize potential risks to human health and ecological receptors." In other words, the selection of the MDL was consistent with the document's overall protective objectives, while at the same time was not understood to represent the final word with respect to characterization of exposure or the potential for effects.

As indicated in Exhibit 4-2, in mid- and late-stage NRDA documents (e.g., injury and other technical reports), the prevalence of substitution with the full DL was lower. Moreover, these two documents tended to use substitution with the full DL in limited ways. HRNRT (2016) used the DL when averaging results from laboratory split samples but not for calculating summary statistics across samples. GDNR et al. (2006), developed as part of a cooperative assessment, calculated means of fish tissue concentrations substituting the DL for NDs and also plotted ND sediment data at the DL; however, in terms of determining injury to surface water in Lake Hartwell, the authors write that

because the DL for PCBs exceeded water quality criteria, the data "cannot be used for injury determination."

More commonly, authors of mid- and later-stage NRDA documents substituted ND results with half the DL or with zero (Exhibit 4-2). For instance, NOAA (2013) is a late-stage document: it is a response to public comments on the Supplement to the Draft Restoration Plan and Programmatic Environmental Impact Statement generated as part of the Duwamish Waterway NRDA. This document references analyses (injury quantification) undertaken in other documents. NOAA (2013) was developed in the context of early settlement and states: "For non-detect values where the detection limit exceeds the service loss levels, no injury is assigned and the concentration is assumed to be close to zero for purposes of the geographic interpolation. Use of this assumption may result in an underestimation of injury, and illustrates the care taken by trustees to minimize the likelihood of overestimating injury."

EcoChem and Geosphere (2002) is another document developed in the context of a settlement proposal. This document is an appendix to the "Hylebos Waterway NRDA Settlement Proposal Report," issued for public review by The Commencement Bay Natural Resource Trustees for purposes of settling natural resource damage liability relating to the Hylebos Waterway. EcoChem and Geosphere (2002) develops an allocation of responsibility for natural resource injuries in the Hylebos Waterway across multiple parties, doing so in part by defining injury footprints associated with different contaminants. In this case, the authors appear to have used substitution with half the DL: for one contaminant (pentachlorophenol[66]), the document notes that half the DL value exceeded the injury threshold resulting in a "ubiquitous injury footprint." With the exception of one small area, this footprint was deemed to not be allocable.

Schein et al. (2015) presents a field study of sediment and soil chemistry to support injury evaluation at the Anniston site in Alabama, where PCBs are the primary contaminant of concern. As such, this document was developed at a more intermediate point in the NRDA process. In this case, the authors calculated total PCB concentrations, and PEC-quotients (i.e., concentrations divided by a probable effects concentration (PEC) threshold), and calculated summary statistics. ND values were treated as half the DL; however, "Non-detect values that were above the PEC were screened out and not included in the totals calculations." This document is a relatively uncommon example in that the authors have a section of their report that provides a more extensive description of the reason behind the selected approach to handling NDs. Referencing a variety of other literature, they write: "A number of investigators have evaluated the implications of applying various procedures for estimating the concentrations of contaminants of concern from less than detection limit data. While there is no consensus on which data censoring methods should be used in various applications, the simplest methods tend to be used most frequently, including deletion of ND values or substitution of a constant, such as zero, the detection limit, or one-half the detection limit (USACE 1995)." Because only 17

[66] Pentachlorophenol was one of many contaminants evaluated as part of this NRDA.

percent of the results were NDs, the authors chose the substitution method using ½ the DL, which was "consistent with the guidance developed by USACE (1995)."

MacDonald et al. (2003) is another example of a document that provides an explanation for its choice of a ND substitution value. In this document, the authors assess injury to human use of fishery resources in the Grand Calumet River, Indiana Harbor Canal, and related areas. The authors used different approaches for NDs depending on the specific analysis. Chemicals of concern include PCBs, oil and related compounds, and metals. The authors state: "Less than detection limit data was treated in several ways, depending on the guidance that has been provided in conjunction with the chemical benchmarks for sediment chemistry and tissue chemistry." For example, in calculating total concentrations of contaminants in sediments, NDs were assigned a value of half the detection limit except where the detection limit was greater than the selected chemical benchmark. In that case, the greater than detection limit value was not used in the calculation of the total concentration or in the assessment of injury. However, for fish tissue samples, "non-detect results data and low level detects were substituted with zero, in accordance with the guidance provided by USFDA (2001)[67] to facilitate comparison with the tolerance of action levels." In contrast, "less than detection limit data for tissue chemistry were assigned a value of one-half of the detection to facilitate comparison with the thresholds used to develop the Indiana FCAs (Anderson et al. 1993)."

4.2.2 TABULATION

Several documents avoided substitution because the document's purposes were served with simple tabulation instead. For instance, in the Hudson River (NY) PCB surface water injury report (HRNRT 2008), the requirements of injury determination were clearly met despite the presence of NDs (about 20 percent) in the dataset. Determining injury did not require calculation of summary statistics or quantitative analyses beyond sample counts; consequently, the Hudson River Trustees could avoid substitution or other more complex approaches.

Similarly, the goal of HRNRT (2013) was to communicate with the public about the general types of exposure data available. Charts with non-parametric values (restricted to detect samples) were included, with ND counts provided in each chart's notes.

It is also common practice for NDs to be identified in tables as part of a range of presented values (e.g., "n.d. to 5 ng/g"); available NRDA documents include many such examples.

4.2.3 WHEN DETECTION LIMITS EXCEED THRESHOLDS

Among the more challenging situations occurs when detection limits exceed relevant injury or effects thresholds, and different authors have taken very different approaches to this situation. HRNRT (2008) and HRNRT (2013) (mid-stage NRDA documents)

[67] The current version of this document (fourth edition) is USFDA (2011). For certain deleterious substances, when calculating totals (e.g., total DDTs, or total aldrin/dieldrin levels) to make comparisons of measured concentrations in fish tissues with action levels, the concentrations below specified numeric values are not counted—i.e., are treated as zero.

avoided the problem by using tabulations/counts. NOAA (2013), developed for settlement, assumed injury to be zero in such cases. Other documents explicitly or implicitly excluded such data from their calculations (EcoChem and Geosphere, 2002; MacDonald et al., 2003; Schein et al., 2015; Latta, 2012). Here again, the authors' choices appears to be in significant part a function of their objectives, which are influenced by the stage of the assessment. Because these are clearly difficult situations to decide how to proceed, they may be particularly suitable for sensitivity analyses (see Chapter 5, Recommendations).

4.2.4 INTERVIEWS

Internal interviews with senior IEc NRDA staff confirm the above, major observations based on available literature: i.e., that the issue of how to handle NDs has generally been addressed through technically simple approaches (enumeration or simple substitution). In short, although we cannot say that more sophisticated statistical approaches have never been applied in any NRDA, it is unequivocally the case that up to this point in time such approaches have not been widely adopted.

4.3 CONCLUSIONS: PAST TREATMENTS OF NON-DETECTS IN NRDA

Both a review of publicly available NRDA documents and interviews with senior professionals working in NRDA have confirmed that trustees have, by and large, addressed NDs in their datasets using one of two general approaches:

- Proportion (e.g., 4 out of 10 values were NDs), or
- Simple substitution (i.e., substituting NDs with either zero, half the DL, or the DL).

More sophisticated statistical approaches, such as those discussed in Chapter 3, are not in widespread use. In some cases, tabulation of NDs was sufficient to meet the documents' objectives. Tabulation is straightforward to implement and avoids potentially more controversial decisions, such as the investment of resources into a more sophisticated statistical approach, and/or the choice of what value to use, for a simple substitution approach.

Where tabulation does not suffice (e.g., when it is necessary to calculate total concentrations or to generate summary statistics), the authors' choice of substitution value appears to be at least in part a function of the document's objectives, which in turn are influenced by the stage in the assessment process.

Where a document's purpose was explicitly protective of the environment (in the sense of preferring to overestimate rather than underestimate the potential for adverse effects), and particularly in earlier stages of an NRDA, substitution using the DL has sometimes been preferred (e.g., NOAA and Ridolfi, 2006), although not universally so (e.g., Office of the Colorado Attorney General et al. (2007) is a PAS and used half the detection limit). In contrast, in the context of settlement, trustees are at times willing to make certain concessions, which, in at least one example, included willingness to forgo injury where concentrations of contaminants were below detection (NOAA 2013). More commonly,

the document's objectives do not express an explicit preference for over- or under-estimation of effects. In many such instances, half the DL is a common simple substitution choice (e.g., Krausmann 1999, EcoChem and Geosphere 2002, NRTSLRE 2013); overall, among those examples where substitution was employed, half the DL was the most common choice. The main exception is for the specific purpose of tabulating total concentrations (e.g., ΣPCBs, tPAHs), in which zero has been the most common substitution. It is also the case that trustees did not adopt a one-size-fits-all approach to substitution: in a number of documents, the authors selected one substitution approach for one type of analysis and a different substitution value for another analysis. Finally, several circumstances lead authors to exclude ND data from specific analyses. These circumstances include when detection limits exceed a relevant injury threshold (EcoChem and GeoSphere, 2002; MacDonald et al., 2003; Schein et al., 2015) and in one case, when a principal components analysis[68] was to be undertaken (USFWS et al., 1999a).

[68] Principal component analysis (PCA) is a statistical technique used to combine large numbers of variables into a smaller number of variables (or components) that are not linearly correlated with one another. It can be used to bring out patterns of similarity and difference between datasets based on these combinations of multiple variables and is sometimes used in environmental forensic investigations to help associate samples with potential sources of contaminants.

CHAPTER 5 | RECOMMENDATIONS

NRDA decisions often rely on computations involving concentration data that contain ND results, i.e., censored results whose exact magnitudes are not known but which are reported as being below specified limits. In such computations, ND values can be treated by various methods including substitution, as well as parametric and non-parametric statistical methods. Each specific type of treatment affects the computational results and hence has the potential to influence the ultimate NRDA decisions. Based on a review of available literature, the following recommendations are provided with respect to the treatment of NDs in the context of NRDA.

1) Every effort must be made to assign the appropriate and consistent DL values to the censored NDs within the LANL database. These efforts should be coordinated with the assistance of LANL contract laboratories and regulatory community. Use of correct DL values is expected to have substantial effects on computed NRDA-related statistics, especially for those datasets having large proportions of NDs.[69]

2) For all radionuclide results, and for "estimated" (J-flagged) inorganic and organic results, use of the uncensored value in NRDA related statistical computations is recommended. Although J-flagged results do not meet the data quality requirements of analytical laboratories, treating them as censored data is sub-optimal. [70]

3) Considering recent precedent established during the *Deepwater Horizon* NRDA studies and with typical practice in other NRDAs, when calculating total concentrations of a group of chemically-related constituents, such as total PAHs, individual NDs should be replaced with zero to avoid cumulative overestimation of effects.

4) When analyzing concentration results associated with individual constituents, the following recommendations are provided:

[69] The recommendations provided in the report pertain to samples for which DLs are provided. Treatment of samples where metadata are incomplete is beyond the scope of this report, as is conducting further review of database information to evaluate the extent to which appropriate DL values are specified in different fields within *Intellus*. We do note, however, that existing data are being evaluated as part of a separate task. Evaluation of existing data will include an evaluation of data completeness, as well as quality, and the Trustees have already established a Quality Management Plan (i.e., Appendix B of LANLTC 2014) that gives general guidance related to the use of data with differing levels of quality. The recommendations in this report are intended to complement, not replace, that guidance.

[70] ITRC (2013) states "In general, a measured concentration (detect) greater than the MDL but less than the reporting limit only reliably demonstrates the chemical is present in the sample at some concentration significantly greater than that of a method blank. Nevertheless, it is generally preferable to utilize qualified detections at their measured values in statistical evaluations — despite their greater analytical uncertainty — rather than treating them as censored values reported to a reporting limit (that is, less than values)."

- For purposes of calculating summary statistics, the recommended methods are listed in Exhibit 5-1. Because we expect that in the large majority of NRDA datasets, the pattern of DLs will not be random (but rather will be clustered towards the low end of the concentration range), ½ DL substitution, followed by the calculation of the desired summary statistics, will be the preferred approach in most cases.

EXHIBIT 5-1 RECOMMENDED NON-DETECT TREATMENTS FOR SUMMARY STATISTICS COMPUTATIONS

PROPORTION OF NON-DETECTS	PATTERNS OF DETECTION LIMITS			
	DLs MAINLY OCCUR WITHIN LOW CONCENTRATIONS		RANDOM OCCURRENCE OF DLs	
	Reliable Distribution Assumptions	Unreliable Distribution Assumptions	Reliable Distribution Assumptions	Unreliable Distribution Assumptions
<15%	½ DL substitution			
15-50%	½ DL substitution		ROS or MLE	Kaplan Meier
50-70%	½ DL substitution		Kaplan Meier	
>70%	Use alternative summary statistics, such as the proportion of detects or the proportion of exceedances			

ROS: regression on order statistics.
MLE: maximum likelihood estimation.

- For performing NRDA-related statistical comparisons, the treatments listed in Exhibit 5-2 are recommended. If ½ DL substitution treatment is selected, appropriate comparative tests (DON, 2002, Section 4.2) should be considered after NDs are substituted. As noted above, because we expect that for the large majority of NRDA datasets, the pattern of detection limits will not be random, ½ DL substitution followed by an appropriate comparison test will be the preferred approach in most cases.

EXHIBIT 5-2 RECOMMENDED NON-DETECT TREATMENTS FOR STATISTICAL COMPARISONS

PROPORTION OF NON-DETECTS	PATTERNS OF DETECTION LIMITS	
	DLs MAINLY OCCUR WITHIN LOW CONCENTRATIONS	RANDOM OCCURRENCE OF DLs
<15%	½ DL substitution followed by appropriate comparison tests*	
15-50%	½ DL substitution followed by appropriate non-parametric comparison tests**	Generalized Wilcoxon for unpaired samples; Paired Prentice-Wilcoxon for paired samples
>50%	Test of proportions	

* For large, normally-distributed and/or low-variance datasets, parametric tests, such as Student's t test, are appropriate. Otherwise, non-parametric tests are recommended.
** Examples of non-parametric comparison tests include the Slippage, Quantile and Wilcoxon Rank Sum tests.

- For performing NRDA-related correlation or regression analyses, treatments listed in Exhibit 5-3 are recommended. If ½ DL substitution is selected, appropriate methods for correlation/regression should be identified after NDs are substituted. As noted above, because we expect that for the large majority of NRDA datasets, the pattern of DLs will not be random (but rather will be clustered towards the low end of the concentration range), in practice, ½ DL substitution followed by an appropriate correlation/regression procedure, will be the preferred approach in most cases.

EXHIBIT 5-3 RECOMMENDED NON-DETECT TREATMENTS FOR CORRELATION AND REGRESSION ANALYSES

PROPORTION OF NON-DETECTS	PATTERNS OF DETECTION LIMITS	
	DLs MAINLY OCCUR WITHIN LOW CONCENTRATIONS	RANDOM OCCURRENCE OF DLs
<15%	½ DL substitution followed by appropriate procedures*	
15-80%	½ DL substitution followed by appropriate non-parametric procedures**	Modified Kendall's tau for correlation; Akritas-Theil-Sen line for regression
>80%	Phi coefficient method for correlation; Regression analysis is not recommended	

* For large, normally-distributed and/or low-variance datasets, parametric procedures, such as Pearson's *r* for correlation and linear regression, are appropriate. Otherwise, non-parametric procedures are recommended.
** Examples of non-parametric procedures include Kendall's tau for correlation and Theil-Sen line[71] for regression.

[71] Note that Theil-Sen line is not the same as the Akritas-Theil-Sen line; the latter is a modified version of Theil-Sen line when some of data are censored.

Although the above tables provide guidance on how to handle datasets with NDs, under certain circumstances, the Trustees may wish to undertake more than one analysis of the data to explore the effects of different methods on the ultimate result. However, before pursuing such sensitivity analyses, the applicability of alternative approaches should be demonstrated. Specifically, the underlying assumptions associated with the selected alternative approach should be fully met.

Sensitivity analyses are most likely to be appropriate when multiple approaches are equally likely to be applicable. For example, when proportion of non-detects is intermediate (i.e., neither very low nor very high), and when a dataset includes both high and low detection limits throughout the range of results, the Trustees might consider implementing methods listed in the right-hand portions of Tables 5-1 through 5-3, as applicable, instead of ½ DL substitution.

That being said, the Trustees should not feel obligated to conduct such sensitivity analyses, particularly when they have reason to believe that doing so is unlikely to materially affect the result. For example, if it is clear that substitution with the full DL (not a recommended approach due to its inherent high-bias) would result in exposure that is below relevant thresholds of concern, then conducting more complicated ROS, MLE, or KM analyses for sensitivity analysis purposes may not be warranted.

REFERENCES

Aboueissa, A. E.-M. A.; Stoline, M.R. (2004) Estimation of the mean and standard deviation from normally distributed singly-censored samples. *Environmetrics* 15(7): 659-673.

