Ann. Occup. Hyg.

doi:10.1093/annhyg/met037

Tests for an Upper Percentile of a Lognormal Distribution Based on Samples with Multiple Detection Limits and Sample-Size Calculation

KALIMUTHU KRISHNAMOORTHY[1]*, THOMAS MATHEW[2] and ZHAO XU[3]

[1]Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70508-1010, USA; [2]Department of Mathematics and Statistics, University of Maryland, Baltimore County, Baltimore, MD 21250, USA; [3]Statacorp LP, College Station, TX 77845, USA

Received 16 December 2012; in final form 16 May 2013; Accepted 24 May 2013

The problem of determining sample size for testing an upper percentile of a lognormal distribution based on samples with multiple detection limits is considered. Two tests, the signed likelihood ratio test and another test based on a pivotal statistic, are outlined. These tests are very satisfactory in controlling type I error rates and comparable in terms of powers. Procedures and R codes for calculating sample sizes for these tests to attain a specified power are given. It is noted that for guaranteeing a given power, increased sample size is necessary due to the presence of detection limits, and the required sample size goes up as the proportion of non-detects goes up. It is also noted that in the multiple detection limit scenario, sample-size determination does not require knowledge of the proportions of non-detects that are expected to be below the individual detection limits; rather, what is required is a knowledge of the overall percentage of non-detects that is to be expected in the entire sample. Sample-size calculation is illustrated using a practical situation.

Keywords: maximum likelihood estimates; non-central *t* distribution; pivotal test; powers; signed likelihood ratio test

Abbreviations: z_p, the 100p percentile of the standard normal distribution; Φ, distribution function of the standard normal random variable; $\exp(\mu)$, the geometric mean of a lognormal distribution; $\exp(\sigma)$, the geometric standard deviation of a lognormal distribution; ξ_p, the 100p percentile of the $N(\mu,\sigma^2)$ distribution, $\xi_p = \mu + z_p\sigma$, ξ_p^0 is a specified value of ξ_p, e.g., $\xi_p^0 = \ln(\text{OEL})$; DL_i, the detection limit for the ith laboratory method or sampling device; $i = 1,\ldots,k$; p_i, probability of a non-detect for the ith device or method, $i = 1,\ldots,k$; m_i, the number of non-detects below DL_i; $i = 1,\ldots,k$; m, $\sum_{i=1}^{k} m_i$, the total number of non-detects; $(\hat{\mu},\hat{\sigma})$, the MLE of (μ,σ) based on a log-transformed random sample with log-transformed detection limits $\ln\text{DL}_1,\ldots,\ln\text{DL}_k$; $\hat{\sigma}_{\xi_p}$, the MLE of σ under the constraint $\xi_p = \mu + z_p\sigma$; $t_{m;1-\alpha}(\delta)$, $100(1-\alpha)$ percentile of the non-central t distribution with degrees of freedom m and the non-centrality parameter δ; $\delta^* = \xi_p^0 - \xi_p/\sigma$, the scaled difference between the specified value and the true value of the 100p percentile of log-transformed exposure distribution (δ^*) is independent of the unit of measurements)

INTRODUCTION

The detection limit (DL) problem has received considerable attention in the industrial hygiene literature due to its common occurrence among exposure data and due to the challenges it presents while performing the data analysis. It is now well recognized that replacing the values below a DL with a constant (such as one-half of the DL) can result in a test procedure with elevated type I error rates or a confidence interval with inaccurate confidence; the issue has recently been brought to

*Author to whom correspondence should be addressed.
Tel: +337-482-5283; fax: +337-482-5346;
e-mail: krishna@louisiana.edu

focus by the commentary by Helsel (2010), the editorial by Ogden (2010), and by the discussions and examples in the second edition of the book by Helsel (2012). The DL problem has mostly been addressed in the case of a single DL, even though some researchers have also addressed the multiple DL scenario; see Krishnamoorthy and Xu (2011) and the references therein for some recent contributions. In the industrial hygiene context, most of the literature on DLs deals with the problem of estimating the relevant parameters, for example, the arithmetic mean, geometric mean, and the geometric standard deviation of the lognormal exposure distribution. The development of confidence intervals and hypotheses tests, as well as a rigorous study of their accuracy, has been undertaken only recently; see Krishnamoorthy *et al.* (2009, 2011) and Krishnamoorthy and Xu (2011). An issue that has been overlooked is that of sample-size determination. To simply say, does the presence of DLs require increased sample size to guarantee a specified power for a test, compared with the case of a sample with no non-detects? How does the presence of multiple DLs affect the power of a test and the sample size required to guarantee a given power? These are the issues investigated in this study.

Although we focus on exposure data following a lognormal distribution, the sample-size calculation described in the sequel is applicable for any distribution that has one-to-one relation with the normal distribution. For instance, if exposure data follow a gamma distribution, then cube root transformation can be used to determine the sample size (see Krishnamoorthy and Xu, 2011). Furthermore, the problem that we have addressed is that of testing an upper percentile of the exposure profile. This is one of the basic strategies recommended in the consensus standard published by the American Industrial Hygiene Association (see Mulhausen and Damiano, 1998; Ignacio and Bullock, 2006). Specifically, we have considered the problem of testing if an upper percentile is below an occupational exposure limit (OEL) and have rigorously investigated the computation of the sample size required to guarantee a given power (say, 80 or 90%). This is done in the multiple DL scenario. Note that the sample-size calculation has to be based on a test that performs well in terms of type I error probability. Such a test is developed by Krishnamoorthy and Xu (2011) based on a suitable pivotal quantity and the test is explained later in this study. In addition, we have also proposed the likelihood ratio test by K. Krishnamoorthy, T. Mathew and Z. Xu (unpublished data), implemented using simulations rather than using large sample approximations. Both the tests exhibit accurate performance regardless of the sample size and number of DLs. We have also included the algorithms necessary to carry out the tests. The sample-size problem is then investigated in the multiple DL scenario. The only other work that we have seen, where the sample-size issue is addressed under DLs, is the study by Chu *et al.* (2006). The authors have addressed the problem of comparing the geometric means of two lognormal distributions when the samples were below DL values and have investigated the sample-size issue.

Typically, sample-size determination does require some knowledge about the population parameters. In our investigation, what is necessary is a value of the percentile (under the alternative hypothesis) at which the power is required to be 80 or 90%, along with a knowledge of the geometric mean or geometric standard deviation of the exposure population. In addition, some idea about the overall percentage of the data below all the DLs is all that is required and it is not necessary to have information on the proportion of the data below each individual DL. This is a practically useful observation. On the basis of this observation, we provided tables of sample sizes required to attain a power of 0.80 or 0.90 when the proportion of non-detects is known a priori. We also posted an R program (www.ucs.louisiana.edu/~kxk4695) that calculates the power of the proposed test for a given sample size and DLs. Our conclusion is that the presence of DLs does result in increased sample size to guarantee a given power. DLs being a commonly occurring phenomenon in exposure data analysis, we hope our work will be of interest to industrial hygienists who want to determine the required sample size while planning their data collection.

