Research Article

Increased Statistical Efficiency in a Lognormal Mean Model

Grant H. Skrepnek¹ and Ashok Sahai²

¹ College of Pharmacy & Peggy and Charles Stephenson Cancer Center, The University of Oklahoma Health Sciences Center, 1110 North Stonewall Avenue, Oklahoma City, OK 73126-0901, USA

² Department of Mathematics & Statistics, Faculty of Science and Technology, The University of the West Indies, St. Augustine Campus, Debe, Trinidad and Tobago

Correspondence should be addressed to Grant H. Skrepnek; grant-skrepnek@ouhsc.edu

Received 14 October 2013; Revised 27 February 2014; Accepted 13 March 2014; Published 14 April 2014

Academic Editor: Shein-chung Chow

Copyright © 2014 G. H. Skrepnek and A. Sahai. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Within the context of clinical and other scientific research, a substantial need exists for an accurate determination of the point estimate in a lognormal mean model, given that highly skewed data are often present. As such, logarithmic transformations are often advocated to achieve the assumptions of parametric statistical inference. Despite this, existing approaches that utilize only a sample's mean and variance may not necessarily yield the most efficient estimator. The current investigation developed and tested an improved efficient point estimator for a lognormal mean by capturing more complete information via the sample's coefficient of variation. Results of an empirical simulation study across varying sample sizes and population standard deviations indicated relative improvements in efficiency of up to 129.47 percent compared to the usual maximum likelihood estimator and up to 21.33 absolute percentage points above the efficient estimator presented by Shen and colleagues (2006). The relative efficiency of the proposed estimator increased particularly as a function of decreasing sample size and increasing population standard deviation.

1. Introduction

The presence of highly skewed data is commonplace across both basic and applied sciences [1, 2]. In certain instances, the logarithmic transformation of such data may be undertaken with the primary purpose of establishing a normal distribution and improving variance, which may include removing heteroskedasticity in the process, to achieve $N(\mu, \sigma^2)$. Patterson [3] provided seminal work concerning the statistical challenges involved in estimating the population mean following the transformation of data. More recently, Shen et al. [4] proposed an improved efficient minimum risk/relative mean squared error (RMSE) estimator of the lognormal mean, and numerous researchers have addressed the estimation of parameters within this distribution from both a frequentist and Bayesian context [5–9].

In presenting the lognormal distribution from a more fundamental perspective, if Y is a random variable with a lognormal distribution and a mean of $E(Y) = \varsigma$, then $\ln(Y)$ will be normally distributed with mean of μ and variance of

σ^2 . Therefore, Y may also be expressed as $Y \sim \ln(\mu, \sigma^2)$ with a mean of ς , observing that

$$\varsigma = \exp\left(\mu + \frac{\sigma^2}{2}\right). \quad (1)$$

As such, when considering a random sample $y_1, y_2, \dots, y_n$ that is i.i.d. and given $\ln(\mu, \sigma^2)$ with a mean of ς , then $x_i = \ln(y_i)$ is i.i.d. as $N(\mu, \sigma^2)$ for $i = 1, \dots, n$. The following may also be defined:

$$\bar{x} = \sum_{i=1}^{i=n} \left(\frac{x_i}{n}\right) \quad s^2 = \sum_{i=1}^{i=n} \frac{(x_i - \bar{x})^2}{(n-1)}, \quad (2)$$

noting that $\bar{x}$ and s^2 are the maximum likelihood estimators (MLE) for μ and σ^2 , respectively [2]. By applying (1) to (2), the usual estimator (UER) for ς is

$$\text{UER}_{\varsigma} = \exp\left(\bar{x} + \frac{s^2}{2}\right). \quad (3)$$

Addressed previously, Shen et al. [4] proposed a new estimator for UER_{ζ} by minimizing its relative mean square error (RMSE) through an application of the delta method [10], which yields the following when applied to (3):

$$RMSE(UER_{\zeta}) = E \left[\frac{(UER_{\zeta} - \zeta)}{\zeta} \right]^2. \quad (4)$$

Notably, the class of estimators used by the authors was

$$\zeta_c = \exp \left(\bar{x} + \frac{(1/(n+d)) \cdot s^2}{2} \right), \quad (5)$$

with

$$d > (-n). \quad (6)$$

By minimizing the RMSE of the estimators in the class by an order of $1/n^2$, the optimal value of “ d ” was obtained, wherein the optimal estimator, c , in the class was identified as

$$c = \frac{1}{(n+4 + (3 \cdot \sigma^2/2))}. \quad (7)$$

By applying the usual unbiased estimate s^2 of σ^2 , a novel minimum risk/relative mean squared error estimator of ζ was developed:

$$SHEN_{(2006)} = \exp \left\{ \bar{x} + \frac{(n-1)s^2}{2(n+4)(n-1) + 3s^2} \right\}, \quad (8)$$

or, alternatively,

$$RMSE(SHEN_{(2006)}) = E \left[\frac{(SHEN_{(2006)} - \zeta)}{\zeta} \right]^2. \quad (9)$$