Aboueissa, A. E.-M. A.; Stoline, M.R. (2006) Maximum likelihood estimators of population parameters from doubly-left-censored samples. *Environmetrics* 17(8): 811-826.

Akritas, M.G.; Ruscitti, T.F.; Patil, G.P. (1994) 7 Statistical analysis of censored environmental data. *Handbook of Statistics* 12: 221-242.

Allert, Ann., DiStefano, R.J., Fairchild, J.F., Schmitt, C.J., and Brumbaugh, W.G. (2010) Effects of mining-derived metals on riffle-dwelling crayfish and in-situ toxicity to juvenile *Orconectes hylas* and *Orconectes luteus* in the Big River of southeast Missouri, USA. Prepared in cooperation with the Missouri Department of Conservation.

Alpha Analytical Laboratories (2016) Frequently Asked Questions https://alphalab.com/index.php/support-services/faq-frequently-asked-questions

The Anniston PCB Site Trustee Council. (2010). Anniston PCB Site State I Assessment Plan.

American Petroleum Institute (API) (2002) Analytical Detection and Quantification Limits: Survey of State and Federal Approaches. Regulatory Analysis and Scientific Affairs Publication Number 4721. June.

ANSI/HPS Standard N13.30-2011. Performance Criteria for Radiobioassay. Approved December 16, 2011. American National Standards Institute, Inc.

Antweiler, R.C.; H.E. Taylor (2008) Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: I. Summary statistics. *Environ Sci Technol* 42(10): 3732-3738.

Antweiler, R.C.; Taylor, H.E. (2015) Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: II. Group Comparisons. *Environ Sci Technol* 49(22): 13439-13446.

Armbruster, D.A.; Pry, T. (2008) Limit of Blank, Limit of Detection and Limit of Quantitation. *Clin Biochem Rev* 29(i):S49-S52.

Arunajadai, S.; Rauh, V. (2012) Handling covariates subject to limits of detection in regression. *Environ Ecol Stat* 19(3): 369-391.

Asimalowo, A.A.; Day, R.D.; Thaung, K.S. Ambient Air Concentration of Hexavalent Chromium in District of Columbia: Any Health Concern? District Department of the Environment, Washington, DC.

Baccarelli, A.; Pfeiffer, R.; Consonni, D.; Pesatori, A.C.; Bonzini, M.; Patterson, D.G.; Bertazzi, P.A.; Landi, M.T. (2005) Handling of dioxin measurement data in the presence of non-detectable values: Overview of available methods and their application in the Seveso chloracne study. *Chemosphere* 60(7): 898-906.

Ballenberger, N.; Lluis, A.; von Mutius, E.; Illi, S.; Schaub, B. (2012) Novel statistical approaches for non-normal censored immunological data: analysis of cytokine and gene expression data. *PLoS One* 7(10): e46423.

Bárdossy, A. (2011) Interpolation of groundwater quality parameters with some values below the detection limit. *Hydrol Earth Syst Sci* 15(9): 2763-2775.

Barghi, M.; Choi, S.-D.; Kwon, H.-O.; Lee, Y.-S.; Chang, Y.-S. (2016) Influence of non-detect data-handling on toxic equivalency quantities of PCDD/Fs and dioxin-like PCBs: A case study of major fish species purchased in Korea. *Environ Pollut* 214: 532-538.

Baron, M.G., Anderson, M., Beltman, D., Podrabsky, T., Walsh, W., Cacela, D. et al. 1999. Association between PCBs, Liver Lesions, and Biomarker Responses in Adult Walleye (*Stizostedium vitreum vitreum*) Collected from Green Bay, Wisconsin.

Beal, D. (2014) A macro for calculating summary statistics on left censored environmental data using the Kaplan-Meier method. Science Applications International Corporation, Oak Ridge, Tennessee.

Brownstone, D.; Valletta, R. (2001) The bootstrap and multiple imputations: harnessing increased computing power for improved statistical tests. *J Econ Perspect* 15(4): 129-141.

Buccianti, A.; Nisi, B.; Martin-Fernandez, J.A.; Palarea-Albaladejo, J. (2014) Methods to investigate the geochemistry of groundwaters with values for nitrogen compounds below the detection limit. *J Geochem Explor* 141: 78-88.

Buhamra, S. S. (1998) The analysis of VOCs survey data from residences in Kuwait. *Environmetrics* 9(3): 245.

Bursian, S.J; Kern, J.; Remington, R.E.; Link, J.E.; Fitzgerald, S.D. (2013) Dietary exposure to fish from the upper Hudson River, New York, USA: Effects on reproduction and offspring growth and mortality. *Environmental Toxicology* 32(4):780-793.

Busschaert, P.; Geeraerd, A.H.; Uyttendaele, M.; Van Impe, J.F. (2011) Hierarchical Bayesian analysis of censored microbiological contamination data for use in risk assessment and mitigation. *Food Microbiol* 28(4): 712-719.

Cano-Sancho, G.; Marín, S.; Ramos, A.J.; Sanchis, V. (2012) Exposure assessment of T2 and HT2 toxins in Catalonia (Spain). *Food Chem Toxicol* 50(3-4): 511-517.

Chen, H.; Quandt, S.A.; Grzywacz, J.G.; Arcury, T.A. (2010) A Distribution-Based Multiple Imputation Method for Handling Bivariate Pesticide Data with Values below the Limit of Detection. *Environ Health Perspect* 119(3): 351-356.

Chen, H.; Quandt, S.A.; Grzywacz, J.G.; Arcury, T.A. (2013) A Bayesian multiple imputation method for handling longitudinal pesticide data with values below the limit of detection. *Environmetrics* 24(2): 132-142.

Chowdhury, F.; Gulshan, J. (2012) Comparison of estimation methods for left censored data. International Conference on Statistical Mining for Bioinformatics, Health Agriculture, and Environment, Bangladesh.

Clean Water Act Methods Update Rule for the Analysis of Effluent. (2015). 80 Fed. Reg. 8956 (February 19, 2015) (to be codified at 40 C.F.R. Part 136).

Colorado Department of Public Health and Environment (CDPHE) (2014). *Practical Quantitation Limitation Guidance Document*. Water Quality Control Division, Laboratory Services Division. Original date July 2008.

Consolvo, J.; Sukosky, J.M. (2011) Monitoring and Compliance: Radiological lab results don't have to be confusing. *Opflow*: July.

Croghan, C.W.; Egeghy, P.P. (2003) Methods of dealing with values below the limit of detection using SAS. Presented at Southeastern SAS User Group, St. Petersburg, FL, September 22-24, 2003.

Currie, L.A. (1968) Limits for qualitative detection and quantitative determination. Application to radiochemistry. *Anal Chem* 40(3): 586-593.

Daniel, D.L. (2015) A Case Study Perspective on Working with ProUCL and a State Environmental Agency in Determining Background Threshold Values. *Int J Environ Res Public Health* 12(10): 12905-12923.

de Gavelle, E.; de Lauzon-Guillain, B.; Charles, M.-A.; Chevrier, C.; Hulin, M.; Sirot, V.; Merlo, M.; Nougadère, A. (2016) Chronic dietary exposure to pesticide residues and associated risk in the French ELFE cohort of pregnant women. *Environ Int* 92: 533-542.

Delistraty, D.A.; Laflamme, D.M. (2001) Influence of toxic equivalency factor scheme and method for treating non-detect values on soil dioxin levels. *Toxicol Environ Chem* 80(1-2): 67-81.

Department of Navy (DON) (2002) Guidance for Environmental Background Analysis, Volume I: Soil. NFESC User's Guide UG-2049-ENV, Naval Facilities Engineering Command, Washington, DC.

EcoChem, Inc. and Geosphere. 2002. Natural Resources Damage Allocation of Injuries to Natural Resources in the Hylebos Waterway. Prepared for NOAA Damage Assessment Center, U.S. Department of Justice.

Efron, B. (1992) Missing data, imputation, and the bootstrap. Technical Report No. 153, Division of Biostatistics, Stanford University, Stanford, CA.

Floit, S.B.; Mahoney, L.A.; Batey, J.C.; Petroff, D.M. (1996) Evaluation of the Use of Substitution Methods to Represent Nondetect Data. Superfund Risk Assessment in Soil Contamination Studies: Second Volume: ASTM International.

Forth, H.P.; Morris, J.M.; Cacela, D. (2015) *Explanation of Analytes Included in the Total Polycyclic Aromatic Hydrocarbon Sums Used by the Deepwater Horizon Natural Resource Damage Assessment Toxicity Group*. DWH NRDA Toxicity TWG, Boulder, CO https://pub-dwhdatadiver.orr.noaa.gov/dwh-ar-documents/952/DWH-AR0155464a.pdf

Frome, E.L.; Watkins, J.W. (2004) Statistical analysis of data with non-detectable values. ORNL/TM-2004/146, Oak Ridge, TN.

Frome, E.L.; Wambach, P.F. (2005) Statistical methods and software for the analysis of occupational exposure data with non-detectable values. ORNL/TM-2005/52, Oak Ridge, TN.

Fusek, M.; Michâlek, J.; Vâvrovâ, M. (2015) Evaluation of Contamination Data with Non-detects using Censored Distributions. *Fresenius Environmental Bulletin* 24(11c): 4165-4172.

Ganser, G.H.; Hewett, P. (2010) An Accurate Substitution Method for Analyzing Censored Data. *J Occup Environ Hyg* 7(4): 233-244.

Gardner, M. (2012) Improving the interpretation of 'less than' values in environmental monitoring. *Water and Environment Journal* 26(2): 285-290.

Gauch H. G. (2003) *Scientific Method in Practice*. Cambridge University Press, 2003.

Georgia Department of Natural Resources (GDNR), South Carolina Department of Natural Resources, South Carolina Department of Health and Environmental Control, United States Army Corps of Engineers, United States Fish and Wildlife Service 2006. Lake Hartwell Restoration and Compensation Determination Plan: Final. With Assistance from Exponent, Industrial Economics, Inc., and Schlumberger Technology Corporation.

Gibbons, R.D.; Coleman, D.E. (2001) Statistical methods for detection and quantification of environmental contamination. New York: John Wiley & Sons.

Ginevan, M.E.; Splitstone, D.E. (2002) Bootstrap upper bounds for the arithmetic mean of right-skewed data, and the use of censored data. *Environmetrics* 13(5-6): 453-464.

Gochfeld, M.; Burger, J.; Vyas, V. (2005) Statistical analysis of data sets with values below detection limits. Consortium for Risk Evaluation with Stakeholder Participation, 3.

Grima, J.; Luque-Espinar, J.A.; Mejía-Gómez, J.Á.; Rodríguez, R. (2014) Analysis of Groundwater Monitoring Data Sets with Non-Detect Observations: Application to the Plana de Sagunto (Valencia, Spain) Groundwater Body. Engineering and Physical Sciences Research Council (EPSRC) Centre for Doctoral Training, Mathematics of Planet Earth, Springer: 507-512.

He, J. (2013) Mixture model based multivariate statistical analysis of multiply censored environmental data. *Adv Water Resour* 59: 15-24.

Hefley, T.J.; Tyre, A.J.; Baasch, D.M.; Blankenship, E.E. (2013) Nondetection sampling bias in marked presence-only data. Ecology and Evolution 3(16): 5225-5236.

Helsel, D.R. (1990) Less than obvious: statistical treatment of data below the detection limit. *Environ Sci Technol* 24(12): 1767-1774.

Helsel, D.R. (2004) Statistical Analysis of Nondetects in Biogeochemical Data. Crustal Imaging and Characterization, USGS, Denver Federal Center, MS 964, Denver, CO.

Helsel, D.R. (2005) Nondetects and Data Analysis: Statistics for Censored Environmental Data. John Wiley & Sons, Hoboken, New Jersey.

Helsel, D.R. (2006) Fabricating data: how substituting values for nondetects can ruin results, and what can be done about it. *Chemosphere* 65(11): 2434-2439.

Helsel, D.R. (2009) Much ado about next to nothing: incorporating nondetects in science. *Ann Occup Hyg* 54(3): 257-262.

Helsel, D.R. (2010) Summing nondetects: incorporating low-level contaminants in risk assessment. *Integr Environ Assess Manag* 6(3): 361-366.

Helsel, D.R. (2012) Statistics for Censored Environmental Data Using Minitab and R (2nd ed.). Hoboken, New Jersey: John Wiley & Sons.

Helsel, D.R.; Cohn, T.A. (1988) Estimation of Descriptive Statistics for Multiply Censored Water Quality Data. *Water Resour Res* 24(12): 1997-2004.

Helsel, D.R.; R.M. Hirsch (1992) Statistical Methods in Water Resources. Elsevier, Amsterdam.

Helsel, D.R.; Lee, L. (2008) NADA for R. A contributed package for censored environmental data. United States Geological Survey (USGS), Reston, VA.

Hewett, P.; Ganser, G.H. (2007) A comparison of several methods for analyzing censored data. *Ann Occup Hyg* 51(7): 611-632.

Higgins, R.M.; Danilowicz, B.S.; Brophy, D.; Geffen, A.J.; McGowan, T.; Gillanders, B. (2013) Influence of the limit of detection on classification using otolith elemental signatures. *Can J Fish Aquat Sci* 70(6): 922-929.

Hoffman, H.J.; Johnson, R.E. (2015) Pseudo-likelihood estimation of multivariate normal parameters in the presence of left-censored data. *J Agric Biol Environ Stat* 20(1): 156-171.

Hoogerbrugge, R.; Liem, A.K.D. (2000) How to handle non-detects. *Organohalogen Compd* 45: 13-16.

Hornung, R.W.; Reed, L.D. (1990) Estimation of average concentrations in the presence of nondetectable values. *Journal of Applied Occupational and Environmental Hygiene* 5(1):46-51.

Hudson River Natural Resource Trustees (HRNRT). 2007. Work Summary and Data Report: Collection of Bats from the Hudson River, New York: 2001 and 2002 Samples. Public Release Version.

Hudson River Natural Resource Trustees (HRNRT). 2008. Injury Determination Report: Hudson River Surface Water Resources. Public Release Version. December.

Hudson River Natural Resource Trustees (HRNRT). 2013. PCB Contamination of the Hudson River Ecosystem: Compilation of Contamination Data Through 2008. January.

Hudson River Natural Resource Trustees (HRNRT). 2016. Data Report for the Characterization of Sediment Chemistry, Sediment Toxicity, and Benthic Invertebrate Community Structure for PCB-contaminated Sediments from the Upper Hudson River, New York. Public Release Version. October 13.

Huston, C.; Juarez-Colunga, E. (2009) Guidelines for Computing Summary Statistics for Data-Sets Containing Non-Detects. Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC.

Huybrechts, T.; Thas, O.; Dewulf, J.; Van Langenhove, H. (2002) How to estimate moments and quantiles of environmental data sets with non-detected observations? A case study on volatile organic compounds in marine water samples. *J Chromatogr A* 975(1): 123.

Huynh, T.; Ramachandran, G.; Banerjee, S.; Monteiro, J.; Stenzel, M.; Sandler, D.P.; Engel, L.S.; Kwok, R.K.; Blair, A.; Stewart, P.A. (2014) Comparison of Methods for Analyzing Left-Censored Occupational Exposure Data. *Ann Occup Hyg* 58(9): 1126-1142.

Huynh, T.; Quick, H.; Ramachandran, G.; Banerjee, S.; Stenzel, M.; Sandler, D.P.; Engel, L.S.; Kwok, R.K.; Blair, A.; Stewart, P.A. (2016) A Comparison of the β-Substitution Method and a Bayesian Method for Analyzing Left-Censored Data. *Ann Occup Hyg* 60(1): 56-73.

Illinois Natural Resource Trustee Council (2006) Assessment plan for the former Indian Refinery NPL Site. June.

Imaizumi, Y.; Suzuki, N.; Shiraishi, H. (2006) Bootstrap methods for confidence intervals of percentiles from dataset containing nondetected observations using lognormal distribution. *J Chemometr* 20(1-2): 68-75.

Industrial Economics, Inc. and CH2M HILL. 2002. Montrose Settlements Restoration Program: Fish Sampling Plan. Prepared for NOAA Damage Assessment Center and the Montrose Settlements Restoration Program.

International Union of Pure and Applied Chemistry (IUPAC) (1997) *Compendium of Analytical Nomenclature. Definitive Rules.* Prepared by Inezédy, J; Lengyel,T; Ure, AM. Third Edition.

International Union of Pure and Applied Chemistry (IUPAC) Compendium of Chemical Terminology, 2nd ed. (the "Gold Book"). Compiled by A. D. McNaught and A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). XML on-line

corrected version: http://goldbook.iupac.org (2006-) created by M. Nic, J. Jirat, B. Kosata; updates compiled by A. Jenkins. ISBN 0-9678550-9-8. doi:10.1351/goldbook. Last update: 2014-02-24; version: 2.3.3. See specifically: http://goldbook.iupac.org/L03540.html

Interstate Technology and Regulatory Council (ITRC) (2013) Groundwater Statistics for Monitoring and Compliance, Statistical Tools for the Project Life Cycle. GSMC-1. Accessed April 2017 at http://www.itrcweb.org/gsmc-1/.