MAXIMUM LIKELIHOOD ESTIMATORS

The starting point of the procedures that we shall develop is the computation of the maximum likelihood estimators (MLEs). Here, we consider only lognormally distributed exposure data. The lognormal distribution is known to be a very good approximation to model the distribution

of exposure samples in many practical scenarios. We recall that exposure measurements follow a lognormal distribution if the log-transformed measurements follow a normal distribution. We shall denote by μ and σ^2 the mean and variance of the log-transformed data; the corresponding normal distribution will be denoted by $N(\mu,\sigma^2)$. Thus, $\exp(\mu)$ and $\exp(\sigma)$, are, respectively, the geometric mean and geometric standard deviation of the exposure data.

Consider a random sample of n exposure measurements, assumed to follow a lognormal distribution, subject to k DLs. Here, k represents, for example, different instruments or methods used for obtaining the exposure data. Let m_i denote the number of non-detects below the ith DL so that $m=\sum_{i=1}^{k} m_i$ is the total number of non-detects. Let X_1, X_2, …, X_{n-m} be the log-transformed data corresponding to the $n-m$ observations that are above the log-transformed DLs. Thus, on the logarithmic scale, we have a random sample of n measurements from a normal distribution, say $N(\mu,\sigma^2)$, with m_i observations below $\ln \mathrm{DL}_i (i=1,...,k), X_1, X_2, ..., X_{n-m}$ being the observations that are greater than the log-transformed DLs (where $m=\sum_{i=1}^{k} m_i$). Define

$$\bar{X}_d = \frac{1}{n-m}\sum_{i=1}^{n-m} X_i \text{ and } S_d^2 = \frac{1}{n-m}\sum_{i=1}^{n-m}(X_i-\bar{X}_d)^2, \quad (1)$$

so that $\bar{X}_d$ and S_d^2 are, respectively, the sample mean and sample variance of the $n-m$ observations $X_1, X_2, ..., X_{n-m}$. The log-likelihood function, after omitting a constant term, can be written as

$$l(\mu,\sigma) = \sum_{i=1}^{k} m_i \ln \Phi(z_i^*) - (n-m)\ln\sigma - \frac{(n-m)(S_d^2+(\bar{X}_d-\mu)^2)}{2\sigma^2}, \quad (2)$$

where $z_i^* = \dfrac{\ln \mathrm{DL}_i - \mu}{\sigma}$, $i=1,...,k$, and $\Phi(\cdot)$ denotes the standard normal cumulative distribution function. The maximum likelihood estimates of μ and σ, say $\hat{\mu}$ and $\hat{\sigma}$, respectively, can be obtained by maximizing the above log-likelihood function. These can be numerically obtained using the bivariate Newton–Raphson iterative method; for details, see Krishnamoorthy and Xu (2011).

Remark 1. All the statistical methods proposed in the literature for the problems involving multiple DLs are based on the assumption that the measurements obtained using different instruments or methods have the same mean and variance. It is possible that the measurements obtained using different methods or instruments have different means and variances, and in that case we are dealing multiple exposure distributions each with a single DL. The problem of exposure assessment in such a set-up is theoretically complex and it is not possible to come up with a single upper percentile of several exposure distributions. Therefore, we shall proceed with the assumption that the measurements obtained by different instruments have the same mean and variance and keep in mind that the results in the sequel are valid only under such assumption.

TESTS AND POWER CALCULATION

Let ξ_p denote the 100pth percentile of the normal distribution $N(\mu,\sigma^2)$. Clearly, $\xi_p = \mu + z_p\sigma$, where z_p is the 100pth percentile of the standard normal distribution. Then, the pth percentile of the original lognormally distributed exposure data is $\exp(\xi_p)$. Suppose we wish to test the hypothesis that the 100pth percentile of the exposure distribution is below an OEL. Formulating this as the alternative hypothesis, we, thus, have the following null hypothesis H_0 and alternative hypothesis H_a:

$$H_0 : \xi_p \geq \xi_p^{\,0} \text{ versus } H_a : \xi_p < \xi_p^{\,0}, \quad (3)$$

where $\xi_p^0 = \ln(\mathrm{OEL})$. The rejection of H_0, based on an appropriate statistical test, leads us to the conclusion that the pth percentile of the exposure distribution is below the OEL. We shall now discuss tests for the above hypothesis, first based on an exposure sample that does not include any non-detects and then on samples subject to k DLs.

A test procedure in the absence of non-detects

Let $\bar{X}$ and S^2, respectively, denote the mean and variance based on a sample of size m from a $N(\mu,\sigma^2)$ distribution. Clearly, the null hypothesis in equation (3) is rejected at a significance level of α if a $100(1-\alpha)\%$ upper confidence limit for ξ_p is less than ξ_p^0, where we recall that $\xi_p = \mu + z_p\sigma$, z_p being the 100pth percentile of the standard normal distribution. An upper confidence limit

for ξ_p is well known in the statistics literature (see Krishnamoorthy and Mathew, 2009, Chapter 2) and is based on the non-central t distribution with non-centrality parameter $z_p\sqrt{n}$ and degrees of freedom $n-1$. Let $t_{n-1;1-\alpha}\left(z_p\sqrt{n}\right)$ denote the $100(1-\alpha)$th percentile of such a non-central t distribution. Then, a $100(1-\alpha)\%$ upper confidence limit for ξ_p is given by $\bar{X}+\frac{1}{\sqrt{n}}t_{n-1;1-\alpha}\left(z_p\sqrt{n}\right)S$ and a test for equation (3) consists of rejecting H_0 if

$$\bar{X}+\frac{1}{\sqrt{n}}t_{n-1;1-\alpha}\left(z_p\sqrt{n}\right)S<\xi_p^0 \text{ equivalently, if}$$
$$\frac{\xi_p^0-\bar{X}}{S}>\frac{1}{\sqrt{n}}t_{n-1;1-\alpha}\left(z_p\sqrt{n}\right). \tag{4}$$

To compute the power function of such a test, we note that

$$\begin{aligned}\frac{\xi_p^0-\bar{X}}{S}&=\frac{\xi_p^0-\xi_p+\xi_p-\bar{X}}{S}\\&=\frac{\left(\xi_p^0-\xi_p+\mu+z_p\sigma-\bar{X}\right)\Big/\left(\sigma/\sqrt{n}\right)}{S\Big/\left(\sigma/\sqrt{n}\right)}\\&\sim\frac{\dfrac{\xi_p^0-\xi_p}{\sigma/\sqrt{n}}+z_p\sqrt{n}+z}{\sqrt{\dfrac{X_{n-1}^2}{n-1}}},\end{aligned}$$

where the standard normal random variable $Z=\sqrt{n}\left(\mu-\bar{X}\right)\big/\sigma$ and the chi-square random variable $X_{n-1}^2=(n-1)S^2/\sigma^2$ are independent. By writing $X\sim Y$, we mean that the two random variables X and Y are identically distributed. Thus, standard distributional results imply that

$$\frac{\xi_p^0-\bar{X}}{S}\sim\frac{1}{\sqrt{n}}t_{n-1}\left(\delta+z_p\sqrt{n}\right),$$
$$\text{with } \frac{\sqrt{n}\left(\xi_p^0-\xi_p\right)}{\sigma}. \tag{5}$$

Here, $t_m(\eta)$ denotes a non-central t distribution with the degrees of freedom m and non-centrality parameter η. The power function, which follows from equations (4) and (5), is given by

$$P\left(t_{n-1}\left(\delta+z_p\sqrt{n}\right)>t_{n-1;1-\alpha}\left(z_p\sqrt{n}\right)\right). \tag{6}$$

Using the result that the non-central $t_m(\eta)$ distribution is stochastically increasing with respect to the non-centrality parameter η, we see that, for a fixed (p, α, n), the above power function is increasing with increasing value of δ. In other words, the power function is increasing with increasing value of δ^*, where

$$\delta^*=\frac{\xi_p^0-\xi_p}{\sigma}. \tag{7}$$

Consequently, the smaller ξ_p is compared with ξ_p^0, the larger the power. After specifying a value of δ^* in equation (7), it is possible to determine the sample size n so that the test provides a given power, say 80%. We shall not go into further details concerning this, since this is available in the literature, and our primary interest is with the exposure data subject to DLs.