Given the prevalence of logarithmic data across research disciplines and the importance in efficiently estimating these data, the purpose of the current study was to derive and assess an approach to obtain statistically efficient estimators of the lognormal mean. More specifically, the objectives involved incorporating more comprehensive information from the sample’s coefficient of variance, $CV = s/(\sqrt{n} \cdot \bar{x})$, following a logarithmic transformation of a resultant sample data in estimating the nontransformed population mean of the original distribution.

2. Preparatory Improvement

Being contingent upon the Rao-Blackwell theorem, any function of a sample mean or sample variance will be a uniformly minimum variance unbiased estimate (UMVUE) or Uniformly minimum mean squared error estimator (UMMSE) of the population if an estimator is an unbiased or minimum mean squared estimator (MMSE). Hence, a usual estimator (UER) as presented in (3) is the UMVUE of $\ln(\zeta)$ (i.e., the population mean of the original lognormal distribution). Furthermore, it should be noted that $(n-1) \cdot s^2/\sigma^2 \sim \chi^2$ distribution with $(n-1)$ degrees of freedom.

Numerous empirical analyses involve conditions wherein small values of the sample estimate of the coefficient of variation are observed. Therein, an alternate estimator appearing in Lovric and Sahai [11], denoted by $g^{\otimes}$, of the population mean, μ , is offered rather than the typical estimator, $\bar{x}$:

$$g^{\otimes} = (\bar{x}) + \frac{\bar{x}}{\left\{ (n \cdot (\bar{x})^2/s^2) - 1 \right\}}. \quad (10)$$

Applying principals outlined in Nikulin [12], the relative efficiency (i.e., a key measure of an estimator’s optimality) of $g^{\otimes}$ versus $\bar{x}$ can be expressed as a percentage as

$$\frac{100}{\eta} = 100 \cdot \left\{ \frac{E(\bar{x} - \mu)^2}{E(g^{\otimes} - \mu)^2} \right\}. \quad (11)$$

The UMVUE of the efficiency ratio, η , as a function of $(\bar{x}, s^2)$ may be determined as follows, given that $(\bar{x}, s^2)$ is a complete sufficient statistic for (μ, σ^2) :

$$\begin{aligned} \eta &= \frac{E(g^{\otimes} - \mu)^2}{E(\bar{x} - \mu)^2} \\ &= \left(\frac{n}{\sigma^2} \right) \cdot E \left[(\bar{x} - \mu) + \left(\frac{\bar{x}}{(n \cdot (\bar{x})^2/s^2) - 1} \right) \right]^2 \end{aligned} \quad (12)$$

$$= 1 + 2J + K,$$

with

$$\begin{aligned} J &= \left(\frac{n}{\sigma^2} \right) \cdot E \left[\frac{\bar{x} \cdot (\bar{x} - \mu)}{(n \cdot (\bar{x})^2/s^2) - 1} \right] \\ &= \left(\frac{n}{\sigma^2} \right) \cdot Es^2 \cdot E \bar{x} \cdot \left[\frac{\bar{x} \cdot (\bar{x} - \mu)}{(n \cdot (\bar{x})^2/s^2) - 1} \right] \\ &= \left(\frac{n \cdot (n/2\pi\sigma^2)^{1/2}}{\sigma^2} \right) \cdot Es^2 \\ &\quad \cdot \left[\int_{-\infty}^{+\infty} \frac{\bar{x}}{(n \cdot (\bar{x})^2/s^2) - 1} \cdot (\bar{x} - \mu) \right. \\ &\quad \cdot \exp \left\{ \frac{-n \cdot (\bar{x} - \mu)^2}{2\sigma^2} \right\} d\bar{x} \left. \right] \\ &= - \left(\frac{n}{2\pi\sigma^2} \right)^{1/2} \cdot Es^2 \\ &\quad \cdot \left[\int_{-\infty}^{+\infty} \frac{\bar{x}}{(n \cdot (\bar{x})^2/s^2) - 1} \cdot (d/d\bar{x}) \right. \\ &\quad \cdot \exp \left\{ \frac{-n \cdot (\bar{x} - \mu)^2}{2\sigma^2} \right\} d\bar{x} \left. \right]. \end{aligned} \quad (13)$$