Ito, T.; Kato, T.; Takagishi, K.; Okabe, S.; Sano, D. (2015) Bayesian modeling of virus removal efficiency in wastewater treatment processes. Water Sci Technol 72(10): 1789-1795.

Jain, R.B. (2016) On the consequence of substituting maximum likelihood estimates for the observations below the limit of detection. *Chemosphere* 144: 2044-2051.

Jin, Y.; Hein, M.J.; Deddens, J.A.; Hines, C.J. (2011) Analysis of lognormally distributed exposure data with repeated measures and values below the limit of detection using SAS. *Ann Occup Hyg* 55(1): 97-112.

Kafatos, G.; Andrews, N.; McConway, K.J.; Farrington, P. (2013) Regression models for censored serological data. *J Med Microbiol* 62(1): 93-100.

Kaplan, E.L. and Meier, P. (1958) Nonparametric Estimation from Incomplete Observations. Journal of the American Statistical Association 53(282): 457-481.

Kato, T.; Miura, T.; Okabe, S.; Sano, D. (2013) Bayesian modeling of enteric virus density in wastewater using left-censored data. *Food and environmental virology* 5(4): 185-193.

Keith, L.H.; Crummett, W.; Deegan Jr., J.; Libby, R.A.; Taylor, J.K.; Wentler, G. (1983) Principles of environmental analysis. *Anal Chem* 55(14):271-279.

Khokan, M.R.; Bari, W.; Khan, J.A. (2013). Weighted maximum likelihood approach for robust estimation: Weibull model. Dhaka Uhiv. J. Sci 61(2):153-156.

Kirchner, G; Steiner, M; Zahringer, M. (2009) A new approach to estimate nuclide ratios from measurements with activities close to background. *J Environ Radioact* 100(6): 484-488.

Krall, J.; Simpson, C.; Peng, R. (2015) A model-based approach for imputing censored data in source apportionment studies. *Environ Ecol Stat* 22(4): 779-800.

Krausmann, J.D. (1999) Data Report on a Reconnaissance-level Assessment for Avian Injury in Commencement Bay. Commencement Bay, Washington. Report No EC/DA98-1: Environmental Contaminants Program, Natural Resource Damage Assessment. April.

Kreinovich, V.; Longpré, L.; Starks, S.A.; Xiang, G.; Beck, J.; Kandathi, R.; Nayak, A.; Ferson, S.; Hajagos, J. (2007) Interval versions of statistical techniques with applications to environmental analysis, bioinformatics, and privacy in statistical databases. *J Comput Appl Math* 199(2): 418-423.

Krishnamoorthy, K.; Mallick, A.; Mathew, T. (2009) Model-Based Imputation Approach for Data Analysis in the Presence of Non-detects. *Ann Occup Hyg* 53(3): 249-263.

Krishnamoorthy, K.; Xu, Z. (2011) Confidence Limits for Lognormal Percentiles and for Lognormal Mean Based on Samples with Multiple Detection Limits. *Ann Occup Hyg* 55(5): 495-509.

Kroll, C.N.; Stedinger, J.R. (1999) Development of regional regression relationships with censored data. *Water Resour Res* 35(3): 775-784.

Lambert, D.; Peterson, B.; Terpenning, I. (1991) Nondetects, detection limits, and the probability of detection. *J Am Stat Assoc* 86(414): 266-277.

Latta, B.C. 2012. 2007 Channel Islands Peregrine Falcon Study, Final Report. Prepared for the U.S. Fish and Wildlife Service, Carlsbad, CA. Project No. 9820002.

LeFrancois, M.; Poeter, E. (2009) Use of Observations below Detection Limit for Model Calibration. *Ground Water* 47(2): 228-236.

Leith, K.F.; Bowerman, W.W.; Wierda, M.R.; Best, D.A.; Grubb, T.G.; Sikarske, J.G. (2010) A comparison of techniques for assessing central tendency in left-censored data using PCB and p,p'DDE contaminant concentrations from Michigan's Bald Eagle Biosentinel Program. *Chemosphere* 80(1): 7-12.

Levitan, D.M.; Schreiber, M.E.; Seal, R.R.; Bodnar, R.J.; Aylor, J.G. (2014) Developing protocols for geochemical baseline studies: An example from the Coles Hill uranium deposit, Virginia, USA. *Appl Geochem* 43: 88-100.

Li, S.; Batterman, S.; Su, F.C.; Mukherjee, B. (2013) Addressing extrema and censoring in pollutant and exposure data using mixture of normal distributions. *Atmos Environ* 77: 464-473.

Long, G.L.; Winefordner, J.D. (1983) Limit of detection: A closer look at the IUPAC definition. *Analy Chem* 55(7):712A-724A.

Los Alamos National Laboratory Natural Resources Trustee Council (LANLTC). 2014. Final Natural Resource Damage Assessment Plan for Los Alamos National Laboratory. February, 2014. Available at: www.lanlnrda.org.

Lubin, J.H.; Colt, J.S.; Camann, D.; Davis, S.; Cerhan, J.R.; Severson, R.K.; Hartge, P. (2004) Epidemiologic Evaluation of Measurement Data in the Presence of Detection Limits. *Environ Health Perspect* 112(17): 1691-1696.

MacDonald, D.D., Smorong, D.E., Lindskoog, R.A., and Ingersoll, C.G. 2003. As Assessment of Injury to Human Uses of Fishery Resources in the Grand Calumet River and Indiana Harbor Canal, the Grand Calumet River Lagoons, and Indiana Harbor and the Nearshore Areas of Lake Michigan. In Association with Industrial Economics, Inc. Prepared for U.S. Fish and Wildlife Service Bloomington Field Office.

Michigan Department of Environmental Quality (MDEQ), Michigan Department of Attorney General, U.S. Fish and Wildlife Service (USFWS), National Oceanic

and Atmospheric Administration (NOAA). 2000. Preassessment Screen: Kalamazoo River Environment Site, Michigan. Prepared by Stratus Consulting, Inc. under subcontract to DLZ Corporation.

Michigan Department of Environmental Quality (MDEQ), Michigan Department of Attorney General, U.S. Fish and Wildlife Service (USFWS), National Oceanic and Atmospheric Administration (NOAA). 2005. Stage I Assessment Report, Volume 1—Injury Assessment: Kalamazoo River Environment: Final Report. Prepared by Stratus Consulting, Inc.

Millard, S.P.; Deverel, S.J. (1988) Nonparametric statistical methods for comparing two sites based on data with multiple nondetect limits. *Water Resour Res* 24(12): 2087-2098.

Mosby, D.E., Weber, J.S., and Klahr, F. 2009. Final Phase I Damage Assessment Plan for Southeast Missouri Lead Mining District: Big River Mine Tailings Superfund Site, St. Francois County and Viburnum Trend Sites, Reynolds, Crawford, Washington, and Iron Counties. Prepared for State of Missouri, Missouri Department of Natural Resources, U.S. Fish and Wildlife Service, U.S. Department of the Interior.

Multi-Agency Radiation Survey and Site Investigation Manual (MARSSIM) (2000) NUREG-1575, Rev. 1. EPA 40-R-97-016, Rev. 1. DOE/EH-0624, Rev. 1. August.

Munoz, G.; Giraudel, J.-L.; Botta, F.; Lestremau, F.; Dévier, M.-H.; Budzinski, H.; Labadie, P. (2015) Spatial distribution and partitioning behavior of selected poly- and perfluoroalkyl substances in freshwater ecosystems: A French nationwide survey. *Sci Total Environ* 517: 48-56.

Myung, I.J. (2003) Tutorial on maximum likelihood estimation. *J Math Psychol* 47(1): 90-100.

Newman, M.C.; Dixon, P.M.; Looney, B.B.; Pinder, J.E. (1989) Estimating Mean and Variance for Environmental Samples With Below Detection Limit Observations. *Water Resour Bull* 25(4): 905-916.

New Mexico Office of Natural Resources Trustee. 2010. Preassessment Screen and Determination: Rio Algom Mines and Quivira Mill Site, McKinley County, New Mexico.

Newton, E.; Rudel, R. (2007) Estimating Correlation with Multiply Censored Data Arising from the Adjustment of Singly Censored Data. *Environ Sci Technol* 41(1): 221-228.

National Oceanic and Atmospheric Administration (NOAA). (2013) Final Lower Duwamish River NRDA Restoration Plan and Programmatic Environmental Impact Statement. Appendix H. Prepared on behalf of the Lower Duwamish River Natural Resource Damage Assessment Trustee Council.

National Oceanic and Atmospheric Administration and U.S. Environmental Protection Agency (NOAA and EPA) (2007) 2002-2004 Southern California coastal marine fish contaminants survey. June.

National Oceanic and Atmospheric Administration National Ocean Service Office of Response and Restoration (NOAA) and Ridolfi Inc. (2006) Final Data Report for the Vieques Island Biota Sampling Project: Vieques Island, Puerto Rico.

National Oceanic and Atmospheric Administration, National Ocean Service Office of Response and Restoration and U.S. Fish and Wildlife Service (NOAA and USFWS). (2007) Diamond Alkali Superfund Site Natural Resource Damage Assessment Plan. Public Review Draft.

Natural Resource Trustee for the State of Colorado (NRTSC). 2007. Natural Resource Damage Assessment Plan for the Rocky Mountain Arsenal, Commerce City, Colorado.

Natural Resource Trustees of the St. Lawrence River Environment (NRTSLRE). 2013. St. Lawrence River Environment Natural Resource Damage Assessment: Restoration and Compensation Determination Plan and Environmental Assessment. Revised. Prepared with assistance from DelVecchio, R., Unsworth, R., Leggett, C., Industrial Economics, Inc. May.

Natural Resource Trustees of the Tennessee Valley Authority Kingston Fossil Plant Coal Ash Release Natural Resource Damage Assessment (NRTTVA). 2015. Tennessee Valley Authority Kingston Fossil Plant Coal Ash Release Natural Resource Damage Assessment: Restoration and Compensation Determination Plan. Prepared with assistance from DelVecchio, R., Fydenkevez, J., Timberlake, T., Industrial Economics, Inc.

NIST/SEMATECH e-Handbook of Statistical Methods, Section 1.3.5.13, http://www.itl.nist.gov/div898/handbook/eda/section3/eda35d.htm, 2013.

Nysen, R.; Faes, C.; Ferrari, P.; Verger, P.; Aerts, M. (2015) Parametric and semi-nonparametric model strategies for the estimation of distributions of chemical contaminant data. *Environ Ecol Stat* 22(2): 423-444.

Oblinger Childress, C.J., Foreman, W.T., Connor, B.F. & Maloney, T.J. (1999) New Reporting Procedures Based on Long-Term Method Detection Levels and Some Considerations for Interpretations of Water-Quality Data Provided by the U.S. Geological Survey National Water Quality Laboratory. Open-File Report 99-193, U.S. Geological Survey, Reston, VA.

Office of the Colorado Attorney General. 2007. Preassessment Screen Determination. Prepared on behalf of, and in consultation with, Natural Resources Trustees for the State of Colorado. January 12.

Orton, T.G.; Rawlins, B.G.; Lark, R.M. (2009) Using measurements close to a detection limit in a geostatistical case study to predict selenium concentration in topsoil. *Geoderma* 152(3-4): 269-282.

Palarea-Albaladejo, J.; Martín-Fernández, J.A.; Olea, R.A. (2014) A bootstrap estimation scheme for chemical compositional data with nondetects. *J Chemometr* 28(7): 585-599.

Palmerton Natural Resource Trustee Council. 2006. Palmerton Zinc Pile Superfund Site Natural Resource Damage Assessment Plan. February.

Palmerton Natural Resource Trustee Council 2007. Data Report for the Scoping Study on Metal Contaminant Levels in Forest Soils and Concurrent Habitat Evaluation for the Palmerton Zinc Natural Resource Damage Assessment, Palmerton, Pennsylvania. January 30.

Pearmain, S. (2015) The importance of statistical data analysis and environmental risk assessment modelling in determining the significance of soil and groundwater contamination. 1st International Conference on Energy, Environment and Climate Changes, Mauritius.

Peto, R. and Peto, J. (1972) Asymptotically efficient rank invariant test procedures (with discussion). *Journal of the Royal Statistical Society Series A* 135:185-206.

Pita, G.L.; Francis, R.; Liu, Z.; Mitrani-Reiser, J.; Guikema, S.; Pinelli, J.P. (2011) Statistical Tools for Populating/Predicting Input Data of Risk Analysis Models. Vulnerability, Uncertainty, and Risk: pp. 468-476.

Popovic, M., Nie, H., Chettle, D. R. & McNeill, F. E. (2007). Random left censoring: A second look at bone lead concentration measurements. *Physics in Medicine and Biology* 52(17): 5369-5378.

Prentice, R.L. (1978) Linear rank tests with right-censored data. *Biometrika* 65: 167-179.

Prentice, R.L. and Marek, P. (1979) A qualitative discrepancy between censored data rank tests. *Biometrics* 35: 861-867.

Qian, S.S.; Schulman, A; Koplos, J; Kotros, A; Kellar, P (2004) A Hierarchical Modeling Approach for Estimating National Distributions of Chemicals in public Drinking Water Systems. *Environ Sci Technol* 38(4): 1176-11825.

Rieger, J.T. (2013) HEIS Sample, Result, and Sampling Site Data Dictionary. Prepared for the U.S. Department of Energy Assistant Secretary for Environmental Management. CH2M Hill. HNF-38155, Revision 1. January.

Roberts, A.D., Mosby, D., Weber, J., Besser, J., Hundley, J., McMurray, S., and Faiman, S.N.D. An Assessment of Freshwater Mussel (Bivalvia: Margaritiferidae and Unionidae) Populations and Heavy Metal Sediment Contamination in the Big River, Missouri.

Rouhani, S.; Baker, M.C.; Steinhoff, M.; Zhang, M.; Oehrig, J.; Zelo, I.J.; Emsbo-Mattingly, S.D.; Nixon, Z.; Willis, J.M.; and Hester; M.W. (2016) Nearshore Exposure to *Deepwater Horizon* Oil. *Marine Ecology Progress Series*. In press.

Sarver, K.M. (2015). Method Detection Limit (MDL) Procedure: EPA Proposed Revision. Cox-Colvin & Associates. https://www.coxcolvin.com/method-detection-limit-mdl-procedure-epas-proposed-revision/

Schein, A., MacDonald, D.D., Prencipe, H.J., Wainright, Sinclair, J.A., Theel, H.J., and Lorentz, W.P. (2015). Supplemental Investigation of Sediment and Soil Chemistry in the Vicinity of the Anniston PCB Site, Anniston, Alabama: Data Report. Prepared for Karen Marlowe, U.S. Fish and Wildlife Service.

Schmoyeri, R.L.; Beauchamp, J.J.; Brandt, C.C.; Hoffman Jr, F.O. (1996) Difficulties with the lognormal model in mean estimation and testing. *Environ Ecol Stat* 3(1): 81-97.

Scott, G.I.; Thompson, R.E.; Voit, E.O. (2000) Statistical modeling of sediment and oyster PAH contamination data collected at a South Carolina estuary (complete and left-censored samples). *Environmetrics* 11(1): 99.

Sharma, M.; McBean, E.A.; Thomson, N. (1995) Maximum Likelihood Method for Parameter Estimation in Linear Model with below-Detection Data. *J Environ Eng* 121(11): 776.

She, N. (1997.) Analyzing censored water quality data using a non-parametric approach. *Journal of the American Water Resources Association* 33(3): 615-624.

Shoari, N.; Dube, J.-S.; Chenouri, S. (2015) Estimating the mean and standard deviation of environmental data with below detection limit observations: Considering highly skewed data and model misspecification. *Chemosphere* 138: 599-608.

Shoari, N.; Dubé, J.-S.; Chenouri, S. (2016) On the use of the substitution method in left-censored environmental data. *Hum Ecol Risk Assess* 22(2): 435-446.

Shorten, P.R.; Pleasants, A.B.; Soboleva, T.K. (2006) Estimation of microbial growth using population measurements subject to a detection limit. *Int J Food Microbiol* 108(3): 369-375.

Silva, E.; Mendes, M.; Ribeiro, L.; Cerejeira, M. (2012) Exposure assessment of pesticides in a shallow groundwater of the Tagus vulnerable zone (Portugal): a multivariate statistical approach (JCA). *Environ Sci Pollut Res Int* 19(7): 2667-2680.

Singh, A. (2006) On the computation of a 95% upper confidence limit of the unknown population mean based upon data sets with below detection limit observations. Prepared for the U.S. Environmental Protection Agency. EPA/600/R-06/002. March 2006.

Sinha, P.; Lambert, M.B.; Trumbull, V.L. (2006) Evaluation of statistical methods for left-censored environmental data with nonuniform detection limits. *Environ Toxicol Chem* 25(9): 2533-2540.

Slymen, D.J.; de Peyster, A. (1994) Hypothesis testing with values below detection limit in environmental studies. *Environ Sci Technol* 28(5): 898.