Test procedures in the presence of non-detects

We shall consider two test procedures: a test based on an approximate pivotal quantity and a likelihood-based test procedure.

A pivotal test. Suppose samples include multiple DLs $\mathrm{DL}_1,\ldots,\mathrm{DL}_k$. Let $\hat{\mu}$ and $\hat{\sigma}$ denote, respectively, the MLEs of μ and σ based on log-transformed samples. Recently, Krishnamoorthy and Xu (2011) showed that

$$Q_p=\frac{\xi_p-\hat{\mu}}{\hat{\sigma}}\sim\frac{z_p-\hat{\mu}^*}{\hat{\sigma}^*}=Q_p^*, \text{ approximately,} \tag{8}$$

where $\hat{\mu}^*$ and $\hat{\sigma}^*$ are the MLEs based on a sample of size n from a standard normal distribution with DLs $\mathrm{DL}_i^*=(\ln\mathrm{DL}_i-\hat{\mu})/\hat{\sigma}$. That is, $\hat{\mu}^*$ and $\hat{\sigma}^*$ are the values of μ and σ, respectively, which maximize the log-likelihood function in equation (2) with $X_1,X_2,\ldots,X_{n-m}$ being a sample from a $N(0,1)$ distribution with DLs $\mathrm{DL}_i^*=(\ln\mathrm{DL}_i-\hat{\mu})/\hat{\sigma}$. An approximate 100 $(1-\alpha)\%$ upper confidence limit for $\xi_p=\mu+z_p\sigma$, based on the above distributional result, is given by

$$\hat{\mu}+Q_{p;1-\alpha}^*\hat{\sigma}, \tag{9}$$

where $Q_{p;1-\alpha}^*$ is the $100(1-\alpha)$ percentile of $Q_p^*=(z_p-\hat{\mu}^*)/\hat{\sigma}^*$. Note that $Q_{p;1-\alpha}^*$ can be estimated by simulating samples from the standard normal distribution $N(0,1)$; see Algorithm 1.

The test on the basis of the upper confidence limit in equation (9) rejects the null hypothesis in equation (3) when

$$\hat{\mu}+Q^*_{p;1-\alpha}\hat{\sigma}<\xi^0_p, \text{ equivalently, } \frac{\xi^0_p-\hat{\mu}}{\hat{\sigma}}>Q^*_{p;1-\alpha}. \quad (10)$$

We shall refer to this test as the pivotal test because the confidence limit in equation (9) is based on the approximate pivotal quantity in equation (8).

Algorithm 1. Let m_i denote the number of non-detects below the DL DL_i in a set of n_i measurements that were made by the ith sampling device (or laboratory method) with DL DL_i, $i=1,\ldots,k$. Compute the MLEs $\hat{\mu}$ and $\hat{\sigma}$ based on log-transformed sample and set $\mathrm{DL}^*_i=\frac{\ln\mathrm{DL}_i-\hat{\mu}}{\hat{\sigma}}$, $i=1,\ldots,k$.

1. Generate a sample of size n_i from $N(0,1)$, $i=1,\ldots,k$. Discard the observations from the ith sample that are less than DL^*_i, $i=1,\ldots,k$. Let m_i denote the number of observations below DL^*_i, $i=1,\ldots,k$, so that $m=\sum_{i=1}^k m_i$.
2. Compute the MLEs $\hat{\mu}^*$ and $\hat{\sigma}^*$ using the sample in step 1.
3. Set $Q^*_p=\frac{z_p-\hat{\mu}^*}{\hat{\sigma}^*}$, where z_p is the $100p$ percentile of $N(0,1)$ distribution.
4. Repeat steps 1–4 for a large number of times, say, M.
5. The $100(1-\alpha)$ percentile of these Q^*_ps generated above is a Monte Carlo estimate of $Q^*_{p;1-\alpha}$.

An R program to compute $Q^*_{p;1-\alpha}$ along with a help file is posted at www.ucs.louisiana.edu/~kxk4695.

The signed likelihood ratio test. The signed likelihood ratio test (SLRT) is equivalent to the usual likelihood ratio test. In addition to the MLEs $\hat{\mu}$ and $\hat{\sigma}$, it requires the MLE of the parameters computed under the constraint specified in the null hypothesis in equation (3). Since H_0 states that $\mu+z_p\sigma=\xi^0_p$, we have $\mu=\xi^0_p-z_p\sigma$. In other words, under H_0, we only need to compute the MLE of σ, say, $\sigma_{\xi^0_p}$, and the MLE of μ (under H_0) is then given by $\mu_{\xi^0_p}=\xi^0_p-z_p\sigma_{\xi^0_p}$. The MLE $\hat{\sigma}^2_{\xi^0_p}$ of σ is referred to as its constrained MLE since it is obtained under the constraint specified in H_0. It is readily verified that $\sigma_{\xi^0_p}$ is the value of σ that maximizes the log-likelihood function

$$l\left(\xi^0_p,\sigma^2\right)=\sum_{i=1}^k m_i\ln\Phi\left(z_{i\xi^0_p}\right)-(n-m)\ln(\sigma)$$
$$-\frac{n-m}{2\sigma^2}\left(s^2_d+\left(\bar{x}_d-\xi^0_p+z_p\sigma\right)^2\right), \quad (11)$$

where $z_{i\xi^0_p}=\frac{\ln\mathrm{DL}_i-\xi^0_p}{\sigma}+z_p$, and $\bar{x}_d$ and s^2_d are the observed values of $\bar{X}_d$ and S^2_d, respectively, defined in equation (1). Computational details on the constrained MLE can be found in a study by K. Krishnamoorthy, T. Mathew and Z. Xu (unpublished data).

Let $\hat{\sigma}^2_{\xi^0_p}$ denote the constrained MLE of σ^2 specified above and let $\hat{\xi}_p=\hat{\mu}+z_p\hat{\sigma}$. The SLRT statistic is written as

$$R_p=\mathrm{sign}\left(\hat{\xi}_p-\xi^0_p\right)\left[2\left\{l\left(\hat{\xi}_p,\hat{\sigma}^2\right)-l\left(\xi^0_p,\hat{\sigma}^2_{\xi^0_p}\right)\right\}\right]^{1/2}, \quad (12)$$

where $\mathrm{sign}(x)=1$ if $x>0$, and is -1 otherwise. The above SLRT statistic R_p follows the standard normal distribution asymptotically. This is a large sample result and the simulation study by K. Krishnamoorthy, T. Mathew and Z. Xu (unpublished data) indicated that the SLRT based on this large sample, normal approximation is not satisfactory even for large samples. Alternatively, arguing as in a study by Krishnamoorthy and Xu (2011), we can estimate the percentiles of R_p using simulated samples from the standard normal distribution; see Algorithm 2 given below. Let $R^*_{p;\alpha}$ denote the 100α percentile of the distribution of R_p so obtained, when $\xi_p=\xi^0_p$. The SLRT rejects H_0 in equation (3) when $R_p<R^*_{p;\alpha}$. The power of the SLRT is given by

$$P\left(R_p<R^*_{p;\alpha}\right), \quad (13)$$

where R_p in equation (12) is a function of the MLEs $\hat{\mu}$ and $\hat{\sigma}$ and the constrained MLE $\hat{\sigma}^2_{\xi^0_p}$ based on a random sample from the $N(\mu,\sigma^2)$ distribution. The percentile $R^*_{p;\alpha}$ can be estimated using the following algorithm.

