Environmental and Ecological Statistics **3**, 81–97 (1996)

Difficulties with the lognormal model in mean estimation and testing

R. L. SCHMOYER[1], J. J. BEAUCHAMP[1], C. C. BRANDT[1] and F. O. HOFFMAN, JR[2]

[1]*Oak Ridge National Laboratory, Computer Science and Mathematics Division, P.O. Box 2008, Oak Ridge, TN 37831-6367, USA*

[2]*Senes Oak Ridge Center for Risk Analysis, Inc., 677 Emory Valley Road, Oak Ridge, TN 37831, USA*

Received November 1994. Revised August 1995

A frequent assumption in environmental risk assessment is that the underlying distribution of an analyte concentration is lognormal. However, the distribution of a random variable whose log has a t-distribution has infinite mean. Because of the proximity of the standard normal and t-distribution, this suggests that a distribution such as the gamma or truncated normal, with smaller right tail probabilities, might make a better statistical model for mean estimation than the lognormal. In order to assess the effect of departures from lognormality on lognormal-based statistics, we simulated complete lognormal, truncated normal, and gamma data for various sample sizes and coefficients of variation. In these cases, departures from lognormality were not easily detected with the Shapiro–Wilk test. Various lognormal-based estimates and tests were compared with alternate methods based on the ordinary sample mean and standard error. The examples were also considered in the presence of random left censoring with the mean and standard error of the product limit estimate replacing the ordinary sample mean and standard error. The results suggest that in the estimation of or tests about a mean, if the assumption of lognormality is at all suspect, then lognormal-based approaches may not be as good as the alternative methods.

Keywords: gamma distribution, left censoring, product limit estimate, risk assessment, truncated normal distribution

1. Introduction

The purpose of this paper is to demonstrate that in estimating or testing hypotheses about a mean, (i) the lognormal distribution may be too heavy tailed to be a reasonable statistical model, and (ii) alternatives may be better than lognormal-based methods. The focus on mean estimation stems partly from US Environmental Protection Agency guidance (EPA, 1989, Section 6.4), which specifies that risks (or hazards) should be characterized in terms of mean contaminant concentrations. Because non-parametric methods tend to focus on percentiles rather than means, the statistical methods used in these problems tend to be parametric, premised, for example, on normality or lognormality. Furthermore, because of the expense of laboratory analysis, as well as administration, field sampling, data validation, etc., sample sizes are often small, sometimes less than 5. When the sample size is small, there are only a few significance levels at which exact, non-randomized non-parametric tests can be run, and the smallest level is usually too large. This hinders non-parametric approaches, but also casts doubt on the adequacy of large sample approximations used in the parametric approaches.

Lognormality is probably assumed more often than normality for the distribution of analyte concentrations, in part because concentrations cannot be negative. Also, if concentrations are products of many slight dilutions, they will be approximately lognormal (see, for example, Ott, 1990, 1995). But generally many physical processes work to shape a contaminant distribution, and measurement errors affect distributions of actual observed variables, adding what is likely to be normal (rather than lognormal) components to the mix. Thus, there is usually no physical basis for lognormality, normality, or any other distribution.

Furthermore, although large sample and central limit approximations apply to the log-scale sample mean and variance, when the underlying distribution is not lognormal, lognormal-based estimates of the (arithmetic) mean will generally be biased–even asymptotically. The ordinary sample mean, on the other hand, is unbiased, so the central limit theorem applies to it directly. This suggests that unless there is a physical basis for lognormality, the ordinary sample means and t-tests may be better than lognormal-based statistics for mean estimation and testing.

Perhaps the most basic problem in risk assessment is to test

$$H_0 : \mu \geq \text{AL} \text{ against } H_1 : \mu < \text{AL} \tag{1}$$

where AL denotes an action level, and μ, a true mean concentration of a particular analyte. If, at a certain site, this hypothesis is rejected at level α, one can be confident at level 1-α that the mean analyte concentration at the site is below AL. This hypothesis is particularly appropriate for sites at which contamination is suspected. We will take testing it to be the basic problem of this paper.

As prescribed by EPA (1989, Section 6.4.1), the test should be performed by rejecting H_0 if and only if a 95% upper confidence bound (UCB) for μ is less than AL. Thus the test is performed at the 0.05 level of significance. The incorrect rejection of H_0 could lead to human health or environmental loss due to exposure to contamination; incorrect acceptance could lead to unnecessary costs of remediation, institutional controls, etc. Implicitly, the AL should be low enough that the 0.05 level is reasonable. If the AL is believed to be a truly dangerous level, then the specified significance level should be much lower than 0.05.

For example, the following concentrations of caesium-137 (pCi/g) were measured in bluegills taken from Popular Creek, an environmentally impacted tributary of the Clinch River in Tennessee (DOE, 1995):

0.0042, 0.0319, 0.0463, 0.0622, 0.0688* 0.0726, 0.0754, 0.0865, 0.0950, 0.1050* and 0.1089*.

Here the * denotes 'non-detect' (left-censoring) and the value is then the detection limit. If detection limits are substituted for the censored observations (a frequently suggested conservative approach in risk assessment), the sample mean concentration is 0.0688 pCi/g with a 95% UCB of 0.0861 pCi/g. The screening level (10^{-6} risk) AL is 0.062 pCi/g. Therefore, H_0 is not rejected in this case. This example will be used later in the paper to illustrate several other approaches to handling censored data.

In Section 2 of this paper we discuss instabilities of the (true) mean when a lognormal distribution is perturbed. Lognormal means are extremely sensitive to perturbations. In analogy with the lognormal distribution we will say that X has log-$t(k)$ distribution if

$$\log(X) = \nu + \tau T \tag{2}$$

where T has a t-distribution with k degrees of freedom, and ν and $\tau > 0$ are location and scale parameters. If X is lognormal with log-scale mean ν and variance τ^2, then $E(X) = e^{\nu+\tau^2/2}$. But if instead X is log-t (any k) with the same ν and τ, then (we show) $E(X)$ does not even exist.

As we discuss in Section 2, because the right tail of the log-t distribution is so heavy, log-t random

variables are occasionally extremely large. Because of the similarity between the normal and t distributions, the lognormal distribution may itself be too heavy tailed, and departures from lognormality in the direction of less right tail weight—for which means are sure to exist—may occur in practice.

In Sections 3–5 we discuss simulated examples that show lognormal-based statistics might not be as good as the ordinary sample mean and t-test, if there are complete (i.e. uncensored) data, or as good as means or tests computed from the product limit estimate (PLE; Kaplan and Meier, 1958), when there is random left censoring. (The PLE is described in Section 5.) The lognormal-based estimators and tests do not perform as well as the alternative approaches when the underlying distribution is either truncated normal or gamma, which have lighter tails than the lognormal. We simulated lognormal, truncated normal, and gamma data, for sample sizes $n = 5$, 10, and 20, and coefficients of variation (CV) = 0.1, 0.5, 1, 2 and 5, with complete data and with three different random left-censoring schemes. Departures from lognormality were not easily detected in the complete cases with the Shapiro–Wilk test (Shapiro and Wilk, 1965), though detection became more likely as the CV increased.