State of Colorado. 2004. Habitat Equivalency Analysis for Terrestrial and Aquatic Resources at the Rocky Mountain Arsenal: Final. Prepared by Stratus Consulting, Inc.

Stoline, M.R. (1993) Comparison of two medians using a two-sample lognormal model in environmental contexts. *Environmetrics* 4(3): 323-339.

Succop, P.A.; Clark, S.; Chen, M.; Galke, W. (2004) Imputation of data values that are less than a detection limit. *J Occup Environ Hyg* 1(7): 436-441.

Thompson, M.L. (2003) Linear regression with Type I interval-and left-censored response data. *Environ Ecol Stat* 10(2): 221-230.

U.S. Army Corps of Engineers (USACE). (1995) Guidelines for statistical treatment of less than detection limit data in dredged sediment evaluations. EEDP-04-23. Waterways Experiment Station. Vicksburg, Mississippi.

U.S. Department of Energy (USDOE). (1998). The Department of Energy Laboratory Accreditation Program for Radiobioassay. DOE-STD-1112-98. December.

U.S. Department of Energy (USDOE). (2008a). DOE Standard: Internal Dosimetry. DOE-STD-1121-2008. Change Notice No. 1. October.

U.S. Department of Energy (USDOE). (2008b). Radiation Protection Programs Guide for use with Titel 10, Code of Federal Regulations, Part 835, Occupational Radiation Protection. DOE G 441.1-1C. May.

U.S. Department of Energy (USDOE). (2013a) Quality Systems for Analytical Services Revision 2.9. Retrieved from http://www.p2s.com/wp-content/uploads/QSAS-Rev-2.9.pdf.

U.S. Department of Energy (USDOE). (2013b). Good Practices for Occupational Radiological Protection in Plutonium Facilities. DOE-STD-1128-2013. April.

U.S. Department of Energy (USDOE). (2014). Hanford Analytical Services Quality Assurance Requirements Document. Volume 4: Laboratory Technical Requirements. DOE/RL-96-68, Rev. 4. September.

U.S. Environmental Protection Agency (1989) Method 1620: Metals by Inductively Coupled Plasma Atomic Emission Spectroscopy and Atomic Absorption Spectroscopy. Office of Water. Draft. 821-B-89-100. September.

U.S. Environmental Protection Agency (1991) Chemical Concentration Data Near the Detection Limit. EPA/903/8-91/001, Hazardous Waste Management Division, Philadelphia, PA.

U.S. Environmental Protection Agency (2000) Assigning Values to Non-Detected/Non-Quantified Pesticide Residues in Human Health Food Exposure Assessments. Office of Pesticide Programs, Washington, DC.

U.S. Environmental Protection Agency (2003) Method 8000D, Determinative Chromatographic Separations https://www.epa.gov/sites/production/files/2015-12/documents/8000d.pdf

U. S. Environmental Protection Agency (2004) Revised Assessment of Detection and Quantitation Approaches. EPA-821-B-04-005, October 2004.

U.S. Environmental Protection Agency (2006a) Data Quality Assessment: Statistical Methods for Practitioners. EPA/240-B-06/003. February 2006.

U.S. Environmental Protection Agency (2006b) IDL- MDL- PQL, What the "L" is going on? What does all this alphabet soup really mean? Region III Quality Assurance, Revision No.:2.5, MDL Factsheet, March 17, 2006. https://www.epa.gov/sites/production/files/2015-06/documents/whatthel.pdf

U.S. Environmental Protection Agency (2007) Report of the Federal Advisory Committee on Detection and Quantitation Approaches and Used in Clean Water Act programs, December 2007. https://www.epa.gov/sites/production/files/2015-10/documents/detection-quant-faca_final-report_2012.pdf

U.S. Environmental Protection Agency (2009) Statistical analysis of groundwater monitoring data at RCRA facilities: Unified guidance. EPA 530/R-09-007, Office of Resource Conservation and Recovery, Washington, DC.

U.S. Environmental Protection Agency (2010) Detection Limit/Quantitation Limit Summary Table. https://www.epa.gov/sites/production/files/2015-06/documents/mdlmql-toolbox-final-oct2010.pdf.

U.S. Environmental Protection Agency (2011) A laboratory study of procedures evaluated by the Federal Advisory Committee on detection and quantitation approaches and uses in Clean Water Act Programs. Office of Water (4303T), Washington, DC, including supplement. https://www.epa.gov/sites/production/files/2015-10/documents/laboratory-study-faca-report_2011.pdf and https://www.epa.gov/sites/production/files/2015-10/documents/supplement-lab-study-report_2011.pdf.

U.S. Environmental Protection Agency (2013) ProUCL Version 5.0.00 Technical Guide, EPA/600/R-07/041. Office of Research and Development, Washington, DC.

U.S. Environmental Protection Agency (2015a) EPA Contract Laboratory Program Statement of Work for Inorganic Superfund Methods Multi-Media, Multi-Concentration ISM02.3 Analytical Services Branch (ASB), Washington, DC. https://www.epa.gov/clp/epa-contract-laboratory-program-statement-work-inorganic-superfund-methods-multi-media-multi-0.

U.S. Environmental Protection Agency (2015b) EPA Contract Laboratory Program Statement of Work for Organic Superfund Methods Multi-Media, Multi-Concentration SOM02.3. Analytical Services Branch (ASB), Washington, DC. https://www.epa.gov/clp/epa-contract-laboratory-program-statement-work-organic-superfund-methods-multi-media-multi-0.

U.S. Environmental Protection Agency (2015c) ProUCL Version 5.1.002 Technical Guide, EPA/600/R-07/041. Office of Research and Development, Washington, DC. October.

U.S. Fish and Wildlife Service (USFWS), U.S. Department of Interior (DOI), U.S. Department of Justice (DOJ). 1999a. Injuries to Surface Water Resources, Lower

Fox River/Green Bay Natural Resource Damage Assessment. Prepared by Stratus Consulting, Inc.

U.S. Fish and Wildlife Service (USFWS), U.S. Department of Interior (DOI), U.S. Department of Justice (DOJ). 1999b. PCB Pathway Determination for the Lower Fox River/Green Bay Natural Resource Damage Assessment: Final Report. Prepared by Stratus Consulting, Inc.

U.S. Fish and Wildlife Service (USFWS). 2003. Preassessment Screen for the Chino, Tyrone, and Morenci Mine Sites, Grant County, New Mexico, and Morenci, Arizona. Prepared by Stratus Consulting, Inc.

U.S. Fish and Wildlife Service (USFWS) and Hagler Bailly Consulting. 1996. Assessment Plan: Lower Fox River/Green Bay NRDA. Prepared by Hagler Bailly Consulting Inc., Boulder, CO, for U.S. Fish and Wildlife Service, U.S. Department of Justice, U.S. Department of the Interior. August.

Uh, H.-W.; Hartgers, F. C.; Yazdanbakhsh, M.; Houwing-Duistermaat, J.J. (2008). Evaluation of regression methods when immunological measurements are constrained by detection limits. *BMC Immunol* 9(1): 59.

Verbovšek, T. (2011) A comparison of parameters below the limit of detection in geochemical analyses by substitution methods. *RMZ Mater Geoenvironment* 58: 393-404.

Vosnakis, K.A.S.; Perry, E.; Madsen, K.; Bradley, L.J.N. (2009) Background Versus Risk-Based Screening Levels - An Examination of Arsenic Background Soil Concentrations in Seven States. *International Journal of Soil, Sediment and Water* 2(2): 1-20.

Wang, W.; Wells, M.T. (2000) Estimation of Kendall's tau under censoring. *Statistica Sinica* 10(2000:1999-1215.

Wisconsin Department of Natural Resources (WDNR) Laboratory Certification Program (1996) Analytical Detection Limit Guidance & Laboratory Guide for Determining Method Detection Limits. PUBL-TS-056-96. April.

Weiss, J., Unsworth, R.E., and Ruder, E. 1997. Assessment Plan for the Natural Resource Damage Assessment of the Grand Calumet River, Indiana Harbor Ship Canal, Indiana Harbor, and Associated Lake Michigan Environments. Prepared for U.S. Fish and Wildlife Service and National Park Service, U.S. Department of the Interior; Department of Environmental Management and Department of Natural Resources, State of Indiana.

Wendelberger, J.; Campbell, K. (1994). Non-detect data in environmental investigations. Los Alamos National Laboratory, MS-F600, Los Alamos, NM.

Wells, G.; Russ, CW (2011) Why use Signal-to-Noise as a Measure of MS Performance When it is Often Meaningless/ Technical Overview. Agilent Technologies.

Wu, H.; Chen, Q.; Ware, L.B.; Koyama, T. (2012) A Bayesian approach for generalized linear models with explanatory biomarker measurement variables subject to detection limit: an application to acute lung injury. *J Appl Stat* 39(8): 1733-1747.

Yu, X.; Liu, P.; Min, J.; Chen, Q. (2009) Regression on order statistics and its application in estimating nondetects for food exposure assessment. *Journal of hygiene research* 38(1): 89-91.

Yuan, Y. C. (2010) Multiple imputation for missing data: Concepts and new development (Version 9.0). SAS Institute, Inc., Rockville, MD.

Zhang, H. (2013) Significance of Nondetects in the Mapping of Soil Contaminants. TRITA-LWR Degree Project 13:03, Royal Institute of Technology (KTH), Stockholm, Sweden.

Zhang, Z.; Lennox, W.C.; Panu, U.S (2004) Effect of percent non-detects on estimation bias in censored distributions. *J Hydrol (Amst)* 297(1-4): 74-94.

Zhao, Y.; Frey, H.C. (2006) Uncertainty for Data with Non-Detects: Air Toxic Emissions from Combustion. *Hum Ecol Risk Assess* 12(6): 1171-1191.

Zoffoli, H.J.O.; Varella, C.A.A.; do Amaral-Sobrinho, N.M.B.; Zonta, E.; Tolon-Becerra, A. (2013) Method of median semi-variance for the analysis of left-censored data: Comparison with other techniques using environmental data. *Chemosphere* 93(9): 1701-1709.

APPENDIX A | SUMMARY OF REVIEWED STATISTICAL REFERENCES

REFERENCE	SUMMARY NOTES
Aboueissa, A. E.-M. A.; Stoline, M.R. (2004) Estimation of the mean and standard deviation from normally distributed singly-censored samples. Environmetrics 15(7): 659-673.	A new computer algorithm for obtaining the Cohen (1959) maximum likelihood estimates of μ and σ is provided which does not require auxiliary tables. Closed form estimates of the mean and standard deviation obtained under a new replacement method are given for normally distributed left-censored samples.
Aboueissa, A. E.-M. A.; Stoline, M.R. (2006) Maximum likelihood estimators of population parameters from doubly-left-censored samples. Environmetrics 17(8): 811-826.	Estimators of the parameters are derived for left-censored data having two detection limits: DL1 and DL2 assuming an underlying normal distribution. Two different approaches for calculating the maximum likelihood estimates (MLE) are given and examined.
Akritas, M.G.; Ruscitti, T.F.; Patil, G.P. (1994) 7 Statistical analysis of censored environmental data. Handbook of Statistics 12: 221-242.	Various methods for parameter estimation are surveyed, including simple substitution of detection limits; maximum likelihood estimators; and probability plotting. Estimation of location difference in the 2-sample case is presented in the framework of extensions of the nonparametric Hodges-Lehmann estimator. Various regression methods for censored data are discussed, including maximum likelihood; Buckley-James; least absolute deviations; and Theil-Sen regression.
Antweiler, R.C.; H.E. Taylor (2008) Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: I. Summary statistics. Environ Sci Technol 42(10): 3732-3738.	For datasets with less than 70% censored data, the best technique overall for determination of summary statistics was the nonparametric Kaplan-Meier technique. ROS and the two substitution methods (1/2DL or random number between zero and DL) were adequate alternatives. At high degrees of censoring (greater than 70% censored data), no technique provided good estimates of summary statistics. Maximum likelihood techniques were found to be far inferior to all other treatments except substituting zero or the detection limit value to censored data.
Antweiler, R.C.; Taylor, H.E. (2015) Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: II. Group Comparisons. Environ Sci Technol 49(22): 13439-13446.	For low degrees of censoring (<25% in each group), the Generalized Wilcoxon (GW) technique and substitution of √2/2 times the detection limit gave overall the best results. For moderate degrees of censoring, MLE worked best, but only if the distribution could be estimated to be normal or log-normal prior to its application; otherwise, GW was a suitable alternative. For higher degrees of censoring (each group >40% censoring), no technique provided reliable estimates of the true probability
Arunajadai, S.; Rauh, V. (2012) Handling covariates subject to limits of detection in regression. Environ Ecol Stat 19(3): 369-391.	Propose the use of the generalized gamma distribution to estimate imputed values for the non-detects

REFERENCE	SUMMARY NOTES
Asimalowo, A.A.; Day, R.D.; Thaung, K.S. Ambient Air Concentration of Hexavalent Chromium in District of Columbia: Any Health Concern? District Department of the Environment, Washington, DC.	Substituting non-detect with MDL is very conservative. The true statistical reflection of ambient concentration is considered to be maximum likelihood estimation.
Baccarelli, A.; Pfeiffer, R.; Consonni, D.; Pesatori, A.C.; Bonzini, M.; Patterson, D.G.; Bertazzi, P.A.; Landi, M.T. (2005) Handling of dioxin measurement data in the presence of non-detectable values: Overview of available methods and their application in the Seveso chloracne study. Chemosphere 60(7): 898-906.	May produce dependable results even when 50-70% of the observations are non-detects and can be performed using commonly available statistical software. Therefore, suggest that distribution-based multiple imputation be the preferred method to analyze environmental data when substantial proportions of observations are non-detects.
Ballenberger, N.; Lluis, A.; von Mutius, E.; Illi, S.; Schaub, B. (2012) Novel statistical approaches for non-normal censored immunological data: analysis of cytokine and gene expression data. PLoS One 7(10): e46423.	For non-normally distributed censored data traditional means such as the Kaplan-Meier method or the generalized Wilcoxon test are described. Tobit regression on ranks meets these requirements and can be used for adjustment for covariates and potential confounding in large and complex immunological datasets.
Bárdossy, A. (2011) Interpolation of groundwater quality parameters with some values below the detection limit. Hydrol Earth Syst Sci 15(9): 2763-2775.	A mixed maximum likelihood approach is used to estimate the marginal distributions of the parameters. After removal of the marginal distributions the next step is the maximum likelihood estimation of the parameters of the spatial dependence including taking those values below the detection limit into account.
Barghi, M.; Choi, S.-D.; Kwon, H.-O.; Lee, Y.-S.; Chang, Y.-S. (2016) Influence of non-detect data-handling on toxic equivalency quantities of PCDD/Fs and dioxin-like PCBs: A case study of major fish species purchased in Korea. Environ Pollut 214: 532-538.	Substitution is not a suitable method to address non-detect (ND) data and can result in significant errors. The use of KM method is preferable for average TEQs, and if the use of KM is not applicable, substitution by zero is preferred.
Beal, D. (2014) A macro for calculating summary statistics on left censored environmental data using the Kaplan-Meier method. Science Applications International Corporation, Oak Ridge, Tennessee.	Kaplan-Meier has been shown to provide more robust estimates of the mean and standard deviation of left censored data than other methods such as simple substitution and maximum likelihood estimates.
Brownstone, D.; Valletta, R. (2001) The bootstrap and multiple imputations: harnessing increased computing power for improved statistical tests. J Econ Perspect 15(4): 129-141.	The bootstrap and multiple imputations are two techniques that can enhance the accuracy of estimated confidence bands and critical values.

