Review

Not peer-reviewed version

How to Analyze Censored Concentration Data Using Modern Statistical Methods of Survival Analysis: Background and Nonparametric Methods

James N. McNair [*], Daniel Frobish, Isabelle Ciarrocchi [†], Richard R. Rediske

Posted Date: 25 March 2026

doi: 10.20944/preprints202603.1922.v1

Keywords: reporting limits; censored concentration data; left censoring; double censoring; nonparametric survival analysis; Turnbull survival function; pairwise comparisons; homogeneity test; monotonic trend test

 doi:10.20944/preprints202603.1922.v1

Review

How to Analyze Censored Concentration Data Using Modern Statistical Methods of Survival Analysis: Background and Nonparametric Methods

James N. McNair [1,*], Daniel Frobish [2], Isabelle Ciarrocchi [2,†], and Richard R. Rediske [1]

[1] Robert B. Annis Water Resources Institute, 740 West Shoreline Dr., Muskegon, MI 49441, USA
[2] Department of Statistics, Grand Valley State University, Mackinac Hall, 1 South Campus Drive, Allendale, MI 49401, USA
* Correspondence: mcnairja@gvsu.edu
† Current address: Perrigo Company, 515 Eastern Ave., Allegan, MI 49010, USA.

Abstract

Quantitative analytical methods for measuring concentrations of chemical substances in aquatic systems typically have acceptable accuracy and precision only for an intermediate range of analyte concentrations. Outside this range, the uncertainty of concentration estimates is too high to justify reporting them as valid measurements for use in statistical analyses. Therefore, concentration estimates falling below the lower reporting limit (LRL) typically are reported as the LRL, along with a code indicating that the measured values fell below the LRL. Such data are called *left-censored data*. Similarly, concentration estimates falling above the upper reporting limit (URL) typically are reported as the URL, along with a code indicating that the measured values exceeded the URL. Such data are known as *right-censored data*. Censored data violate assumptions underlying most parametric statistical methods, such as *t*-tests, regression analysis, and analysis of variance. We briefly review various statistical methods that have been employed for analyzing censored concentration data, then review in greater detail some modern statistical survival-analysis methods that have become available in standard software within the last 10 years and can be applied to concentration data with both left- and right-censored values. Methods are illustrated with real data.

Keywords: reporting limits; censored concentration data; left censoring; double censoring; non-parametric survival analysis; Turnbull survival function; pairwise comparisons; homogeneity test; monotonic trend test

1. Introduction

Studies of aquatic systems in ecology, environmental science, and environmental engineering commonly measure concentrations of chemical contaminants, nutrients, microorganisms, or genetic markers. The purpose of these studies is often to characterize spatial patterns or temporal trends of specific analytes, identify potential sources of contamination, or assess compliance with water-quality standards. A common nuisance in such studies is that a significant proportion of the concentration measurements fall below the lower reporting limit (LRL) for the analytical method being used. These measurements are by definition too uncertain to be treated as valid concentration estimates in statistical analyses; one can only be confident that the actual concentrations to which they correspond are less than the LRL. For that reason, their numerical values typically are not reported or used in statistical analyses. They are instead reported only as being less than the LRL, along with the numerical value of the LRL. In statistics, such values are referred to as *left-censored data*.

Some of the commonly used analytical methods also have an upper reporting limit (URL) that is low enough to be encountered in practice; the Colilert-18® enzyme-substrate method for quantifying coliform bacteria is an example. Like concentration measurements that fall below the LRL, those that exceed the URL are by definition too uncertain to be treated as valid concentration estimates in

 doi:10.20944/preprints202603.1922.v1

statistical analyses; one can only be confident that the actual concentrations corresponding to these values are greater than the URL. Such measurements typically are reported only as being greater than the URL, along with the numerical value of the URL. In statistics, such values are referred to as *right-censored data*.

Data sets that include a mix of valid and censored concentrations are the norm in field studies of aquatic systems, with left-censoring being particularly common. For example, McNair et al. [1] used paired Colilert-18 and qPCR measurements of *E. coli* concentrations in split samples from bathing beaches in the state of Michigan (USA) to compare the two types of data with respect to their usefulness for characterizing *E. coli* concentration distributions and detecting differences between such distributions for beaches on inland lakes, rivers, and the Laurentian Great Lakes. At one of the beaches (Wolverine Lake), 2.4% of the Colilert-18 concentrations in 2019 were right-censored and none were left-censored, while 32.8% of the qPCR concentrations were left-censored and none were right-censored. At another beach (Ross Lake), 14.8% of the Colilert-18 concentrations in 2019 were right-censored and none were left-censored, while 25.9% of the qPCR concentrations were left-censored and none were right-censored. In the combined 2019 and 2020 data for all beaches and sampling dates (3,205 pairs of observations), 2.8% of the Colilert-18 concentrations were left-censored and 0.3% were right-censored, while 52.2% of the qPCR concentrations were left-censored and none were right-censored.

In sharp contrast to the prevalence of censored concentration data in field studies, standard parametric statistical methods such as *t*-tests, least-squares regression, and analysis of variance require that the data to be analyzed consist entirely of valid concentration measurements. These are the statistical methods that would normally be used to assess potential differences between sampling sites or dates or to compare concentrations with water quality standards if all the data were valid measurements, but they are not appropriate for data sets that include a significant proportion of censored values. How, then, can such data sets be rigorously analyzed?

Historically, two different ways of approaching this problem have been popular with environmental scientists and engineers. The most common approach is to replace censored concentrations with fictitious values fabricated in some convenient way (e.g., by arbitrarily replacing each $<$ LRL value with LRL/2) and then proceed with statistical analysis as if all the data were valid measurements. The other approach is simply to treat concentration estimates below the LRL or above the URL as if they were valid measurements. The rationale for the latter approach is that, while values outside the reporting limits do indeed have unacceptably high uncertainty, one can argue that they are nevertheless likely to be closer to the true concentrations than are the fictitious values fabricated by the former approach.

Both of these traditional approaches to handling censored concentration data are clearly unsatisfactory and scientifically indefensible. Beginning in the 1980s, various authors drew attention to this fact. Helsel [2] provides a good summary of evidence that the two historical approaches to handling censored concentration data just mentioned can lead to incorrect or misleading conclusions, and also provides a good survey of alternative statistical methods for handling such data that were available in standard statistical software in the early 2000s.

The purpose of the present paper is to provide an updated survey of modern statistical methods of survival analysis that are appropriate for analyzing concentration data that may include a mix of valid, left-censored, right-censored, and possibly interval-censored values, and to provide examples of how to implement these methods using R [3] and SAS [4] statistical software. Though certain classical statistical methods can be applied to data that include censored values (e.g., by partitioning observed concentrations into two or more discrete classes, with $<$ LRL and $>$ URL values included in separate classes), we restrict attention to methods from the specialized statistical discipline of survival analysis that are specifically designed for data sets that may include multiple types of censored values.

Various authors have previously advocated the use of traditional methods of survival analysis (e.g., the Kaplan-Meier estimator of the survival curve, estimates of its point-wise confidence limits, and estimates of distribution quantiles and their confidence intervals) that were originally developed

 doi:10.20944/preprints202603.1922.v1

for applications in medical research and are applicable only to right-censored data. By contrast, most of the older environmental studies that produced censored concentration data used detection limits as the sole basis for censoring (an inappropriate practice that we strongly discourage; see Section 2), so left-censoring was the only form of censoring that was possible. Use of traditional methods of survival analysis for such data requires reversing or "flipping" the concentration scale to transform left-censored values into right-censored values [2]. But flipping cannot resolve the problem if both left-censored and right-censored values are present, as will sometimes be the case if proper methods of quantitative analytical chemistry are used to establish both lower and upper reporting limits.

In contrast to previous accounts of survival-analysis methods for concentration data, we will focus on modern methods that do not require flipping the concentration data, and that can accommodate data where the censored values may include any mix of left-censored, right-censored, and interval-censored values. The present paper restricts attention to nonparametric methods, which are widely available in standard statistical software (we plan to publish a future paper dealing with semiparametric methods, including the Cox proportional hazards model and the accelerated failure-time model, but versions of these methods that can handle doubly-censored or interval-censored data are not yet available in standard statistical software). The methods we review permit one to estimate the probability distribution function and its pointwise confidence intervals for the concentration of a given analyte at an individual sampling site or date, estimate concentration quantiles and their 95% confidence intervals, perform one-sided and two-sided pairwise comparisons of samples from different sites or dates, test the null hypothesis of homogeneity for samples from multiple sites or dates, and perform tests for monotonic trends across samples from multiple sites or dates.

The remainder of this paper is organized as follows. Section 2 reviews some background information from quantitative analytical chemistry that explains how censored concentration data arise. Section 3 provides a brief overview of various methods that have been used in the past for analyzing censored data. Section 4 reviews some important concepts and terminology from statistical survival analysis. Section 5 presents an overview of the main nonparametric methods of survival analysis that can be applied to censored concentration data for one, two, or multiple sites or dates. Section 5 also includes examples where each statistical method is applied to real concentration data using R software (examples using SAS software are included in the online supporting information). We conclude with a general discussion in Section 6.

2. How Do Censored Concentration Data Arise?

Methods of quantitative analytical chemistry typically have a working range of concentrations of the target analyte, within which the reliability of concentration estimates is adequate and outside of which it is not. The lower and upper boundaries of this range serve as reporting limits for the method, meaning that only concentrations lying between the two limits are sufficiently reliable to justify reporting them as numerical concentrations and using them in statistical analyses. In laboratory studies, one can often dilute or concentrate samples that lie outside the reporting limits and re-analyze them. But in large-scale field studies and monitoring programs, this usually is either not necessary (e.g., if the goal is to assess compliance with a regulatory standard, and the standard lies well within the reporting limits), not feasible (e.g., if funding levels do not permit re-analysis), or not possible (e.g., if the sample analysis time exceeds the sample holding time, as is often the case when monitoring *E. coli* levels with the Colilert-18 enzyme-substrate method). In such cases, values lying below the lower reporting limit (LRL) are reported simply as less than the LRL, along with the numerical value of the LRL. Similarly, values lying above the upper reporting limit (URL) are reported simply as greater than the URL, along with the numerical value of the URL. This common data-reporting practice ensures that any data set that includes unreliable measurements lying outside the reporting limits will contain censored observations.

The specific estimation methods and numerical criteria for the LRL and URL vary for different types of analytical methods, and often vary among different authors and standards organizations

 doi:10.20944/preprints202603.1922.v1

for the same analytical method. But because of the importance of these limits as the source of censoring for concentration data, it will be useful to briefly consider an example that illustrates some of the main reasons they are necessary, while largely avoiding the specific numerical criteria that vary among authors and standards organizations (see [5–7] for additional details). For this purpose, we will consider the large class of analytical methods that estimate concentrations using a calibration curve. Examples of such methods that are commonly used to analyze samples from aquatic systems include molecular absorption spectroscopy, gas chromatography-mass spectrometry (GC-MS), liquid chromatography-tandem mass spectrometry (LC-MS/MS), inductively coupled plasma mass spectrometry (ICP-MS), and real-time quantitative polymerase chain reaction analysis (qPCR). In all these cases, the calibration curve relates instrument response (signal intensity, peak area, ion counts, threshold cycle number) to known standard concentrations. The curve is typically constructed using multiple standards across a defined concentration range.

For the many methods that rely on them, calibration curves play a central role in defining the statistical boundaries of environmental measurement data. Before any statistical treatment of censored values occurs, analytical chemistry has already imposed quantitative boundaries through the calibration model. These constraints establish the lower and upper limits within which measured concentrations are considered quantitatively reliable. Thus, the calibration curve is not merely a laboratory artifact; it is the mechanism that defines the measurable domain of a dataset. For this reason, the development and validation of the calibration curve should be viewed as the first statistical decision in any environmental measurement program that requires one, because it defines the quantitative limits within which environmental data can be interpreted defensibly.

In mathematical language, a calibration curve is a strictly monotonic function $y = S(x)$ (in most cases increasing with analyte concentration, but in some cases decreasing) that maps the concentration x of a target analyte to the intensity y of a signal from an analytical instrument (Figure 1A, blue curve). Once fitted to data from a series of known analyte concentrations, the calibration curve can be inverted to yield a strictly monotonic function $x = S^{-1}(y)$ (Figure 1B, blue curve) that can be used to estimate unknown concentrations of the analyte in field samples by measuring the intensity of the instrument signal.

A common example is the degree to which light of a specific wavelength is absorbed by a dissolved analyte (as measured with a spectrophotometer), over a fixed path length, as a function of analyte concentration. According to Beer's law, the proportion of transmitted light (i.e., transmittance) should decline exponentially with increasing analyte concentration, so the negative logarithm of transmittance (i.e., absorbance) should be directly proportional to the analyte concentration. But as Skoog et al. ([8], p. 306) note, systematic deviations from direct proportionality commonly occur, especially at high concentrations but sometimes at low concentrations, as well. Thus, the relationship between analyte concentration and absorbance — the calibration curve — is reliably linear only over an intermediate range of concentrations.

As an example, Fritz and Schenk ([9], p. 350) make the following specific recommendation: "Spectrophotometric measurement of absorbance (or transmittance) is very inaccurate at both very low and very high readings. For this reason, the concentration of absorbing substance should always be adjusted until the absorbance is in the range 0.10–1.00 (or to 1.50 for some precision spectrophotometers)." But as noted above, when conducting large-scale field studies with limited funding, dilution or concentration of samples is often either not necessary (e.g., if the purpose of the study is to assess compliance with a numerical regulatory criterion that is well within the limits of quantification without dilution), not feasible, or not possible. In such cases, left-censored, right-censored, or doubly-censored data may result, depending on the range of field concentrations.