In the complete-data simulations, we compared the performances of (i) the lognormal uniformly most powerful unbiased (UMPU) test of Land (1971) with the ordinary t-test, for the hypothesis H_0 against H_1; and (ii) the lognormal minimum variance unbiased estimator (MVUE) of Finney (1941) with the ordinary sample mean. Methods based on the direct substitution of the log-scale sample mean and variance were also considered. As optimal (i.e. UMPU or MVUE) lognormal-based procedures are unknown for randomly left-censored data (cf. Cohen, 1991, Chapter 6), in the censored cases, the direct method was compared to PLE-based estimates and tests.

On the basis of an analytical comparison of variances, Shimizu (1988) shows that for complete lognormal data and a range of parameters, the lognormal MVUE is not much more efficient than either the maximum likelihood estimate (MLE) or the ordinary sample mean. Our simulations suggest a similar conclusion, but for distributions in addition to the lognormal, smaller sample sizes, a wider range of CVs, and the direct substitution estimate rather than the MLE. (The direct estimate uses the unbiased log-scale variance estimate, whereas the MLE uses the MLE of the variance.)

Our perspective in interpreting simulation results in this paper is that neither the underlying distribution nor the CV will be known exactly. Because of arguments in Sections 2 and 3, we tend to give distributions like the truncated normal or gamma higher weight than the lognormal, and we give small to moderate CVs less weight than large ones. Of course this interpretation is inappropriate if lognormality is certain or CV is known to be large. (What constitutes a large CV is discussed in Section 3.)

Simulation results are presented in Sections 4 and 5. The simulations show that lognormal-based estimators are often too large. The ordinary sample mean or PLE mean seem to be better than the MVUE or lognormal direct estimate (in terms of mean squared error) when the underlying distribution is truncated normal or gamma, and are nearly as good as the MVUE when the data are lognormal. The lognormal-based estimators are especially bad when the underlying distribution is gamma and the CV is large.

The simulations also show that for very large CVs, both the lognormal-based tests and the t-tests or PLE-based tests can break down. The lognormal-based tests tend to be extremely conservative when the underlying distribution is not lognormal, whereas the t-test or PLE-based tests tend to be extremely anti-conservative. In any particular context, which test is best depends on the losses of Type I or II errors. With this in mind, we believe that, in general, the UMPU and direct tests tend to be too conservative: the t-tests or PLE-based tests are more powerful and yet maintain, approximately, the nominal Type I error rate.

2. Ill-posed nature of the lognormal model

The normal and t-distributions are well known to be similar, especially when k, the t-distribution degrees of freedom, is large. (Chu, 1956, showed that the relative error in using the normal distribution to approximate the t-distribution is uniformly smaller than 1/k.) The t-distribution has tails that are slightly heavier than the normal's, but has moments of all orders up to k-1 (Johnson and Kotz, 1970). Thus, although the t-distribution is most often used to construct tests and confidence intervals, in applications where the normal distribution is a reasonable model for concentrations (or other phenomena), for some sufficiently large k, the t-distribution should also be reasonable. Similarly, if the lognormal distribution is reasonable, then the log-$t(k)$, as defined in (2), should be also.

But the mean of the log-$t(k)$ distribution is infinite. (The mean is

$$\int_{-\infty}^{\infty} e^{\nu+\tau t} \frac{\Gamma((k+1)/2)}{(k\pi)^{1/2}\Gamma(k/2)} \left(1+\frac{t^2}{k}\right)^{-(k+1)/2} dt$$

The integrand approaches infinity as $t \to \infty$. Hence the integral is divergent.) Because of the proximity of (log)normal and (log)t-distributions, estimating a lognormal mean is thus ill-posed.

The log-t distribution has no mean because its right tail is too heavy. This causes log-t random variables to be occasionally very large. The density of the log-$t(k)$ distribution is

$$(\tau x)^{-1} \frac{\Gamma((k+1)/2)}{(k\pi)^{1/2}\Gamma(k/2)} \left(1+\left(\frac{\log(x)-\nu}{\tau}\right)^2/k\right)^{-(k+1)/2}$$

As $x \to \infty$,the log of this density is $O(-\log(x))$. Let f_N be the density of a $N(\nu, \tau^2)$ distribution truncated at 0. Then

$$f_N(x) = \frac{\phi[(x-\nu)/\tau]}{\tau\Phi(\nu/\tau)} \tag{3}$$

where ϕ and Φ are the standard normal density and distribution functions. Let f_G denote the single-parameter gamma density with parameter γ. Then $f_G(x) = x^{\gamma-1}e^{-x}/\Gamma(\gamma)$, for $x > 0$. Also let f_L be the lognormal density, $f_L(x) = (x\tau)^{-1}\phi[(\log(x)-\nu)/\tau]$. As $x \to \infty$, $\log[f_N(x)]$ is $O(-x^2)$, $\log[f_G(x)]$ is $O(-x)$, and $\log[f_L(x)]$ is $O(-(\log(x))^2)$. Thus in terms of tail weight,

log$-t$> lognormal > gamma > truncated normal.

The gamma distribution has been considered for modelling environmental concentrations (Gilbert, 1987), as has the normal and thus the truncated normal. As indicated in the next section, the truncated normal and gamma distributions both have means. All this suggests that at least in some applications, because of its heavy right tail, the lognormal density might not make a good model for mean estimation. Actually, for modelling concentrations, any of the above distributions is too heavy-tailed, because concentrations are bounded from above. That is, there are only so many milligrams of A that will fit into a litre of B or a kilogram of C. This does not imply that each of the distributions is an inadequate model, but it does suggest that distributions with lighter right tails might be better.

3. Simulated examples

In the last section we gave a theoretical reason to suspect that in practice true underlying distributions may depart from the lognormal. Even if they do, however, lognormal-based estimates and

tests might be good, if they are robust to the departures that occur. But in this and the next two sections we demonstrate with simulation results that lognormal-based statistics are not robust. In particular, for truncated normal or gamma distributions, ordinary t-tests and sample means, or their PLE-based analogues, are considerably better.

Without loss of generality, we will make the mean of our lognormal, truncated normal, and gamma distributions equal to 1. We take the CV (or equivalently, since the mean is 1, the standard deviation) to be 0.1, 0.5, 1, 2 and 5. Sample sizes in the simulations are 5, 10 and 20, which reflect the small to moderate sample sizes typical in many environmental risk problems.

By first-order Taylor approximation, the standard deviation of $\log_{10}(X)$ is approximately $\text{CV} \times \log_{10}(\text{e})$. For each specified CV, we thus determine the corresponding approximate log-scale standard deviation as in Table 1.

If four standard deviations represents a normal range, then, on the base-10 log scale, four standard deviation represents the normal range in orders of magnitude. Thus CV = 5 corresponds to a range of about 8.7 orders of magnitude, which in most practical applications would be unrealistically large. Even a CV of 2 is very large. Nevertheless, the table is based on approximation, and the case CV = 5 makes it easy to see trends in power, mean squared error, etc. as a function of CV, so here we take CV = 5 as the upper extreme.