REFERENCE	SUMMARY NOTES
Buccianti, A.; Nisi, B.; Martin-Fernandez, J.A.; Palarea-Albaladejo, J. (2014) Methods to investigate the geochemistry of groundwaters with values for nitrogen compounds below the detection limit. J Geochem Explor 141: 78-88.	Since nitrogen species are often affected by the presence of numerous data below the detection limit, their role was investigated by considering different imputation methods.
Buhamra, S. S. (1998) The analysis of VOCs survey data from residences in Kuwait. Environmetrics 9(3): 245.	Since the data are left censored and contain extreme values, traditional methods of analysis are inappropriate and, instead, a nonparametric method due to Akritas (1992, Statistics and Probability Letters, 13, 209-221) was used.
Busschaert, P.; Geeraerd, A.H.; Uyttendaele, M.; Van Impe, J.F. (2011) Hierarchical Bayesian analysis of censored microbiological contamination data for use in risk assessment and mitigation. Food Microbiol 28(4): 712-719.	Distributions are fit using Bayesian analysis, and results are compared to results obtained with a methodology based on maximum likelihood estimation and the non-parametric bootstrap method. The Bayesian model is also extended hierarchically to estimate the effects of the individual elements of a covariate.
Cano-Sancho, G.; Marín, S.; Ramos, A.J.; Sanchis, V. (2012) Exposure assessment of T2 and HT2 toxins in Catalonia (Spain). Food Chem Toxicol 50(3-4): 511-517.	Three different approaches were considered to handle the left censored data: (1) a substitution method, (2) a parametric method using the maximum likelihood estimation (MLE) and (3) a non-parametric method using the Kaplan–Meier (KM) estimator. Accuracy and reliability of the statistic estimates were assessed building the related confidence intervals using a pseudo-parametric bootstrap method.
Chen, H.; Quandt, S.A.; Grzywacz, J.G.; Arcury, T.A. (2010) A Distribution-Based Multiple Imputation Method for Handling Bivariate Pesticide Data with Values below the Limit of Detection. Environ Health Perspect 119(3): 351-356.	Simple substitution may bias parameter estimation. In contrast, multiple imputation (MI) methods yield valid and robust parameter estimates and explicit imputed values for variables that can be analyzed as outcomes or predictors. The distribution-based MI method is a valid and feasible approach to analyze bivariate data with values <LOD, especially when explicit values for the nondetections are needed.
Chen, H.; Quandt, S.A.; Grzywacz, J.G.; Arcury, T.A. (2013) A Bayesian multiple imputation method for handling longitudinal pesticide data with values below the limit of detection. Environmetrics 24(2): 132-142.	Bayesian multiple imputation estimates performed well in most settings, and [the authors] recommend the use of this valid and feasible approach to analyze multivariate data with values < LOD.
Chowdhury, F.; Gulshan, J. (2012) Comparison of estimation methods for left censored data. International Conference on Statistical Mining for Bioinformatics, Health Agriculture, and Environment, Bangladesh.	The present study aims to compare a nonparametric (Kaplan-Meier estimator), a parametric (Maximum Likelihood estimator) and a semi-parametric method (the Regression on Ordered Statistics (ROS) method) using simulated data under different censoring schemes, different sample sizes and different distributions.
Croghan, C.W.; Egeghy, P.P. (2003) Methods of dealing with values below the limit of detection using SAS. Presented at Southeastern SAS User Group, St. Petersburg, FL, September 22-24, 2003.	The extrapolation and maximum likelihood estimate techniques have smaller error rates than all the standard replacement techniques. Although more computational, these methods produce more reliable descriptive statistics.

REFERENCE	SUMMARY NOTES
Currie, L.A. (1968) Limits for qualitative detection and quantitative determination. Application to radiochemistry. Anal Chem 40(3): 586-593.	The questions of signal detection and signal extraction in analytical chemistry and nuclear chemistry were reexamined and three limiting levels, L_C, L_D and L_Q, were defined with exact equations and series of working formulae.
Daniel, D.L. (2015) A Case Study Perspective on Working with ProUCL and a State Environmental Agency in Determining Background Threshold Values. Int J Environ Res Public Health 12(10): 12905-12923.	Efforts were made to address: inappropriate outlier detection, upper tolerance limit (UTL) calculations based on gamma distributions when non-detects were present, and inappropriate use of nonparametric UTL formulas.
de Gavelle, E.; de Lauzon-Guillain, B.; Charles, M.-A.; Chevrier, C.; Hulin, M.; Sirot, V.; Merlo, M.; Nougadère, A. (2016) Chronic dietary exposure to pesticide residues and associated risk in the French ELFE cohort of pregnant women. Environ Int 92: 533-542.	To handle left-censored data: a lower-bound scenario (LB), where undetected results were set to zero, and an upper-bound scenario (UB), where undetected results were set to the detection limit if the substance was expected to be found in food and zero if it was not. A better management of left-censored data and more sensitive analyses of the main food contributors might help to refine the UB exposure and risk assessments.
Delistraty, D.A.; Laflamme, D.M. (2001) Influence of toxic equivalency factor scheme and method for treating non-detect values on soil dioxin levels. Toxicol Environ Chem 80(1-2): 67-81.	The purpose of this study was to evaluate the influence of four toxic equivalency factor (TEF) schemes (i.e., human/mammalian [Tl, T2], fish [T3], bird [T4]) and three substitution methods for treating non-detect (ND) data (i.e., ND = 0 [Nl], ND = 0.5 DL [N2], ND = DL [N3] where DL = detection limit) on polychlorinated dibenzo-p-dioxin (PCDD) and dibenzofuran (PCDF) toxic equivalent (TEQ) concentrations in agricultural soils in Washington state (USA).
Efron, B. (1992) Missing data, imputation, and the bootstrap. Technical Report No. 153, Division of Biostatistics, Stanford University, Stanford, CA.	A substantial theory of imputation has been developed to estimate a parameter of interest 0 in a missing data situation. Here we bring bootstrap methods to bear on the question of assigning confidence intervals for 0. Nonparametric bootstrap intervals based on a missing-data estimator 0 give convenient and accurate answers.
Floit, S.B.; Mahoney, L.A.; Batey, J.C.; Petroff, D.M. (1996) Evaluation of the Use of Substitution Methods to Represent Nondetect Data. Superfund Risk Assessment in Soil Contamination Studies: Second Volume: ASTM International.	Four substitution methods for handling nondetect data were applied: 1) one-half the SQL 2) One fifth the contract required quantitation limit 3) zero; and 4) the use of a random number generator to represent non-detect data with values ranging between zero and SQL.
Frome, E.L.; Wambach, P.F. (2005) Statistical methods and software for the analysis of occupational exposure data with non-detectable values. ORNL/TM-2005/52, Oak Ridge, TN.	1. quantities based on the maximum likelihood method for randomly left censored lognormal data 2. mean exposure level and UCL are obtained using the product limit estimate 3. Upper percentile (UTL) is obtained using a nonparametric approach
Frome, E.L.; Watkins, J.W. (2004) Statistical analysis of data with non-detectable values. ORNL/TM-2004/146, Oak Ridge, TN.	1. quantities based on the maximum likelihood method for randomly left censored lognormal data 2. mean exposure level and UCL are obtained using the product limit estimate 3. Upper percentile (UTL) is obtained using a nonparametric approach

REFERENCE	SUMMARY NOTES
Fusek, M.; Michâlek, J.; Vâvrovâ, M. (2015) Evaluation of Contamination Data with Non-detects using Censored Distributions. Fresenius Environmental Bulletin 24(11c): 4165-4172.	The method of maximum likelihood considering doubly left-censored samples is applied, which allows for a better evaluation of the obtained experimental data than commonly used methods where all values below the detection limits are replaced by a constant.
Ganser, G.H.; Hewett, P. (2010) An Accurate Substitution Method for Analyzing Censored Data. J Occup Environ Hyg 7(4): 233-244.	The ß- substitution method produced results comparable to the MLE method and is considerably easier to calculate, making it an attractive alternative. In terms of bias it is clearly superior to the commonly used LOD/2 and LOD/ √ 2 substitution methods.
Gardner, M. (2012) Improving the interpretation of 'less than' values in environmental monitoring. Water and Environment Journal 26(2): 285-290.	Substitution methods are acknowledged to be biased. With the aim of promoting a more technically sound approach to dealing with 'less than' data, a supplementary spreadsheet tool is supplied to provide the reader with ready introductory access to a simple way to apply maximum likelihood methods.
Gibbons, R.D.; Coleman, D.E. (2001) Statistical methods for detection and quantification of environmental contamination. New York: John Wiley & Sons.	Different methods for deriving detection limits and quantitation limits were discussed in Part I, including single-concentration-based methods and calibration-based methods. Various statistical methods for analyzing environmental data were also discussed in Part II, including comparing to regulation standards and background, detecting trend and correlation, and analyzing censored data.
Ginevan, M.E.; Splitstone, D.E. (2002) Bootstrap upper bounds for the arithmetic mean of right-skewed data, and the use of censored data. Environmetrics 13(5-6): 453-464.	Using Monte Carlo simulation, we show that a bootstrap upper bound is a much better approximation to the true upper bound on the population arithmetic mean. Finally, we present a bootstrap procedure for use when the data are left censored by detection/quantitation limits and discuss Monte Carlo results that support the use of this procedure when as much as one-half of the sample consists of censored observations.
Gochfeld, M.; Burger, J.; Vyas, V. (2005) Statistical analysis of data sets with values below detection limits. Consortium for Risk Evaluation with Stakeholder Participation, 3.	Despite its limitations, the default method of using half the MDA, widely used for inorganic and organic contaminants, can provide a useful representation, for some radiologic data as well. As the proportion of non-detects increases the reliability of any comparison decreases.
Grima, J.; Luque-Espinar, J.A.; Mejía-Gómez, J.Á.; Rodríguez, R. (2014) Analysis of Groundwater Monitoring Data Sets with Non-Detect Observations: Application to the Plana de Sagunto (Valencia, Spain) Groundwater Body. Engineering and Physical Sciences Research Council (EPSRC) Centre for Doctoral Training, Mathematics of Planet Earth, Springer: 507-512.	Annex IV of the Groundwater Directive sets that all measurements below the quantitation limit have to be substituted by half of the value of the highest quantitation limit, except for total pesticides. Censored estimation techniques, like Kaplan-Meier or Robust Regression have proved to be helpful in checking compliance with threshold values.
He, J. (2013) Mixture model based multivariate statistical analysis of multiply censored environmental data. Adv Water Resour 59: 15-24.	Gaussian mixture model (GMM) To cope with the censored data with multiple DLs, an expectation-maximization (EM) algorithm in a multivariate setting is developed

REFERENCE	SUMMARY NOTES
Hefley, T.J.; Tyre, A.J.; Baasch, D.M.; Blankenship, E.E. (2013) Nondetection sampling bias in marked presence-only data. Ecology and Evolution 3(16): 5225-5236.	We developed a marked inhomogeneous Poisson point process model that accounted for nondetection and aggregation behavior in animals and tested our methods on simulated data. Weighted likelihood methods can be used to correct for nondetection if estimates of the probability of detection are available.
Helsel, D.R. (1990) Less than obvious: statistical treatment of data below the detection limit. Environ Sci Technol 24(12): 1767-1774.	For small amounts of censoring and one reporting limit, Kendall's robust line can be fit to the data. For moderate censoring or multiple reporting limits, tobit regression can be performed. For more severe censoring of the dependent variables, logistic regression is appropriate. When response and explanatory variables contain severe censoring, contingency tables can be performed.
Helsel, D.R. (2004) STATISTICAL ANALYSIS OF NONDETECTS IN BIOGEOCHEMICAL DATA. Crustal Imaging and Characterization, USGS, Denver Federal Center, MS 964, Denver, CO.	Substituting one-half the detection limit and performing traditional statistical tests is an overly-simplistic practice which results in significant errors in interpretation.
Helsel, D.R. (2006) Fabricating data: how substituting values for nondetects can ruin results, and what can be done about it. Chemosphere 65(11): 2434-2439.	Two decades of research has shown that this fabrication of values produces poor estimates of statistics, and commonly obscures patterns and trends in the data. Substituting values for nondetects should be used rarely, and should generally be considered unacceptable in scientific research.
Helsel, D.R. (2009) Much ado about next to nothing: incorporating nondetects in science. Ann Occup Hyg 54(3): 257-262.	The author conducted a comprehensive review of literatures and textbook published on incorporating non-detects into statistical analyses, and made four general suggestions.
Helsel, D.R. (2010) Summing nondetects: incorporating low-level contaminants in risk assessment. Integr Environ Assess Manag 6(3): 361-366.	KM estimates are far less affected by the least precise data than are estimates computed using substitution. No assumptions about the distribution of data (whether they follow a normal or other distribution) need be made.
Helsel, D.R. (2012) Statistics for Censored Environmental Data Using Minitab and R (2nd ed.). Hoboken, New Jersey: John Wiley & Sons.	Use of interval-censored methods for treating true nondetects as lower than and separate from values between the detection and quantitation limits ("remarked data"). Shows why substitution methods fail.
Helsel, D.R.; Cohn, T.A. (1988) Estimation of Descriptive Statistics for Multiply Censored Water Quality Data. Water Resour Res 24(12): 1997-2004.	Probability plotting and maximum likelihood methods perform substantially better than simple substitution procedures now commonly in use. Probability plotting methods are more robust than maximum likelihood methods to misspecification of the parent distribution and their use should be encouraged in the typical situation where the parent distribution is unknown.
Helsel, D.R.; Lee, L. (2008) NADA for R. A contributed package for censored environmental data. United States Geological Survey (USGS), Reston, VA.	The typical solution of substituting one-half the detection limit and proceeding with regression, t-tests, etc., has repeatedly been shown to be inaccurate. Instead, these data can be effectively interpreted using survival analysis techniques more traditionally applied to right-censored data. Methods include censored maximum likelihood (ML), Kaplan-Meier, and the Akritas version of Kendall's robust line that is applicable (unlike ML) to doubly-censored data.
Hewett, P.; Ganser, G.H. (2007) A comparison of several methods for analyzing censored data. Ann Occup Hyg 51(7): 611-632.	No single method was unequivocally superior across all scenarios, although nearly all of the methods excelled in one or more scenarios. Overall, only the MLE- and LPR-based methods performed well across all scenarios, with the robust versions generally showing less bias than the standard versions when challenged with a contaminated lognormal distribution and multiple LODs

REFERENCE	SUMMARY NOTES
Higgins, R.M.; Danilowicz, B.S.; Brophy, D.; Geffen, A.J.; McGowan, T.; Gillanders, B. (2013) Influence of the limit of detection on classification using otolith elemental signatures. Can J Fish Aquat Sci 70(6): 922-929.	Estimating actual values of non-detected concentrations, through imputation, is now a plausible and simply applied strategy and should be considered as an alternative to constant replacement. We compare a robust regression approach to estimating values for non-detects alongside a variety of constant replacement methods.
Hoffman, H.J.; Johnson, R.E. (2015) Pseudo-likelihood estimation of multivariate normal parameters in the presence of left-censored data. J Agric Biol Environ Stat 20(1): 156-171.	We propose a pseudo-likelihood method utilizing pairs of variables that provides MLEs of mean and unstructured covariance parameters corresponding to a multivariate normal or lognormal distribution in the presence of left-censored data.
Hoogerbrugge, R.; Liem, A.K.D. (2000) How to handle non-detects. Organohalogen Compd 45: 13-16.	Several imputation strategies have been evaluated on the basis of an artificially censored (by increasing the LOD) dataset of dioxin concentrations in cow's milk (1989-1990).
Huston, C.; Juarez-Colunga, E. (2009) Guidelines for Computing Summary Statistics for Data-Sets Containing Non-Detects. Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC.	The methods presented in this document are: 1: substitution; 2: Kaplan-Meier, as part of nonparametric methods; 3: lognormal model based on maximum likelihood estimation; 4: and robust regression on order statistics, which is a semiparametric method.
Huybrechts, T.; Thas, O.; Dewulf, J.; Van Langenhove, H. (2002) How to estimate moments and quantiles of environmental data sets with non-detected observations? A case study on volatile organic compounds in marine water samples. J Chromatogr A 975(1): 123.	Several parametric and robust parametric approaches based on the maximum likelihood principle and probability-plot regression method were evaluated for the estimation of the mean, standard deviation, median and interquartile range. Methods with the least distributional assumptions, such as the robust bias-corrected restricted maximum likelihood method, perform best for estimating the mean and standard deviation, while both parametric and robust parametric techniques can be used for quantiles.
Huynh, T.; Quick, H.; Ramachandran, G.; Banerjee, S.; Stenzel, M.; Sandler, D.P.; Engel, L.S.; Kwok, R.K.; Blair, A.; Stewart, P.A. (2016) A Comparison of the ß-Substitution Method and a Bayesian Method for Analyzing Left-Censored Data. Ann Occup Hyg 60(1): 56-73	Suggest the use of Bayesian methods if the practitioner has the computational resources and prior information, as the method would generally provide accurate estimates and also provides the distributions of all of the parameters, which could be useful for making decisions in some applications.
Huynh, T.; Ramachandran, G.; Banerjee, S.; Monteiro, J.; Stenzel, M.; Sandler, D.P.; Engel, L.S.; Kwok, R.K.; Blair, A.; Stewart, P.A. (2014) Comparison of Methods for Analyzing Left-Censored Occupational Exposure Data. Ann Occup Hyg 58(9): 1126-1142.	ß-substitution method generally performed as well or better than the ML and K-M methods in most simulated lognormal and mixed lognormal distribution conditions. The ML method was suitable for large sample sizes (N ≥ 30) up to 80% censoring for lognormal distributions with small variability (geometric standard deviation = 2–3). The K-M method generally provided accurate estimates of the arithmetic mean when the censoring was <50% for lognormal and mixed distributions. The accuracy and precision of all methods decreased under high variability (geometric standard deviation = 4 and 5) and small to moderate sample sizes (N < 20) but the ß-substitution was still the best of the three methods.