Algorithm 2. For a given sample of size n with DLs $\mathrm{DL}_1,\ldots,\mathrm{DL}_k$, let n_i denote the size of the sample analyzed by the i th laboratory, $i=1,\ldots,k$ so that $\sum_{i=1}^k n_i=n$, where we also note that $n_i\geq m_i$ for each i.

1. Compute the MLEs $\hat{\mu}$, $\hat{\sigma}^2$ and for a given ξ_p^0, compute the constrained MLE $\hat{\sigma}^2_{\xi_p^0}$; we refer to K. Krishnamoorthy, T. Mathew and Z. Xu (unpublished data) for details regarding the computation of these. Set $\hat{\mu}_0 = \xi_p^0 - z_p \sigma_{\xi_p^0}$.
2. Generate a sample of size n_i from $N\left(\hat{\mu}_0, \hat{\sigma}^2_{\xi_p^0}\right)$, $i = 1,\ldots,k$.
3. Discard the observations from the i th sample, which are less than $\ln\mathrm{DL}_i$, $i = 1,\ldots,k$. Let m_i denote the number of observations below $\ln\mathrm{DL}_i$, $i = 1,\ldots,k$, so that $m = \sum_i^k m_i$.
4. Based on the sample in step 3, compute the MLEs $\tilde{\mu}$ and $\tilde{\sigma}^2$ by maximizing equation (2) and the constrained MLE $\tilde{\sigma}^2_{\xi_p^0}$ by maximizing equation (11).
5. Compute $R^* = \mathrm{sign}\left(\tilde{\xi}_p - \xi_p^0\right)\left[2\left\{l\left(\tilde{\xi}_p, \tilde{\sigma}^2\right) - l\left(\xi_p^0, \tilde{\sigma}^{*2}_{\xi_p^0}\right)\right\}\right]^{1/2}$, where $\tilde{\xi}_p = \tilde{\mu} + z_p\tilde{\sigma}$.
6. Repeat steps 1–4 for a large number of times, say, M.
7. The 100α percentile of these 10 000 R^* is an estimate of $R^*_{p;\alpha}$ defined in equation (13).

TYPE I ERROR AND POWER STUDIES

The power of the pivotal test is given by $P\left(\frac{\xi_p^0 - \hat{\mu}}{\hat{\sigma}} > Q^*_{p;1-\alpha}\right)$, where $\hat{\mu}$ and $\hat{\sigma}$ are the MLEs based on a random sample from the $N(\mu,\sigma^2)$ distribution, and $Q^*_{p;1-\alpha}$ is defined below equation (9). Following the arguments of Krishnamoorthy and Xu (2011), the power function can be expressed as

$$P\left(\frac{\xi_p^0 - \hat{\mu}}{\hat{\sigma}} > Q^*_{p;1-\alpha}\right) \simeq P\left(\frac{\delta^* + z_p - \hat{\mu}^*}{\hat{\sigma}^*} > Q^*_{p;1-\alpha}\right), \tag{14}$$

where δ^* is defined in equation (7) and $\left(\hat{\mu}^*, \hat{\sigma}^*\right)$ is as defined in equation (8). It is clear from the power function in equation (14) that the power of the pivotal test depends on the parameters only via $\delta^* = \left(\xi_p^0 - \xi_p\right)/\sigma$. However, it is difficult to derive the power function explicitly and simulation studies should be carried out to understand its power properties. The powers of the pivotal test can be estimated using the following algorithm.

Algorithm 3. For given values of $(n, p, \alpha, \mathrm{DL}_1,\ldots, \delta^*, \xi_p^0, \sigma)$:

1. Calculate z_p and $\mu = \xi_p^0 - (\delta^* + z_p)\sigma$.
2. Generate a sample of size n from a $N(\mu,\sigma^2)$ distribution with DLs $\ln\mathrm{DL}_1,\ldots,\ln\mathrm{DL}_k$.
3. Compute the MLEs $\hat{\mu}$ and $\hat{\sigma}$ based on the sample in step 2 and set $Q_p = \left(\xi_p^0 - \hat{\mu}\right)/\hat{\sigma}$.
4. Set $\mathrm{DL}_i^* = \ln\mathrm{DL}_i - \hat{\mu}/\hat{\sigma}$, $i = 1,\ldots,k$. Use Algorithm 1 with $(\hat{\mu}, \hat{\sigma}, \mathrm{DL}_i^*)$ and $M = 5000$ runs to find $Q^*_{p;1-\alpha}$.
5. Repeat steps $2-4$ for N times.
6. The percentage of these N Q_p s that are greater than $Q^*_{p;1-\alpha}$ is an estimate of the exact power on the left-hand side of equation (14).

The powers of the SLRT can be estimated using an algorithm similar to Algorithm 3. Note that the power calculation in Algorithm 3 involves a total of $M \times N$ simulation runs. Based on our experience, $M = 2500$ and $N = 5000$ would produce reasonably accurate estimates of the power. However, Algorithm 3 with these choices of M and N involves $2500 \times 5000 = 17\,500\,000$ runs so it could be very time consuming. An alternative calculation based on the approximation in equation (14) is much faster and is described in Algorithm 4.

Algorithm 4. For given values of $(n, p, \alpha, \mathrm{DL}_1,\ldots, \mathrm{DL}_k, \delta^*, \xi_p^0, \sigma)$:

1. Calculate z_p and $\mu = \xi_p^0 - \left(\delta^* + z_p\right)\sigma$ and set $\mathrm{DL}_i^* = \frac{\ln\mathrm{DL}_i - \mu}{\sigma}$, $i = 1,\ldots,k$.
2. Generate a sample of size n from a $N(0,1)$ distribution with DLs $\mathrm{DL}_i^*,\ldots,\mathrm{DL}_k^*$.
3. Compute the MLEs $\hat{\mu}^*$ and $\hat{\sigma}^*$ based on the generated sample in step 2. Set $Q_p^* = (z_p - \hat{\mu})/\hat{\sigma}^*$ and $Q = (\delta^* + z_p - \hat{\mu})/\hat{\sigma}^*$.
4. Repeat steps 2 and 3 for 10 000 times.
5. Find the $100(1-\alpha)$ percentile of 1000 Q_p^* s, and call it $Q^*_{p;1-\alpha}$. The percentage of these 10 000 Q s that are greater than $Q^*_{p;1-\alpha}$ is an estimate of the approximate power on the right-hand side of equation (14).