The true underlying CV can be estimated, but when n is small, CV estimates can be unreliable. For example, in 5000 of 10 000 simulated lognormal trials with $n = 10$ and a true CV of 2, the CV estimate (sample standard deviation divided by sample mean) was less than 1.13. In the statistical approaches we consider in our simulations, CV is taken to be unknown.

The lognormal distribution has mean $e^{\nu+\tau^2/2}$, variance $e^{2\nu+\tau^2}(e^{\tau^2}-1)$, and thus $\text{CV} = (e^{\tau^2}-1)^{1/2}$. For a mean of 1 and specified CV, $\tau^2 = \log(1+\text{CV}^2)$, and $\nu = -\tau^2/2$. We can thus determine the lognormal ν and τ for specified CV (and mean 1).

Let Γ_γ denote a gamma random variable with single parameter γ. Both the mean and variance of Γ_γ are γ, and so its CV is $\gamma^{-1/2}$. For a specified CV, we thus take $\gamma = 1/\text{CV}^2$. To make the mean 1, we divide the Γ_γ variate by γ, which does not affect the CV.

If X has the truncated normal density (3), then, by direct integration, the constraint $\text{E}(X) = 1$ reduces to

$$0 = (\theta - \tau^{-1})(2\pi)^{1/2}\Phi(\theta) + e^{-\theta^2/2} \tag{4}$$

where $\theta = \nu/\tau$. Also by direct integration

$$\text{E}(X^2) = \tau^2\left[\theta^2 + \frac{1}{\Phi(\theta)}\left[\left(\frac{2}{\pi}\right)^{1/2}\theta e^{-\theta^2/2} + \frac{1}{2}\left(1 + G_{3/2}\left(\frac{\theta^2}{2}\right)\right)\right]\right]$$

where $G_{3/2}$ denotes the cumulative distribution function of a gamma random variable with

Table 1. Approximate standard deviations for various CV's

CV *of* X	*Approximate standard deviation of* $\log_{10}(X)$	*Four standard deviations* $\log_{10}(X)$ *(approximate)*
0.1	0.043	0.17
0.5	0.217	0.87
1.0	0.434	1.7
2.0	0.869	3.5
5.0	2.17	8.7

parameter 3/2. By integration by parts, $G_{3/2}(\theta^2/2) = -(2/\pi)^{1/2}\theta e^{-\theta^2/2} + 2\Phi(\theta) - 1$, and

$$E(X^2) = \tau^2\left[1 + \theta^2 + \frac{\theta\phi(\theta)}{\Phi(\theta)}\right] \tag{5}$$

Notice for $E(X) = 1$, that $CV = (E(X^2) - 1)^{1/2}$. Subject to (4), θ is a function of τ. For any given τ, θ can be determined using (4) and, for example, Newton's method. Then, for that τ and that θ, (5) can be evaluated to determine CV. Using bisection, τ and thus θ can be determined for any specified CV. In this way we computed the truncated normal τ and θ (and $\nu = \theta\tau$) for each CV.

Gamma, truncated normal and lognormal densities are illustrated in Fig. 1. For CV = 0.1, the densities are all very similar, and all are close to N(1, 0.01). This is clear for the truncated normal, because the variance is so small, and for the lognormal, because $X \approx \log(X) + 1$ when X is close to 1 (or see Shimizu and Crow, 1988, for a much stronger approximation). For the Γ_γ distribution, the similarity can be seen by considering characteristic functions: The characteristic function of the Γ_γ distribution scaled by $1/\gamma$ is $(1 - it/\gamma)^{-\gamma}$, which behaves like $\exp(it - t^2/(2\gamma))$ as $\gamma \to \infty$, or equivalently, like $\exp(it - t^2CV^2/2)$ as $CV = \gamma^{-1/2} \to 0$. The characteristic function of the $N(1, \tau^2)$ distribution is $\exp(it - t^2\tau^2/2)$, or equivalently, $\exp(it - t^2CV^2/2)$.

4. Complete-data simulation results

We simulated 10 000 sets of data for each of the sample sizes, CVs, and distributions given in Section 3. For each simulated data set, we computed (i) level 0.01, 0.05, and 0.10 Shapiro–Wilk tests for lognormality; (ii) the ordinary t-test of H_0 ($\mu \geq AL$) against the alternative H_1, and, for the same hypotheses, the UMPU test of Land (1971); and (iii) the ordinary sample mean and the lognormal MVUE of Finney (1941).

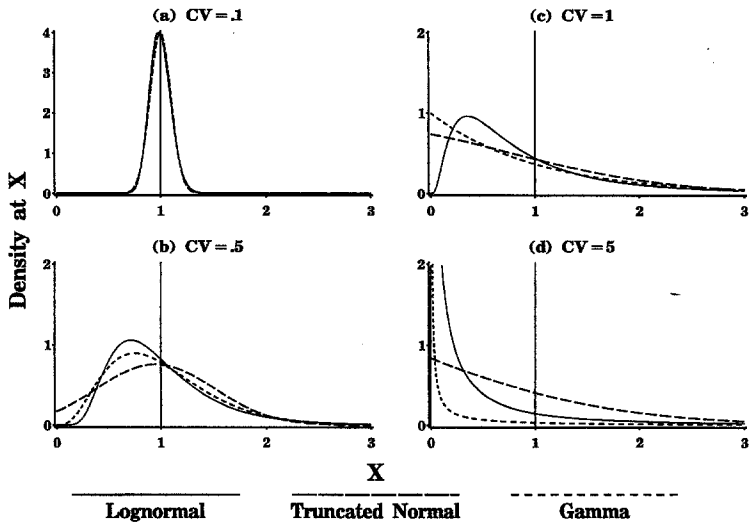

Figure 1. Three densities with mean = 1 and various CVs. The lognormal, truncated normal, and gamma densities with CV = 0.1, 0.5, 1, 2 and 5 are used in the simulations to compare lognormal-based methods with the ordinary sample mean and t-test. The CV = 2 densities are not shown, but can be inferred approximately by interpolation.

The UMPU test requires numerical methods to compute, which we accomplished using a Fortran program written by Land, Greenberg, Hall, and Drzyzgula (obtained from Charles E. Land, Radiation Epidemiology Branch, National Cancer Institute, EPN 408, Bethesda, MD 20892). The MVUE calculations, the Shapiro–Wilk tests, and all other calculations besides the UMPU tests were done in SAS.

The Shapiro–Wilk test results are in Table 2. Standard errors of these results range from 0.0008 (for $\hat{p} = 0.007$) to 0.005 (for $\hat{p} = 0.5023$). But in spite of the sampling error, it is clear that (i) for $n = 5$, the departures from lognormality are virtually undetectable, at least with the Shapiro–Wilk test (for example, for the gamma distribution $n = 5$ and CV = 5, lognormality was not rejected more than 75% of the time, even at the 0.10 significance level); and (ii) for larger n, the detection probabilities increase, but in all cases there are appreciable probabilities of not detecting the departure. And yet, as we will show, there are appreciable differences in the performance of lognormal-based and alternative inferential procedures under these distributions.