An analytical method used to create a calibration curve typically is adequately sensitive to analyte concentration only over an intermediate range of concentrations, with sensitivity (slope of the calibration curve, Figure 1A) markedly declining at concentrations below and above this range. The insensitive portions of the curve appear as "shoulders" at low and high concentrations

 doi:10.20944/preprints202603.1922.v1

(Figure 1A, blue curve). The sensitive part of the calibration curve typically is well approximated by a straight line (Figure 1A, dashed red line), possibly after transforming the analyte concentration or instrument signal intensity or both. Linearity at intermediate concentrations often is supported by a mechanistic physicochemical theory (e.g., Beer's law), which breaks down at low and high concentrations where the calibration curve becomes distinctly nonlinear. Also, to the best of our knowledge, the rigorous statistical theory of calibration (including estimation of confidence intervals for predicted concentrations) has been adequately developed only for linear calibration curves ([10], section 4.6; [11]; [12], section 15.3). For these reasons, the LRL must be no smaller than the lower limit of the concentration interval where the calibration curve is well approximated by a straight line, and the URL must be no larger than the upper limit of that interval (Figure 1A, gray arrows).

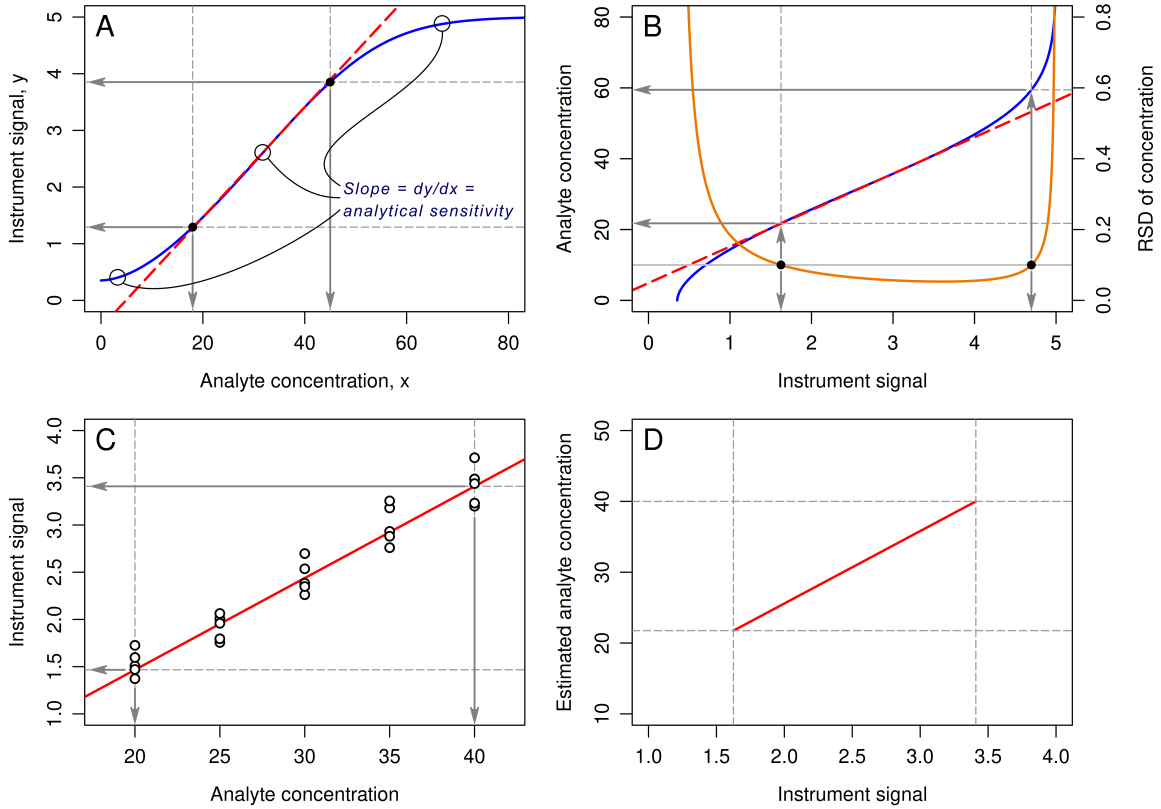

Figure 1. Hypothetical calibration curve and its inverse. **A**: A nonlinear calibration curve $y = S(x)$ (solid blue line) and linear approximation to its central portion (dashed red line). Vertical and horizontal arrows indicate the limits of analyte concentration and instrument signal intensity for which the linear approximation is considered satisfactory. Analytical sensitivity at concentration x is the slope of the calibration curve at x. Based loosely on Figure 3A of the Eurachem lab guide [7]. **B**: Inverse function $x = S^{-1}(y)$ (solid blue line) of the nonlinear calibration curve in panel A and its linear approximation (dashed red line). Orange curve is the relative standard deviation (RSD) σ/μ of estimated concentration, a unitless measure of uncertainty. Solid gray horizontal line is the maximum level of uncertainty of concentration estimates that is considered acceptable, commonly chosen as 0.1 [5]. Arrows indicate the implied lower and upper limits of quantification in the signal domain (horizontal axis) and concentration domain (vertical axis). **C**: Linear calibration curve $y = \hat{\beta}_0 + \hat{\beta}_1 x$ (solid red line) from panel A, fitted to calibration data (circles) for five standard concentrations; $\hat{\beta}_0$ and $\hat{\beta}_1$ are least-squares estimates of the intercept and slope parameters. Arrows indicate the minimum and maximum standard concentrations (horizontal axis) and corresponding maximum and minimum instrument signal intensities (vertical axis). **D**: The inverted form $x = (y - \hat{\beta}_0)/\hat{\beta}_1$ of the fitted linear calibration curve in panel C, clipped to the ranges of analyte concentration and instrument signal intensity for which the linear approximation and level of uncertainty are acceptable and analyte concentrations lie between the minimum and maximum standards.

An additional reason for using only the linear portion of the calibration curve is that, when the curve is inverted for use in estimating analyte concentrations from measured values of instrument

 doi:10.20944/preprints202603.1922.v1

signal intensity, the shoulders at low and high instrument signal intensities become steeply sloped regions where concentration estimates are highly sensitive to variation in instrument signal intensity (Figure 1B, blue curve). When signal intensity measurements are converted to estimated concentrations, these steep nonlinear portions of the inverted curve greatly magnify the uncertainty of concentration estimates, as indicated by the relative standard deviation (RSD, called the coefficient of variation in statistics) (Figure 1B, orange curve). The degree of magnification is approximately proportional to the local slope of the inverted curve: the steeper the slope is, the more sensitive the inverted curve is to error in the signal intensity and hence the greater the error in predicted analyte concentration. The result is unacceptably high uncertainty of concentration estimates for signal intensities in the nonlinear parts of the inverted curve, with the acceptable limit of uncertainty commonly being chosen as RSD = 0.1 (Figure 1B, solid gray horizontal line). The concentrations corresponding to the lower and upper limits of this interval of acceptable measurement uncertainty are the lower and upper limits of quantification (LLOQ, ULOQ) (Figure 1B, gray arrows).

As noted above, the linear part of the calibration curve is fitted to calibration data consisting of measured values of instrument signal intensity for replicates of a series of standards with known concentrations of the target analyte, using an appropriate statistical regression method (Figure 1C). This procedure yields estimates of the slope and intercept parameters of the linear regression model. The inverse function is then easily determined, yielding a function that allows one to estimate unknown analyte concentrations in samples, based on measured values of the instrument signal intensity (Figure 1D). The associated statistical theory for linear calibration (e.g., [11]) also allows one to estimate 95% confidence intervals for estimated concentrations. But because the fitted calibration curve and its inverse function are only valid for the range of standard concentrations employed, the minimum and maximum standard concentrations and their associated minimum and maximum instrument signal stengths impose another set of limits on how small the LRL can be and how large the URL can be (Figure 1C, gray arrows).

Taking all of the above considerations into account, the LRL would be chosen as the maximum of the lower limit of linearity, the lowest standard concentration, and the LLOQ, while the URL would be chosen as the minimum of the upper limit of linearity, the highest standard concentration, and the ULOQ (Figure 1D). Only measured concentrations x such that LRL $\leqslant x \leqslant$ URL are reported as numerical concentrations. Values below the LRL are reported simply as the LRL, along with a code indicating that the measured value was below the LRL. Similarly, values above the URL are reported as the URL, along with a code indicating that the measured value was above the URL. The exact format for recording this information in a spreadsheet or database depends jointly on which statistical methods and which statistical software (e.g., R or SAS) will be used to analyze the data (see Section 4.2). But the key point is that, because sample concentration estimates less than the LRL or greater than the URL are not reported as numerical concentrations, data sets containing such values will be left-censored by the LRL and right-censored by the URL. The entire process of sample collection, sample preparation, sample analysis, estimation of analyte concentrations with an inverse calibration curve, data reporting subject to reporting limits, and statistical analysis of the resulting censored data is summarized schematically in Figure 2 (adapted from Figure 1 of [5]).

Finally, we note that older studies of contamination in aquatic systems often ignore the quantification limits of concentration measurements and focus only on the limit of detection (LOD), which is then used as the sole reporting limit. Definitions of the LOD vary, but it is essentially a measure of the lowest instrument signal intensity (and corresponding estimate of analyte concentration) that one can be reasonably confident is too high to have been produced by a sample that does not contain the target analyte [5–7]. An important fact is that the LLOQ typically is greater than the LOD. For example, Eurachem ([7], pp. 24–25) notes that as a rule of thumb, the LOD is roughly 3 times the standard deviation of the blanks, while the LLOQ is roughly 10 times the standard deviation of the blanks and hence about 3.3 times the LOD. The LOD is useful for providing evidence that the analyte is or is not present in a sample at a concentration high enough for the analytical method to reliably detect it, but it

 doi:10.20944/preprints202603.1922.v1

does not provide evidence that the concentration can be measured with acceptable uncertainty. That evidence is provided by the LLOQ, ULOQ, and (for methods that employ a calibration curve) the limits of linearity of the calibration curve and the maximum and minimum calibration standards. It follows that the LOD typically is irrelevant in determining reporting limits. The relationships between the LOD, LRL, and URL are shown schematically in Figure 3.

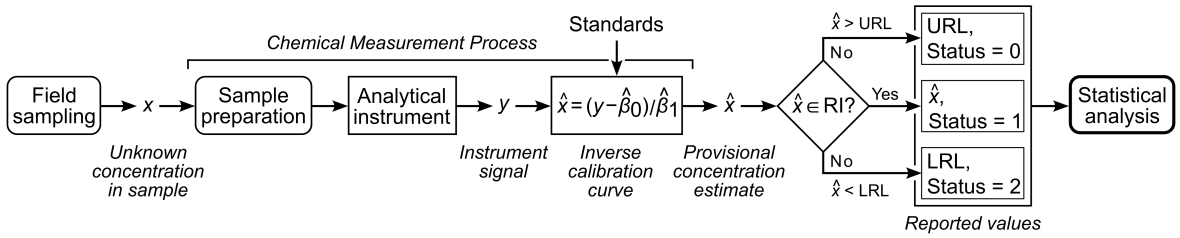

Figure 2. Schematic representation of the entire process of sample collection, preparation, and analysis, estimation of analyte concentration with an inverse calibration curve, data reporting subject to reporting limits, and statistical analysis of the resulting censored data. The combination of sample preparation, use of an analytical instrument to produce an instrument signal whose intensity is a strictly monotonic function of the unknown analyte concentration, and conversion of the instrument signal to a concentration estimate is what Currie [5] calls the *chemical measurement process*. Notation: $\hat{\beta}_0$ and $\hat{\beta}_1$ are least-squares estimates of the intercept and slope parameters of the calibration curve; LRL and URL are the lower and upper reporting limits; RI is the reporting interval [LRL, URL]. The status codes shown for reporting data are those used in R's `interval` format (see Section 4.2).

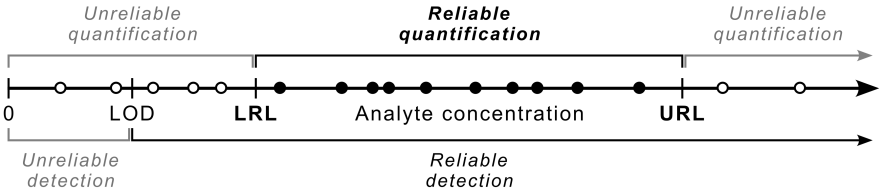

Figure 3. Relationships between the limit of detection (LOD), lower reporting limit (LRL), and upper reporting limit (URL). Dots on the concentration axis represent data; values between the reporting limits are shown as filled black dots, while values outside the reporting limits are shown as open dots.

3. Examples of Methods for Analyzing Censored Concentration Data

Helsel [2] and Shoari [13] review various rigorous and not-so-rigorous methods that have been used in the past by researchers in environmental science and engineering to analyze censored concentration data. In this section, we outline a few of the most commonly-used methods to illustrate the variety of alternative approaches.

Before proceeding, however, it is necessary to point out a problem with the term "sample" as traditionally applied to concentration data by environmental scientists and engineers and to indicate how we will resolve it in this paper. Environmental scientists and engineers commonly use this term to mean two distinctly different things: (1) an individual volume of water or other environmental material collected in the field and (2) a set of several such volumes collected individually from a particular sampling site that are intended to be representative of the site. The first usage often appears when describing field sampling methods or laboratory analytical methods, while the second often appears when describing statistical methods or presenting results of statistical analysis. To avoid using the same term in these two conflicting ways, we need a different term for one of the meanings. Following Hinkelmann and Kempthorne [14], we will use the term "sampling unit" for the first meaning and the term "sample" for the second. With this terminology, a sample consists of multiple sampling units, the number of sampling units is the sample size, and each concentration measurement is obtained from a single sampling unit.

 doi:10.20944/preprints202603.1922.v1

3.1. Deleting Censored Concentrations

One approach to handling censored data is to delete them or, slightly less flagrantly, to delete all data from sampling locations where some of the data are censored. This approach eliminates the problem of deciding how to extract valid information from the censored data and also mitigates the problem of violating assumptions of standard statistical methods such as ordinary least-squares regression and analysis of variance (these statistical problems are briefly discussed in Section 3.2). Though it is not difficult to find examples of published studies that use this approach, it should never even be considered. For example, left-censored values will be the lowest concentrations in a data set if there is a single LRL, and will be among the lowest if the data include subsets with different LRLs. It should be obvious that deleting these values will bias any inferences about the mean or median concentration at a given sampling site or time and therefore will invalidate statistical comparisons between different sites or times.