In explaining the aforementioned in more detail, particularly the development of term J to include the terms Es^2 and

$E\bar{x}$, suppose $\phi(\bar{x}, s^2)$ is a function of the sample mean $\bar{x}$ and the sample variance s^2 for a random sample from a normal population. Therefore, $\bar{x}$ and s^2 would be known as having independent sampling distributions. Consequently, the expression $E\{\phi(\bar{x}, s^2)\}$ may be regarded as a “two-phase” exercise; in the first phase, $E\bar{x}\{\cdot\}$ may be viewed as the expectation with respect to the random variable $\bar{x}$, treating the silenced (i.e., pseudo) relative variable s^2 as a constant, $Es^2\{\dots\}$. Subsequently, in the second phase, the random variable s^2 also has an expectation of being viewed as $Es^2\{\dots\}$. Cumulatively, therefore, $E(\phi(\cdot)) = E_{\bar{x}, s^2}\{\phi(\cdot)\} = E_{s^2}[E_{\bar{x}}\{\phi(\cdot)\}]$. Applied specifically to (12), the integration by parts may be detailed accordingly to ultimately yield $J = Es^2 \cdot E\bar{x}[j] = E[j]$:

$$\begin{aligned}
 E_{\bar{x}} \left[\frac{\bar{x}(\bar{x} - \mu)}{\{(n(\bar{x})^2/s^2) - 1\}} \right] \\
 &= - \left(\frac{n}{\{n/2\pi\sigma^2\}^{1/2}} \right) \\
 &= \int_{-\infty}^{\infty} \frac{\bar{x}(\bar{x} - \mu)}{\{(n(\bar{x})^2/s^2) - 1\}} \\
 &\quad \times (\bar{x} - \mu) \exp \left\{ \frac{-n(\bar{x} - \mu)^2}{2\sigma^2} \right\} d\bar{x} \\
 &= \int_{-\infty}^{\infty} \frac{\bar{x}(\bar{x} - \mu)}{\{(n(\bar{x})^2/s^2) - 1\}} \left(\frac{d}{d\bar{x}} \right) \\
 &\quad \cdot \left\{ \exp \left\{ \frac{-n(\bar{x} - \mu)^2}{2\sigma^2} \right\} d\bar{x} \right\}.
 \end{aligned} \tag{14}$$

Notably, the other term vanishes by the well-known properties of definite integral, $\int_{-A}^A \{\text{integrand}\} d\bar{x}$, as the integrand is the odd function of $\bar{x}$. Subsequently, the remainder of the derivation follows by way of integration by parts as

$$J = Es^2 \cdot E\bar{x}[j] = E[j], \tag{15}$$

with

$$\begin{aligned}
 j &= - \left[\left\{ \left(\frac{n \cdot (\bar{x})^2}{s^2} \right) - 1 \right\} \cdot \left\{ \left(\frac{n \cdot (\bar{x})^2}{s^2} \right) + 1 \right\}^{-2} \right] \\
 &= [(u - 1) \cdot (u + 1)^{-2}],
 \end{aligned} \tag{16}$$

as

$$u = \frac{n \cdot (\bar{x})^2}{s^2}. \tag{17}$$

Again, it is important to note that, independent of $\bar{x}$, $(n - 1) \cdot s^2/\sigma^2$ approximates a χ^2 distribution with $(n - 1)$ degrees of freedom. Therefore, again applied to (12),

$$\begin{aligned}
 K &= \left(\frac{n}{\sigma^2} \right) \cdot Es^2 \cdot E\bar{x} \cdot \left[(\bar{x})^2 \cdot \left(\frac{n \cdot (\bar{x})^2}{s^2} - 1 \right)^{-2} \right] \\
 &= \left(\frac{\left\{ \left((n - 1)/2\sigma^2 \right)^{(n-1)/2} \cdot (1/\Gamma(n - 1)/2) \right\}}{\sigma^2} \right) \cdot E\bar{x} \\
 &\quad \cdot \left[\int_0^{+\infty} \left\{ \left(\frac{(\bar{x})^2}{s^2} \right) \cdot \left(\frac{n \cdot (\bar{x})^2}{s^2} - 1 \right)^{-2} \right\} \right. \\
 &\quad \cdot \left. \left((s^2)^{(n-1)/2} \right) \cdot \exp \left\{ \frac{-(n - 1) \cdot s^2}{2\sigma^2} \right\} ds^2 \right].
 \end{aligned} \tag{18}$$