REFERENCE	SUMMARY NOTES
Imaizumi, Y.; Suzuki, N.; Shiraishi, H. (2006) Bootstrap methods for confidence intervals of percentiles from dataset containing nondetected observations using lognormal distribution. J Chemometr 20(1-2): 68-75.	'lognormal, parametric, bootstrap-t' method (LN-P-Bt)—the most reliable method—95% of the upper confidence limits of the 95th percentiles were from 1.7 to 2.1 times greater than the 95th percentiles directly predicted from the actual sample, and the logarithmic ranges for 90% of the bootstrap confidence intervals varied from 3.0 to 4.4.
Interstate Technology and Regulatory Council (ITRC) (2013) Groundwater Statistics for Monitoring and Compliance, Statistical Tools for the Project Life Cycle. GSMC-1.	Tarone-Ware two-sample test as an alternative to the t-test. Mann-Kendall trend test (estimate sample statistics). Kaplan-Meier method for calculating an upper confidence limit on the mean). Impute an estimated value for each nondetect prior to further statistical analysis.
Ito, T.; Kato, T.; Takagishi, K.; Okabe, S.; Sano, D. (2015) Bayesian modeling of virus removal efficiency in wastewater treatment processes. Water Sci Technol 72(10): 1789-1795.	We modeled the probabilistic distribution of virus removal efficiency in a wastewater treatment process with a Bayesian approach, and investigated how many detect samples in influent and effluent are necessary for accurate estimation.
Jain, R.B. (2016) On the consequence of substituting maximum likelihood estimates for the observations below the limit of detection. Chemosphere 144: 2044-2051.	Use of MLE procedures with multiple imputations to replace observations below the LOD has been recommended. The use of more than one multiply imputed variable in a regression model was not found to be of concern. The results show that the use of multiple imputations does not generate additional variabilities in the estimates of these statistics beyond tolerable statistical noise. However, when the percent observations in the data are relatively high, there is some possibility of obtaining disparate results.
Jin, Y.; Hein, M.J.; Deddens, J.A.; Hines, C.J. (2011) Analysis of Lognormally Distributed Exposure Data with Repeated Measures and Values below the Limit of Detection Using SAS. Ann Occup Hyg 55(1): 97-112.	MLE method resulted in less bias and performed well even for censoring up to 80%, whereas the substitution method resulted in considerable bias.
Kafatos, G.; Andrews, N.; McConway, K.J.; Farrington, P. (2013) Regression models for censored serological data. J Med Microbiol 62(1): 93-100	The results showed that the simple substitution and deletion methods worked reasonably well for low proportions of data censored (<20 %). However, in general, the censored regression method gave estimates closer to the truth than the other methods examined under different scenarios, such as types of equations used and violation of regression assumptions. Interval-censored regression produced the least biased estimates for assay data resulting from dilution series. Censored regression produced the least biased estimates in comparison with the other methods examined. Moreover, the results suggest using interval-censored regression methods for assay data resulting from dilution series.
Kato, T.; Miura, T.; Okabe, S.; Sano, D. (2013) Bayesian modeling of enteric virus density in wastewater using left-censored data. Food and environmental virology 5(4): 185-193.	We applied a Bayesian model that is able to model both the detected data (detects) and non-detects to simulated left-censored datasets of enteric virus density in wastewater.
Kirchner, G.; Steiner, M.; Zahringer, M. (2009) A new approach to estimate nuclide ratios from measurements with activities close to background. J Environ Radioact 100(6): 484-488.	Using Bayesian statistics, a method is presented which allows statistical inference on nuclide ratios taking into account both prior knowledge and all information collected from the measurements. It is shown that our method allows quantitative conclusion to be drawn if counts of single isotopes are low or become even negative after background subtraction

REFERENCE	SUMMARY NOTES
Krall, J.; Simpson, C.; Peng, R. (2015) A model-based approach for imputing censored data in source apportionment studies. Environ Ecol Stat 22(4): 779-800.	When estimating the complete data distribution, these commonly applied methods to adjust censored data (deletion/substitution) perform poorly compared with model-based imputation methods. Model-based multiple imputation frequently leads to better source estimation.
Kreinovich, V.; Longpré, L.; Starks, S.A.; Xiang, G.; Beck, J.; Kandathi, R.; Nayak, A.; Ferson, S.; Hajagos, J. (2007) Interval versions of statistical techniques with applications to environmental analysis, bioinformatics, and privacy in statistical databases. J Comput Appl Math 199(2): 418-423.	"Reverse" Kaplan-Meier was the only method inherently able to deal with censored data with multiple LODs, and may be the most accurate since it avoids data manipulation needed for use with other commonly used statistical methods.
Krishnamoorthy, K.; Mallick, A.; Mathew, T. (2009) Model-Based Imputation Approach for Data Analysis in the Presence of Non-detects. Ann Occup Hyg 53(3): 249-263.	Two imputation approaches are investigated in the paper: one uses approximate maximum likelihood estimates (MLEs) of the parameters and a second approach uses simple ad hoc estimates that were developed for the specific purpose of imputations. Only the MLE-based approach is satisfactory for large sample sizes
Krishnamoorthy, K.; Xu, Z. (2011) Confidence Limits for Lognormal Percentiles and for Lognormal Mean Based on Samples with Multiple Detection Limits. Ann Occup Hyg 55(5): 495-509.	The proposed methods are based on the maximum likelihood estimates. They perform well with respect to coverage probabilities as well as power and are applicable to small sample sizes. The proposed approaches are also applicable for finding confidence limits for the percentiles of a gamma distribution.
Kroll, C.N.; Stedinger, J.R. (1999) Development of regional regression relationships with censored data. Water Resour Res 35(3): 775-784.	Ordinary least squares and adding a small constant to all at-site quantile estimates performed poorly compared to the use of a Tobit model, which is a maximum likelihood estimator (MLE) procedure that represents the below threshold estimates as a range from zero to the threshold level.
Lambert, D.; Peterson, B.; Terpenning, I. (1991) Nondetects, detection limits, and the probability of detection. J Am Stat Assoc 86(414): 266-277.	The measurements that are likely to be reported as nondetect can be described by a new concept, the probability of acceptance, where acceptance means that a measurement passes the requirements for being reported as a detect. The 90th percentile of the probability of acceptance curve is a reasonable upper bound or censoring limit for a measurement reported as nondetect. The probability of acceptance also suggests the complexity of the data analysis task.
LeFrancois, M.; Poeter, E. (2009) Use of Observations below Detection Limit for Model Calibration. Ground Water 47(2): 228-236.	It is common to either delete or substitute values for nondetect observations for use in model calibration, but this practice can bias the estimated parameter values and the model predictions. We propose use of the censored-residual approach to including nondetect values as observations for calibration. In this approach, residuals are calculated as the detection limit minus the simulated value when the simulated value exceeds the detection limit, and the residual is assigned a value of zero when the simulated value is below the detection limit.

REFERENCE	SUMMARY NOTES
Leith, K.F.; Bowerman, W.W.; Wierda, M.R.; Best, D.A.; Grubb, T.G.; Sikarske, J.G. (2010) A comparison of techniques for assessing central tendency in left-censored data using PCB and p,p'DDE contaminant concentrations from Michigan's Bald Eagle Biosentinel Program. Chemosphere 80(1): 7-12.	Summary statistics were calculated using (1) the '0.0001' near-zero method of substitution, (2) substitution with '1/2*DL', (3) multiple imputation, and (4) Kaplan-Meier estimation. Median was used for comparison. Several analytical options for datasets with non-detect observations are available. The general consensus is that substitution methods ((1) and (2)) can produce biased summary statistics, especially as levels of substitution increase.
Levitan, D.M.; Schreiber, M.E.; Seal, R.R.; Bodnar, R.J.; Aylor, J.G. (2014) Developing protocols for geochemical baseline studies: An example from the Coles Hill uranium deposit, Virginia, USA. Appl Geochem 43: 88-100.	Regression on order statistics was used to handle non-detect data.
Li, S.; Batterman, S.; Su, F.C.; Mukherjee, B. (2013) Addressing extrema and censoring in pollutant and exposure data using mixture of normal distributions. Atmos Environ 77: 464-473.	Dirichlet process mixture of normals has advantages by characterizing uncertainty around the number of components, and by providing a formal assessment of uncertainty for all model parameters through the posterior distribution. The method adapts to a spectrum of departures from standard model assumptions and provides robust estimates of the exposure density even under censoring due to MDL.
Lubin, J.H.; Colt, J.S.; Camann, D.; Davis, S.; Cerhan, J.R.; Severson, R.K.; Hartge, P. (2004) Epidemiologic Evaluation of Measurement Data in the Presence of Detection Limits. Environ Health Perspect 112(17): 1691-1696.	Truncated data methods (e.g., Tobit regression) and multiple imputation offer two unbiased approaches for analyzing measurement data with detection limits. If interest resides solely on regression parameters, then Tobit regression can be used. If individualized values for measurements below detection limits are needed for additional analysis, such as relative risk regression or graphical display, then multiple imputation produces unbiased estimates and nominal confidence intervals unless the proportion of missing data is extreme.
Millard, S.P.; Deverel, S.J. (1988) Nonparametric statistical methods for comparing two sites based on data with multiple nondetect limits. Water Resour Res 24(12): 2087-2098.	The best overall test, in terms of maintained α level, is the normal scores test based on a permutation variance. In cases where the α level is maintained, however, the Peto-Prentice statistic based on an asymptotic variance performs as well or better.
Munoz, G.; Giraudel, J.-L.; Botta, F.; Lestremau, F.; Dévier, M.-H.; Budzinski, H.; Labadie, P. (2015) Spatial distribution and partitioning behavior of selected poly- and perfluoroalkyl substances in freshwater ecosystems: A French nationwide survey. Sci Total Environ 517: 48-56.	Robust ROS method was preferred for descriptive statistics computation while the Akritas–Theil–Sen estimator was used for regression and correlation analyses
Myung, I.J. (2003) Tutorial on maximum likelihood estimation. J Math Psychol 47(1): 90-100.	Unlike least-squares estimation which is primarily a descriptive tool, MLE is a preferred method of parameter estimation in statistics and is an indispensable tool for many statistical modeling techniques, in particular in non-linear modeling with non-normal data.

REFERENCE	SUMMARY NOTES
Newman, M.C.; Dixon, P.M.; Looney, B.B.; Pinder, J.E. (1989) Estimating Mean and Variance for Environmental Samples With Below Detection Limit Observations. Water Resour Bull 25(4): 905-916.	Substitution or deletion methods provide poor estimates of the mean and variance of censored samples. Restricted maximum likelihood estimates are easily computable, are less biased and more accurate than the other estimators when the parent distribution is known. The robustness of regression of normal scores after log transformation of the data suggests that these methods are the most effective when a parent distribution cannot be identified.
Newton, E.; Rudel, R. (2007) Estimating Correlation with Multiply Censored Data Arising from the Adjustment of Singly Censored Data. Environ Sci Technol 41(1): 221-228.	Instead of using the multiply censored data directly, cp.mle2 relies on ML estimates of the covariance of the singly censored laboratory data. With increasing levels of censoring, most of the correlation methods are highly biased. The simple substitution methods in general tend toward zero if singly censored and one if multiply censored. ck.taub tends toward zero. Least biased is cp.mle2, however, it has higher variance than some of the other estimators. Overall, cs.det performs the worst and cp.mle2 the best.
Nysen, R.; Faes, C.; Ferrari, P.; Verger, P.; Aerts, M. (2015) Parametric and semi-nonparametric model strategies for the estimation of distributions of chemical contaminant data. Environ Ecol Stat 22(2): 423-444.	We focus on several families of distributions that are related to the log-normal distribution in some direct or indirect way, and that are parametric or semi-nonparametric extensions of the log-normal distribution: the log-skew-normal, the log-t, the log-skew-t, the Weibull, the gamma, the generalized-gamma, and the semi-nonparametric estimator of Zhang and Davidian (Biometrics 64(2):567-669).
Orton, T.G.; Rawlins, B.G.; Lark, R.M. (2009) Using measurements close to a detection limit in a geostatistical case study to predict selenium concentration in topsoil. Geoderma 152(3-4): 269-282.	We found that the Bayesian approach based on soft data resulted in smoother maps, reduced the errors of the predictions, and provided a better representation of the associated uncertainty.
Palarea-Albaladejo, J.; Martín-Fernández, J.A.; Olea, R.A. (2014) A bootstrap estimation scheme for chemical compositional data with nondetects. J Chemometr 28(7): 585-599.	A bootstrap scheme is devised that handles nondetects by adding an imputation step within the resampling process and conveniently propagates their associated uncertainly. Results suggest that nondetect bootstrap based on model-based imputation is generally preferable. A robust approach based on isometric log-ratio transformations appears to be particularly suited in this context.
Pearmain, S. (2015) The importance of statistical data analysis and environmental risk assessment modelling in determining the significance of soil and groundwater contamination. 1st International Conference on Energy, Environment and Climate Changes, Mauritius.	It is considered bad practice to simply delete non-detects from a dataset and risk assessors frequently substitute a fraction of the detection limit (i.e., L/2) for each non-detect. Recent research however indicates that this substitution of values can undermine robust statistical assessment and other more robust alternatives such as the use of hypothesis tests, correlation coefficients and regression equations should be considered when addressing non-detects.
Pita, G.L.; Francis, R.; Liu, Z.; Mitrani-Reiser, J.; Guikema, S.; Pinelli, J.P. (2011) Statistical Tools for Populating/Predicting Input Data of Risk Analysis Models. Vulnerability, Uncertainty, and Risk: pp. 468-476.	This paper uses Bayesian Belief Networks and Classification and Regression Trees to populate the missing information inside a database based on the structure of the available data.

REFERENCE	SUMMARY NOTES
Qian, S.S.; Schulman, A.; Koplos, J.; Kotros, A.; Kellar, P. (2004) A Hierarchical Modeling Approach for Estimating National Distributions of Chemicals in public Drinking Water Systems. Environ Sci Technol 38(4): 1176-11825.	Presents a Bayesian-based hierarchical model for estimating the national distribution of the mean concentrations. The model, which assumes log- normality, was evaluated using simulated datasets generated from a series of Weibull distributions to illustrate the robustness of the model. The Bayesian method is able to quantify the uncertainty in the estimated cumulative density function.
Schmoyeri, R.L.; Beauchamp, J.J.; Brandt, C.C.; Hoffman Jr, F.O. (1996) Difficulties with the lognormal model in mean estimation and testing. Environ Ecol Stat 3(1): 81-97.	In the presence of random left censoring, the product limit estimate replaced the ordinary sample mean and standard error.
Scott, G.I.; Thompson, R.E.; Voit, E.O. (2000) Statistical modeling of sediment and oyster PAH contamination data collected at a South Carolina estuary (complete and left-censored samples). Environmetrics 11(1): 99.	The Weibull almost always provides a better fit to the data than the lognormal distribution. Methods based on the underlying distribution of the data give more consistent results than those obtained by commonly used substitution methods.
Sharma, M.; McBean, E.A.; Thomson, N. (1995) Maximum Likelihood Method for Parameter Estimation in Linear Model with below-Detection Data. J Environ Eng 121(11):	Parameter estimates of the model were obtained using the following three procedures: (1) the NAG-15 routine for maximization of a likelihood function; (2) the proposed algorithm for the equivalent LS method; and (3) the modified iterative least squares method.
Shoari, N.; Dube, J.-S.; Chenouri, S. (2015) Estimating the mean and standard deviation of environmental data with below detection limit observations: Considering highly skewed data and model misspecification. Chemosphere 138: 599-608.	ROS, GROS, and MLE under gamma distribution are generally robust to model misspecifications regardless of skewness, sample size, and censoring percentage. Since the characteristics of environmental data (e.g., type of distribution and skewness) are unknown a priori, we suggest using MLE based on gamma distribution, rROS and GROS.
Shoari, N.; Dubé, J.-S.; Chenouri, S. (2016) On the use of the substitution method in left-censored environmental data. Hum Ecol Risk Assess 22(2): 435-446.	Although the performance of the substitution-based method improves as the censoring percentage decreases, its performance still depends on the population's distributional characteristics. Caution must be taken in using the substitution method when analyzing censored environmental data.
Shorten, P.R.; Pleasants, A.B.; Soboleva, T.K. (2006) Estimation of microbial growth using population measurements subject to a detection limit. Int J Food Microbiol 108(3): 369-375.	We develop a maximum likelihood estimation procedure for determining the mean and variance in microbial population size from microbial population measurements subject to a detection limit.
Silva, E.; Mendes, M.; Ribeiro, L.; Cerejeira, M. (2012) Exposure assessment of pesticides in a shallow groundwater of the Tagus vulnerable zone (Portugal): a multivariate statistical approach (JCA). Environ Sci Pollut Res Int 19(7): 2667-2680.	Using joint correspondence analysis was still possible to establish relations between pesticides and their temporal trend in a case study where there were more than 80% of data censored.