Note that Algorithm 4 involves only 10 000 runs so it is much less time consuming than Algorithm 3. Furthermore, our extensive simulation studies indicated that the results based on Algorithms 3 and 4 are very similar. Therefore, we shall use Algorithm 4 for power calculation in the sequel.

Table 1. Type I error rates and powers of the pivotal test and the SLRT (in parentheses) for $H_0 : \xi_{0.95} \geq \xi^0_{0.95}$ versus $H_\alpha : \xi_{0.95} < \xi^0_{0.95}$ as a function of (μ, σ) at 5% significance level, two DLs.

Type I error rates when $\xi^0_{0.95} = 2$, $\mu = \xi^0_{0.95} - \sigma(\delta^* + z_{0.95})$, $\delta^* = (\xi_{0.95} - \xi^0_{0.95})/\sigma = 0$

(p_1, p_2)	$n=10$			$n=20$			$n=30$		
	σ			σ			σ		
	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
(0.2,0.3)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)
(0.3,0.4)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)
(0.3,0.5)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)
(0.5,0.6)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)	0.05(0.05)
(0.7,0.8)	0.06(0.05)	0.05(0.05)	0.05(0.06)	0.05(0.05)	0.04(0.05)	0.05(0.06)	0.05(0.05)	0.05(0.05)	0.05(0.05)
Uncensored	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05

Powers when $\xi^0_{0.95} = 2$, $\mu = \xi^0_{0.95} - \sigma(\delta^* + z_{0.95})$, $\delta^* = (\xi_{0.95} - \xi^0_{0.95})/\sigma = 1$

(p_1, p_2)	$n=10$			$n=20$			$n=30$		
	σ			σ			σ		
	0.5	1	1.5	0.5	1	1.5	0.5	1	1.5
(0.2,0.3)	0.35(0.34)	0.36(0.37)	0.36(0.35)	0.66(0.64)	0.67(0.67)	0.66(0.66)	0.85(0.85)	0.84(0.84)	0.84(0.84)
(0.3,0.4)	0.33(0.34)	0.33(0.34)	0.32(0.34)	0.61(0.62)	0. 63(0.61)	0.63(0.61)	0.81(0.81)	0.81(0.80)	0.82(0.80)
(0.3,0.5)	0.32(0.31)	0.31(0.33)	0.32(0.31)	0.61(0.59)	0.63(0.61)	0.61(0.57)	0.80(0.81)	0.79(0.79)	0.80(0.80)
(0.5,0.6)	0.27(0.27)	0.26(0.27)	0.27(0.28)	0.54(0.53)	0.52(0.55)	0.51(0.52)	0.74(0.76)	0.73(0.74)	0.74(0.72)
(0.7,0.8)	0.28(0.27)	0.27(0.28)	0.28(0.27)	0.40(0.43)	0.40(0.44)	0.41(0.41)	0.63(0.61)	0.61(0.62)	0.60(0.61)
Uncensored	0.41	0.41	0.41	0.74	0.74	0.74	0.90	0.90	0.90

Powers when $\xi^0_{0.95} = 3$, $\mu = \xi^0_{0.95} - \sigma(\delta^* + z_{0.95})$, $\delta^* = (\xi_{0.95} - \xi^0_{0.95})/\sigma = 1$

(p_1, p_2)	$n=10$			$n=20$			$n=30$		
	μ			μ			μ		
	1	2	2.5	1	2	2.5	1	2	2.5
(0.2,0.3)	0.35(0.35)	0.36(0.35)	0.35(0.36)	0.67(0.65)	0.66(0.65)	0.68(0.64)	0.85(0.84)	0.85(0.84)	0.85(0.84)
(0.3,0.4)	0.33(0.36)	0.31(0.34)	0.34(0.35)	0.65(0.64)	0.61(0.62)	0.62(0.60)	0.82(0.81)	0.82(0.81)	0.80(0.80)
(0.3,0.5)	0.33(0.32)	0.31(0.33)	0.31(0.32)	0.61(0.57)	0.62(0.62)	0.61(0.62)	0.81(0.81)	0.81(0.81)	0.80(0.80)
(0.5,0.6)	0.25(0.28)	0.28(0.28)	0.27(0.29)	0.53(0.56)	0.54(0.54)	0.53(0.53)	0.72(0.75)	0.72(0.75)	0.73(0.74)
(0.7,0.8)	0.28(0.25)	0.27(0.26)	0.29(0.26)	0.40(0.42)	0.39(0.44)	0.41(0.43)	0.62(0.63)	0.59(0.62)	0.62(0.62)
Uncensored	0.41	0.41	0.41	0.74	0.74	0.74	0.90	0.90	0.90

$\xi_{0.95}$, true unknown 95th percentile to be tested; $\xi^0_{0.95}$, specified value of the 95th percentile; p_i, proportion of non-detects below DL_i; $\ln \mathrm{DL}_i = \mu + z_{p_i}\sigma$; uncensored: sample with no non-detects, i.e., $(p_1, p_2) = (0,0)$; $z_{0.95}$, 95th percentile of a standard normal.

It should also be noted that the powers become equal to the type I error rates when $\delta^* = (\xi_p - \xi_p^0)/\sigma = 0$. In order to compare the type I error rates of the pivotal test and the SLRT, we estimated the type I error rates of both tests and reported them in the first part of Table 1 for sample sizes 10, 20, and 30, with two DLs, as indicated in Table 1. Estimated type I error rates in Table 1 are very close to the nominal level 0.05 for all the cases considered, which indicate that both tests control the type I error rates very satisfactorily.

We estimated the powers of the pivotal test and the SLRT using Algorithm 4 and presented them in Table 1. For convenience and ease of tabulation, we presented the powers after specifying values for the proportions of non-detects $p_1,\ldots,p_k$. Recall that DL_i and p_i are related through the equation $\ln \mathrm{DL}_i = \mu + z_{p_i}\sigma$, $i = 1,\ldots,k$. The powers were estimated as a function of (μ,σ) while $\delta^* = (\xi_p^0 - \xi_p)/\sigma$ is fixed. Note that for a fixed δ^*, $\mu = \xi_p^0 - (\delta^* + z_p)\sigma$. We computed the powers for different values of σ while δ^* is fixed at the value 1 and $\xi_p^0 = 2$, and presented them in the second part of Table 1. Examination of powers clearly indicate that the powers of the SLRT do not depend on individual values of σ and μ (that produce $\delta^* = 1$). Specifically, we see that when $(p_1,p_2) = (0.2,0.3)$, $n = 10$, $\xi_p^0 = 2$, and $\delta^* = 1$, the powers for various values of σ are approximately the same. These small differences among the powers for various values of σ, while p_i s and δ^* are fixed, could be due to simulation errors. The values in the third part of Table 1 are the estimated powers at $\xi_p^0 = 3$ and $\delta^* = 1$. We once again note that the powers in the third part of Table 1 are close to the corresponding ones in the second part; see, for example, the powers in the row corresponding to $(p_1,p_2) = (0.7,0.8)$ in the second part and those in the third part of Table 1. Thus, it is clear that for a fixed $(n,(p_1,p_2),\alpha)$, the powers of both tests are comparable and they depend on the parameters only via $\delta^* = (\xi_p^0 - \xi_p)/\sigma$.