Furthermore

$$\begin{aligned}
 \frac{d}{ds^2} \cdot (s^2)^{(n-1)/2} \cdot \exp \left\{ \frac{-(n - 1) \cdot s^2}{2\sigma^2} \right\} \\
 &= \left\{ \frac{(n - 1)}{2\sigma^2} \right\} \cdot \sigma^2 \cdot (s^2)^{(n-3)/2} \cdot \exp \left\{ \frac{-(n - 1) \cdot s^2}{2\sigma^2} \right\} \\
 &\quad - (s^2)^{(n-1)/2} \left\{ -\frac{(n - 1)}{2\sigma^2} \right\} \cdot \exp \left\{ -\frac{(n - 1) s^2}{2\sigma^2} \right\}.
 \end{aligned} \tag{19}$$

By applying (19) within (18), the following may be obtained:

$$\begin{aligned}
 K &= \left(\frac{\left\{ \left((n - 1)/2\sigma^2 \right)^{(n-1)/2} \cdot (1/\Gamma(n - 1)/2) \right\}}{n} \right) \\
 &\quad \cdot E(\bar{x}) \left[\int_0^{+\infty} \left\{ \left(\frac{n \cdot (\bar{x})^2}{s^2} \right) \cdot \left(\frac{n \cdot (\bar{x})^2}{s^2} - 1 \right)^{-2} \right\} \right. \\
 &\quad \cdot \left. \left((s^2)^{(n-3)/2} \right) \cdot \exp \left\{ \frac{-(n - 1) \cdot s^2}{2\sigma^2} \right\} ds^2 \right] \\
 &\quad - \left[\frac{2 \cdot \left\{ \left((n - 1)/2\sigma^2 \right)^{(n-1)/2} \cdot (1/\Gamma(n - 1)/2) \right\}}{n \cdot (n - 1)} \right] \\
 &\quad \cdot (\bar{x}) \left[\int_0^{+\infty} \left\{ \left(\frac{n \cdot (\bar{x})^2}{s^2} \right) \cdot \left(\frac{n \cdot (\bar{x})^2}{s^2} - 1 \right)^{-2} \right. \right. \\
 &\quad \cdot \left. \left(\frac{d}{ds^2} \right) \cdot \left((s^2)^{(n-1)/2} \right) \right. \\
 &\quad \cdot \left. \left. \exp \left\{ \frac{-(n - 1) \cdot s^2}{2\sigma^2} \right\} ds^2 \right\} \right].
 \end{aligned} \tag{20}$$

Again through the application of integration by parts,

$$\begin{aligned}
 K &= E(s^2) \cdot E(\bar{x}) \left[\left(\frac{n \cdot (\bar{x})^2}{s^2} \right) \cdot \left(\frac{n \cdot (\bar{x})^2}{s^2} - 1 \right)^{-2} \right] \\
 &\quad - \left\{ \frac{2 \cdot \left\{ \left((n-1)/2\sigma^2 \right)^{(n-1)/2} \cdot (1/\Gamma(n-1)/2) \right\}}{n \cdot (n-1)} \right\} \\
 &\quad \cdot E(\bar{x}) \left[\int_0^{+\infty} \left\{ \left(\frac{n \cdot (\bar{x})^2}{s^2} \right) \cdot \left(\frac{n \cdot (\bar{x})^2}{s^2} - 1 \right)^{-2} \right. \right. \\
 &\quad \cdot \left(\frac{d}{ds^2} \right) \cdot \left((s^2)^{(n-1)/2} \right) \\
 &\quad \cdot \exp \left\{ \frac{-(n-1) \cdot s^2}{2\sigma^2} \right\} ds^2 \left. \right] \\
 &= E \left[\left(\frac{n \cdot (\bar{x})^2}{s^2} \right) \cdot \left(\frac{n \cdot (\bar{x})^2}{s^2} - 1 \right)^{-2} \right] - \left\{ \frac{2}{(n-1)} \right\} \\
 &\quad \cdot E \left[\left(\frac{n \cdot (\bar{x})^2}{s^2} \right) \cdot \left(\frac{n \cdot (\bar{x})^2}{s^2} + 1 \right) \cdot \left(\frac{n \cdot (\bar{x})^2}{s^2} - 1 \right)^{-3} \right] \\
 &= E \left[\left(\frac{n \cdot (\bar{x})^2}{s^2} \right) \cdot \left(\frac{n \cdot (\bar{x})^2}{s^2} - 1 \right)^{-2} - \left\{ \frac{2}{(n-1)} \right\} \right. \\
 &\quad \cdot \left[\left(\frac{n \cdot (\bar{x})^2}{s^2} \right) \cdot \left(\frac{n \cdot (\bar{x})^2}{s^2} + 1 \right) \right. \\
 &\quad \cdot \left. \left. \left(\frac{n \cdot (\bar{x})^2}{s^2} - 1 \right)^{-3} \right] \right], \tag{21}
 \end{aligned}$$