3.2. Data-Fabrication Methods

The most common approach to handling censored concentration data in the past has been to replace censored observations with fabricated values, then analyze the edited data set using standard statistical methods for noncensored data such as *t*-tests, least-squares regression models, or analysis of variance. For purposes of illustration, we discuss two examples of the various types of data-fabrication methods: replacing each $<$ LRL value with LRL/2, and replacing all $<$ LRL values by fabricated values derived from a regression model based on quantiles or order statistics of an assumed parametric probability distribution.

3.2.1. One-Half LRL

Replacement of censored values by one-half the LRL is, in our experience, by far the most common method of handling left-censored concentration data in applied environmental studies. If, for example, there is a single LRL and 20% of the concentrations in a data set comprising 100 observations fall below it, then after replacement, 20% of the data will share exactly the same numerical value (= LRL/2), and this value will have no objective support in the data. Aside from the completely arbitrary nature of this method, the resulting large number of ties in the revised data set clearly is not consistent with a normal distribution or with any other continuous distribution. Therefore, parametric statistical methods that assume the data come from a normal distribution, as well as nonparametric methods that simply assume the data come from a continuous distribution, will fail any valid assessment of their assumptions. The same will be true if observations less than the LRL are replaced with a value other than LRL/2, such as zero, LRL, or LRL/$\sqrt{2}$, all of which have been employed in the past [2,15–17]. Because any single replacement value for left-censored data will be both fabricated and arbitrary, and also because replacement of all left-censored data with the same value will produce a large number of ties when the proportion of left-censored data is significant, we strongly discourage the use of this method.

3.2.2. Methods Based on Quantiles or Order Statistics

An obvious and serious problem with data-fabrication methods that substitute the same value for every left-censored observation is that any significant number of censored observations will result in a sufficient number of ties to invalidate most standard statistical methods. This fact stimulated interest in data-fabrication methods that spread the fictitious concentrations over the interval between 0 (or $-\infty$, for log-transformed data) and the LRL instead of concentrating them on a single value (Figure 4). Here we mention two methods of this type: one based on quantiles and the other on order statistics. These methods are more complicated than the LRL/2 method just described, so to conserve space, we describe them only briefly here (Shumway et al. [18] provide a good concise overview). Importantly, both methods depend critically on the choice of a specific parametric probability distribution (typically a normal or lognormal distribution) as the distribution from which the data were sampled. The basic idea behind the methods is to fit the chosen probability distribution to the noncensored data and then

 doi:10.20944/preprints202603.1922.v1

assign fabricated values to the censored data, based on the theoretical quantiles or expected values of the order statistics for the chosen distribution. These methods avoid the problem of assigning the same value to all censored data, but they suffer from two serious shortcomings: the values assigned to the censored data are fabricated, and especially when a nontrivial proportion of the data are censored, the missing portion of the concentration distribution precludes an adequate assessment of the appropriateness of the chosen probability distribution.

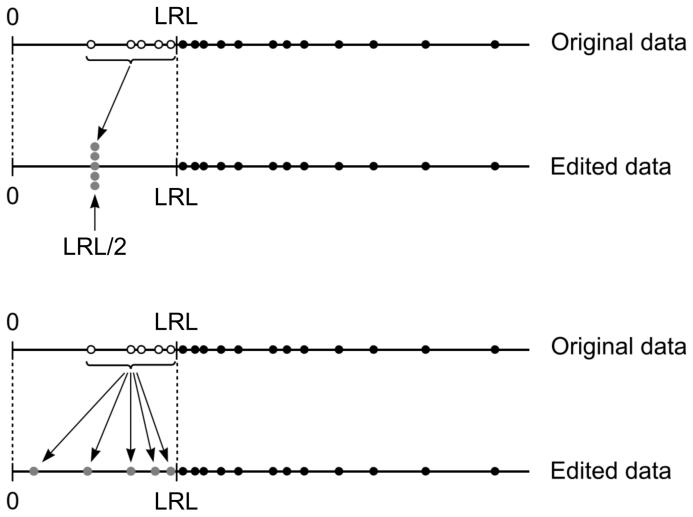

Figure 4. Two types of data-fabrication methods. The examples show two ways of handling a set of data comprising n observed concentrations, of which c are left-censored (open circles) and the remaining $n - c$ are noncensored (filled black circles). The most common way to handle censored observations is to replace each with half the LRL, yielding an edited data set with c tied values at LRL/2 (top example, filled gray circles). An alternative approach is to employ a statistical method that uses the $n - c$ noncensored data to estimate the expected values of the first c quantiles or order statistics of the full data set, based on an assumed parametric probability distribution, and then to replace the censored observations with these fabricated values (bottom example, filled gray circles).

3.3. Ignoring the Reporting Limits

The simplest approach to dealing with data that are below the LRL or above the URL is to ignore the problem and use all the estimated concentrations as if they were valid concentration measurements, including readings that are outside the reporting limits and therefore are known to have unacceptably high levels of measurement error. As mentioned in the Introduction, the original rationale for this approach is that, while concentration measurements outside the range of quantification admittedly have unacceptably high imprecision and/or bias, they may nevertheless be closer to the actual concentrations than are the purely fictitious values created by data-fabrication methods. In our experience, this approach is not nearly as common as replacing censored values with half the LRL.

Helsel ([2], section 3.3) discusses the problems inherent in this method. Gilbert ([19], p. 178) states that this method produces biased estimates of the mean and variance of the analyte concentration. Antwiler and Taylor [15] found that it performed poorly in estimating the mean, quartiles, and standard deviation compared to various data-fabrication methods, parametric maximum-likelihood estimators, and nonparametric survival-analysis methods. However, George et al. [17], based on results of a simulation study, concluded that ignoring the reporting limits is the best approach. But because their simulation study does not include a plausible representation of measurement error, the excessive levels of measurement bias and imprecision that characterize data outside the reporting limits in real data sets—the two main problems that reporting limits are designed to mitigate—are absent from the simulated data. Reporting limits were therefore prevented from playing their intended role in filtering out highly biased and imprecise data (because no such data were present), and it is therefore

 doi:10.20944/preprints202603.1922.v1

not surprising that statistical analyses were not degraded by treating data outside the reporting limits as valid measurements.

Our own view is that concentration estimates that lie outside the reporting limits, and that are therefore known to be contaminated with unacceptably high levels of measurement error, should not be used in statistical analyses, even if their numerical values are available. A guiding principle of rigorous statistical analysis of censored data is to use only information about the data that is known with high confidence. Survival-analysis methods applied to properly censored concentration data are fully consistent with this principle. They use numerical measurements only for data whose bias and imprecision are known to be acceptably low. Censored values are included in analyses, but only information about them that is known with high confidence is employed; namely, that the true concentration corresponding to each censored concentration estimate lies below a particular numerical value (the LRL) or above a particular numerical value (the URL). These methods are statistically rigorous and have an extensive and proven record of successful application in medical research involving censored time-to-event data.

3.4. Partitioning Concentrations into Discrete Classes

An obvious way to handle data that include censored values is to partition the concentration scale into two or more discrete classes, one of which contains all concentrations less than the LRL as a subset and another that contains all concentrations greater than the URL as a subset. In our experience, the most common example of this approach occurs in monitoring or assessment studies where the goal is to determine whether the concentration of a particular analyte exceeds a threshold value (e.g., a water-quality standard or some other management target). Typically, it is possible to choose or adapt an analytical method such that the LRL and URL for every sample unit analyzed brackets the threshold value (if the URL is greater than any concentration observed, it can be treated as infinite). Every sample unit analyzed can then be classified as either exceeding the threshold ("success") or not ("failure"), and standard nonparametric statistical methods for binomial success probabilities or proportions can be used to rigorously analyze the data. The binomial parameter here represents the probability that the concentration measured in a randomly chosen sample unit will exceed the threshold value, but for most purposes it can be interpreted as the theoretical proportion of a large number of sample units that would exceed the threshold.

Several kinds of statistical analyses can be performed using such data, with the appropriate methods depending on sample size and in some cases on whether the estimated success probability is very close to 0 or 1. Examples include estimating the success probability and its 95% confidence interval, estimating the difference between success probabilities and its 95% confidence interval for two sites or dates, and testing the null hypothesis that the success probabilities for two sites or dates are the same versus the two-sided or appropriate one-sided alternative hypothesis. Particularly good estimators for confidence intervals of a binomial success probability are the Wilson interval and the Agresti-Coull interval [20,21]. For large samples, good estimators for confidence intervals of the difference between two success probabilities are the Newcombe hybrid score interval and the Agresti-Caffo interval [22–24]; for small samples, exact intervals can also be computed (e.g., [25,26]). For testing the null hypothesis of no difference between the success probabilities for two sites or dates against one-sided or two-sided alternative hypotheses, classic large-sample tests are available [23,27], but small-sample tests are often preferable, with the best methods being the mid-P version of the exact conditional binomial test and various versions and modifications of Barnard's exact unconditional binomial test [23,27–29].

3.5. Survival-Analysis Methods

The fact that the statistical methods of survival analysis are specifically designed to accommodate censored data distinguishes them from most other statistical methods, but this is not their only distinguishing feature. Another fundamental difference is that survival analysis focuses mainly on characterizing and comparing entire probability distributions rather than measures of location (central

 doi:10.20944/preprints202603.1922.v1

tendency). As discussed in the next section, probability distributions can be characterized in several mathematically equivalent ways, including the probability distribution function, complementary probability distribution function (survivor function), and probability density function. Rigorous methods are also available for estimating distribution quantiles and their confidence limits, and for assessing effects of covariates, but the main emphasis is always on characterizing and comparing entire distributions.

Three main types of statistical methods for characterizing and comparing probability distributions are available in survival analysis: nonparametric, semiparametric, and fully parametric. We briefly review all three of these approaches in the present section. The remainder of this paper, however, focuses entirely on nonparametric methods.

Nonparametric methods of survival analysis avoid assuming any specific probability distribution from which observations are sampled. The main tools employed for concentration data that include a mix of valid, left-censored, right-censored, and/or interval-censored observations are the Turnbull estimator of the complementary distribution function, various nonparametric tests of homogeneity and pairwise differences such as the log-rank test and its various generalizations, and nonparametric tests for monotonic trends. We discuss these tools in more detail in Section 5. In the Supplementary Materials, we provide examples of R and SAS code to illustrate application of these methods to data.

Semiparametric methods also make no assumption about the mathematical form of the probability distribution from which observations are assumed to be sampled, but they do assume a parametric form for the dependence of this distribution on categorical or quantitative covariates. The best-known semiparametric tools are the Cox proportional hazards model, the accelerated failure-time model, and various extensions of both. Modern forms of these methods that can be applied to the full range of types of censored concentration data commonly encountered in studies of aquatic systems (i.e., left-censored, right-censored, doubly-censored, and interval-censored data) are still being actively developed by survival-analysis statisticians. As mentioned in the Introduction, we plan to address semiparametric methods in a separate paper, both because versions of these methods suitable for doubly-censored and interval-censored data are not yet available in standard statistical software and because semiparametric methods require lengthier exposition than do nonparametric methods.

Fully parametric methods of survival analysis require one to assume a specific mathematical form for the probability distribution from which observations are assumed to be sampled. Methods that make use of covariates also require one to assume a specific parametric form for the dependence of the assumed distribution on the covariates. Applications of these methods to censored time-to-event data in medical research have shown that the statistical results often depend strongly on the particular probability distribution chosen, while the data typically do not provide sufficient evidence to adequately justify any particular choice [30]. For that reason, modern statistical survival analysis in medical research almost always uses nonparametric and semiparametric methods, and we suggest that these are likely to be the most useful methods for environmental concentration data, as well.

4. Basic Concepts and Terminology

4.1. Functions for Specifying Probability Distributions

The specialized statistical discipline of survival analysis deals with time-to-event data where some of the event times are censored in the sense that one only knows they are greater than some value (e.g., the final observation time in an experiment), less than some value (e.g., the first observation time in an experiment), or between two values (e.g., successive observation times). In biological and medical applications, where the term "survival analysis" originated, the event of interest might be (for example) the time until an individual organism dies, a seed germinates, or an embryo transitions from one developmental stage to the next. In engineering applications, the event of interest might be the break-down or failure of a machine or electrical system, and in this context, the same statistical discipline is often referred to as reliability analysis or failure-time analysis. The broader term, time-to-event analysis, emphasizes the conceptual unity of these applications and reduces the terminological

 doi:10.20944/preprints202603.1922.v1

confusion created by applying methods with names referring to survival or failure times to, for example, seed germination times.

In all of these traditional applications of survival analysis, the focus is on the time until an event occurs. But if one takes a step back and considers more generally the basis of these statistical methods, it becomes clear that the underlying concepts are not restricted to processes unfolding in time. In reality, the fundamental concepts simply involve properties of probability distributions, and the discipline of survival analysis deals with methods for characterizing and comparing those distributions when some of the data are censored. Thus, there is no reason why survival analysis cannot be applied to data comprising concentrations of chemical compounds or microorganisms where some of the observations are censored (e.g., below the LRL). In our experience, however, the traditional terminology of survival analysis acts as an impediment to acceptance of these methods for applications involving concentrations instead of event times, because the picturesque terms typically used (e.g., survivor function, failure-time distribution function, hazard function) only make intuitive sense if the data are generated by a process that acts over time and terminates the life of an organism or a machine. Therefore, we will suggest alternative terms that, while less vivid, have the advantage of not suggesting associations that are incompatible with applications to chemical or microbial concentration data.

We assume that concentration data from any given sampling site or date are sampled from a probability distribution, and that the true values of all concentrations (not the estimates generated by an analytical instrument) lie in the interval $[0, \infty)$. For technical reasons, we assume the probability is zero that the true value of the analyte concentration is identically zero, since we wish to exclude probability distributions that have an "atom" of probability at a concentration of zero. Implicitly, then, we assume that "everything is everywhere", so that any contaminant of interest will be present even at pristine sites but at concentrations that might be so low as to be undetectable by available chemical or microbiological analytical methods.