Here, we would like to point out that the missing mechanism when DLs are present is 'missing at random' and not 'missing completely at random'; the latter scenario results in a reduced sample size consisting of uncensored data. How does the powers of the tests compared when we have reduced number of observations resulting from the presence of DLs and that resulting from data missing completely at random? Such a power comparison can be done using the power values reported in Table 1. For example, consider a sample of size 20 with two DLs and $(p_1,p_2) = (0.5,0.6)$. Then, the expected number of non-detects is 11 and the expected number of detected observations is 9 in the sample. The power values in such a scenario will be larger compared with that based on an uncensored sample of size 9. This should be clear by comparing the power values in Table 1 corresponding to an uncensored sample of size 10 (we did not compute them for a sample of size 9). We notice that an uncensored sample of size 10 gives a smaller power compared with a sample of size 20 with two DLs and $(p_1,p_2) = (0.5,0.6)$. The conclusion is that data missing at random due to the presence of DLs is more informative than data missing completely at random.

SAMPLE-SIZE CALCULATION

We shall now address sample-size determination in the context of testing the hypotheses in equation (3). The problem is of practical relevance since it is important to know if increased sample size is necessary due to the presence of DLs and if so by how much. We first note that although the pivotal test and the SLRT are comparable in terms of type I error rates and powers, the pivotal test has an edge over the SLRT because the former does not require additional computation of the constrained MLE of σ^2. Therefore, we shall describe the sample-size calculation only for the pivotal test. As noted earlier, the power of the pivotal test depends only on δ^* and not on the individual values that make up δ^* or ξ_p^0. Thus, it is enough to calculate the sample sizes for various values of δ^* and the proportions of non-detects.

In Table 2, sample sizes are given when one, two, or three DLs are anticipated in the sample. These sample sizes were calculated to attain a power close to 0.80 and 0.90 for testing if the 95th percentile of an exposure distribution is less than a specified value, when the test is carried out at a level of 5%. As the sample size for the censored case must be larger than the one for the uncensored case, we first computed the sample size required for the uncensored case using equation (6). Using this sample size as a starting value, we found the required sample size for the censored case by trial-error method; more details with an illustrative example are given in the next section. We observe from Table 2 that the sample size to attain the specified power increases with increasing proportion(s) of non-detects and decreases with increasing scaled difference δ^*.

Table 2. Sample sizes for the pivotal test to attain a specified power for testing $H_0 : \xi_{0.95} \geq \xi^0_{0.95}$ versus $H_\alpha : \xi_{0.95} < \xi^0_{0.95}$ at the 5% significance level.

$\xi^0_{0.95} = 1.5$; single detection limit

	Power = 0.80						Power = 0.90					
	δ^*						δ^*					
p_1	0.5	0.75	1	1.25	1.5	2.0	0.5	0.75	1	1.25	1.5	2.0
0	73	37	23	16	13	9	99	49	30	21	16	11
0.2	83	41	27	19	15	10	109	54	34	25	19	13
0.4	91	47	30	22	19	12	123	61	39	28	22	15
0.6	105	56	38	27	22	17	138	70	47	34	27	20
0.7	112	61	41	31	25	19	149	78	53	39	31	23
0.8	121	67	46	36	29	22	159	86	58	45	38	27
(p_1, p_2)	0.5	0.75	1	1.25	1.5	2.0	0.5	0.75	1	1.25	1.5	2.0
(0,0)	73	37	23	16	13	9	97	49	30	21	16	11
(0.2,0.3)	83	41	27	20	15	11	110	56	34	25	20	14
(0.3,0.4)	87	45	29	21	16	12	117	61	38	27	21	15
(0.3,0.5)	92	48	30	23	17	13	121	62	41	29	22	15
(0.5,0.6)	102	53	35	26	21	15	130	68	45	33	23	18
(0.7,0.8)	115	63	43	32	25	19	150	78	53	40	33	24
(p_1, p_2, p_3)	0.5	0.75	1	1.25	1.5	2.0	0.5	0.75	1	1.25	1.5	2.0
(0,0,0)	73	37	23	16	13	9	97	49	30	21	16	11
(0.1,0.2,0.3)	80	41	26	20	15	10	110	55	35	24	19	13
(0.3,0.4,0.5)	92	45	31	22	18	13	120	62	39	28	22	16
(0.4,0.5,0.6)	93	50	33	25	20	14	127	65	42	31	25	17
(0.5,0.6,0.7)	102	55	36	28	21	15	137	69	47	34	27	19
(0.6,0.7,0.8)	108	58	40	30	25	17	145	76	51	37	30	22

$\xi_{0.95}$, true unknown 95th percentile to be tested; $\xi^0_{0.95}$, specified value of the 95th percentile; p_i, proportion of non-detects below DL_i; $\ln \mathrm{DL}_i = \mu + z_{p_i}\sigma$.

Finally, an important observation from Table 2 is that the sample size is mainly affected by the overall percentage of non-detects and not by the number of DLs. For instance, we see from Table 2 that when $(p_1, p_2, p_3) = (0.3, 0.4, 0.5)$ so that the average proportion of non-detects is 0.4, the required sample sizes to attain a power of 0.80 (or 0.90) are approximately the same as those corresponding to a single DL and proportion of non-detects $p_1 = 0.4$; see the values in Table 2 under 'single DL', $p_1 = 0.4$ and power = 0.80 (or 0.90). These sample sizes are also approximately the same as the ones for the case of two DLs with $(p_1, p_2) = (0.3, 0.5)$ so that the average proportion of non-detects is once again 0.4. Similarly, by comparing the sample sizes for $(p_1, p_2, p_3) = (0.6, 0.7, 0.8)$ with the corresponding ones for $p_1 = 0.7$ (in single DL case), we see that the sample sizes to attain a specified power is mainly affected by the overall percentage of non-detects. This finding suggests that information on the overall percentage of non-detects is enough to determine the sample size to attain a specified power; see also the illustrative example in the next section.

In view of the above findings, we calculated sample sizes required to attain powers 0.80 and 0.90 for various values of scaled difference δ^* ranging from 0.5 to 2.2 and overall proportion of non-detects denoted by p_a, ranging from 0 to 0.85. Note that the uncensored case corresponds to $p_a = 0$. These sample sizes are reported in Table 3 for power 0.80 and in Table 4 for power 0.90. As an illustration, suppose it is desired to determine the sample size to attain a power of 0.80 when $\delta^* = 1.1$ and the overall proportion of non-detects is 0.45; the required sample size is 28. If no non-detect is anticipated, then the required sample size is 20.