or, expressed differently, noting that $K = E(k)$:

$$\begin{aligned}
 k &= \left\{ \frac{n \cdot (\bar{x})^2}{s^2} \right\} \cdot \left\{ \left(\frac{n \cdot (\bar{x})^2}{s^2} \right) - 1 \right\}^{-2} \\
 &\quad - \left\{ \left(\frac{2}{(n-1)} \right) \cdot \frac{n \cdot (\bar{x})^2}{s^2} \right\} \cdot \left\{ \left(\frac{n \cdot (\bar{x})^2}{s^2} \right) + 1 \right\} \\
 &\quad \cdot \left\{ \left(\frac{n \cdot (\bar{x})^2}{s^2} \right) - 1 \right\}^{-3}. \tag{22}
 \end{aligned}$$

As such, the UMVUE relating to the statistical relative efficiency of $g^{\otimes}$ versus $\bar{x}$, given that $\eta = 1 + 2J + K$ (hence, $\hat{\eta} = 1 + 2j + k$) from (12), and including (16) and (19), would be derived from

$$\hat{\eta} = 1 + 2j + k = 1 - 2$$

$$\cdot \left\{ \left[\left(\frac{n \cdot (\bar{x})^2}{s^2} \right) + 1 \right] \cdot \left[\left(\frac{n \cdot (\bar{x})^2}{s^2} \right) - 1 \right]^{-2} \right\}$$

$$\begin{aligned}
 &+ \left(\left\{ \left[\frac{n \cdot (\bar{x})^2}{s^2} \right] \cdot \left[\left(\frac{n \cdot (\bar{x})^2}{s^2} \right) - 1 \right]^{-3} \right\} \right. \\
 &\quad \cdot \left. \left\{ (n-3) - \left(\frac{n \cdot (\bar{x})^2}{s^2} \right) \cdot (n+1) \right\} \right) \\
 &\quad \times ((n-1))^{-1} \\
 &= 1 - \left[\left(\left\{ \left[\left(\frac{n \cdot (\bar{x})^2}{s^2} \right) - 1 \right]^{-3} \right\} \right. \right. \\
 &\quad \cdot \left[(n-3) \cdot \left(\frac{n \cdot (\bar{x})^2}{s^2} \right)^2 \right] \\
 &\quad + \left[(n-3) \cdot \left(\frac{n \cdot (\bar{x})^2}{s^2} \right) \right] \\
 &\quad \cdot \left. \left. -2 \cdot (n-1) \right\} \right) \\
 &\quad \times ((n-1))^{-1} \Big], \tag{23}
 \end{aligned}$$

and, thus,

$$\frac{100}{\eta} = 100 \cdot \frac{E(\bar{x} - \mu)^2}{E(g^{\otimes} - \mu)^2} > 100\%, \tag{24}$$

if $0 < \hat{\eta} < 1$, $0 < 1 + 2j + k < 1$, or $0 < \{(n \cdot (\bar{x})^2/s^2)^2 \cdot (n-3) + (n \cdot (\bar{x})^2/s^2) \cdot (n-3) - 2 \cdot (n-1)\}$ per (17), as $(n \cdot (\bar{x})^2/s^2) > 1$ for all $(\bar{x})^2 > s^2/n$, with the coefficient of variation of $\bar{x} < 1$ per (16), or if $(n \cdot (\bar{x})^2/s^2) > (n+1)/(n-3)$.

Given the aforementioned, the alternate estimator defined in (10), $g^{\otimes}$, would be a more efficient estimator of the normal population mean, μ . It is important to also note that this proposed estimator could be expressed as a function of the square of sample coefficient of variation:

$$g^{\otimes} \cong (\bar{x}) \cdot \left[1 + \left(\frac{s^2}{n \cdot (\bar{x})^2} \right) \right]. \tag{25}$$

3. The Improved Lognormal Mean Estimator

As previously noted, the purpose of this research investigation was to improve and test the estimator proposed by Shen (2006), presented initially in (8),

$$\text{SHEN}_{(2006)} = \exp \left\{ \bar{x} + \left[\frac{(n-1) \cdot s^2}{2 \cdot (n+4)} \right] + 3 \cdot s^2 \right\}. \tag{26}$$

Offered in (25), the development of the proposed estimator, $g^{\otimes}$, draws upon (10) and, through substitution from ((8), (26)), the following is obtained:

$$g^{\otimes} = \bar{x} + \frac{s^2}{(n \cdot \bar{x})} = \bar{x} + \left(1 + \frac{s^2}{n(\bar{x})^2} \right). \tag{27}$$