The criteria for evaluating the tests of H_0 were size (i.e. maximum null-hypothesis rejection probability) and power as a function of AL. The primary criterion for the point estimates was mean squared error (MSE), but mean absolute error (MAE) was also considered. Statistics were tallied at the end of each set of 10 000 trials to arrive at overall assessments of performance. MSEs and MAEs were calculated along with their standard errors. Because squared errors for different estimates computed from the same data are correlated, paired differences between the squared errors for the ordinary sample mean and MVUE were also calculated. In a few cases, additional simulation trials were performed to reduce standard errors, so that the sampling errors for MSE differences would be negligible. Analysis of binomial sampling errors and the observed proportions of correct decisions made by each particular test (in the 10 000 trials) indicates that the sampling errors for test power are also negligible.

The MVUE is $\exp(\hat{\nu})g(\hat{\tau}^2/2)$, where $\hat{\nu}$ and $\hat{\tau}^2$ are the log-scale sample mean and unbiased variance estimate, and

$$g(t) = 1 + \frac{n-1}{n}t + \frac{(n-1)^3}{n^2 2!}\frac{t^2}{n+1} + \frac{(n-1)^5}{n^3 3!}\frac{t^3}{(n+1)(n+3)} + \cdots$$

(Finney, 1941, also expresses g in terms of a Bessel function.) Because it requires numerical methods, Land's UMPU test is more difficult to describe than any of the other procedures discussed here. We therefore refer the reader to Land (1971) for a complete discussion of that test.

Table 2. Proportions of Shapiro–Wilk test rejections in 10 000 simulation trials.

		$n = 5$			$n = 10$			$n = 20$		
		Test size			*Test size*			*Test size*		
CV	*Distribution*	*0.01*	*0.05*	*0.10*	*0.01*	*0.05*	*0.10*	*0.01*	*0.05*	*0.10*
0.1	Gamma	0.0070	0.0364	0.0871	0.0100	0.0456	0.0954	0.0120	0.0530	0.1031
	Truncated normal	0.0079	0.0396	0.0940	0.0125	0.0528	0.1120	0.0187	0.0755	0.1345
0.5	Gamma	0.0088	0.0422	0.0980	0.0174	0.0734	0.1374	0.0439	0.1241	0.1975
	Truncated normal	0.0251	0.0910	0.1712	0.1347	0.2683	0.3650	0.4014	0.5788	0.6571
1.0	Gamma	0.0132	0.0614	0.1264	0.0590	0.1453	0.2265	0.1644	0.3148	0.4179
	Truncated normal	0.0180	0.0733	0.1431	0.0828	0.2003	0.2920	0.2579	0.4533	0.5615
2.0	Gamma	0.0303	0.1019	0.1862	0.1536	0.3146	0.4302	0.4476	0.6614	0.7492
	Truncated normal	0.0171	0.0669	0.1319	0.0700	0.1828	0.2847	0.2313	0.4464	0.5231
5.0	Gamma	0.0420	0.1345	0.2363	0.2169	0.4309	0.5591	0.6277	0.8291	0.8950
	Truncated normal	0.0183	0.0653	0.1343	0.0674	0.1798	0.2650	0.2141	0.3979	0.5023

In addition to the ordinary sample mean, MVUE, t-test and UMPU test, we also computed direct-substitution point estimates and tests in the simulations. The direct estimates are obtained by substituting $\hat{\nu}$ and $\hat{\tau}^2$ into the expression $\exp(\nu + \tau^2/2)$ for the mean. The direct test, originally due to D. R. Cox, is discussed (and referred to as 'direct') by Land (1972). The MVUE of the variance of $\hat{\nu} + \hat{\tau}^2/2$, it turns out is, $\hat{\tau}^2/n + \hat{\tau}^4/(2(n+1))$, and an approximate $1-\alpha$ UCB for the mean is thus $\exp(\hat{\nu} + \hat{\tau}^2/2 + t^{\alpha}_{n-1}[\hat{\tau}^2/n + \hat{\tau}^4/(2(n+1))]^{1/2})$,where t^{α}_{n-1} is the α quantile of the t-distribution with $n-1$ degrees of freedom. The direct estimates and tests are easy to compute, which made them more attractive before electronic computers than they are today. In simulations, Land found the direct test to be the best alternative to the UMPU test.

Because the direct estimates are biased, even under lognormality, the MVUE can be regarded as a refinement of them. Similarly, Land's UMPU tests are a refinement of the direct tests. Thus, for complete lognormal data, the direct methods are inferior to the MVUE and UMPU approaches. For data other than lognormal, however, it is possible the direct approaches could be better. Thus we also consider the direct estimates and tests.

Results for the tests are presented as power curves in Fig. 2. These are for the case $n = 10$. The results for $n = 5$ and 20 are similar, the differences between the two power curves being bigger for $n = 5$ and smaller for $n = 20$. In view of Fig 1a, it is not surprising that for CV = 0.1, the behaviour

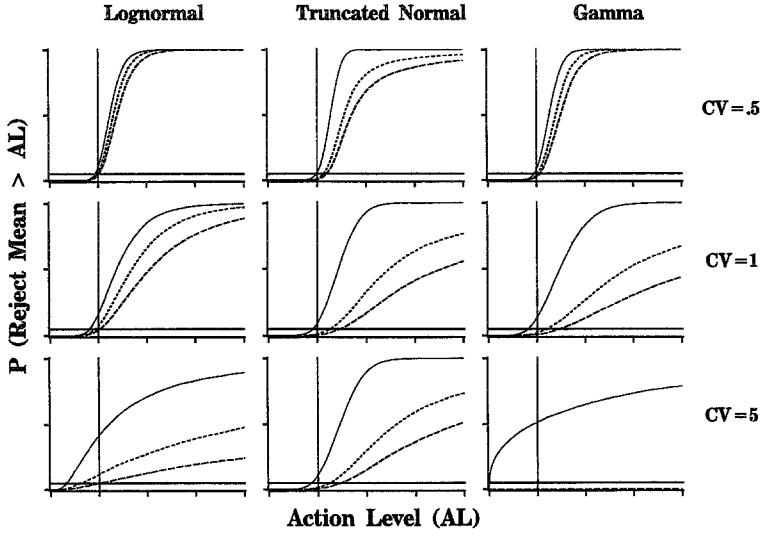

Figure 2. Power curves for $n = 10$, complete data (for 10 000 simulation trials). The curves give power estimates as a function of AL for the t-test (solid line), the UMPU test (long dashes), and the direct test (short dashes). Each horizontal axis extends from 0 to 4 with reference line at 1 (mean). Each vertical axis extends from 0 to 1 with reference line at 0.05 (nominal α). The power curves for $n = 5$ and 20 are similar; the differences between the UMPU and t-test curves are larger for $n = 5$ and smaller for $n = 20$.

of the tests was essentially the same for all three distributions. For this reason, the CV = 0.1 results are not shown in Fig. 2. The CV = 2 power curves are not illustrated either, but can be inferred approximately by interpolation.

For CV > 0.1, the UMPU test was conservative (actual size less than nominal size) except when the distribution was lognormal, whereas the t-test was anti-conservative, appreciably for the gamma and lognormal distributions with CV = 5. On the other hand the UMPU test was much less

powerful than the t-test of the same nominal size, even when the distribution was lognormal. The direct test power curve was always between the other two, but tended to be closer to the curve of the UMPU test.