We will use capital letters (e.g., X) to denote random variables representing concentrations, and lower-case letters (e.g., x) to denote observed or measured values of random variables. Four functions related to continuous random variables are important in survival analysis, with their traditional names being the probability distribution function, survivor function, probability density function, and hazard function (Table 1). In order to understand the statistical methods of survival analysis, it is necessary to be familiar with these functions, which we now briefly discuss.

Table 1. Probability functions for a continuous random variable.

Name for time-to-event data	Name for concentration data
Probability distribution function	Probability distribution function (PDF)
Survivor function	Complementary probability distribution function (cPDF)
Probability density function	Probability density function (pdf)
Hazard function	Attenuation function (AF)

The *probability distribution function* (PDF) of concentration X, denoted $F(x)$, is a dimensionless function that tells us the probability that the value of random variable X is less than or equal to x. That is,

$$F(x) = \Pr\{X \leqslant x\}, \tag{1}$$

where x typically has dimensions of Mass $\cdot$ Volume^{-1} and its possible values lie in the interval $[0, \infty)$. (It is sometimes desirable to analyze concentration data after log-transforming them, in which case possible values of x lie in the interval $(-\infty, \infty)$; to simplify our exposition, we will assume concentration data have not been log-transformed and therefore lie in the interval $[0, \infty)$.) We will restrict attention to distribution functions $F(x)$ that are differentiable and for which the limit of $F(x)$ as x approaches zero from above is zero; i.e., $\lim_{x \downarrow 0} F(x) = 0$.

Be aware that, whereas "probability distribution function" is the term used for $F(x)$ in probability theory (with the more general term "distribution function" being used in measure theory for functions

 doi:10.20944/preprints202603.1922.v1

that have nothing to do with probability), it is common in applied statistics to refer to $F(x)$ as the *cumulative* probability distribution function. But as probability theorists dutifully point out (e.g., [31], p. 179), the qualifier "cumulative" is redundant because a distribution function is cumulative by definition.

The *complementary probability distribution function* (cPDF) of concentration X, denoted $\mathscr{F}(x)$, is a dimensionless function with the same domain as $F(x)$ that represents the probability that $X > x$. It is simply the complement of $F(x)$ and is given by

$$\mathscr{F}(x) = 1 - F(x). \tag{2}$$

In survival analysis, $\mathscr{F}(x)$ is called the survivor function (or survival function), but we find this term confusing when dealing with concentration data and therefore usually employ the generic term from probability theory.

The *probability density function* (pdf), denoted $f(x)$, is the derivative of the PDF with respect to its argument x. In symbols,

$$f(x) = F'(x) = -\mathscr{F}'(x), \tag{3}$$

where the prime ($'$) denotes differentiation with respect to x. It follows from Equation (3) that $F(x)$ can be expressed in terms of $f(x)$ as

$$F(x) = \int_0^x f(\xi)\mathrm{d}\xi. \tag{4}$$

Note that both $F(x)$ and $\mathscr{F}(x)$ are dimensionless, while $f(x)$ has the same dimensions as $1/x$.

The final function related to random variable X that is important to be familiar with is usually called the hazard function in survival analysis and reliability analysis and denoted by $h(x)$. However, we feel the term "hazard function" is confusing when dealing with concentration data, so we will propose an alternative. We have been unable to think of a similarly picturesque term for applications to concentration data, so we will propose an alternative that, like probability density function, probability distribution function, and complementary probability distribution function, does not have a strong connection to any particular application and therefore should at least not be confusing.

Before stating the alternative term we propose, it will be helpful to provide a generic definition of the function $h(x)$ that is called the hazard function in survival analysis. For probability distributions whose distribution function is differentiable, we may define $h(x)$ by

$$h(x) := \frac{f(x)}{\mathscr{F}(x)} = -\frac{\mathscr{F}'(x)}{\mathscr{F}(x)}. \tag{5}$$

Because $\mathscr{F}(x)$ is dimensionless, it is clear from Equation (5) that $h(x)$ is dimensional, with the same dimensions as $f(x)$.

It follows from Equation (5) that $\mathscr{F}$ obeys the differential equation,

$$\frac{\mathrm{d}\mathscr{F}}{\mathrm{d}x} = -h(x)\mathscr{F}(x), \quad x \geq 0. \tag{6}$$

The solution, subject to initial condition $\mathscr{F}(0) = 1$, is

$$\mathscr{F}(x) = \exp\left(-\int_0^x h(\xi)\mathrm{d}\xi\right). \tag{7}$$

This result and Equations (2)–(5) show that functions $F(x)$, $\mathscr{F}(x)$, $f(x)$, and $h(x)$ are mathematically equivalent, in the sense that knowledge of one of them implies the other three. However, each function has a different use in survival analysis, with cPDF or survivor function $\mathscr{F}(x)$ and hazard function $h(x)$ playing particularly important roles in the statistical theory.

Equation (7) shows that $h(x)$ determines the rate at which $\mathscr{F}(x)$ decreases with increasing x. In survival applications, this rate of decrease is the per-capita mortality rate, while in reliability analysis,

 doi:10.20944/preprints202603.1922.v1

it is the per-machine failure rate. In both of these time-to-event applications, $h(x)$ is often referred to by the less application-specific but still picturesque term, "hazard function", as noted above. But while picturesque terms like mortality rate, failure rate, and hazard function are useful as mnemonic devices and in stimulating intuition in survival and reliability applications, they become confusing in applications to other types of data, such as seed germination times (where seed survival and germination are biologically distinct phenomena) or chemical concentrations. Therefore, bearing in mind that $h(x)$ determines the rate at which $\mathscr{F}(x)$ decreases with increasing x, we suggest the term *attenuation function* (AF) as a reasonable application-independent term.

Examples illustrating the role of $h(x)$ in attenuating $\mathscr{F}(x)$ as x increases are shown in Figure 5. The figure shows plots of $h(x)$ (left panel) and $\mathscr{F}(x)$ (right panel) for the Weibull distribution. The AF and cPDF for this distribution are given by

$$h(x) = rs(rx)^{s-1}, \quad \mathscr{F}(x) = e^{-(rx)^s}, \quad x \geq 0, \tag{8}$$

where $s > 0$ is the shape parameter and $r > 0$ is the rate parameter. Three examples of each function are shown in Figure 5. All examples have shape parameter $s = 3$, but values of the rate parameter differ ($r = 1.0, 1.5, 2.0$). Note that as r increases, the $h(x)$ curves increase faster with increasing x, causing the $\mathscr{F}(x)$ curves to attenuate faster with increasing x.

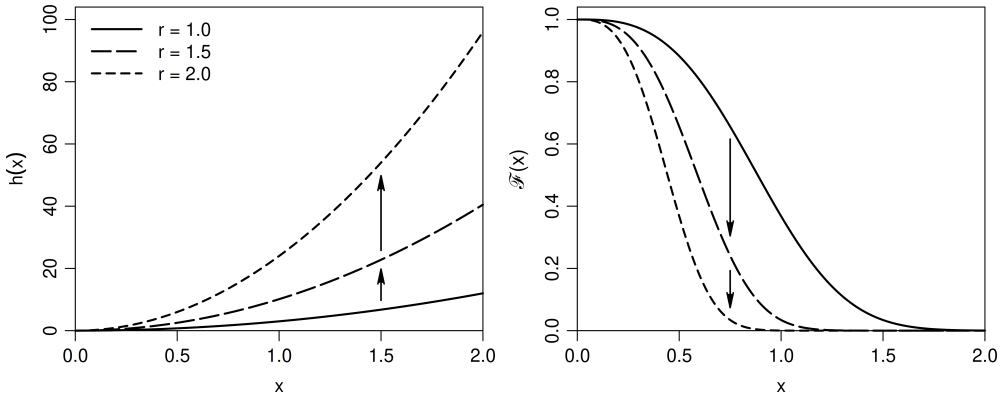

Figure 5. Three examples of AF $h(x)$ (left panel) and cPDF $\mathscr{F}(x)$ for the Weibull distribution given by Equation (8), all with shape parameter $s = 3$. The rate parameter has values $r = 1.0$, 1.5, and 2.0. Note that increasing the rate parameter (indicated by arrows) causes $h(x)$ to increase faster and $\mathscr{F}(x)$ to attenuate faster as x increases.

4.2. Censored Data

Survival analysis is unique among statistical disciplines in that it is specifically designed to handle censored data. In laboratory studies that produce time-to-event data (e.g., seed germination tests, bioassay tests with mortality as an endpoint), it is common to employ a fixed study duration that is long enough to ensure that most test subjects experience the event of interest (e.g., germination, mortality) but not long enough to ensure that all do. For test subjects that have not experienced the event by the end of the experiment, the actual event time T is unknown; all that is known is that T is greater than the termination time T_{end} of the experiment. Thus, there is a "veil" covering the time axis to the right of T_{end}, preventing us from seeing the portion $(T_{\text{end}}, T]$ of event time T that is beyond T_{end}. Such data are called *right-censored* data. Traditional statistical methods such as t-tests and analysis of variance — or even the usual maximum-likelihood estimator for the mean — cannot be applied to such data, because they require a valid measurement for every observation.

With concentration data, censoring occurs most commonly when a measured concentration is less than the LRL of the analytical method being employed. Many analytical methods for environmental samples also have a URL that may be low enough so that field concentrations sometimes exceed it. Examples include the Colilert-18® method for quantifying coliform bacteria, and analytical methods that employ a calibration curve based on a fixed range of standards. For such methods, a concentration

 doi:10.20944/preprints202603.1922.v1

measurement will be censored if it is greater than the URL. LRLs and URLs are set by statistical methods and ensure that all reported concentration estimates have acceptably low bias and imprecision.

An important difference between censored survival times and censored concentrations is that, whereas no estimate of the true survival time is available for a censored event time, an estimate with unacceptably high uncertainty will be available for a censored concentration value if all concentration estimates, regardless of their quality, are reported. Thus, in statistical analyses of censored concentration data, a temptation exists to employ data of unacceptably high bias and imprecision that does not exist for time-to-event data. Two fundamental principles of rigorous statistical methods for analyzing censored data are to avoid fabricating data and to use only information from available data that is known with high confidence, even if additional information is available that is known to be unreliable.

When a concentration estimate x is less than the LRL, we know that its numerical value is contaminated with too much measurement error and bias to justify treating the value as reliable and using it in statistical analyses; all we can be confident of is that $x <$ LRL. In this case, the concentration is said to be *left-censored*. Similarly, if a concentration estimate x exceeds the URL, all we can be confidenbt of is that $x >$ URL and the concentration is said to be *right-censored*. Data sets that include a mix of valid, left-censored, and right-censored concentrations are said to be *doubly-censored*, though each censored observation is either left- or right-censored.

Another type of censoring that is important to know about is *interval censoring*. A measured concentration is (strictly) interval-censored if we know only that it lies between two values L and R, with $L < R$. In reality, all concentration data are either left-, right-, or interval-censored, because concentrations can be measured to only a relatively small number of significant digits. But this complication is universally ignored in statistical analyses, and we know of no commonly-used analytical method that produces concentration estimates that would be treated as strictly interval-censored in practice if their values were between the LRL and URL. Nevertheless, there are two reasons why it is important to be aware of the term "interval-censored". First, in the typical case where the LOD for an analytical method is less than the LLOQ, some authors (e.g., [2]) suggest treating concentrations between the LOD and the LLOQ as interval-censored, and only values below the LOD as left-censored (we disagree with this approach, because all concentration estimates below the LRL are too uncertain to treat as valid numerical measurements). The second and more important reason is that this term is commonly used in statistical software and the literature in a broad sense that includes valid, left-censored, right-censored, and strictly interval-censored data as special cases. Thus, statistical methods for interval-censored data typically are intended for application to data sets that are *allowed but not required* to include strictly interval-censored values, as well as valid and possibly left- and/or right-censored values.

In recording data from an empirical study where some of the concentrations are censored, it is very important to use a format for the data that is compatible with statistical software that will be used to analyze the data and is also simple and intuitive for laboratory personnel to use when entering the data. The vast majority of time-to-event data sets include a mix of valid (usually called "exact" in the statistical literature) and right-censored data, and as a result, the proper way to record such data is highly standardized. By contrast, most environmental data sets consisting of measured concentrations include a mix of valid and left-censored data, and many include right-censored data, as well. Formats for recording such data are not fully standardized in the statistical literature or software, so it is important to find out what format is compatible with the particular software one intends to use.

Based on our own experience, it is also important to consider that students and laboratory technicians who often produce and record the data find certain formats counter-intuitive and difficult to work with. A basic law of data entry is that the more logical and intuitive the format for entering data is, the less frequent recording errors will be. Data recorded in any of the common formats can be converted to any of the others by a few lines of computer code, so there is no need to sacrifice convenience and accuracy in recording data to achieve the data format required by a particular software

 doi:10.20944/preprints202603.1922.v1

program. As long as all required information is recorded, one can use different formats for data entry and data analysis.

Two rather different methods for characterizing doubly-censored data are encountered in statistical software, both of which require that two values be recorded. One method characterizes each observation using a concentration (or time) and a status code, while the other characterizes each observation using the two endpoints of a concentration (or time) interval. We will call these the value-plus-code and interval-endpoints approaches. With the value-plus-code approach (which R calls the `interval` format), the recorded concentration is either a valid concentration, the LRL (implying the concentration is left-censored at that value), or the URL (implying the concentration is right-censored at that value), and the status code tells us which. For example, with the value-plus-code data format in R, status codes 0, 1, and 2 mean that the recorded value is right-censored, valid, and left-censored. With the interval-endpoints approach (which R calls the `interval2` format, and which is the default format in SAS), each observation is viewed as an interval specified by its left and right endpoints. If the two endpoints are the same, the value of the shared endpoints is a valid concentration. If the left endpoint is missing (indicated by logical constant `NA` in R and by a blank field in SAS) and the right endpoint is numeric, then the concentration is left-censored and the right endpoint is the LRL. If the right endpoint is missing and the left endpoint is numeric, then the concentration is right-censored and the left endpoint is the URL. Figure 6 illustrates some alternative data coding schemes for the value-plus-code and interval-endpoints approaches to characterizing doubly-censored data. As this figure suggests, data sets that are composites of data from different labs, analysis dates, technicians, and so on typically have a variety of different LRL and URL values, so these limits must be recorded for each sampling unit analyzed.