Describing the approach through which $g^{\otimes}$ from (10) was developed to the expression in (25) and further to (27),

$$\begin{aligned}
 g^{\otimes} &= (\bar{x}) + \frac{\bar{x}}{\{(n \cdot (\bar{x})^2 / s^2) - 1\}} \\
 &= (\bar{x}) \left[\frac{(n \cdot (\bar{x})^2 / s^2)}{\{(n \cdot (\bar{x})^2 / s^2) - 1\}} \right] \\
 &= (\bar{x}) \left[1 + \left\{ \frac{n \cdot (\bar{x})^2}{s^2} - 1 \right\}^{-1} \right] \\
 &= (\bar{x}) \left[1 + \frac{s^2}{n \cdot (\bar{x})^2} \left\{ 1 - \frac{s^2}{n \cdot (\bar{x})^2} \right\}^{-1} \right] \\
 &= (\bar{x}) \left[1 + \frac{s^2}{n \cdot (\bar{x})^2} \left\{ 1 + \frac{s^2}{n \cdot (\bar{x})^2} - \frac{s^4}{n^2 \cdot (\bar{x})^2} + 0 \left(\frac{1}{n^4} \right) \right\} \right] \\
 &= (\bar{x}) \left[1 + \frac{s^2}{n \cdot (\bar{x})^2} + 0 \left(\frac{1}{n^2} \right) \right] \\
 &\simeq (\bar{x}) \left[1 + \frac{s^2}{n \cdot (\bar{x})^2} \right] \text{ up to a term of } 0 \left[\frac{1}{n} \right].
 \end{aligned} \tag{28}$$

To illustrate, if $1/n^2$ is even $1/(31)^2$, the term is negligible.

To derive an efficient estimator of the normal variance using the sample coefficient of variance, the sixth iteration efficient estimator of the normal variance presented in Lovric and Sahai [11] may be applied as

$$E_6 ER(\sigma^2) = 6M_1 \cdot G_6^{**} + \frac{M_2 \cdot G^*}{(M_1 + M_2)}, \tag{29}$$

where

$$\begin{aligned}
 M_1 &= \frac{2 \cdot (n-1)}{(n+1)^2}, \\
 M_2 &= \left(\left[2 \cdot \left(\frac{s^2}{n \cdot (\bar{x})^2} \right) \right] \cdot \left[2 + \left(\frac{s^2}{n \cdot (\bar{x})^2} \right) \right] \right. \\
 &\quad \cdot \left[1 + \left(\frac{s^2}{n \cdot (\bar{x})^2} \right) \right]^2 \Bigg) \\
 &\quad \times \left(3 \cdot \left[1 + \left(\frac{s^2}{n \cdot (\bar{x})^2} \right)^2 - 2 \right]^2 \right)^{-1}, \\
 G_6^{**} &= (s^2) \cdot \left\{ \frac{1 + (s^2/n \cdot (\bar{x})^2)}{3 \cdot [1 + (s^2/n \cdot (\bar{x})^2)^2 - 2]} \right\} \\
 G^* &= (s^2) \cdot \left\{ \frac{(n-1)}{(n+1)} \right\}.
 \end{aligned} \tag{30}$$

Consequently, the proposed lognormal mean estimator from the current investigation, $GS_{(2014)}$, is

$$GS_{(2014)} = \exp \left\{ (g^{\otimes}) + E_6 ER \left(\frac{\sigma^2}{2} \right) \right\}. \tag{31}$$

4. Empirical Simulation Study

4.1. Study Methodology. To compare the proposed estimator, $GS_{(2014)}$, to the existing efficient estimator, $SHEN_{(2006)}$, and the usual maximum likelihood estimator, UER, a simulation study was undertaken using various sample sizes (i.e., $n = 31, n = 41, n = 51, n = 71, n = 101, n = 121, n = 151$, and $n = 201$) and population mean values (i.e., $\mu = 0.30, \mu = 0.35, \mu = 0.40, \mu = 0.45, \mu = 0.50, \mu = 0.55$, and $\mu = 0.60$), with a fixed variance of $\sigma^2 = 0.36$ (i.e., $\sigma = 0.60 \geq \mu$). Some 11,000 iterations across sample sizes were conducted using Matlab 2010b [The Mathworks Inc., Natick, MA], drawing randomly from a population of $N(\mu, \sigma^2)$. Comparisons of $GS_{(2014)}$ versus UER and $SHEN_{(2006)}$ versus UER were drawn in percentage terms via relative efficiencies, $RelEff_{\%}$, as

$$\begin{aligned}
 RelEff_{\%} \{ (SHEN_{(2006)}) \text{ versus } (UER) \} \\
 &= 100 \cdot \frac{RMSE(UER)}{RMSE(SHEN_{(2006)})},
 \end{aligned} \tag{32}$$