In any particular context, which test is best depends on the losses of type I or II. Size must be balanced against power, which saves resources. Since the nominal test size here, 0.05, is presumed adequate, an actual size of 0.10 or 0.15 may not be too serious. (It would be if the nominal size was 0.001, for example.) Larger than nominal sizes would also be more acceptable if the rejection probability is very small when, say, AL = 1/2.

In interpreting the results, we assume that neither the underlying distribution nor the CV is known exactly. Because of the arguments in Section 2 and 3, we tend to give the truncated normal and gamma results higher weight than the lognormal, and a CV of 0.5 or 1 more weight than a CV of 5. Of course this interpretation is inappropriate if lognormality is certain or CV is known to be large. With this in mind, we believe that the UMPU and direct tests tend to be too conservative; t-tests are better. For example, for $n = 10$ and CV = 1, see Table 3. The size of the t-test extends to 0.166 in the lognormal case, but the power of the UMPU test at AL = 2 is less than 0.5 in the lognormal case, and less than 0.2 for the other distributions. The power of the t-test is at least 0.8.

For the gamma distribution with CV = 5, all of the tests break down. For example, for the gamma distribution and $n = 20$, the UMPU and direct tests did not reject at all in 10 000 trials for AL = 1 or AL = 2. The t-test rejected 40% of the time for AL = 1.

Even for the lognormal distribution, if CV < 1, the t-test is only moderately anti-conservative. For example, for the lognormal distribution with $n = 5$, see Table 4. (The power of the t-test actually increases in CV for CV greater than about 2. As CV increases, the lognormal density becomes highly skewed (Fig. 1d), so that nearly all observations–and thus nearly all sample means and standard errors–are extremely small.)

Table 5 contains average MSE differences for the ordinary sample mean and MVUE for complete data. We do not consider direct-estimate results in Table 5, because their performance is so bad. The problem is instability, because $\exp(\hat{\nu} + \hat{\tau}^2/2)$ is occasionally anomalously large. For example, for 10 000 truncated normal trials with $n = 10$ and CV = 0.5, the largest ordinary sample mean was 1.63, the largest MVUE was 14.6, but the largest direct estimate was 96.7. The occasional large direct estimates result in large mean squared errors. For example, for 10 000 gamma trials with $n = 10$ and CV = 1 we got an MSE estimate (std. err.) of 14.8 (7.6) for the direct estimate, 0.48 (0.06) for the MVUE, and 0.098 (0.002) for the ordinary sample mean. In fact, for lognormal data and large CV, the direct estimates seem to be inferior even to the ordinary sample mean: for $n = 10$ and CV = 5 and 100 000 lognormal trials, we got an MSE estimate (std. err.) of 17.3(3.1) for the direct estimate, 2.7 (0.46) for the ordinary sample mean, and 1.18 (0.04) for the MVUE. Only for CV = 0.1 did the direct estimates do nearly as well as the others.

Since the mean here is 1, in interpreting MSEs we considered an absolute error of less than 10^{-2},

Table 3. Size and power estimates for n = 10 and CV = 1

	Size estimate (for AL = 1*)*			*Power estimate for AL* = 2		
Distribution	*t-test*	*Direct test*	UMPU *test*	*t-test*	*Direct test*	UMPU *test*
Lognormal	0.166	0.080	0.049	0.852	0.663	0.475
Truncated normal	0.094	0.013	0.006	0.930	0.329	0.170
Gamma	0.134	0.021	0.008	0.830	0.265	0.129

Table 4. Size and power estimates for the lognormal distribution with n = 5

	Size estimate (for AL = 1*)*			*Power estimate for* AL = 2		
CV	*t-test*	*Direct test*	UMPU *test*	*t-test*	*Direct test*	UMPU *test*
0.1	0.060	0.054	0.049	1.000	1.000	1.000
0.5	0.117	0.075	0.048	0.918	0.783	0.584
1.0	0.193	0.098	0.050	0.714	0.447	0.234
2.0	0.307	0.120	0.050	0.653	0.308	0.127
5.0	0.459	0.145	0.049	0.692	0.254	0.089

or an MSE of less than 10^{-4} to be of no practical importance. Similarly, a difference between two MSEs of less than 10^{-4} was also regarded as unimportant. For CV = 0.1, none of the differences exceeded 5×10^{-7}, and so results for CV = 0.1 are not presented in Table 5.

The MSE differences mostly increase in magnitude as CV increases. For CV $\leq$ 1, when the underlying distribution is lognormal, the MSE of the sample mean is nearly as low as the MSE of the MVUE, but for CV = 5 the differences are appreciable. For example, for lognormal data with $n = 20$, when CV = 5, the MSE is 0.93 for the sample mean and 0.48 for the MVUE, and when CV = 2, the MSE is 0.18 for the sample mean and 0.14 for the MVUE. For the truncated normal or gamma distributions, when CV > 0.1, the ordinary sample mean has smaller MSE than the MVUE (as seen by the negative differences). For the gamma distribution with CV = 5, the MVUE breaks down completely. This is due to the high density near zero, in this case (see Fig. 1d), which in turn translates to left-skewness on the log-scale. Of course very small concentrations do sometimes occur in practice.

Although not shown here, complete-data MAEs were also examined. The differences among the MAEs were not as big as those among the MSEs, and the MAE in several lognormal cases was in

Table 5. Average differences of squared errors, MVUE from sample mean (10 000 simulation trials)[a]

CV	*Distribution*	$n = 5$ *Difference*	$n = 10$ *Difference*	$n = 20$ *Difference*
0.5	Lognormal	2.85×10^{-4}	1.67×10^{-4}	6.35×10^{-5}
	Truncated normal	-1.81×10^{-2}	-5.23×10^{-2}	-3.18×10^{-2}
	Gamma	-1.07×10^{-3}	-9.16×10^{-4}	-5.91×10^{-4}
1.0	Lognormal	9.35×10^{-3}	4.72×10^{-3}	2.17×10^{-3}
	Truncated normal	-2.62×10^{-1}	-3.18×10^{-1}	-1.936×10^{-1}
	Gamma	-3.18×10^{-1}	-3.84×10^{-1}	-2.21×10^{-1}
2.0	Lognormal[b]	1.14×10^{-1}	7.74×10^{-2}	3.39×10^{-2}
	Truncated normal	-1.93×10^{-1}	-2.12×10^{-1}	-1.95×10^{-1}
	Gamma	-4.10×10^{12}	-4.57×10^{14}	-1.76×10^{13}
5.0	Lognormal[c]	1.74×10^{0}	1.49×10^{0}	4.45×10^{-1}
	Truncated normal	-3.98×10^{-1}	-3.07×10^{-1}	-2.67×10^{-1}
	Gamma[d]	–	–	–

[a] All differences significant at level 0.05 or less, except for the gamma distribution with CV = 2 or 5.
[b] Based on 100 000 trials for $n = 5$.
[c] Based on 100 000 trials for $n = 5$ and 10.
[d] Overflow–difference negative and huge.

fact larger for the MVUE than the sample mean, but in general, conclusions based on MAE and MSE were the same. Unless the assumption of lognormality is certain, the ordinary sample mean seems best.