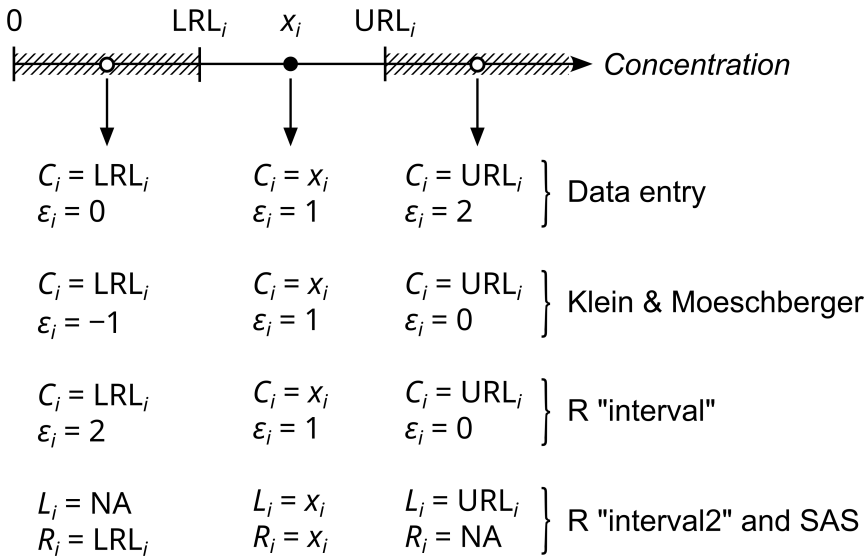

Figure 6. Four alternative coding schemes for doubly-censored data. The horizontal line at the top represents the domain of all possible concentrations; concentrations x_i lying between corresponding limits LRL$_i$ and URL$_i$ are measurable (filled black circle), while concentrations lying below the LRL$_i$ or above the URL$_i$ are not (open circles in shaded portions of the concentration domain). C_i is a recorded concentration, ε_i is a status code, and L_i and R_i are the recorded left and right endpoints of a data interval. "Data entry" is a coding scheme we have found convenient for laboratory technicians to use when entering data (the status codes are intuitive because their order follows the pattern of left-censored on the left, valid in the middle, and right-censored on the right along the concentration axis), which are then converted via computer to the scheme required by a particular statistical function or procedure. "Klein & Moeschberger" is a coding scheme mentioned by Klein and Moeschberger ([32], p. 71). The `interval` and `interval2` schemes are used by various R survival-analysis functions; the `interval2` scheme is also used by SAS but with missing values represented by blank fields instead of by logical constant `NA`.

 doi:10.20944/preprints202603.1922.v1

The same two approaches can be used if strictly interval-censored data are present, but with a slight extension for the value-plus-code approach. Strictly interval-censored data require two concentration values (instead of only one) plus a status code when the value-plus-code approach is used. In R's `interval` format, the first concentration value must be numeric and is used in statistical analysis regardless of the value of the status code; the second concentration value is ignored unless the status code is 3 (which means the observation is strictly interval censored and lies between the first and second recorded concentrations), in which case it must be greater than the first concentration. With the interval-endpoints format, strictly interval-censored data are implied whenever both endpoints are numeric and the first is less than the second.

As a very simple illustration, the first several lines in a spreadsheet containing data from two stream sampling sites might be set up as shown in Figure 7. The left and middle examples show the two common ways to format censored data for R, while the right table shows standard SAS formatting. The `site` column in each example shows two hypothetical sampling sites (`S1` and `S2`), which we. With R's `interval` formatting (left example), each entry in the `conc` column contains a numerical concentration (in mg L^{-1}, say), which represents either a valid concentration, the LRL for a left-censored concentration, or the URL for a right-censored concentration, as determined by the corresponding integer code in the `status` column (0: URL, 1: valid concentration, 2: LRL). With R's `interval2` formatting (middle example), each entry in the `left` and `right` columns contains either a concentration or R's logical constant `NA` ("not available") for missing data, both of which are treated as numerical values by R and therefore can be included in the same vector or the same column of a data frame. SAS formatting is essentially the same as R's `interval2` format, except that spreadsheet cells for missing data are simply left blank. Note that these concentration data consist of one left-censored value (with LRL = 0.03 mg L^{-1}), one right-censored value (with URL = 2.50 mg L^{-1}), and four valid concentrations. Also shown in these examples is a column labeled `turb`, which represents turbidity (in nephelometric turbidity units, or NTU). This is an example of a quantitative explanatory variable or covariate, which are often used in semiparametric regression models. If there were additional explanatory variables, their values for different sample units would be included in additional columns, with one column for each variable.

R "interval"

site	conc	status	turb
S1	0.12	1	11.2
S1	0.03	2	10.5
S1	0.15	1	12.3
S2	1.33	1	31.7
S2	2.50	0	29.8
S2	1.75	1	32.4

R "interval2"

site	left	right	turb
S1	0.12	0.12	11.2
S1	NA	0.03	10.5
S1	0.15	0.15	12.3
S2	1.33	1.33	31.7
S2	2.50	NA	29.8
S2	1.75	1.75	32.4

SAS

site	left	right	turb
S1	0.12	0.12	11.2
S1		0.03	10.5
S1	0.15	0.15	12.3
S2	1.33	1.33	31.7
S2	2.50		29.8
S2	1.75	1.75	32.4

Figure 7. Examples showing three common ways to format data in a spreadsheet for export as a CSV file to be imported into R (left and middle examples) or SAS (right example). Each example has a header row containing names of the fields in the data records. *Left:* R's `interval` format. Fields are `site` (sampling site), `conc` (concentration), `status` (status code), and `turb` (turbidity, an explanatory variable). Status codes are 0: reported `conc` is the URL for a right-censored concentration, 1: reported `conc` is a valid concentration, and 2: reported `conc` is the LRL for a left-censored concentration. *Middle:* R's `interval2` format. Fields are `site`, `left` (left endpoint of the concentration interval), `right` (right endpoint of the concentration interval), and `turb`. `NA` is a logical constant indicating that the value is missing. *Right:* SAS format. The same format as in the middle example, except that endpoint cells with a missing value are blank instead of containing `NA`.

An important censoring-related distinction in all types of survival analysis is whether the limits of quantification are fixed or are adjusted to ensure that a predefined proportion of the samples yield valid concentration measurements (or a predefined proportion of test subjects experience the event of interest). For concentration data, it usually is appropriate to assume fixed censoring (known as *Type I censoring* in survival analysis), and we will do so throughout this paper. In studies that produce

 doi:10.20944/preprints202603.1922.v1

time-to-event data, it is sometimes desirable to run an experiment until a fixed number or proportion of test subjects experience the event of interest instead of running it for a fixed total duration. Because the censoring time is random rather than fixed for this type of censoring (known in survival analysis as *Type II censoring*), somewhat different statistical methods are sometimes required for their analysis.

Another important distinction we must mention is between *informative* and *non-informative* censoring. A full explanation of this distinction requires a technical detour that is inappropriate for this paper (we refer the interested reader to the book by Kalbfleisch and Prentice ([33], section 3.2)), but the basic idea is rather simple: For left-censored concentration data, the numerical value of the LRL or URL must not be correlated with the median or other location measure of the probability distribution from which concentrations in the data set were sampled. This means that the numerical value of the LRL or URL must provide no information about whether the true median concentration at a sampling site is high or low, or whether the true median concentration at one site is higher or lower than at another site. This condition would be violated if, for example, sampling sites with unusually high levels of a particular contaminant also had unusually high levels of a second contaminant that interfered with analysis of the first in a way that increased the LRL. In this paper, we will restrict attention to non-informative censoring, both because this is the most common type in applications and because most statistical methods of survival analysis currently available in software packages require it.

5. Nonparametric Survival-Analysis Methods

In this section, we review modern nonparametric methods of survival analysis that can accommodate the full range of censoring types likely to be encountered in concentration data from studies of aquatic systems and that are currently available in standard statistical software. But before we begin, it will be prudent to address a common misconception regarding nonparametric statistical methods in general.

In our experience, scientists and engineers often assume nonparametric statistical methods are inherently much less powerful than parametric methods and therefore should be avoided unless there is no defensible alternative. By contrast, statisticians who design and analyze medical research studies that produce censored time-to-event data almost always use distribution-free nonparametric or semiparametric methods, and the same is true of statisticians working on various other types of applications. For example, Conover ([34], p. 2) states: "Nonparametric methods have become essential tools in the workshop of the applied scientist who needs to do statistical analyses. When the price for making a wrong decision is high, applied scientists are very concerned that the statistical methods they are using are not based on assumptions that appear to be invalid, or are impossible to verify." Hollander et al. ([27], p. *xiii*) go further and state that "the nonparametric approach is the preferred methodology for statisticians" — except, of course, for certain types of designed experiments that produce sufficiently large quantities of uncensored data with statistician-friendly properties, so the simple distributional assumptions of standard parametric methods are appropriate and can be convincingly verified.

Hollander et al. ([27], pp. 1–2) provide a list of 10 advantages of nonparametric statistical methods, including the following (which we paraphrase) that are relevant to survival-analysis methods that can be applied to concentration data:

- Nonparametric methods are much less sensitive to outliers than are parametric methods.
- Nonparametric methods make fewer assumptions that must be verified regarding properties of the underlying populations from which the data are obtained. In particular, no specific probability distribution is assumed for concentration data. These methods are therefore applicable in many cases where parametric methods are invalid because their distributional assumptions are clearly untenable, or where these assumptions cannot be convincingly assessed.
- In cases where the probability distribution assumed by a parametric method is consistent with properties of the sampled population (regardless of whether this consistency can be convincingly

 doi:10.20944/preprints202603.1922.v1

demonstrated), nonparametric methods often are nearly as efficient as parametric methods. This means that they require only slightly larger sample sizes to achieve the same statistical power.

Regarding the last two points, Lehmann ([35], p. *viii*) states: "The feature of nonparametric methods mainly responsible for their great popularity (and to which they owe their name) is the weak set of assumptions required for their validity. Although it was believed at first that a heavy price in loss of efficiency would have to be paid for this robustness, it turned out, rather surprisingly, that the efficiency of the Wilcoxon tests and other nonparametric procedures holds up quite well under the classical assumption of normality and that these procedures may have considerable advantages in efficiency (as well as validity) when the assumption of normality is not satisfied."

Having hopefully dispelled the notion that nonparametric statistical methods are inherently much less powerful than parametric methods, we turn now to our review of nonparametric survival-analysis methods that can be applied to concentration data that may include left-censored, right-censored, and interval-censored values. Three main types of problems can be addressed with these methods:

1. characterizing the distribution of concentrations for an individual study site or date,
2. pairwise comparison of concentration distributions for different sites or dates, and
3. testing homogeneity and detecting monotonic trends in concentration distributions for three or more sites or dates.

We discuss these three problems in turn in this section.

As background, we note that previous expositions of survival analysis methods for censored concentration data that we are aware of focus mainly on data where the only type of censoring is left-censoring. They then rely on the device of "flipping" concentrations (i.e., reversing the concentration scale) to transform left-censored data to right-censored data (e.g., [2]). This kluge was introduced by Ware and Demets in 1976 [36] and makes it possible to employ the older methods of survival analysis that assume right-censoring is the only type of censoring present. Flipping, however, cannot resolve the problem posed by doubly-censored and interval-censored data, and it is conceptually awkward (though statistically valid) when only left-censoring is present.

Both R and SAS now have functions and procedures that permit most of the key nonparametric methods of survival analysis to be applied to data sets containing any mixture of valid, left-censored, right-censored, and interval-censored data on the original concentration scale. We will restrict attention to these newer methods, thereby avoiding the need to discuss data flipping and the "un-flipping" that is required in order to present and interpret results intelligibly.

In view of our emphasis on the latest survival-analysis methods, we strongly recommend that only software developed by research statisticians who specialize in survival analysis and that is intended for application to time-to-event data in medical research should be used for analyzing concentration data. The main reason is that the statistical discipline of survival analysis is both highly nuanced and rapidly evolving, and all of the leading-edge software is being created specifically for applications in medical research (usually in the form of R packages). By learning to use this state-of-the-art software, researchers and practioners interested in applications involving concentration data will know how to record their data in a way that is consistent with accepted practice in the statistical discipline of survival analysis (including proper coding of censored and uncensored observations) and will always be able to access the latest and widest range of expertly developed survival-analysis methods as new methods are implemented in the software.

5.1. Characterizing Concentration Distributions

Nonparametric survival-analysis methods yield quantitative estimates of the PDF, cPDF, and corresponding pointwise confidence intervals for an individual site or date without requiring one to assume a specific probability distribution for the analyte concentration. They also yield estimates of the median and other quantiles and their 95% (or other) confidence intervals. These methods can be viewed as extensions of the simple Kaplan-Meier estimator of the cPDF and are the only type of

 doi:10.20944/preprints202603.1922.v1

survival-analysis method that both accommodates all major types of censoring and has been available in standard statistical software for more than 10 years.

Regarding the median and other quantiles, we note that the mean and standard deviation are almost never estimated in medical research studies involving censored time-to-event data. There are two reasons for this, both of which apply to environmental concentration data. The first is that the classical estimators of the mean and standard deviation employed in parametric statistics cannot be employed when censored values are present (unless fabricated values are substituted for censored data, which we discourage), because they require valid numerical estimates of all observed concentrations. The second reason is that the data typically come from distributions with pronounced positive skew, and the median rather than the mean is the preferred measure of location for such distributions (e.g., [37], p. 30). Alternative measures of dispersion that are more appropriate than the standard deviation for censored data from positively skewed distributions include the interquartile range (the difference between the third and first quartiles, or 75-th and 25-th percentiles) and the semi-interquartile range (half the interquartile range) (e.g., [37], p. 31). Thus, unless one is conducting an applied study in which statistical methods are rigidly imposed by a governmental regulatory agency, one should characterize the location and dispersion of the distribution underlying a set of censored concentration data for a single sampling site or date by using the median (and 95% confidence interval) and interquartile or semi-interquartile range.