$$\begin{aligned}
 RelEff_{\%} \{ (GS_{(2014)}) \text{ versus } (UER) \} \\
 &= 100 \cdot \frac{RMSE(UER)}{RMSE(GS_{(2014)})},
 \end{aligned}$$

as, per (4) and (9),

$$RMSE(UER) = E \left[\frac{(UER - \varsigma)}{\varsigma} \right]^2, \tag{33}$$

$$RMSE(SHEN_{(2006)}) = E \left[\frac{(SHEN_{(2006)} - \varsigma)}{\varsigma} \right]^2, \tag{34}$$

$$RMSE(GS_{(2014)}) = E \left[\frac{(GS_{(2014)} - \varsigma)}{\varsigma} \right]^2. \tag{35}$$

Therein, actual RMSEs of all estimators in (33), (34), and (35) were calculated as an average across each of the simulation's 11,000 iterations.

4.2. Results. Presented in Table 1, the relative efficiencies of both the proposed estimator, $GS_{(2014)}$, and the existing efficient estimator, $SHEN_{(2006)}$, recorded improvements to the usual maximum likelihood estimator, UER. In general, relative efficiencies for both efficient estimators increased as a function of lower sample size, though the proposed estimator, $GS_{(2014)}$, was also more efficient at lower population standard deviations. Across 74 of the 77 analytic categories (96%), $GS_{(2014)}$ noted higher relative efficiencies than $SHEN_{(2006)}$, with the absolute difference being most pronounced at lower sample sizes plus lower population standard deviations. To illustrate, the greatest absolute percentage difference favoring

TABLE 1: Relative efficiencies of $\text{SHEN}_{(2006)}$ and $\text{GS}_{(2014)}$ estimators across varying sample sizes and population standard deviations.

Sample size Estimators	Relative efficiency versus the usual estimator, UER [in %]						
	Population standard deviation						
	$\mu = 0.30$	$\mu = 0.35$	$\mu = 0.40$	$\mu = 0.45$	$\mu = 0.50$	$\mu = 0.55$	$\mu = 0.60$
$n = 31$							
SHEN ₍₂₀₀₆₎	108.14	108.06	107.89	107.39	108.71	108.11	108.60
GS ₍₂₀₁₄₎	129.47	125.70	123.76	119.88	117.23	113.86	112.19
$n = 41$							
SHEN ₍₂₀₀₆₎	106.89	106.38	106.10	105.29	106.84	105.52	106.10
GS ₍₂₀₁₄₎	126.19	123.80	118.30	113.86	112.55	108.98	108.00
$n = 51$							
SHEN ₍₂₀₀₆₎	104.79	104.36	104.38	104.58	104.40	104.57	104.70
GS ₍₂₀₁₄₎	124.74	118.70	114.05	110.72	108.10	106.59	105.63
$n = 61$							
SHEN ₍₂₀₀₆₎	104.25	104.29	103.85	104.26	104.57	103.93	103.43
GS ₍₂₀₁₄₎	120.78	115.46	111.35	108.80	107.33	105.20	103.76
$n = 71$							
SHEN ₍₂₀₀₆₎	103.20	103.49	104.04	103.50	103.32	103.27	103.32
GS ₍₂₀₁₄₎	117.87	112.91	110.03	107.06	105.23	104.06	103.33
$n = 81$							
SHEN ₍₂₀₀₆₎	103.51	103.75	102.76	102.93	102.70	103.08	103.08
GS ₍₂₀₁₄₎	115.47	111.51	107.52	105.69	104.00	103.53	102.96
$n = 91$							
SHEN ₍₂₀₀₆₎	103.02	102.89	102.83	102.69	102.20	103.09	103.02
GS ₍₂₀₁₄₎	113.48	109.63	106.91	104.88	103.11	103.39	102.80
$n = 101$							
SHEN ₍₂₀₀₆₎	102.53	102.27	102.41	102.30	102.65	102.40	102.27
GS ₍₂₀₁₄₎	111.65	108.05	105.76	104.03	103.43	102.46	101.86
$n = 121$							
SHEN ₍₂₀₀₆₎	101.87	101.54	102.37	102.29	102.07	101.75	101.94
GS ₍₂₀₁₄₎	109.28	105.84	104.93	103.60	102.44	101.49	101.45
$n = 151$							
SHEN ₍₂₀₀₆₎	101.58	102.15	102.03	101.67	101.84	101.55	101.72
GS ₍₂₀₁₄₎	107.08	105.40	103.83	102.39	102.04	101.25	101.25
$n = 201$							
SHEN ₍₂₀₀₆₎	101.11	101.33	101.29	101.10	101.18	101.05	101.28
GS ₍₂₀₁₄₎	104.82	103.39	102.28	101.38	101.08	100.64	100.83