REFERENCE	SUMMARY NOTES
Sinha, P.; Lambert, M.B.; Trumbull, V.L. (2006) Evaluation of statistical methods for left-censored environmental data with nonuniform detection limits. Environ Toxicol Chem 25(9): 2533-2540.	Recommendations: datasets with 15 to 50% nondetected samples-log-probit regression method and use of Chebyshev theorem to estimate 95% upper confidence limits; datasets with 51 to 80% nondetected samples-bounding method and use of Chebyshev theorem to estimate 95% upper confidence limits.
Slymen, D.J.; de Peyster, A. (1994) Hypothesis testing with values below detection limit in environmental studies. Environ Sci Technol 28(5): 898.	Presents regression models for analyzing data from an experimental design when some values are below detection limit using a readily available statistical package.
Stoline, M.R. (1993) Comparison of two medians using a two-sample lognormal model in environmental contexts. Environmetrics 4(3): 323-339.	A test procedure is derived for comparing medians in a lognormal two-sample context where some data may be left-censored owing to non-detects. The EM algorithm is used to obtain maximum likelihood estimates of the parameters.
Succop, P.A.; Clark, S.; Chen, M.; Galke, W. (2004) Imputation of data values that are less than a detection limit. J Occup Environ Hyg 1(7): 436-441.	Imputation of the low lead loadings was accomplished by substituting the value associated with the median percentile below each laboratory's method detection limit. A correlation of r = 0.50 was calculated between the predicted and reported dust lead loadings, with only slight bias (2.9%) in the predicted values. Results suggest that analytical laboratories should provide a numerical result for all analyzed samples, with a "flag" of those values below their detection limit, since these results may be more accurate than any imputed value, particularly those provided by the commonly used method of dividing the minimum detection limit by the square root of 2.
Thompson, M.L. (2003) Linear regression with Type I interval-and left-censored response data. Environ Ecol Stat 10(2): 221-230	We develop and evaluate a maximum likelihood approach to linear regression analysis. The maximum likelihood approach represents only a moderate increase in power, but we show that the bias in substitution estimates may be substantial.
U.S. Department of Energy (USDOE). (2013a) Quality Systems for Analytical Services Revision 2.9. Retrieved from http://www.p2s.com/wp-content/uploads/QSAS-Rev-2.9.pdf.	This document establishes a single, integrated quality assurance (QA) program for analytical laboratories who support the U.S. DOE operations. This unified QA program helps harmonize analytical data quality requirements across various Federal agencies.
U.S. Environmental Protection Agency (1991) Chemical concentration data near the detection limit. EPA/903/8-91/001, Hazardous Waste Management Division, Philadelphia, PA.	Describes statistical methods to estimate concentrations below the detection limit as technically superior to the three substitution methods mentioned. These statistical methods are effective only for datasets having a high proportion of detects (typically, greater than 50%).
U.S. Environmental Protection Agency (2000) Assigning Values to Non-Detected/Non-Quantified Pesticide Residues in Human Health Food Exposure Assessments. Office of Pesticide Programs, Washington, DC	In general, the Office of Pesticide Programs recommends use of a default value of ½ the Limit of Detection (LOD) or ½ the Limit of Quantitation (LOQ) for commodities which have been treated but for which no detectable residues are measured.

REFERENCE	SUMMARY NOTES
U. S. Environmental Protection Agency (2004) Revised Assessment of Detection and Quantitation Approaches. EPA-821-B-04-005, October 2004.	EPA re-assessed a number of procedures on deriving detection limit and quantitation limit and concluded that no single pair of detection and quantitation limit procedures perfectly meets EPA's six evaluation criteria. Among all the procedures evaluated, MDL and ML procedures are the closest to meeting the six criteria. To improve, EPA proposed modest revisions to the MDL procedure and to codify an ML definition and procedure.
U.S. Environmental Protection Agency (2009) Statistical analysis of groundwater monitoring data at RCRA facilities: Unified guidance. EPA 530/R-09-007, Office of Resource Conservation and Recovery, Washington, DC.	Simple substitution recommended only if no more than 10-15% of the sample observations are ND Censored estimation technique (Kaplan-Meier or Robust Regression on Order Statistics [ROS]) recommended if detection frequency is no less than 50% For lower detection frequencies, Tarone-Ware two-sample test or Kruskal-Wallis test are recommended
U.S. Environmental Protection Agency (2011a) A Laboratory Study of Procedures Evaluated by the Federal Advisory Committee on Detection and Quantitation Approaches and Uses in Clean Water Act Programs. Office of Water (4303T), Washington, DC.	FACDQ Single Lab Procedure v2.4, the FACDQ Single Lab Procedure v2.4T and the modified LCMRL procedure. Censored methods = less than 50% of the method blanks analyzed yield a numerical result (regardless of detection or other reporting limits) and meet qualitative identification criteria. The primary difference between version 2.4 and version 2.4T of the FACDQ procedure is the use of a prediction limit based on a t statistic to replace the tolerance limit (specified as "k" in the procedure) in the calculation of detection limits for uncensored methods. Version 2.4T also includes specific precision and accuracy measurement quality objectives (MQOs) which were used for this study, as well as several changes to improve the clarity and organization of the procedure.
U.S. Environmental Protection Agency (2011b) Supplement to report entitled, "A laboratory study of procedures evaluated by the Federal Advisory Committee on detection and quantitation approaches and uses in Clean Water Act Programs." https://www.epa.gov/sites/production/files/2015-10/documents/supplement-lab-study-report_2011.pdf.	To assess the appropriateness of using FACDQ's procedures on generating reliable DL and QL, EPA selected two commonly used analytical methods and tested them in six laboratories to compare FACDQ's DL/QL and EPA's MDL/ML. The results were discussed in this report: both the FACDQ DL/QL and the EPA MDL/ML generally met the acceptable range for the measurement quality objectives for false negative rate, relative standard deviation, and mean recovery. However, FACDQ's DL is inherently more stringent against false positives than EPA's MDL, resulting in a higher detection limit value for an analyte to be considered present in a sample.
Uh, H.-W.; Hartgers, F. C.; Yazdanbakhsh, M.; Houwing-Duistermaat, J.J. (2008). Evaluation of regression methods when immunological measurements are constrained by detection limits. BMC Immunol 9(1): 59.	The deletion and extrapolation by regression on order statistics methods gave biased parameter estimates. The single substitution method underestimated variances, and logistic regression suffered loss of power...tobit regression performed well when the proportion of nondetects was less than 30%, and that taken together the multiple imputation method performed best.

REFERENCE	SUMMARY NOTES
U.S. Environmental Protection Agency. (2013) ProUCL Version 5.0. 00 Technical Guide-Statistical Software for Environmental Applications for Data Sets with and without Nondetect Observations. EPA: Washington, WA, USA. U.S. Environmental Protection Agency (2015c) ProUCL Version 5.1.002 Technical Guide, EPA/600/R-07/041. Office of Research and Development, Washington, DC. October.	ProUCL 5.0 computes upper limits using KM estimates in gamma (lognormal) UCL and UTL equations provided the detected observations in the left-censored dataset follow a gamma (lognormal) distribution. Some poor performing commonly used and cited methods such as the DL/2 substitution method and H-statistic based UCL computation method have been incorporated in ProUCL for historical reasons, and research and comparison purposes.
Verbovšek, T. (2011) A comparison of parameters below the limit of detection in geochemical analyses by substitution methods. RMZ Mater Geoenvironment 58: 393-404	A large dataset of generated values with normal and lognormal distributions was tested for different percent of censoring from 1% to 50%, plus the censored data of five selected geochemical parameters. Results indicate that the best substitution method is by LOD/√2, as it produces the smallest errors.
Vosnakis, K.A.S.; Perry, E.; Madsen, K.; Bradley, L.J.N. (2009) Background Versus Risk-Based Screening Levels - An Examination of Arsenic Background Soil Concentrations in Seven States. International Journal of Soil, Sediment and Water 2(2): 1-20.	This paper appears to simply have used ½ the detection limit for non-detects without much discussion
Wendelberger, J.; Campbell, K. (1994). Non-detect data in environmental investigations. Los Alamos National Laboratory, MS-F600, Los Alamos, NM.	One of the most commonly used replacement methods is to substitute each nondetect value by half of its detection limit. Other commonly used replacement values are zeros or the detection limits.
Wu, H.; Chen, Q.; Ware, L.B.; Koyama, T. (2012) A Bayesian approach for generalized linear models with explanatory biomarker measurement variables subject to detection limit: an application to acute lung injury. J Appl Stat 39(8): 1733-1747.	We propose a Bayesian approach for generalized linear models with explanatory variables subject to lower, upper, or interval DLs. In both real and simulation studies, we compared the proposed Bayesian approach to four commonly used methods in a logistic regression model with explanatory variable measurements subject to DL.
Yu, X.; Liu, P.; Min, J.; Chen, Q. (2009) Regression on order statistics and its application in estimating nondetects for food exposure assessment. Journal of hygiene research 38(1): 89-91.	The results show that ROS method performs better obviously than substitution methods for being robust and convenient for posterior analysis.

REFERENCE	SUMMARY NOTES
Yuan, Y. C. (2010) Multiple imputation for missing data: Concepts and new development (Version 9.0). SAS Institute, Inc., Rockville, MD.	Multiple imputation provides a useful strategy for dealing with datasets with missing values. Instead of filling in a single value for each missing value, Rubin's (1987) multiple imputation procedure replaces each missing value with a set of plausible values that represent the uncertainty about the right value to impute.
Zhang, H. (2013) Significance of Nondetects in the Mapping of Soil Contaminants. TRITA-LWR Degree Project 13:03, Royal Institute of Technology (KTH), Stockholm, Sweden.	Statistical analysis methods for nondetects involve substitution by half of the detection limit (DL/2), maximum likelihood estimation (MLE), Kaplan-Meier and regression on ordered statistics (ROS).
Zhang, Z.; Lennox, W.C.; Panu, U.S (2004) Effect of percent non-detects on estimation bias in censored distributions. J Hydrol (Amst) 297(1-4): 74-94.	To incorporate non-detects in the estimation process, a simple substitution by the MDL (method detection limit) and the maximum likelihood estimation method are routinely implemented as standard methods by US-EPA laboratories. This paper utilizes a mathematical approach to derive the bias functions and resulting bias curves are developed to investigate the censored samples from a variety of probability distributions such as normal, log-normal, gamma, and Gumbel distributions
Zhao, Y.; Frey, H.C. (2006) Uncertainty for Data with Non-Detects: Air Toxic Emissions from Combustion. Hum Ecol Risk Assess 12(6): 1171-1191.	The estimated means of the censored dataset by conventional methods are usually biased. Maximum likelihood estimation (MLE) and bootstrap simulation have been demonstrated as a statistically robust method to quantify variability and uncertainty of censored datasets and can provide asymptotically unbiased mean estimates.
Zoffoli, H.J.O.; Varella, C.A.A.; do Amaral-Sobrinho, N.M.B.; Zonta, E.; Tolon-Becerra, A. (2013) Method of median semi-variance for the analysis of left-censored data: Comparison with other techniques using environmental data. Chemosphere 93(9): 1701-1709.	In general, the simple substitution and deletion methods showed biased performance, with exceptions for L/2, Inter and L/[radic]2 methods that can be used with caution under specific conditions. In general, the SemiV method and other parametric methods showed similar performances and were less biased than other methods. The SemiV method is a simple and accurate procedure that can be used in the analysis of datasets with less than 50% of left-censored data.

APPENDIX B | NRDA DOCUMENTS AND THEIR HANDLING OF NON-DETECTS

SITE, STATE	CONTAMINANT(S)	NATURAL RESOURCE(S)	NRDA CONTEXT	TYPE(S) OF ANALYSIS	APPROACH TO HANDLING NDS	REFERENCE	NOTES
Anniston, AL	PCBs	surface water, groundwater, sediment, biological resources	Stage I Assessment Plan	Graphical display of PCBs concentrations in sediments by location; for groundwater presents a range of results; purpose was to determine if the site had been exposed to PCBs	Surface water: A count of samples below the analytical DL is presented on the graph separately. Groundwater: Use ND as the lower bound of a range (.e. samples showed ND to 7400 µg/L)	The Anniston PCB Site Trustee Council 2010	All data used are from earlier studies
Anniston, AL	PCBs, total metals, TOC, percent solids	sediment	Field study of sediment and soil chemistry to support injury evaluation	Total PCB concentrations, calculating PEC-quotients (i.e., concentrations divided by a threshold), and summary statistics	Non-detect values were treated as half the detection limit; "Non-detect values that were above the PEC were screened out and not included in the totals calculations."	Schein et al. 2015	
Commencement Bay, WA	multiple	birds	Data report on a reconnaissance-level avian injury assessment	Calculating TEQs; summary statistics	Substitution with ½ DL for TEQ calculation, and DL for summary statistics.	Krausmann 1999	
Fox River, WI	PCBs	surface water	Assessment plan	PCB concentration	ND listed in tables	USFWS and Hagler Bailly Consulting, 1996	These were summaries of previous sampling done at the site, not sampling designed specifically for the NRDA
Fox River, WI	PCBs	surface water	Injury determination; includes some elements of the injury quantification phase (e.g., pathway)	--Calculating tPCBs (sum of congeners, sum of Aroclors) --Chart of individual data points, percentile (x) versus concentration (y)	For summations of individual congeners: substitute ND with 0. For summations of Aroclors: substitute ND with DL. In charts of individual data points, NDs are plotted at the DL and identified using symbology.	USFWS et al., 1999a	
Fox River, WI	PCBs	walleye salmon	Technical report	Individual PCB congener concentrations and total PCBs	Presented as ND in tables for the single sample in which individual congeners are measured; Does not state how NDs are handled in the calculation of total PCBs.	Baron et al. 1999	

Fox River, WI	PCBs	sediment, surface water, fish	Pathway determination	--Principal components analysis (sediments) --k-means cluster analysis (sediments)	Congener concentrations are first expressed as the percent of tPCBs, "where tPCBs are defined as the sum of detectable congeners" (i.e., non-detect congeners are substituted with zero.) Furthermore, "To avoid biases that could be associated with very low PCB concentrations and subsequent high analytical variability, we omitted samples which had total PCB concentrations < 1ug/kg or had fewer than 10 detectable congeners. In addition, specific congeners were omitted from the analysis if the median percent composition for that congener was <1% in all regions"	USFWS et al., 1999b	
Fox River, WI	PCBs	surface water	Injury determination; includes some elements of the injury quantification phase (e.g., pathway)	--Calculating tPCBs (sum of congeners, sum of Aroclors) --Chart of individual data points, percentile (x) versus concentration (y)	For summations of individual congeners: substitute ND with 0 . For summations of Aroclors: substitute ND with DL In charts of individual data points, NDs are plotted at the DL and identified using symbology.	USFWS et al., 1999a	
Grand Calumet River, IN	PCB	sediment	Assessment plan	Summary statistics	Lists transects where samples were below detection limit	Weiss et al. 1997	
Grand Calumet River, IN	PCB, pesticides	sediment, fish	Injury determination: Human uses of fishery resources technical report; includes some elements of injury quantification	Calculating total concentrations; concentrations compared to benchmark levels and tolerance action levels; summary statistics	In calculating total concentrations of COPCs in sediments, assigned 1/2 detection limit except where the detection limit was greater than the selected chemical benchmark. In that case the greater than detection limit value was not used in the calculation of the total concentration or in the assessment of injury to human uses of fishery resources. For tissue samples, less than detection limit data and low level	MacDonald et al. 2003	The second file is just appendices, mostly with extended results.

					detects were treated as zero in accordance with the guidance provided by USFDA (2001) to facilitate comparison with the tolerance of action levels. However, "By comparison, less than detection limit data for tissue chemistry were assigned a value of one-half of the detection to facilitate comparison with the thresholds used to develop the Indiana FCAs (Anderson et al. 1993). When the detection limit was greater than the selected benchmark for fish tissue chemistry, then the result was not used in the assessment of injury to human uses of fishery resources."		
Hudson River, NY	PCBs	bats	Data report	--Calculating tPCBs --Summary statistics	Substitution with zero	HRNRT 2007	
Hudson River, NY	PCBs	surface water	Injury determination report	Percentage of samples exceeding water quality standards/criteria	Substitution with zero	HRNRT 2008	
Hudson River, NY	PCBs	mink	Injury study	--Calculating ∑PCBs and TEQs --Summary statistics	Substitution with 1/2 DL.	Bursian et al. 2013	The authors state "The choice of assigned value (0, one-half detection limit, or detection limit) had no substantive influence on ΣPCB congener or TEQ concentrations based on a quantitative assessment."
Hudson River, NY	PCBs	sediments	Compilation of contaminant data for the public	Summary statistics	Percent ND is tabulated in chart's notes	HRNRT 2013	
Hudson River, NY	PCBs	sediments	Compilation of contaminant data for the public	Summary statistics	Percent ND is tabulated in chart's notes	HRNRT 2013	
Hudson River, NY	PCBs	sediments	Compilation of contaminant data for the public	Summary statistics	Percent ND is tabulated in chart's notes	HRNRT 2013	