5. Non-detects

Detection limits pose another dimension of complexity in estimating or testing hypotheses about means. With complete data, if n is large or CV is low, then any reasonable method will do well. When there is censoring, however, a statistical method should also account for it. Even with large n or little natural variation, inaccurate accounting for censoring can lead to a poor estimate or test. In this section we consider a few simulated examples like the previous ones–with the same underlying distributions and sample sizes–but with non-detects.

Detection limits for environmental concentrations are affected by instrumentation and calibration, analytical methods (e.g. dilutions), and natural factors (e.g. soil moisture). Sometimes they are fixed, but in our experience, they are more often random. When there is random left censoring, the entire lognormal-based statistical approach is an open question. UMPU tests or MVUEs, for example, have not been generalized to the case of non-detects. A likelihood ratio approach might be taken to test H_0 under lognormality, but perhaps because the null hypothesis is composite in both the log-scale mean and variance, an exact likelihood ratio test that accommodates random left censoring has not yet been derived, as far as we know, and we do not pursue such a test here.

With censoring, direct estimates can be computed by plugging MLEs into the expression, $\nu + \tau^2/2$, and direct tests can be derived, as in the complete case but using Taylor expansions and the asymptotic covariance matrix of the MLEs to get a standard error for the estimate $\hat{\nu} + \hat{\tau}^2/2$. When there are all detects, this approach reduces the direct method except for some n to $n - 1$ adjustments. Of course the direct method is approximate even without censoring.

The product limit estimator (PLE) of Kaplan and Meier (1958) is a censoring-adjusted non-parametric estimator of the cumulative distribution function. For possibly left-censored concentrations, it is defined as follows. Suppose that detects occur at L distinct concentrations, $c_1 < \cdots < c_L$. Let d_j be the number of detects at c_j, and let n_j be the sum of the number of detects or non-detects whose respective value or detection limit is less than or equal to c_j. Then the PLE at x is 1 for $x \geq c_L$, and

$$\prod_{j:c_j > x} \frac{n_j - d_j}{n_j}$$

for $c_1' \leq x < c_L$, where c_1' is the smaller of c_1 and the smallest detection limit of the non-detects. For $0 \leq x \leq c_1'$, the PLE is either 0 or undefined, depending on whether c_1' is a detect or detection limit, but we chose the value 0 in both cases.

The mean of the PLE can be used as a censoring-adjusted mean estimate. Kaplan and Meier derived an approximate standard error of the PLE mean, which can be used for hypothesis tests or confidence bounds for the mean. Kaplan and Meier suggested incorporating the factor $d/(d-1)$, where d is the number of detects, into the standard error, and we also make that adjustment. With that adjustment, the PLE mean and standard error reduce to the usual mean and standard error when there is no censoring.

Consider again the caesium-137 data of Section 1. The PLE mean for this data is 0.0567 pCi/g with a 95% UCB of 0.0758 pCi/g. The lognormal-direct estimate of the mean is 0.0640 with a 95% UCB of 0.1223 pCi/g. These results are quite different from the PLE estimate and UCB, especially considering the AL of 0.062. This illustrates the sometimes conservative nature of the lognormal approach.

We have seen in the complete case that, unless CV is small, the lognormal-direct estimator is not as good as the ordinary sample mean (the lognormal-direct estimator is unstable, and the test seems to lack power). So it should come as no surprise if the direct methods are also poor relative to the PLE methods when there are non-detects. However, when the underlying distribution is lognormal or nearly so, the direct methods, have the additional advantage of accounting, approximately, for the non-detects. Thus it is conceivable that, when there is censoring, the direct methods might do better relative to alternative approaches than they do in the complete case. For this reason we also consider the direct methods in simulated examples with censoring.

For our simulations, we assumed there are at least two detects in each simulated sample. If there are zero detects, the lognormal MLE problem is degenerate (likelihood maximized where variance is zero and log-scale mean is anything less than the minimum detection limit). If there is only one detect, validity of all of the approximations should also be in doubt. (Also, if the single detect happens to be smaller than all of non-detects' detection limits, then the MLE again occurs where the variance is zero.)

We also assumed that censoring is independent of concentration. For each simulated observation we first generated a pseudorandom lognormal, truncated normal, or gamma variate X. Two of the X's were immediately taken as observations. For each remaining X, an independent random 'detection limit' was generated. If the detection limit exceeded X, the observation was set to the detection limit and called censored; otherwise the observation was set to X. Three levels of censoring were simulated by making the detection limits either U(0, 1), U(0.5, 1.5),or U(2, 3), where U(a, b) denotes a random variable that is uniform on the interval [a, b].

Because there seems to be no omnibus goodness-of-fit test for lognormality when there is random left censoring, we did not perform a goodness-of-fit test in the censored-data simulations. We already know that our truncated normal and gamma departures from lognormality are difficult to detect with complete samples. For censored samples, whatever the goodness-of-fit test, detecting them should be even more difficult.

In the simulations we compared the performances of the direct and PLE estimates and tests. The direct approach was the only lognormal-based method we considered. We used t-quantiles and $n - 1$ degrees of freedom (rather than z-quantiles or any other degree-of-freedom adjustment) in both tests, so the tests would be commensurate. Alternate approaches and fine-tuning the degrees of freedom might affect our conclusions. As in the complete case simulations, the tests and estimators were analysed in terms of power and squared error. All programming was done in SAS. MLEs were computed using the EM algorithm described in Lawless (1982, p. 223).

As in the complete case, the examples behave predictably as n varies. Results for all of the estimators or tests were essentially the same for CV = 0.1, and the direct estimates and tests broke down for gamma data with CV = 5. For CV = 1, the censoring levels achieved for the U(0, 1), U(0.5, 1.5) and U(2, 3) distributions were roughly 25, 50 and 75%, respectively, regardless of the distribution of X. The amounts of censoring varied for the other CV and n, but were generally in these ranges.

Table 6 contains maxima and average squared-error differences for the 10 000 direct and PLE estimates of the mean for the censored simulations with $n = 10$ and CV = 1. The table illustrates the instability of the direct estimate. Occasional, anomalously large point estimates result in huge squared errors and squared error differences. In terms of MSE differences, the PLE mean did better than the direct point estimate in all of the $n = 10$ and CV = 1 simulations except the lognormal cases with U(0.5, 1.5) and U(2, 3) censoring. MSE differences in the other cases are positive, and for the truncated normal and gamma distributions, much larger. Except for CV = 0.1, the conclusion was the same for all n and CV we considered: in our censored examples, the direct estimator, which was not competitive in the complete case, still does not seem competitive.

Figure 3 shows test power curves for censored data with CV = 0.5. Results for U(0, 1) and U(0.5, 1.5) censoring are similar to the uncensored results for CV = 0.5 shown in Fig. 2, though differences

Table 6. Statistics for mean estimates, 10 000 simulation trials with $n = 10$, CV = 1, and censored data.