$\text{SHEN}_{(2006)}$: Shen et al. (2006) [4] estimator; $\text{GS}_{(2014)}$: proposed estimator; UER: usual maximum likelihood estimator; n : sample size; μ : population standard deviation.

the efficiency of $\text{GS}_{(2014)}$ compared to $\text{SHEN}_{(2006)}$ was +21.33 percent at $n = 31$ and $\mu = 0.30$. Comparatively, the maximum difference across the three cases where $\text{SHEN}_{(2006)}$ outperformed the proposed estimator was +0.9 percent at $n = 201$ and $\mu = 0.60$. The relative efficiencies for $\text{GS}_{(2014)}$ ranged from a low of 100.64 percent ($\mu = 0.55$, $n = 201$) to a high of 129.47 percent ($\mu = 0.30$, $n = 31$) compared to a low of 101.05 percent ($\mu = 0.55$, $n = 201$) to a high of 108.71 percent ($\mu = 0.50$, $n = 31$) for the $\text{SHEN}_{(2006)}$ estimator. Collectively, findings support the proposed estimator, $\text{GS}_{(2014)}$, under almost all combinations of sample sizes and population standard deviations compared

to the current efficient estimator and the usual maximum likelihood estimator. In instances where the existing efficient estimator did perform better, the differences were small and perhaps clinically negligible (i.e., absolute differences below 1.0%).

5. Conclusion

Within the context of clinical and other scientific research, a substantial need exists for an accurate determination of the point estimate for a lognormal mean. The transformation of

highly skewed data is often undertaken to achieve assumptions required for parametric statistical inference. Despite this, existing approaches that capture only a sample's mean and variance do not necessarily yield the most efficient estimator. The current investigation developed and tested more efficient point estimators for a lognormal mean model by capturing more complete information within the sample's coefficient of variation. Results of an empirical simulation study across varying sample sizes and population standard deviations indicated relative improvements in efficiency of up to 129.47 percent compared to the usual maximum likelihood estimator and up to 21.33 percentage points above the current efficient estimator. The relative efficiency of the proposed estimator increased particularly as a function of decreasing sample size and increasing population standard deviation.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

- [1] G. H. Skrepnek, "Regression methods in the empirical analysis of health care data," *Journal of Managed Care Pharmacy*, vol. 11, no. 3, pp. 240–251, 2005.
- [2] G. H. Skrepnek, E. L. Olvey, and A. Sahai, "Econometric approaches in evaluating costs and outcomes within pharmacoeconomic analyses," *Pharmaceuticals Policy and Law*, vol. 14, no. 1, pp. 105–122, 2012.
- [3] R. L. Patterson, "Difficulties involved in the estimation of a population mean using transformed sample data," *Technometrics*, vol. 8, pp. 535–537, 1966.
- [4] H. Shen, L. D. Brown, and H. Zhi, "Efficient estimation of log-normal means with application to pharmacokinetic data," *Statistics in Medicine*, vol. 25, no. 17, pp. 3023–3038, 2006.
- [5] I. G. Evans and S. A. Shaban, "A note on estimation in log-normal models," *Journal of the American Statistical Association*, vol. 69, pp. 779–781, 1974.
- [6] A. L. Rukhin, "Improved estimation in lognormal models," *Journal of the American Statistical Association*, vol. 81, pp. 1046–1049, 1986.
- [7] G. H. Skrepnek, "The contrast and convergence of bayesian and frequentist statistical approaches in pharmacoeconomic analysis," *PharmacoEconomics*, vol. 25, no. 8, pp. 649–664, 2007.
- [8] W. J. Padgett and L. J. Wei, "Bayes estimation of reliability for the two-parameter lognormal distribution," *Communications in Statistics—Theory and Methods*, vol. 6, pp. 443–457, 1977.
- [9] X. H. Zhou, "Estimation of the log-normal mean," *Statistics in Medicine*, vol. 17, pp. 2251–2264, 1998.
- [10] G. W. Oehlert, "A note on the delta method," *American Statistician*, vol. 46, pp. 27–29, 1992.
- [11] M. M. Lovric and A. Sahai, "An iterative algorithm for efficient estimation of normal variance using sample coefficient of variation," *InterStat*, article 001, 2011.
- [12] M. S. Nikulin, "Efficiency of a statistical procedure," in *Encyclopedia of Mathematics*, M. Hazewinkel, Ed., Springer, Berlin, Germany, 1992.