Hudson River, NY	PCBs	sediments	Compilation of contaminant data for the public	Summary statistics	Percent ND is tabulated in chart's notes	HRNRT 2013	
Hudson River, NY	PCBs, primarily; other contaminants (metals, mercury, parent and alkylated PAHs, semi-volatile VOCs, organochlorine pesticides, toxaphene) also measured	benthos	Pilot study of injury to benthos	(a) Calculating total concentrations in sediment samples (e.g., tPCBs as sum of homologs, tPAHs, tDDTs) (b) Calculating summary statistics across samples (c) Averaging results from laboratory split samples	(a) and (b) Substitution with 1/2 DL (c) Substitution with the maximum DL	HRNRT 2016	
Hylebos Waterway, WA	Multiple	sediments	An allocation of responsibility for natural resource injuries across multiple parties, in the context of settlement	Defining injury footprints associated with a specific contaminant	Substitution with 1/2 DL appears to have been used. For one contaminant (pentachlorophenol), the 1/2 DL values exceeded the injury threshold resulting in a "ubiquitous injury footprint." With the exception of one small area, this footprint was deemed to not be allocable.	EcoChem and GeoSphere 2002	Appendix to the Hylebos Waterway NRDA Settlement Proposal Report; the appendix's title is "Natural Resource Damage Allocation of Injuries to Natural Resources in the Hylebos Waterway".
Indian Refinery, IL	PAHs	surface water, soil	Assessment plan	Calculating total PAHs	Do not explicitly say, but able to recreate the sum listed in Table 3.2 when ND= 0	Illinois Natural Resource Trustee Council 2006	
Kalamazoo River, MI	PCBs	Surface water; mice	Preassessment screen	--Summary statistics (surface water) --% detect, summary statistics (mice)	Minimum presented as 1/2 DL (surface water) ND listed in table (mice)	MDEQ et al., 2000	
Kalamazoo River, MI	PCBs, PCDDs, PCDFs	multiple	"Stage I" injury assessment	-Total PCBs as sum of 18 congeners (bass) -Total PCBs as sum of 77 congeners (bald eagles) -TCDD-equivalents of PCBs and PCDDs/PCFDs (fish eggs) -Number of samples with and without PCBs detected (residuals and soil samples) -Mean PCB concentration (residuals, soil samples, small mammals, shrews) -Comparison of concentration at site to concentration at reference site (surface	Bass: Substitution with 0 Bald eagles: "The Trustees used a linear regression model to estimate total PCB concentrations [defined here as 77 congeners] from the measured sum of [up to 20] detected congeners" Residuals, sediment, and soil: ND listed in table as part of a range from an earlier study Residuals, sediment, soil, small mammals, shrews: for the mean, substitution with 1/2 detection limit	MDEQ et al., 2005	

				water)	Surface water: plotted as 1/2 detection limit and identified with distinct symbol Fish eggs: calculated TCDD-eqs based only on detected PCB congeners		
Lake Hartwell, SC/GA	PCBs	multiple	Cooperative assessment of injury	(a) Charts of sediment concentration vs. percentile (b) Means	(a, b) ND included as full DL. In addition, for surface water, because DL for PCBs > water quality criteria, they "cannot be used for injury determination"	GDNR et al. 2006	Data are "inconclusive" for establishing injury to Lake Hartwell surface water based on water concentrations only because detection limits are higher than the criteria.
Lower Duwamish, WA	Multiple	not specified	This document is the response to public comments on the Supplement to the Draft Restoration Plan (RP) and Programmatic Environmental Impact Statement (PEIS); it references analyses (injury quantification) undertaken in other documents.	Geographic interpolation	In the context of defining injuries (Appendix C in the RP/Programmatic EIS), for non-detect values where the detection limit exceeds the service loss levels, no injury is assigned and the concentration is assumed to be close to zero for purposes of the geographic interpolation.	NOAA 2013	
Montrose, CA	DDTs, PCBs	fish	Sampling plan developed as part of the restoration program (post-settlement)	Calculating total homologues and total PCBs	Substitution with zero	Industrial Economics and CH2M Hill 2002	Goal was to provide scientifically defensible measures of geographic extent and severity of DDT and PCB contamination in local sports and subsistence fish. This is just the plan for the analysis.

Montrose, CA	DDTs, PCBs	fish	Fish survey report developed as part of the restoration program (post-settlement)	Calculating total PCBs	Substitution with zero	NOAA and EPA 2007	
Montrose, CA	DDTs, PCBs	peregrine falcon	Study report developed as part of the restoration program (post-settlement)	--Calculating total PCBs --Summary statistics (mean)	For calculating total PCBs, substitution with zero . Table 10 lists sample results including means; the presented mean appears to exclude non-detects	Latta 2012	
Palmerton Zinc, PA	metals	surface water	Assessment plan	Maximum concentrations compared to Federal Ambient Water Quality Criterion and Drinking Water Standards	In tables, ND stated to be the "maximum concentration" in some datasets.	Palmerton Natural Resource Trustee Council 2006	
Palmerton Zinc, PA	metals	soil	Scoping study	Mean, max, min of soil concentrations; mean compared to literature phytotoxicity thresholds	ND included in analysis as 0	Palmerton Natural Resource Trustee Council 2007	
Passaic River / Diamond Alkali, NJ	Dioxins and multiple others	multiple	Assessment plan	Detection frequency; summary statistics	Substitution, either with zero or with 1/2 DL. The approach varies sometimes because it depends on the source cited; in other cases, the reason for the differences in approach is not specified.	NOAA and USFWS 2007	
Phelps Dodge, NM/AZ	As, Be, Cd, Cr, Cu, Pb, Mn, Ni, Hg, Zn	sediments, invertebrates	Preassessment screen	--Calculating total concentrations	ND listed in tables	USFWS 2003	
Quivera Mine, NM	sulfate, uranium, radium, selenium	groundwater, soil , soil leachate ; surface water	Preassessment screen and determination	Mean; comparison of mean to water quality standards	Substitution with DL	New Mexico Office of Natural Resources Trustee 2010	
Rocky Mountain Arsenal, CO	dieldrin, aldrin, endrin, DDE, arsenic, mercury, DDT	small birds (to represent the terrestrial environment) and aquatic/semi-aquatic birds (to represent the aquatic environment)	Habitat equivalency analysis	Summary statistics (assumed)	Substitution with 1/2 DL	State of Colorado 2004	Developed on behalf of the State of Colorado, which was acting independently of the Federal Trustees in this matter. Some injuries (e.g., to Lake Ladora) are "driven by exposure point concentrations that are based on nondetect values [as] The detection limits for these media... were relatively high."
Rocky	dieldrin, aldrin,	terrestrial biota	Assessment plan	Summary statistics	Minimum presented as the certified	NRTSC 2007	Developed on behalf of the State

Mountain Arsenal, CO	endrin, DDE, arsenic, mercury				reporting limit (CRL). Means calculated using one-half the CRL, where the analyte was not detected.		of Colorado, which was acting independently of the Federal Trustees in this matter.
Rocky Mountain Arsenal, CO	dieldrin, aldrin, endrin, DDE, arsenic, mercury, DDT	fish	Preassessment screen	Summary statistics	"Mean is calculated when 50 percent or more of samples (n > 2) have detectable contaminant levels. If less than 50 percent of samples have detectable contaminant levels, only the range of values are presented. When calculating the mean, values of 1/2 the detection limit are substituted for 'BDL'"	Office of the Colorado Attorney General et al. 2007	Developed on behalf of the State of Colorado, which was acting independently of the Federal Trustees in this matter.
Southeast Missouri Lead Mining District, MO	cadmium, zinc, lead	sediments	Assessment plan	Percent of samples exceeding various thresholds	"For West Fork, most of the data were collected via XRF. The detection limit for Cd using this method is 40ppm, a value that exceeds the OMOE-Severe threshold of 10ppm. Therefore we calculated threshold exceedances using the raw data, some of which fell below the detection limit but above damage thresholds. We received confirmatory lab analysis for certain samples from West Fork. When confirmatory lab data were available, we used an average of the field-based XRF value and the lab value."	Mosby et al. 2009	
Southeast Missouri Lead Mining District - Big River Mine, MO	Co, Ni, Cu, Zn, Cd, Pb	crayfish, surface water, sediment	USGS administrative (technical) report	Summary statistics	Substitution with 1/2 DL	Allert et al. 2010	
Southeast Missouri Lead Mining District - Big River Mine, MO	Pb, Cd, Zn	freshwater mussel, sediment	Technical report	Percent relative standard deviation (RSD), calculated as SD/Mean x 100	RSD was stated to be "invalid" if one or more of the replicates were below the method detection limit In a table of metal data, "<LOD" was used to indicate a value below the limit of detection	Roberts et al. [undated]	Concentration information was used along with mussel population distribution information to tie contaminants to patterns of distribution
St.	PCBs, PAHs,	multiple	Restoration	--Summary statistics	A couple of tables list some	NRTSLRE 2013	

Lawrence, NY	fluoride, and metals		compensation and determination plan	--Counts of detects/NDs	individual biota samples that had an ND result. Counts of ND samples identified; mean concentrations calculated by substituting 1/2 DL for NDs.		
St. Lawrence, NY	PCBs	groundwater	Injury assessment published as part of RCDP	All values were ND; determination of no injury made	n/a	NRTSLRE 2013	The detection limits were less than or equal to EPA's drinking water standard but were greater than the human health criterion. Groundwater data were from reservation lands; NYSDEC did not elect to pursue a groundwater claim.
St. Lawrence, NY	PCBs, PAHs, fluoride, and metals	multiple	Restoration compensation and determination plan	--Summary statistics --Counts of detects/NDs	A couple of tables list some individual biota samples that had an ND result. Counts of ND samples identified; mean concentrations calculated by substituting 1/2 DL for NDs.	NRTSLRE 2013	
St. Lawrence, NY	PCBs	groundwater	Injury assessment published as part of RCDP	All values were ND; determination of no injury made	n/a	NRTSLRE 2013	The detection limits were less than or equal to EPA's drinking water standard but were greater than the human health criterion. Groundwater data were from reservation lands; NYSDEC did not elect to pursue a groundwater claim.
St. Lawrence, NY	PAHs	sediments	Injury assessment published as part of RCDP	Mean compared to service levels derived from sediment quality guidelines and thresholds from site-specific toxicity tests	ND included in analysis as 1/2 DL	NRTSLRE 2013	
St. Lawrence, NY	PCBs	sediments, fish	Injury assessment published as part of RCDP	Mean compared to service levels thresholds derived from sediment quality guidelines, site-specific toxicity tests, and literature studies	ND included in analysis as 1/2 DL, except when summing total PCBs (if an individual congener or Aroclor was reported as ND, that ND was included as 0 in the sum)	NRTSLRE 2013	
TVA, TN	multiple	sediment, fish	Injury assessment published as part of RCDP/EA	Geomean concentrations compared to sediment quality guidelines and/or literature thresholds	--ND included in analysis as 1/2 DL --Where the majority of samples were ND, even though we included the NDs as 1/2 DL, we did not use those datasets to directly assess injury in the assessment area.	NRTTVA 2015	

Vieques Island, Puerto Rico	Explosive compounds, PCBs, organochlorine pesticides, trace elements	land crab and fiddler crab	Data report and screening-level risk assessment	Calculating tPCBs/tDDTs; statistical summaries; comparisons with thresholds; comparisons between sample locations	Substitution with the MDL	NOAA and Ridolfi 2006	"The primary purpose of the investigation was to characterize concentrations of explosive compounds, polychlorinated biphenyls (PCBs), organochlorine pesticides, and trace elements in land and fiddler crab. In addition, the Agency for Toxic Substances and Disease Registry (ATSDR) has used the land crab data presented in this report to write a Public Health Consultation (PHC)... " The results from the ATSDR and NOAA "will assist the USFWS in determining whether selected refuge areas can be opened to crab harvesting." The file available online does not include the figures, tables, or appendices.

(This page intentionally left blank to facilitate double-sided printing)

APPENDIX C | GLOSSARY

TERM	DEFINITION
Akritas-Theil-Sen line (ATS)	Akritas-Theil-Sen line or ATS (Akritas et al., 1994) refers to a nonparametric procedure for calculating the slope of a regression line involving dataset containing non-detects. For more information see Section 3.3.
Censored data	A data condition in which the value of a measurement or observation is only partially known. If the value is known to fall below a certain level (although the exact value is uncertain), it is described as left-censored. Non-detect analytical results are an example of left-censored data. Right-censored data (not the main subject of this report) are those for which a value is known to fall above a certain level although the exact value is uncertain (e.g., an organism is known to have lived for at least X years but its exact lifespan is not known).
Detection limit (DL)	The lowest result that can be reliably detected and distinguished from the blank sample.
Distribution	A listing or function that shows all the possible values of a variable and their corresponding probabilities of occurrence. Examples of distributions include the normal, log-normal, and gamma distributions, among many others.
Gehan comparative test	This is a nonparametric test for evaluating the significance of differences in the medians of two datasets involving non-detects with varying detection limits. For more information, see DON (2002, Section 4.2.4).
Generalized Wilcoxon method	This is a generalized nonparametric test for evaluating the significance of differences in the medians of two datasets containing non-detects with varying detection limits. For more information, see Section 3.2.
Kaplan-Meier (KM) method	KM is a nonparametric method for construction of the cumulative distribution function (CDF) of a dataset that contains censored data. The constructed CDF in turn can be used to estimate the summary statistics of interest. For more information, see Section 3.1.4.
Kendall's tau method	This is a nonparametric method to calculate correlation between two paired datasets consisting of unique values based on the number of concordant and discordant pairs. A pair of (X_i,Y_i) and (X_j,Y_j) is concordant, when $(X_i-X_j)(Y_i-Y_j)>0$, otherwise discordant. For more information, see Helsel (2005, Chapter 11).
Linear regression	This is a parametric approach for modeling the relationship between a scalar dependent variable and one or more explanatory variables, see Helsel and Hirsch (1992, Chapter 9).

TERM	DEFINITION
Lower confidence limit (LCL)	A LCL is the lower bound of a confidence interval. A confidence interval is the range within which one is confident (at a selected confidence level, such as 95%) that the statistic will occur. For example, the mean of a dataset is often presented along with its 95% confidence interval (i.e., the LCL and UCL values), which together provide a measure of central tendency for the data (i.e., the mean), along with a measure of uncertainty in the mean (i.e., the confidence interval).
Mann- Kendall test	This is a nonparametric method to calculate the significance of Kendall-tau correlation between two paired datasets. For more information, see Helsel (2005).
Maximum likelihood estimation (MLE)	MLE is a method of estimating the parameters of a statistical model, such as the mean of a distribution, by maximizing the likelihood of the occurrence of the observed data. For more information, see Section 3.1.2.
Non-detect	An analytical result that falls below the relevant detection limit (DL).
Non-parametric methods	Statistical methods that do not make any assumption about the distribution of the data to which the method is being applied.
Paired Prentice-Wilcoxon method	This is a nonparametric variation of the Generalized Wilcoxon test that is especially designed for evaluating the significance of differences in paired datasets. For more information, See section 3.2.
Parametric methods	Statistical methods that depend on an assumption about the distribution of the data to which the method is being applied.
Pearson's r method	This is a widely-used parametric method to calculate linear correlation between two paired datasets. For more information, see Helsel (2005, Chapter 11).
Phi coefficient method	This is a nonparametric method to calculate correlation when the dataset contains large proportions of non-detects. This procedure requires conversion of the investigated datasets into binary values. For more information, see Section 3.3.
Quantile test	This is a nonparametric test for evaluating the significance of differences in the numbers of observed values exceeding a given quantile in two datasets. For more information, see DON (2002, Section 4.2.2).
Quantitation limit (QL)	The smallest detectable concentration that can be reliably quantified.
Regression on order statistics (ROS)	ROS is a semi-parametric, imputation technique to construct the cumulative distribution function (CDF), which in turn can be used for estimation of summary statistics of censored data. For more information, see Section 3.1.3.
Slippage Test	This is a nonparametric test for evaluating the significance of differences between two datasets based on the number of observed values in one dataset that exceed the maximum observed value in the other dataset. For more information, see DON (2002, Section 4.2.1).

TERM	DEFINITION
Student's t test	This is a widely used parametric test for evaluating the significance of difference between the mean values of two datasets. Modified versions of this method include the Welch's or Satterwaite's test, when the assumption of equality of variance is removed, see DON (2002, Sections 4.2.5 and 4.2.6)
Substitution or imputation methods	These methods assign surrogate numerical values to ND values which are then treated as equivalent to detected values in subsequent analyses. Typical surrogate values include 0, various fractions of DL, full DL or randomly assigned values between 0 and DL. For more information, see Section 3.1.1.
Theil-Sen line	Theil-Sen line refers to a nonparametric procedure for calculating the slope of a regression line. For more information see Helsel (2005, Chapter 12).
Two-sample test of proportions	This is a nonparametric test for evaluating the significance of differences between proportions of observed values in two datasets exceeding a given criterion. For more information, see DON (2002, Section 4.2.7).
Upper confidence limit (UCL)	A UCL is the upper bound of a confidence interval. A confidence interval is the range within which one is confident (at a selected confidence level, such as 95%) that the statistic will occur. For example, the mean of a dataset is often presented along with its 95% confidence interval (i.e., the LCL and UCL values), which together provide a measure of central tendency for the data (i.e., the mean), along with a measure of uncertainty in the mean (i.e., the confidence interval).
Wilcoxon rank sum test	This is a nonparametric test for evaluating the significance of differences in the medians of two datasets involving non-detects with identical detection limits. For more information, see DON (2002, Section 4.2.3).