Nonparametric estimates of the cPDF for doubly-censored data are usually obtained using an iterative procedure due to Turnbull [38,39] or various more-recent extensions, with data represented in the `interval2` format (or its SAS equivalent) discussed above. Each observation is recorded in one of the three forms $(\texttt{NA}, x_i)$, $[x_i, x_i]$, or $(x_i, \texttt{NA})$, representing left-censored, valid, or right-censored observations, respectively. For valid observations, x_i is a measured concentration, for left-censored observations, $x_i = \text{LRL}_i$, and for right-censored observations, $x_i = \text{URL}_i$. Thus, each observation i includes a single distinct numerical value x_i that specifies one or both interval endpoints, according as the observation is or is not censored.

The x_i values for different observations typically include repeats or ties, meaning that two or more observations share the same value of x_i (usually an LRL or URL). Let n be the total number of observations, and suppose there are $m \leq n$ distinct values among the x_i. Let $x_{[j]}$ denote these distinct values for $j = 1, 2, \ldots, m$, ordered so that $x_{[1]} < x_{[2]} < \cdots < x_{[m]}$. Finally, let $V_{[j]}$ denote the number of valid concentrations with $x_i = x_{[j]}$, $L_{[j]}$ the number of left-censored concentrations with $x_i = \text{LRL}_i = x_{[j]}$, and $R_{[j]}$ the number of right-censored concentrations with $x_i = \text{URL}_i = x_{[j]}$. With this notation, the data may be represented by the set of m 4-tuples $(x_{[j]}, V_{[j]}, L_{[j]}, R_{[j]})$ for $j = 1, 2, 3, \ldots, m$. The kernel of the likelihood function, given the data, is then

$$\prod_{j=1}^{m} f(x_{[j]})^{V_{[j]}} F(x_{[j]})^{L_{[j]}} \mathscr{F}(x_{[j]})^{R_{[j]}}, \tag{9}$$

where $f(\cdot)$ is the pdf, $F(\cdot)$ is the PDF, and $\mathscr{F}(\cdot)$ is the cPDF of the unknown probability distribution from which the x_i are assumed to be sampled.

In parametric survival analysis, functional forms would be specified for the pdf, PDF, and cPDF, and their parameter values would be chosen to maximize the likelihood kernel. Nonparametric approaches avoid choosing any specific parametric probability distribution. As mentioned above in Section 3.5, the main reason for avoiding parametric distributions is that results for censored data often depend strongly on which distribution is chosen, while evidence supporting any particular choice is often inconclusive [30]. The goal then shifts to estimating the numerical values of the PDF or cPDF at distinct concentrations $x_{[j]}$, $j = 1, 2, 3, \ldots, m$, with the value of the function remaining unchanged between these points, as in the classic Kaplan-Meier survivor function for right-censored data.

Unlike the simple case of right-censoring, there is no explicit formula for the nonparametric maximum likelihood estimator of the cPDF when both left- and right-censoring are allowed. However, Turnbull [38,39] proposed an iterative algorithm for obtaining numerical estimates of the values of the

 doi:10.20944/preprints202603.1922.v1

cPDF at the distinct values $x_{[j]}$ (an example of what would later be called the expectation-maximization or EM algorithm), yielding a decreasing step function similar to the Kaplan-Meier survival curve. Turnbull's method and various modifications of it are employed in several statistical functions now available in R and SAS, making it simple for users to estimate the cPDF and PDF, confidence intervals, and quantiles for doubly-censored data sets (technical details underlying Turnbull's method are beyond the scope of this review; see Turnbull's papers [38,39]).

5.1.1. R Example

In R, the cPDF, PDF, and their pointwise confidence intervals for doubly-censored data can be estimated using a combination of the `Surv()` and `survfit()` functions from the `survival` package in essentially the same way these functions and intervals are estimated for traditional right-censored data encountered in time-to-event analysis. Quantiles can be estimated from a `survfit` object using the `quantile()` function from the `survival` package.

To illustrate the use of R software for statistical analysis of censored concentration data, all examples in the present paper utilize a set of 3,205 estimates of the concentration of the fecal coliform bacterium, *Escherichia coli* (*E. coli*), at freshwater bathing beaches across the state of Michigan, USA, in 2019 and 2020. The data we use here were produced for county health departments by multiple laboratories throughout the state as part of Michigan's annual beach monitoring program, using the Colilert-18® enumeration method. Details on sampling sites, dates, and methods, are presented by McNair et al. [1]. The LRL and URL for the data are 1 and 2420 MPN/100 mL (MPN: Most Probably Number) for all samples. Roughly 3% of the data are left-censored and 0.3% are right-censored, with the percentages varying markedly among counties. Of particular importance for the examples presented in the present paper is the fact that every beach was assigned to one of three classes (inland-lake, river, or coastal beach), according to whether the beach was located on an inland lake, a river, or one of the Laurentian Great Lakes. A question of interest is whether *E. coli* concentrations tend to differ among these three types of beach.

Figure 8 shows estimated cPDF (left panel) and PDF (right panel) for coastal beaches, with pointwise 95% confidence intervals. The quartiles correspond to concentrations at which the PDF crosses the dashed horizontal lines at 0.25, 0.50, and 0.75 on the vertical axis. Table 2 shows the estimated quartiles and 95% confidence intervals produced by the `quantile()` function. Confidence intervals for the cPDF, PDF, and quartiles reflect the value of the `conf.type` argument of `survfit()` (actually `survfit.formula()`), which was set to `"log-log"` for consistency with the SAS default.

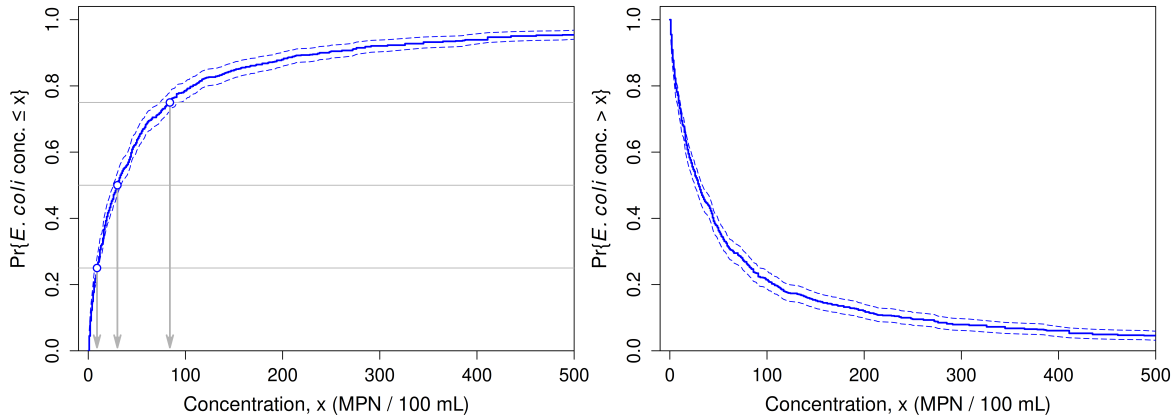

Figure 8. Estimated PDF (left, solid blue curve) and cPDF (right, solid blue curve) and their pointwise 95% confidence limits (dashed blue curves) for *E.coli* concentrations (MPN/100 mL) at coastal beaches in Michigan. In the left panel, horizontal gray lines are drawn at probabilities 0.25, 0.50, and 0.75; the gray arrows indicate the corresponding quantiles (equivalently, the 25th, 50th, and 75th percentiles). Computations were performed with R functions `Surv()`, `survfit()`, and `quantile()`.

 doi:10.20944/preprints202603.1922.v1

Table 2. Estimated quantiles for probabilities 0.25, 0.50, and 0.75 and their 95% confidence limits for *E. coli* concentrations (MPN/100 mL) at coastal beaches in Michigan. The corresponding cPDFs and PDFs are plotted in Figure 8. LCL and UCL denote the estimated lower and upper 95% confidence limits for the quantiles. Quantiles and confidence limits represent *E. coli* concentrations and are in units of MPN/100 mL. Computations were performed with R functions `Surv()`, `survfit()`, and `quantile()`.

Probability	Quantile	LCL	UCL
0.25	9	8	11
0.50	30	26	34
0.75	84	75	93

5.1.2. SAS Example

In SAS, the cPDF, PDF, their pointwise confidence intervals, quantiles, and confidence intervals for the quantiles all can be estimated with the ICLIFETEST procedure. An example demonstrating the estimation of these quantities is included in the online Supplementary Materials.

5.2. Pairwise Comparison of Concentration Distributions

Nonparametric methods of survival analysis include methods that can be used to test hypotheses about potential differences in concentration distributions for different sampling sites or dates when the data contain a mix of valid concentrations and one or more types of censored data. The main tests of interest are tests for pairwise differences between sites, tests of homogeneity for $k \geqslant 3$ sites, and tests for monotonic trends across $k \geqslant 3$ sites. We discuss tests for pairwise differences in the present section; homogeneity tests and tests for monotonic trends are discussed in Section 5.3.

The null and alternative hypotheses for pairwise tests between sites are most naturally stated in terms of the cPDFs rather than the PDFs, since the ordering of sites by cPDF is the same as their ordering by the median when the cPDFs do not cross decisively (see below). The alternative hypothesis can be two-sided or one-sided. Following Oller and Langohr [40], the null and alternative hypotheses (H_0 and H_1) for pairwise tests between two sites $i \neq j$ can be stated as follows:

H_0: $\mathcal{F}_i(x) = \mathcal{F}_j(x)$ for all x
H_1: $\mathcal{F}_i(x) \neq \mathcal{F}_j(x)$ for some x (two-sided)
H_1: $\mathcal{F}_i(x) \geq \mathcal{F}_j(x)$ for all x, with "$>$" for some x (one-sided, "site i greater than site j")
H_1: $\mathcal{F}_i(x) \leq \mathcal{F}_j(x)$ for all x, with "$<$" for some x (one-sided, "site i less than site j").

As usual, the appropriate one-sided alternative must be chosen before examining the data and should be based on objective and rational considerations, such as the location of a known contaminant source relative to different sampling sites.

Both R and SAS provide functions or procedures for performing pairwise tests, based on extensions of the Fleming-Harrington class of tests appropriate for doubly-censored and interval-censored data. This type of test employs two parameters, traditionally denoted ρ and λ, that allow one to weight low, moderate, and high concentrations uniformly or differentially in comparing sites. However, unless one has a compelling *a priori* reason for using differential weighting, we recommend following accepted practice in survival analysis of time-to-event data and employing equal weighting ($\rho = \lambda = 0$), in which case the test is often referred to as a log-rank test.

All of the tests for comparing concentration distributions for discrete groups (e.g., different sampling sites or dates) utilize tests from the Fleming-Harrington family of tests, which are "geared to detect alternative hypotheses where the hazards [AFs] between groups differ but do not cross" ([40], p. 3). To the best of our knowledge, available software packages for handling doubly-censored and interval-censored data do not produce estimates of AFs, so the "no crossing" condition for AFs cannot be directly assessed. A practical indirect way to determine whether there is strong evidence that two AFs, $h_i(x)$ and $h_j(x)$ for distinct groups i and j, cross decisively is to overlay plots of the Turnbull estimates of the PDFs or cPDFs for the two groups and visually assess whether they cross decisively. If so, there is strong evidence that the AFs cross decisively; otherwise, not.

 doi:10.20944/preprints202603.1922.v1

The rationale for this simple visual diagnostic is as follows. If AFs $h_i(x)$ and $h_j(x)$, $i \neq j$, do not cross, then either $h_i(x) \geq h_j(x)$ for all x or $h_i(x) \leq h_j(x)$ for all x. Let $\mathscr{F}_i(x)$ denote the cPDF for any group i. A basic property of continuous probability distributions with support $0 \leq x < \infty$ is that the survival and hazard functions are related by

$$\mathscr{F}_i(x) = \exp\left(-\int_0^x h_i(\xi)d\xi\right) \tag{10}$$

for all x. If $h_i(x) \leq h_j(x)$ for all x, then $\int_0^x h_i(\xi)d\xi \leq \int_0^x h_j(\xi)d\xi$ and Equation (10) implies $\mathscr{F}_i(x) \geq \mathscr{F}_j(x)$ for all x. Similarly, if $h_i(x) \geq h_j(x)$ for all x, then $\mathscr{F}_i(x) \leq \mathscr{F}_j(x)$ for all x. This argument shows that if two hazard functions do not cross, then neither do the corresponding survival functions. By contraposition, if two survival functions *do* cross, then the corresponding hazard functions also must cross, implying that Fleming-Harrington tests are not appropriate. Thus, a simple diagnostic for determining whether Fleming-Harrington tests are appropriate for two groups is to plot the corresponding survival functions and visually determine whether they cross decisively (we say "decisively" because the survival functions being plotted are only estimates, and we want convincing evidence that the unknown true survival functions cross). If they do, we can be confident that the hazard functions also cross and therefore that Fleming-Harrington tests are not appropriate. On the other hand, if the survival functions do not cross, then the evidence is consistent with the hypothesis that the hazard functions do not cross decisively and hence that Fleming-Harrington tests are appropriate.

We note that the same problem arises with Fleming-Harrington tests for right-censored time-to-event data, where the Kaplan-Meier (instead of Turnbull) estimator of the cPDF is imployed. For example, Hosmer et al. [41] state:

> A problem can occur if the estimated survival functions cross one another. This means that, in some time intervals, one group will have a more favorable survival experience, while in other time intervals, the other group will have the more favorable experience. This situation is analogous to having interaction present when applying Mantel-Haenszel methods to a stratified contingency table... Fleming, Harrington, and O'Sullivan [42] proposed a method that addresses the problem by using, as a test statistic, the maximum observed difference between the two survival functions. This test has not been implemented in any software package... For the time being, our only check is via a visual examination of the plot of the Kaplan-Meier estimator for the groups being compared. If we see that the curves cross, then this "interaction" may be present.