Table 3. Sample sizes for the pivotal test to attain a power of 0.80 for testing $H_0 : \xi_{0.95} \geq \xi^0_{0.95}$ versus $H_\alpha : \xi_{0.95} < \xi^0_{0.95}$ at the 5% significance level.

p_a	$\delta^* = (\xi_p - \xi^0_p)/\sigma$															
	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.9	2	2.2
0	73	54	41	33	28	23	20	18	16	14	13	12	11	10	9	8
0.100	77	56	43	36	28	24	22	19	17	16	13	13	12	10	10	9
0.150	80	58	45	36	30	25	23	19	18	16	14	13	12	11	10	9
0.175	80	59	46	38	30	26	23	20	19	16	15	13	13	11	10	9
0.200	83	59	46	38	32	27	23	20	19	17	15	14	13	11	10	10
0.250	84	61	48	40	32	28	24	20	19	18	16	14	13	12	11	10
0.275	86	63	48	40	33	29	25	21	19	18	16	14	14	12	11	10
0.300	86	64	48	41	34	29	25	21	19	18	17	15	14	13	12	10
0.325	88	64	51	41	34	30	25	23	20	18	17	15	14	13	12	11
0.350	90	65	52	42	34	30	25	23	21	19	17	16	15	13	12	11
0.375	90	66	53	42	35	30	26	23	21	19	17	16	15	13	12	11
0.400	91	68	54	43	35	30	27	24	21	19	17	16	15	14	12	11
0.425	94	69	54	44	37	32	28	24	21	20	19	17	16	14	13	12
0.450	95	70	55	44	37	32	28	25	22	20	19	18	16	14	13	12
0.475	96	72	56	45	38	33	29	25	23	21	19	18	17	14	14	13
0.500	97	74	56	47	39	33	29	26	23	21	20	18	17	15	15	13
0.525	100	77	57	47	39	33	29	27	24	22	20	19	18	16	15	13
0.550	101	77	60	48	41	35	31	27	25	23	22	19	18	16	15	14
0.575	103	77	62	49	42	35	31	29	26	24	22	20	19	16	15	14
0.600	105	77	62	50	42	38	33	29	26	24	22	20	19	17	17	15
0.625	107	79	62	51	44	38	33	30	26	24	23	21	20	18	18	16
0.650	109	80	64	54	44	38	35	31	28	25	23	23	21	18	18	16
0.675	111	84	65	55	46	41	35	33	28	26	25	23	22	19	18	16
0.700	112	84	66	55	47	41	37	34	30	28	25	24	22	19	19	17
0.725	114	84	67	57	48	43	37	34	30	29	27	24	24	21	19	18
0.750	117	87	69	57	51	44	37	36	32	30	27	26	24	22	20	19
0.775	119	93	71	60	51	46	40	36	33	31	29	27	25	22	21	20
0.800	121	93	72	62	54	46	42	40	35	33	29	28	26	23	22	20
0.825	122	95	76	65	57	50	45	41	36	34	31	29	27	23	23	21
0.850	128	100	82	70	60	53	50	45	42	35	34	32	30	24	23	21

$\xi_{0.95}$, true unknown 95th percentile to be tested; $\xi^0_{0.95}$, specified value of the 95th percentile; p_a, overall proportion of non-detects.

Remark 2. Obviously, the power should decrease as the percentage of non-detects increases. Specifically, the power of the test based on a sample of size $n = 30$ with proportion of non-detects $p = 0.20$ should be larger than the power based on $n = 30$ and $p = 0.50$. However, the relationship between the power and the censoring intensity is not obvious. For example, we observe from Table 3 that an uncensored sample of size $n = 9$ is required to attain a power of 0.80 when $\delta^* = 2$. If a sample is expected to include about 85% non-detects, then for the same $\delta^* = 2$, a sample of size 23 is required to attain the power of 0.80. Note that a sample of size 23 with $p = 0.85$ is expected to include only $(1 - 0.85) \times 23 \simeq 3$ detected observations, not 9. This implies that our testing method utilizes the information provided by the values of the DLs and the number of non-detects.

AN ILLUSTRATIVE EXAMPLE

We shall now illustrate the sample-size methodology that we have developed by considering the assessment of carbon monoxide levels at a

Table 4. Sample sizes for the pivotal test to attain a power of 0.90 for testing $H_0 : \xi_{0.95} \geq \xi^0_{0.95}$ versus $H_\alpha : \xi_{0.95} < \xi^0_{0.95}$ at the 5% significance level.

p_a	$\delta^* = (\xi_p - \xi^0_p)/\sigma$															
	0.5	0.6	0.7	0.8	0.9	1	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.9	2	2.2
0	99	71	55	44	36	30	26	23	20	18	17	15	14	12	11	10
0.100	104	76	59	47	38	33	28	24	22	20	18	16	15	13	12	11
0.150	108	78	60	48	39	34	29	26	22	20	18	16	16	14	13	11
0.175	108	79	60	48	40	34	29	26	23	20	19	17	16	14	13	12
0.200	109	80	62	49	41	35	30	26	23	21	19	18	16	14	13	12
0.250	114	82	63	51	42	35	30	27	24	21	20	18	16	14	14	12
0.275	115	84	64	51	43	36	32	28	24	23	20	19	17	15	14	12
0.300	116	85	65	52	44	37	32	28	24	23	20	19	18	15	14	13
0.325	117	86	67	53	44	38	32	28	26	23	21	19	18	16	15	13
0.350	119	87	68	54	45	38	33	29	26	24	21	20	19	16	15	13
0.375	120	88	68	55	46	39	34	30	27	24	22	20	19	16	15	14
0.400	123	90	69	56	46	40	34	30	27	25	22	20	19	16	15	14
0.425	125	91	70	57	48	40	35	32	28	25	24	21	19	17	16	15
0.450	127	93	72	58	48	41	35	32	28	26	24	22	20	18	17	15
0.475	129	94	73	59	49	42	37	32	29	27	24	23	20	18	17	15
0.500	130	96	74	60	50	43	38	33	29	27	24	23	22	19	18	16
0.525	133	97	76	61	52	44	38	34	30	27	26	23	22	19	18	16
0.550	135	99	77	63	52	45	39	35	31	28	26	24	22	20	18	16
0.575	137	100	80	63	54	45	40	36	32	28	26	24	23	20	19	17
0.600	139	103	81	65	55	47	41	37	33	31	27	25	23	21	20	18
0.625	142	105	82	66	56	48	42	37	33	31	29	26	25	22	21	18
0.650	144	107	83	68	57	49	43	38	35	32	29	28	25	22	21	19
0.675	145	108	85	70	58	50	44	40	35	33	30	28	26	24	22	20
0.700	148	111	87	72	60	52	45	42	37	35	31	29	28	25	23	21
0.725	151	114	88	73	62	54	47	43	38	36	33	31	29	25	24	22
0.750	153	115	91	75	63	54	49	44	39	36	34	32	30	26	26	23
0.775	156	117	93	77	65	58	51	46	43	38	36	34	31	29	26	24
0.800	159	119	97	79	67	60	53	49	45	42	38	35	34	31	27	26
0.825	161	123	98	82	72	63	59	53	49	45	42	41	36	34	32	26
0.850	176	141	116	99	90	73	71	70	63	62	62	61	53	52	45	27

$\xi_{0.95}$, true unknown 95th percentile to be tested; $\xi^0_{0.95}$, specified value of the 95th percentile; p_a, overall proportion of non-detects.

workplace. The Occupational Safety and Health Administration permissible exposure limit for carbon monoxide is 50 p.p.m. It is desired to test if the 95th percentile of the contaminant distribution is indeed less than 50 p.p.m. In the notations of this study, the hypotheses of interest are

$$H_0 : \xi_{0.95} \geq \xi^0_{0.95} \text{ versus } H_\alpha : \xi_{0.95} < \xi^0_{0.95}, \quad (15)$$

where $\xi_{0.95}$ is the true log-transformed 95th percentile of the contaminant distribution in the workplace and $\xi^0_{0.95} = \ln(50) = 3.912$. An industrial hygienist, based on his preliminary inspection, guesstimates that the 95th percentile of the concentration distribution in his workplace is around 20 p.p.m.; that is, $\xi_{0.95} = \ln(20)$. It is desired to determine the sample size so that the test would reject H_0 in equation (15) with probability (power) 0.90, when the true 95th percentile is 20 p.p.m. On the basis of past data or based on data from a similar workplace, suppose the value of the geometric standard deviation is guessed to be 2.015 p.p.m. Then from equation (7),

$$\delta^* = \frac{\xi_p^0 - \xi_p}{\sigma} = \left(\ln(50) - \ln(20)\right)/\ln(2.015) = 1.309.$$

Since we are computing the power at the value $\xi_{0.95} = \ln(20)$, the value of μ is $\mu = \ln(20) - z_{0.95}\sigma = \ln(20) - 1.645 \times \ln(2.015) = 1.8485$. In other words, the power is to be calculated at the parameter values $\xi_{0.95} = \ln(20)$, $\sigma = \ln(2.015)$, or equivalently, $\mu = 1.8485$, $\sigma = \ln(2.015) = 0.701$.