		Censoring distribution			
Distribution	*Statistic*	*None*	U*(0, 1)*	U*(0.5, 1.5)*	U(2, 3)
Lognormal	PLE max.	3.30	3.29	3.42	4.90
	Direct max.	3.73	3.22	3.42	3.86
	MSE diff.[a]	0.002[b]	0.001[b]	−0.024[b]	−0.016[b]
Truncated normal	PLE max.	2.09	2.51	3.51	3.51
	Direct max.	206.4	49.2	58.1	20.6
	MSE diff.[a]	7.42	0.595[b]	0.593	0.108[b]
Gamma	PLE max.	2.77	2.81	5.5	5.5
	Direct max.	583.5	137.7	789.7	682.4
	MSE diff.[a]	69.9	3.14	63.8	47.1

[a] Average of PLE-from-direct squared-error difference.
[b] Significant at 0.05 level.

between the direct and PLE curves are smaller, especially for the U(0.5, 1.5) censoring. For the U(2, 3) censoring, the direct and PLE power curves are close, but the direct test is slightly better in the vicinity of AL = 2.

Figure 4 is like Fig. 3 but for CV = 1. Again the U(0, 1) and U(0.5, 1.5) results are similar to the uncensored results for CV = 1 in Fig. 2. For the U(2, 3) censoring, the two power curves are nearly coincident for low AL, but diverge with the PLE curve being larger for high AL, so that the PLE test is clearly preferable. Differences between the two curves are much greater for CV = 1 than for CV = 0.5.

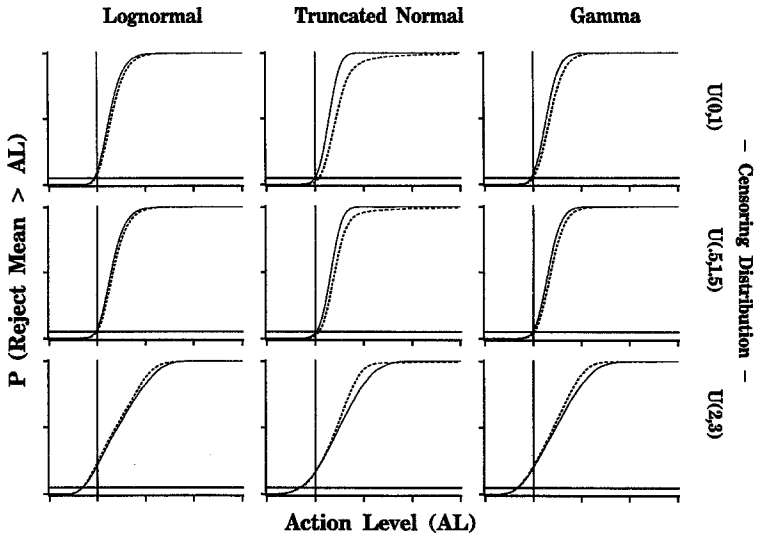

Figure 3. Power curves for $n = 10$, CV = 0.5, censored data (for 10 000 simulation trials). The plots are as in Fig. 2, but here for censored data and the PLE-based test (solid line) and the lognormal-direct test (dashes). The U(0, 1) censoring resulted in 80–5% detects, the U(0.5, 1.5) in 55–60% detects, and the U(2, 3) in 21–2% detects. Power curves for $n = 5$ and 20 are similar; the differences between the tests are larger for $n = 5$ and smaller for $n = 20$.

Power curves for CV = 2 are similar to the curves in Fig. 4, but for the gamma and lognormal distributions, the differences between the direct and PLE curves are greater, as are violations of nominal test size. The power curves for truncated normal data seem fairly stable in CV. Curves for CV = 5 are also similar to the CV = 2 curves, but even more extreme. For the gamma distribution with CV = 5, the PLE test size is far in excess of nominal, and the direct test is almost powerless. For the lognormal distribution with CV = 5, both the PLE and direct tests are anti-conservative.

Interpretation of the power curves is again a matter of opinion. Again, because of the arguments in Sections 2 and 3, we tend to give results for the truncated normal and gamma distributions more weight than for the lognormal, and results for CV of 0.5 or 1 more weight than for CV of 5. For U(0, 1) and U(0.5, 1.5) censoring, the direct test seems too conservative relative to the PLE test. For U(2, 3) censoring, the direct test does slightly better for CV = 0.5, but the PLE test does considerably better for higher CV. In general the PLE test seems better.

6. Conclusions

Our simulations are only examples, by no means all-encompassing. Other distributions could have been considered. For example, we have only considered distributions that, when CV is small, are very close to lognormal (or normal). Alternatives such as the triangular distribution, which for small CV is distinctly different from the normal or lognormal, would also be of interest. (The symmetric triangular density that is non-zero on [0.755, 1.245] has mean 1, CV = 0.1, and is also difficult to detect with the Shapiro–Wilk test with small complete samples).

Other methods might also modify our conclusions. In the complete case, modified t-statistics (Johnson, 1978) and bootstrap variations of them (Sutton, 1993) seem superior to ordinary t-statistics and are thus more attractive than ordinary t-statistics as alternatives to lognormal-based

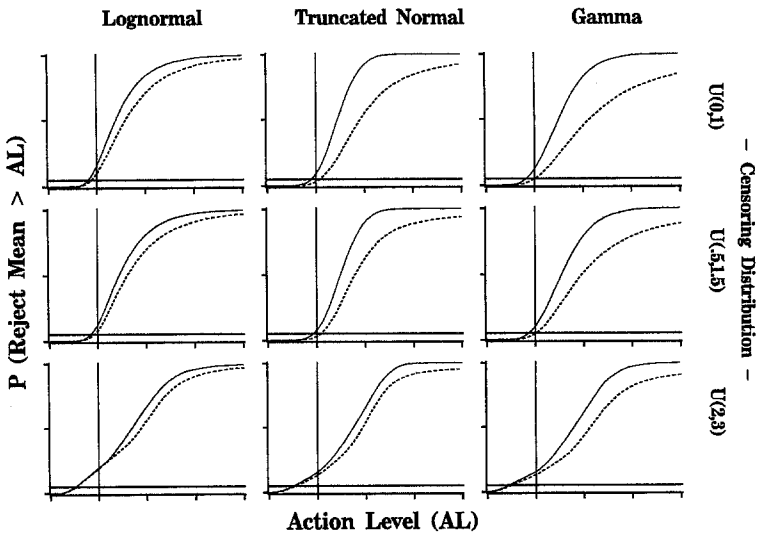

Figure 4. Power curves for $n = 10$, CV = 1, censored data (for 10 000 simulation trials). The plots are as in Fig. 3, but here for CV = 1. The U(0, 1) censoring resulted in 70–5% detects, the U(0.5, 1.5) in 49–54% detects, and the U(2, 3) in 25–7% detects. Power curves for $n = 5$ and 20 are similar; the differences between the tests are larger for $n = 5$ and smaller for $n = 20$.

approaches. Modified t-statistics might well be better with censored data too, though problems like estimating skewness (or adjusting for it in data skewed by the substitution of detection limits for actual values) would have to be worked out. Other variations, even just different degrees-of-freedom adjustments might also modify our conclusions.