Thus, Hosmer et al. [41] recommend the same visual diagnostic that we outlined above: overlay plots of the two estimated cPDFs (or PDFs) and visually determine whether they decisively cross. Kaplan-Meier estimates of the cPDFs are used with right-censored data, and Turnbull estimates with left-censored, doubly-censored, or interval-censored data.

5.2.1. R Example

R's standard `survival` package includes functions that allow one to perform the types of hypothesis tests discussed above for data sets where all censored values are right-censored. Data sets where all censored values are left-censored can also be handled if the data are flipped to reverse the concentration scale. However, data sets that include both left-censored and right-censored (or interval-censored) values cannot be handled by these functions, as flipping is useless for such data.

These limitations of the `survival` package are not a serious concern, because R add-on packages are available that provide an extensive set of statistical tools for dealing with data that include any mix of valid, left-censored, right-censored, and interval-censored data. In our opinion, the most useful of these packages for the types of hypothesis tests addressed in this section is currently the `FHtest` package [40], which includes functions that permit comparison of two or more cPDFs using an extended version of the Fleming-Harrington class of tests. The two key functions for our present

 doi:10.20944/preprints202603.1922.v1

purposes are `FHtesticp()` and `FHtestics()`, which implement the extended Fleming-Harrington test using a permutation distribution and a score-vector distribution, respectively. We are not aware of any guidance on minimum sample sizes, so we suggest employing the `exact=TRUE` option when function `FHtesticp()` is used unless the computational method fails because the sample sizes are too large.

An example of PDFs and cPDFs for *E. coli* concentrations (MPN/100 mL) at inland-lake and coastal beaches in Michigan is shown in Figure 9. Note that the 25th, 50th, and 75th percentiles for inland-lake beaches are all less than the corresponding percentiles for coastal beaches (left panel), and that the cPDF $\mathscr{F}_{\mathrm{I}}(x)$ for inland-lake beaches is less than the cPDF $\mathscr{F}_{\mathrm{C}}(x)$ for coastal beaches for all concentrations $x > 0$ (right panel). In this sense, then, the ordering of concentration distributions by quartiles is the same as the ordering by cPDF. This correspondence between orderings by quartiles and by cPDFs makes cPDFs a more natural choice than PDFs for stating null and alternative hypotheses regarding pairwise differences and monotonic trends in concentration distributions for different sites ordates.

As just noted, visual inspection of Figure 9 reveals that the estimated quartiles for inland-lake beaches are consistently less than those for coastal beaches. But is there statistically sound evidence that the apparent difference between the two concentration distributions is real? First, we are not aware of any plausible connection between processes that might be responsible for producing unusually high or low true concentrations in the field and processes in the laboratory that are responsible for determining the censoring levels (LRL and URL), so it is reasonable to assume that censoring is non-informative. Next, we note that the cPDF for inland-lake beaches is less than that for coastal beaches at all *E. coli* concentrations, so the two PDFs certainly do not cross decisively. In view of these two properties of the data, we used the `FHtesticp()` function with parameters $\rho = 0$ and $\lambda = 0$ to test the null hypothesis of no difference against the two-sided alternative hypothesis, R returns a p-value of 1.5×10^{-6}, which provides strong evidence that the null hypothesis is false and hence that the two distributions are different. If we had a valid *a priori* reason to be interested only in the one-sided alternative with inland-lake beaches less than coastal beaches, the p-value would decrease to 7.5×10^{-7}, which again provides strong evidence that the null hypothesis is false but now supports the alternative that *E. coli* concentrations tend to be lower at inland-lake beaches that at coastal beaches.

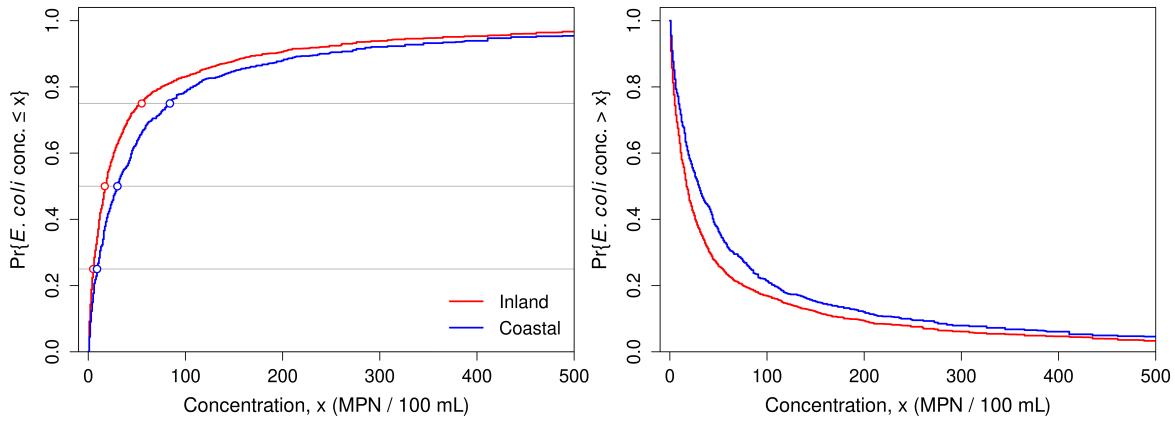

Figure 9. Estimated PDFs (left) and cPDFs (right) for inland-lake and coastal beaches (red and blue lines) at Michigan beaches. In the left panel, horizontal gray lines indicate probabilities 0.25, 0.50, and 0.75; abscissas of the intersections of these lines with the PDFs (indicated by dots) are the corresponding quantiles of the *E. coli* distributions for inland-lake and coastal beaches.

5.2.2. SAS Example

Pairwise tests comparing cPDFs can be performed in SAS with the `ICLIFETEST` procedure. An example is presented in the online Supplementary Materials.

 doi:10.20944/preprints202603.1922.v1

5.3. Tests of Homogeneity and Monotonic Trends in Multiple Concentration Distributions

The purpose of homogeneity tests is to determine if the inevitable numerical differences between observed concentration distributions for $k \geqslant 3$ sites (or dates) can reasonably be attributed merely to chance variation. If not, there is strong evidence that at least two of the distributions exhibit a real difference. Following Oller and Langohr [40], the null and alternative hypotheses for k-site tests of homogeneity can be stated in terms of the cPDFs as:

H_0: $\mathscr{F}_1(x) = \mathscr{F}_2(x) = \cdots = \mathscr{F}_k(x)$ for all x

H_1: $\mathscr{F}_i(x) \neq \mathscr{F}_j(x)$ for some concentration x and pair of sites $i, j \neq i$.

Note that the alternative hypothesis is necessarily two-sided. If the null hypothesis is rejected, pairwise comparisons, each with a two-sided alternative hypothesis, can be run (as described in Section 5.2) for all $k(k-1)/2$ distinct pairs of data groups (sites or dates) to determine which pairs show strong evidence of a difference after adjusting p-values to account for the number of comparisons made (typically using Holm's method).

The purpose of tests for monotonic trends is to determine if there is strong evidence that the concentration distributions for $k \geqslant 3$ sites exhibit a specified monotonic ordering. Again following Oller and Langohr [40], the null and alternative hypotheses can be stated in terms of the cPDFs as:

H_0: $\mathscr{F}_1(x) = \mathscr{F}_2(x) = \cdots = \mathscr{F}_k(x)$ for all x

H_1: $\mathscr{F}_1(x) \geq \mathscr{F}_2(x) \geq \cdots \geq \mathscr{F}_k(x)$ for all x, with "$>$" for some x and pair $i < j$ ("decreasing trend")

H_1: $\mathscr{F}_1(x) \leq \mathscr{F}_2(x) \leq \cdots \leq \mathscr{F}_k(x)$ for all x, with "$<$" for some x and pair $i < j$ ("increasing trend").

The alternative hypothesis is necessarily one-sided but can be either increasing or decreasing. As usual with one-sided tests, the ordering of sites in the putative trend must be chosen before examining the data and should be based on objective and rational considerations.

As an aid to interpreting strictly monotonic trends in cPFDs, Figure 10 shows a hypothetical example with pdfs (top panel) and corresponding cPDFs (bottom panel) for concentrations at three sampling sites. True distributions are shown rather than empirical estimates, so there is no censoring. To further simplify the example, the pdfs have the same shape and are symmetric, so we may use means to characterize their locations. The means increase from site 1 to site 2 to site 3 and hence form an increasing trend; that is, $\mu_1 < \mu_2 < \mu_3$. The corresponding cPDFs have the property that $\mathscr{F}_1(x) < \mathscr{F}_2(x) < \mathscr{F}_3(x)$ for all $x > 0$.

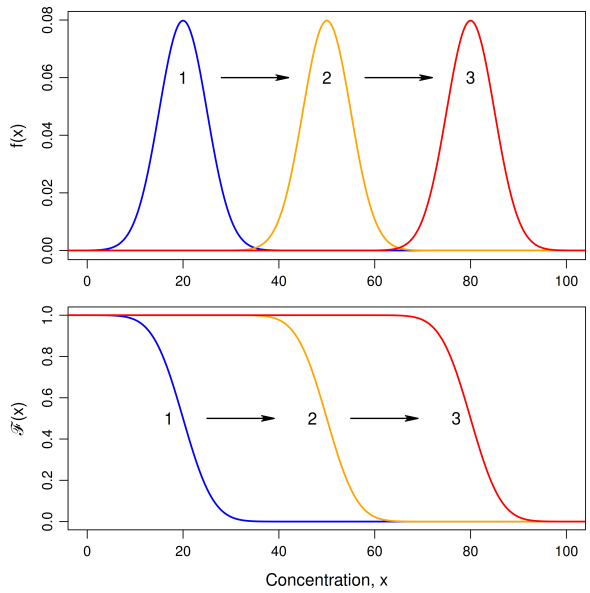

Figure 10. Hypothetical example where the pdfs $f_i(x)$ for concentrations at three sampling sites exhibit an increasing trend from sites 1 to 2 to 3 (top panel), so the site means have the property that $\mu_1 < \mu_2 < \mu_3$. As a result, the cPDFs $\mathscr{F}_i(x)$ (bottom panel) have the property that $\mathscr{F}_1(x) < \mathscr{F}_2(x) < \mathscr{F}_3(x)$ for all $x > 0$.

 doi:10.20944/preprints202603.1922.v1

As in the case of pairwise tests, both R and SAS provide functions or procedures for performing tests of homogeneity and monotonicity for $k \geqslant 3$ sites, based on extensions of the Fleming-Harrington class of tests appropriate for doubly-censored and interval-censored data.

5.3.1. R Example

A particularly useful R package for assessing homogeneity and monotonicity of $k \geqslant 3$ sites is the `FHtest` package [40]. For our present purposes, the two key functions this package provides are the sames ones we mentioned earlier in connection with pairwise tests: `FHtesticp()` and `FHtestics()`. These functions implement an extended version of the Fleming-Harrington test using a permutation distribution and a score-vector distribution, respectively.

Figure 11 shows PDFs (left panel) and cPDFs (right panel) for concentration distributions of *E. coli* concentration at inland-lake, river, and coastal beaches in Michigan. Visual inspection suggests that the distributions are not all the same. We used the `FHtesticp()` function to test the null hypothesis that all three cPDFs are the same against the alternative hypothesis that at least two of them differ, with equal weighting for all concentrations ($\rho = \lambda = 0$). `FHtesticp` returned a p-value of $p = 2.48 \times 10^{-11}$ for the test, which provides strong evidence that the null hypothesis is false and therefore at least two of the distributions differ.

Now suppose there is a valid *a priori* reason for expecting that the cPDFs for the three types of beach either show no monotonic trend (the null hypothesis) or show a monotonically increasing trend in the order, inland-lake → river → coastal. Assuming, for the purpose of this example, that there is such a reason, we used the `FHtesticp()` function to test the null hypothesis of no ordering against the one-sided alternative just stated. `FHtesticp` returned a p-value of $p = 2.73 \times 10^{-12}$ for the test, providing strong evidence that the null hypothesis is false and hence that the monotonic increasing trend is real.

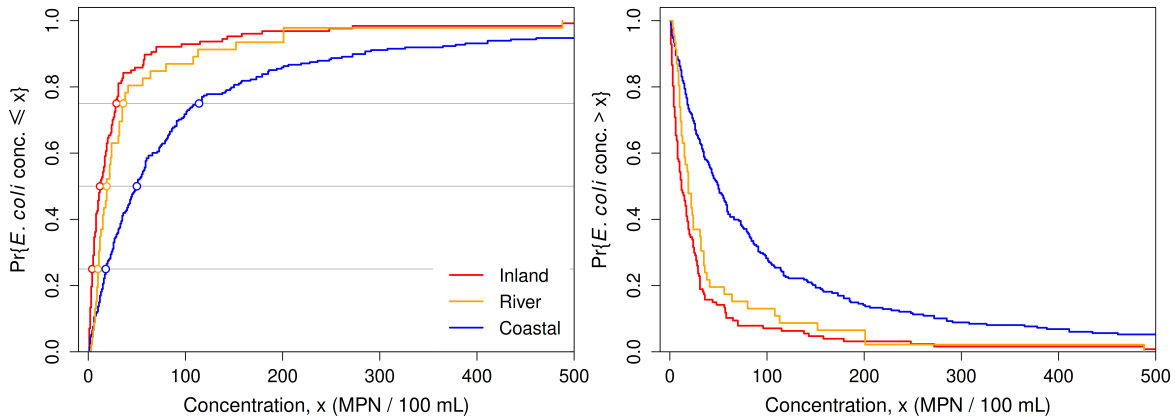

Figure 11. Estimated PDFs (left) and cPDFs (right) for *E. coli* concentrations at inland-lake, river, and coastal beaches in Macomb and St. Clair counties, Michigan. In the left panel, horizontal gray lines indicate probabilities 0.25, 0.50, and 0.75; absissas of the intersections of these lines with the PDFs (indicated by dots) are the corresponding quantiles of the distributions.