For this particular example, if no non-detect is expected, then the required sample size n is determined by the equation (6):

$$P\left(t_{n-1}\left(2.954\sqrt{n}\right) > t_{n-1;.95}\left(1.645\sqrt{n}\right)\right) = .090.$$

The power at $n = 18$ is 0.867; at $n = 19$, it is 0.887; and at $n = 20$, it is 0.904. So if no non-detect is expected, then the required sample size to attain the power of 0.90 is 20.

Suppose the samples will be analyzed by three laboratory methods or devices with the DLs $(\ln DL_1, \ln DL_2, \ln DL_3) = (1.3, 1.5, 1.7)$. Then, the sample sizes and powers can be calculated using Algorithm 4 as shown below.

1. Assume a value for n that is at least 20, say 21. Recall that $\sigma = 0.701$, $\mu = 1.8485$, $z_{0.95} = 1.645$, and $\delta^* = 1.309$.
2. Use Algorithm 4 with
$$DL_1^* = \frac{\ln DL_1 - \mu}{\sigma} = \frac{1.3 - 1.844}{0.701} = -0.7776,$$
$$DL_2^* = \frac{\ln 1.5 - 1.844}{0.701} = 0.4919, \text{ and}$$
$$DL_3^* = \frac{\ln 1.7 - 1.844}{0.701} = -0.2062 \text{ to compute}$$
the power.
3. Repeat step 2 by increasing sample size by 1 unit at a time, until the power becomes close to 0.90.

Our calculation based on Algorithm 4 with 10 000 simulation runs produced the results as given in Table 5. We see from Table 5 that a sample of size 25 would be enough to attain a power of 0.90. An R program for calculating sample size is posted at www.ucs.louisiana.edu/~kxk4695.

The sample size for the above problem can also be obtained from Table 4 as follows. Note that proportion of non-detects from the laboratory 1 is $p_1 = \Phi[(1.3 - \mu/\sigma)] = 0.217$, from laboratory 2 is $p_2 = \Phi[(1.5 - \mu/\sigma)] = 0.310$, and from laboratory 3 is $p_3 = \Phi[(1.7 - \mu/\sigma)] = 0.416$. The average of these proportions is 0.314. Referring to Table 4 with $\delta^* = 1.309 \simeq 1.3$ and $p_a = 0.314 \simeq 0.3$, we find the sample size to be 24. If we take $p_a = 0.314 \simeq 0.325$, then the sample size is 26. Thus, we see that a sample of size 25 is required to attain a power close to 0.90.

DISCUSSION

Proper analysis of exposure data is crucial for setting exposure limits and for drawing valid conclusions regarding the exposure levels. In this context, one of the complicating issues is the presence of DLs. Standard results based on large sample theory are usually not applicable here since typical exposure samples are seldom large due to the cost involved in obtaining the data. Thus, accurate small sample methodology is highly desirable. In particular, it is important to determine the required sample size so as to have methodologies that guarantee a specified level of statistical accuracy. This work has been motivated by all these considerations in the context of exposure data where multiple DLs are anticipated.

If the problem of interest concerns an upper percentile of the exposure profile, then our work can be readily extended to any distribution that has one-to-one relation (at least approximately) with the normal distribution. For example, if the exposure data follow a gamma distribution, our method can be applied after taking cube root transformation of the data (see Krishnamoorthy and Xu, 2011). If an upper

Table 5. Powers of the pivotal test.

	$(\ln DL_1, \ln DL_2, \ln DL_3) = (1.3, 1.5, 1.7)$, $\sigma = 0.701$				
n	21	22	23	24	25
$Q^*_{.95;.95}$	2.518	2.488	2.464	2.439	2.415
Power	0.845	0.862	0.877	0.897	0.907

percentile is below an OEL, then we can conclude that majority of the exposures are below the OEL. Consequently, it is of interest to test if the upper percentile is below the OEL. It is in this context that we have investigated the sample size issue when single or multiple DLs are present. We have developed the necessary algorithm for the sample-size determination and have also provided tables of the sample sizes. Our investigation shows that the required sample size is not affected by the number of DLs; rather, it is affected by the overall proportion of the data below the DLs.

In spite of the extensive industrial hygiene literature on the DL problem, satisfactory confidence intervals and test procedures have been lacking; in fact, it appears that sample size determination has not been addressed at all. It is hoped that this present study will be beneficial to industrial hygienists in terms of drawing their attention to accurate data analysis strategies and in terms of providing sample-size guidelines, when non-detects are to be expected due to the presence of single or multiple DLs.

FUNDING

National Institute for Occupational Safety and Health (R01 OH003628).

Acknowledgement—We are grateful to two reviewers for providing useful comments and suggestions.

REFERENCES

Chu H, Nie L, Cole SR. (2006) Sample size and statistical power assessing the effect of interventions in the context of mixture distributions with detection limits. Stat Med; 25: 2647–57.

Helsel D. (2010) Much ado about next to nothing: incorporating nondetects in science. Ann Occup Hyg; 54: 257–62.

Helsel D. (2012) Statistics for censored environmental data using Minitab and R. 2nd edn. New York: John Wiley.

Ignacio JS, Bullock WH, editors. (2006) A strategy for assessing and managing occupational exposures. 3rd edn. Fairfax, VA: American Industrial Hygiene Association Press.

Krishnamoorthy K, Mallick A, Mathew T. (2009) Model-based imputation approach for data analysis in the presence of non-detects. Ann Occup Hyg; 53: 249–63.

Krishnamoorthy K, Mallick A, Mathew T. (2011) Inference for the lognormal mean and quantiles based on samples with left and right type I censoring. Technometrics; 53: 72–83.

Krishnamoorthy K, Mathew T. (2009) Statistical tolerance regions: theory, applications and computation. Hoboken, NJ: Wiley.

Krishnamoorthy K, Xu Z. (2011) Confidence limits for lognormal percentiles and for lognormal mean based on samples with multiple detection limits. Ann Occup Hyg; 55: 495–509.

Mulhausen JR, Damiano J, editors. (1998) A strategy for assessing and managing occupational exposures. 2nd edn. Fairfax, VA: American Industrial Hygiene Association Press.

Ogden TL. (2010) Handling results below the level of detection. Ann Occup Hyg; 54: 255–6.