Not surprisingly, our analyses indicate that when CV is small, all of our statistics should perform similarly. This is borne out in the simulations.

Our simulated examples clearly forewarn that when CV is large, say CV > 2, all of the approaches considered here can be way off in size or very unpowerful. MSEs can be huge. As indicated in Section 3, however CV < 2 is more likely in practice.

We also argue, in Section 2, that at least where mean estimation is the (presumably reasonable) goal, the true underlying data distribution is likely to have lighter tail weight than the lognormal: if the lognormal is a reasonable model, then the log-t should also be; because the log-t distribution does not have a mean, it is not a reasonable model for mean estimation; therefore, for mean estimation, the lognormal model is at least suspicious. Furthermore, as indicated by Shapiro–Wilk test results, for sample sizes of 20 or less, even substantial departures from lognormality can be very difficult to detect. Thus, for intermediate CV values ($0.1 <$ CV < 2), we can state the following general conclusions.

For our one-sided hypothesis tests: (i) In the complete case, the ordinary t-test, although somewhat anti-conservative, seems to be much more powerful than the UMPU test–so much so, that it is better. (Of course we are not saying this for the lognormal case.) The direct test's power tends to be intermediate between the ordinary t-test and the UMPU test. (ii) With censoring, the PLE-based test seems better than the direct test, though when the censoring is heavy, both the PLE and direct tests seem to be very anti-conservative.

Similarly, if squared error is the estimation criterion: (i) In the complete case the ordinary sample mean seems to do better than the MVUE, except, of course, when the data is lognormal. But even then, it does nearly as well. (ii) In the censored case, the direct estimates seems to be especially unstable. PLE means are better.

The unstable behaviour of the direct point estimates is related to the ill-posed nature of the mean itself. For risk models in which contamination is received and integrated over wide spatial and temporal ranges, the mean is likely to be appropriate. But in many settings a different parameter, for example, a median, or 95 or 99 percentile, may lead to a better indicator of a 'reasonable maximum exposure' (as defined in EPA, 1989). Our results suggest that there is good reason at least to consider alternatives to the mean parametrization of exposure. (See for example Hunt, 1972.)

Acknowledgement

This work was sponsored by Office of Environmental Management, U.S. Department of Energy under contract DE-AC05-84OR21400 with Lockheed Martin Energy Systems, Inc. We thank the referees for many helpful comments and suggestions.

References

Chu, J.T (1956) Errors in normal approximations to the t, τ, and similar types of distribution. *Annals of Mathematical Statistics*, **27**, 780–9.

Cohen, A.C. (1991) *Truncated and Censored Samples–Theory and Applications*. Marcel Dekker, New York.

DOE (1995) Remedial Investigation Report for the Clinch River/Poplar Creek Operable Unit (Technical Draft). Oak Ridge National Laboratory, Oak Ridge, Tennessee.

EPA (1989) *Risk Assessment Guidance for Superfund, Volume 1, Human Health Evaluation Manual (Part A)*. EPA/540/1-89/002, December 1989.

Finney, D.J. (1941) On the distribution of a variate whose logarithm is normally distributed. *Journal of the Royal Statistical Society, Supplement*, **7,** 155–61.

Gilbert, R.O. (1987) *Statistical Methods for Environmental Pollution Monitoring*. Van Nostrand Reinhold, New York.

Hunt, W.F. (1972) The precision associated with the sampling frequency of log-normally distributed air pollutant measurements. *Journal of the Air Pollution Control Association*, **22,** 687–91.

Johnson, N.J. (1978) Modified *t* tests and confidence intervals for asymmetrical populations. *Journal of the American Statistical Association*, **73,** 536–44.

Johnson, N.L. and Kotz, S. (1970) *Continuous Univariate Distributions*–2. Houghton Mifflin, Boston, p.41.

Kaplan, E.L. and Meier, P. (1958) Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, **53,** 457–81.

Land, C.E. (1971) Confidence intervals for linear functions of the normal mean and variance. *Annals of Mathematical Statistics*, **42,** 1187–205.

Land, C.E. (1972) An evaluation of approximate confidence interval estimation methods for lognormal means. *Technometrics*, **14,** 145–58.

Lawless, J.F. (1982) *Statistical Models and Methods for Lifetime Data*. Wiley, New York.

Ott, W.R. (1990) A physical explanation of the lognormality of pollutant concentrations. *Journal of Air and Waste Management Association*, **40,** 1378–83.

Ott, W.R. (1995) *Environmental Statistics and Data Analysis*. CRC Press, Boca Raton, FL.

Shapiro, S.S. and Wilk, M.B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, **52,** 591–611.

Shimizu, K. (1988) Point estimation. In: *Lognormal Distribution: Theory and Applications*, E.L. Crow and K. Shimizu (eds) *Marcel Dekker, New York, pp.27–86*.

Shimizu, K. and Crow, E.L. (1988) History, genesis, and properties. In: *Lognormal Distributions: Theory and Applications*, E.L. Crow, and K. Shimizu, (eds), Marcel Dekker, New York, pp.1–25.

Sutton, C.D. (1993) Computer-intensive methods for tests about the mean of an asymmetrical distribution. *Journal of the American Statistical Association*, **73,** 536–44.

Biographical sketches

Richard Schmoyer is a statistician at the Oak Ridge National Laboratory, and a graduate of the Pennsylvania State University. Over the years he has collaborated with researchers in many fields, with work described, for example, in *Ecotoxicology and Environmental Safety*, the *Journal of Applied Toxicology, Water Resources Research, Technometrics* and the *Journal of the American Statistical Association*.

J.J. Beauchamp is a research staff member in the Statistics Group of the Mathematics Sciences Section of the Computer Science and Mathematics Division at the Oak Ridge National Laboratory (ORNL), and also an adjunct professor of statistics at the University of Tennessee at Knoxville. He received his Ph.D in statistics from Florida State University, and has done consulting and collaboration with researchers within the ORNL Environmental Sciences Division for more than 20 years on statistical issues associated with the evaluation of various impacts on environmental or ecological systems.

Craig Brandt is a computer programmer in the Computer Science and Mathematics Division at the Oak Ridge National Laboratory. He is a graduate of the University of Tennessee. He works in developing and implementing statistical software, in environmental and ecological applications. He

is a member of the American Statistical Association, the Biometric Society, the American Institute of Biological Sciences, and the Ecological Society of America.

Owen Hoffman is president and director of the Center for Risk Analysis, SENES Oak Ridge, Inc. in Oak Ridge, Tennessee. He is widely recognized for his work on evaluating the reliability of predictions made using environmental transfer models, which includes the application of formal procedures for quantitative uncertainty analysis in human health and ecological risk assessment. He is a member of the National Council on Radiation Protection and Measurements, Science Advisory Board of the Environmental Protection Agency, the Health Advisory Panel to the Colorado State Department for Health for Dose Reconstruction at Rocky Flats, and the Advisory Panel to the Center for Disease Control for Dose Reconstruction and Analytical Epidemiology. He has also served as a Chief Scientist to the International Atomic Energy Agency in Vienna for testing exposure assessment models using data sets obtained from world-wide fallout from the Chernobyl accident.