5.3.2. SAS Example

Tests of homogeneity and monotonicity for $k \geqslant 3$ sites (or dates) can be performed in SAS using the `ICLIFETEST` procedure. An example is presented in the online Supplementary Materials.

6. Discussion

Concentration estimates produced by the methods of quantitative analytical chemistry are always uncertain to some degree. A key point is that the level of uncertainty varies markedly with analyte concentration, typically in a U-shaped pattern where uncertainty is unacceptably high for sufficiently low and high concentrations but acceptably low for intermediate concentrations. Only concentration estimates in this intermediate range have measurement uncertainty that is low enough to justify

 doi:10.20944/preprints202603.1922.v1

reporting their numerical values and using these values in statistical analyses. The limits of this intermediate concentration range are the lower reporting limit (LRL) and upper reporting limit (URL), while the range itself may conveniently be referred to as the reporting interval. For concentration measurements outside the reporting interval, all we know with confidence is that the true analyte concentrations are either below the LRL or above the URL. Data sets containing such values will be a mix of valid numerical concentrations (measurements lying within the reporting interval) and censored concentrations (measurements lying outside the reporting interval).

Historically, studies of aquatic systems that acquired and statistically analyzed concentration data often employed a limit of detection (LOD) as the sole reporting limit, ignoring the lower and upper limits of quantification and, for methods relying on a calibration curve, also ignoring the concentration range spanned by calibration standards, the concentration range over which the relationship between analyte concentration and instrument signal intensity is approximately linear, and hence the range of instrument signal intensity over which valid confidence intervals can be estimated for predicted analyte concentrations. If rigorous and informative statistical analyses are to be conducted that go beyond merely determining whether there is strong evidence that some of the measured concentrations exceed numerical water-quality standards, it is important to employ reporting limits based on all properties of the chemical measurement process that affect reliability of the resulting concentration measurements. A properly defined LRL typically will lie well above the LOD, and a properly defined URL will lie well above the LRL.

In studies of aquatic systems, the most common statistical methods for analyzing censored concentration data have employed various methods of data fabrication to assign numerical values to censored concentrations. The resulting mixture of valid and fabricated data are then analyzed with traditional statistical methods that treat all data as reliable. While not statistically rigorous, this approach is perhaps innocuous if only a very small proportion of the data are censored. However, it is not unusual to encounter aquatic systems where a substantial proportion — sometimes more than half — of the data are censored. The data fabrication approach is then clearly unacceptable, and some other approach must be used.

The approach we favor for most statistical analyses of censored concentration data is to use methods developed for time-to-event data in the statitical discipline of survival analysis. This approach uses only information about each observation that is known with acceptably low uncertainty. For left-censored values, this information consists of the LRL and the fact that the measurement was less than the LRL; for right-censored values, it consists of the URL and the fact that the measurement was greater than the URL.

Most methods of survival analysis were developed for medical research studies where the times to occurrence of some meaningful event (e.g., death, cure, relapse of a medical condition) for subjects in each of one or more treatment groups are recorded. For subjects where the event of interest did not occur during the fixed duration D of the study, the numerical value of D is recorded as the event time, along with a code indicating that the actual time was greater than D. The only type of censoring produced by such studies is right censoring, which is why most of the traditional methods of survival analysis were designed specifically for right-censored data.

By contrast, the most prevalent type of censoring in concentration data from aquatic systems is left-censoring. Indeed, in studies that improperly employ an LOD as the only reporting limit, left-censoring is the only form of censoring that can occur. In such cases, the data can be "flipped" by reversing the concentration scale, thereby transforming left-censored values to right-censored values so that traditional statistical methods of survival analysis can be used. While this approach is conceptually awkward, it is statistically defensible. But in cases where properly defined lower and upper reporting limits are employed, both left-censored and right-censored concentrations commonly occur, creating a statistical problem that flipping cannot resolve. What is needed are rigorous statistical methdods that can handle data sets that include both types of censored values, and possibly interval-censored values, as well.

 doi:10.20944/preprints202603.1922.v1

The first such methods were developed by Turnbull [38,39] in the mid 1970s but have only begun to be included in standard statistical software in the last 10 or so years. The main types of nonparametric methods are now available in both R and SAS, and these are the methods we have reviewed here. Table 3 lists these methods, together with references to the sections of this paper where they are discussed.

Table 3. Summary of nonparametric survival analysis methods discussed in this review.

Statistical task	Section
Characterize concentration distributions • Estimate PDF, cPDF, and point-wise confidence intervals • Estimate quantiles and their confidence intervals	5.1
Pairwise comparison of concentration distributions	5.2
Tests of homogeneity and monotonic trends	5.3

Less progress has been made with including semiparametric methods of survival analysis (e.g., the proportional hazards model, the accelerated failure-time model) for doubly-censored and interval-censored data in standard statistical software, though we are aware that statisticians are developing R packages to address this problem at the time of this writing. Statisticians rarely use fully parameteric methods of survival analysis in medical applications, because statistical results usually depend strongly on which parametric probability distribution the data are assumed to be sampled from, but it is rarely possible to convincingly determine which distribution is most appropriate for a particular set of data ([34], p. 2; [30], pp. 2–3). We expect the same to be true of environmental concentration data.

In closing, we reiterate that two fundamental principles of rigorous statistical methods for censored data analysis are to avoid fabricating data and to use only information from available data that is known with high confidence. The latest nonparametric methods of survival analysis provide statistical tools that are consistent with both of these principles, can be applied to data sets that include any combination of left-uncensored, right-censored, and interval-censored values, and do not require flipping the data. It is too soon, however, to recommend that flipping be avoided altogether, because versions of semiparametric regression-like methods (e.g., the Cox proportional hazards model and the accelerated failure-time model) that can properly handle left-censored, doubly-censored, and interval-censored data are not yet available in standard statistical software. These methods are heavily used in statistical analyses of right-censored time-to-event data and greatly expand the types analyses that can be performed. However, the versions available in standard statistical software are restricted to right-censored data. Until versions of these methods become available in standard software that can accommodate a wider variety of censoring types, there will be no alternative to applying the traditional versions for right-censored data to flipped left-censored concentration data. This solution, of course, is hardly satisfactory, because it does not apply to double-censored or interval-censored data sets.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org, `Sample_R_and_SAS_Programs_and_Output.pdf` (read this file first), `Sample_R_program.R`, `Test_data.csv` (for use with the R program), and `Sample_SAS_program.txt` (creates its own simulated data, then analyzes them).

Author Contributions: Conceptualization, J.N.M., D.F.; methodology, J.N.M., D.F.; formal analysis, J.N.M., I.C.; writing—original draft preparation, J.N.M.; writing—review and editing, J.N.M., D.F., R.R.R., and I.C.; visualization, J.N.M. and I.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

 doi:10.20944/preprints202603.1922.v1

Conflicts of Interest: Isabelle Ciarrocchi was a graduate student in the Department of Statistics at Grand Valley State University when all parts of this research except manuscript revision were completed. She is now employed by the Perrigo Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. McNair, J.N.; Rediske, R.R.; Hart, J.J.; Jamison, M.N.; Briggs, S. Performance of Colilert-18 and qPCR for monitoring E. coli contamination at freshwater beaches in Michigan. *Environments* **2025**, *12*, 21.
2. Helsel, D.R. *Statistics for Censored Environmental Data Using Minitab and R*; John Wiley & Sons: Hoboken, New Jersey, 2012.
3. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2025.
4. SAS Institute Inc.. *SAS OnlineDoc 9.1.3*. SAS Institute Inc., Cary, North Carolina, USA, 2004.
5. Currie, L.A. Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995). *Pure and Applied Chemistry* **1995**, *67*, 1699–1723.
6. NCCLS. *Protocols for Determination of Limits of Detection and Limits of Quantitation; Approved Guideline*. National Committee for Clinical Laboratory Standards, Wayne, Pennsylvania, USA, 2004.
7. Eurachem. The fitness for purpose of analytical methods: A laboratory guide to method validation and related topics, 2nd ed., 2014.
8. Skoog, D.; Holler, F.; Crouch, S. *Principles of Instrumental Analysis*; Cengage Learning: Boston, Massachusetts, 2018.
9. Fritz, J.S.; Schenk, G.H. *Quantitative Analytical Chemistry*; Allyn and Bacon, Inc.: Newton, Massachusetts, 1987.
10. Kutner, M.H.; Nachtsheim, C.J.; Neter, J.; Li, W. *Applied Linear Statistical Models*; McGraw-Hill: New York, 2005.
11. Parker, P.A.; Vining, G.G.; Wilson, S.R.; Szarka III, J.L.; Johnson, N.G. The prediction properties of classical and inverse regression for the simple linear calibration problem. *Journal of Quality Technology* **2010**, *42*, 332–347.
12. Montgomery, D.C.; Peck, E.A.; Vining, G.G. *Introduction to Linear Regression Analysis*; John Wiley & Sons: Hoboken, New Jersey, 2021.
13. Shoari, N. Quantitative analysis of left-censored concentration data in environmental site characterization. PhD thesis, École de Technologie Supérieure, 2016.
14. Hinkelmann, K.; Kempthorne, O. *Design and Analysis of Experiments, volume 1: Introduction to Experimental Design*; Vol. 1, John Wiley & Sons: Hoboken, New Jersey, 2008.
15. Antweiler, R.C.; Taylor, H.E. Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets: I. Summary statistics. *Environmental Science & Technology* **2008**, *42*, 3732–3738.
16. Antweiler, R.C. Evaluation of statistical treatments of left-censored environmental data using coincident uncensored data sets. II. Group comparisons. *Environmental Science & Technology* **2015**, *49*, 13439–13446.
17. George, B.J.; Gains-Germain, L.; Broms, K.; Black, K.; Furman, M.; Hays, M.D.; Thomas, K.W.; Simmons, J.E. Censoring trace-level environmental data: Statistical analysis considerations to limit bias. *Environmental Science & Technology* **2021**, *55*, 3786–3795.
18. Shumway, R.H.; Azari, R.S.; Kayhanian, M. Statistical approaches to estimating mean water quality concentrations with detection limits. *Environmental Science & Technology* **2002**, *36*, 3345–3353.
19. Gilbert, R.O. *Statistical Methods for Environmental Pollution Monitoring*; John Wiley & Sons: Hoboken, New Jersey, 1987.
20. Agresti, A.; Coull, B.A. Approximate is better than "exact" for interval estimation of binomial proportions. *The American Statistician* **1998**, *52*, 119–126.
21. Brown, L.D.; Cai, T.T.; DasGupta, A. Interval estimation for a binomial proportion. *Statistical Science* **2001**, *16*, 101–133.
22. Agresti, A.; Caffo, B. Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. *The American Statistician* **2000**, *54*, 280–288.
23. Agresti, A. *Categorical Data Analysis*; John Wiley & Sons: Hoboken, New Jersey, 2013.
24. Fagerland, M.W.; Lydersen, S.; Laake, P. The McNemar test for binary matched-pairs data: Mid-p and asymptotic are better than exact conditional. *BMC Medical Research Methodology* **2013**, *13*, 91.

 doi:10.20944/preprints202603.1922.v1

25. Calhoun, P. *Exact: Unconditional Exact Test*, 2022. R package version 3.2.
26. Shan, G.; Wang, W. *ExactCIdiff: Inductive Confidence Intervals for the Difference Between Two Proportions*, 2022. R package version 2.1.
27. Hollander, M.; Wolfe, D.A.; Chicken, E. *Nonparametric Statistical Methods*; John Wiley & Sons: Hoboken, New Jersey, 2014.
28. Lydersen, S.; Fagerland, M.W.; Laake, P. Recommended tests for association in 2×2 tables. *Statistics in Medicine* **2009**, *28*, 1159–1175.
29. Attwood, K.; Park, S.; Hutson, A.D. Practical and robust test for comparing binomial proportions in the randomized phase II setting. *Pharmaceutical Statistics* **2022**, *21*, 361–371.
30. Anderson-Bergman, C. icenReg: Regression models for interval censored data in R. *Journal of Statistical Software* **2017**, *81*, 1–23.
31. Feller, W. *An Introduction to Probability Theory and Its Applications*; John Wiley & Sons: Hoboken, New Jersey, 1968.
32. Klein, J.P.; Moeschberger, M.L. *Survival Analysis: Techniques for Censored and Truncated Data*; Springer-Verlag: New York, 2003.
33. Kalbfleisch, J.D.; Prentice, R.L. *The Statistical Analysis of Failure Time Data*; John Wiley & Sons: Hoboken, New Jersey, 2002.
34. Conover, W.J. *Practical Nonparametric Statistics*; John Wiley & sons: Hoboken, New Jersey, 1999.
35. Lehmann, E.L. *Nonparametrics: Statistical Methods Based on Ranks*; Holden-Day: San Francisco, 1975.
36. Ware, J.H.; Demets, D.L. Reanalysis of some baboon descent data. *Biometrics* **1976**, pp. 459–463.
37. Collett, D. *Modelling survival data in medical research*; Chapman and Hall/CRC: New York, 2023.
38. Turnbull, B.W. Nonparametric estimation of a survivorship function with doubly censored data. *Journal of the American Statistical Association* **1974**, *69*, 169–173.
39. Turnbull, B.W. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society: Series B (Methodological)* **1976**, *38*, 290–295.
40. Oller, R.; Langohr, K. FHtest: An R Package for the Comparison of Survival Curves with Censored Data. *Journal of Statistical Software* **2017**, *81*, 1–25. https://doi.org/10.18637/jss.v081.i15.
41. Hosmer Jr, D.W.; Lemeshow, S.; May, S. *Applied Survival Analysis: Regression Modeling of Time-To-Event Data*; John Wiley & Sons: Hoboken, New Jersey, 2008.
42. Fleming, T.R.; Harrington, D.P.; O'sullivan, M. Supremum versions of the log-rank and generalized Wilcoxon statistics. *Journal of the American Statistical Association* **1987**, *82*, 312–320.