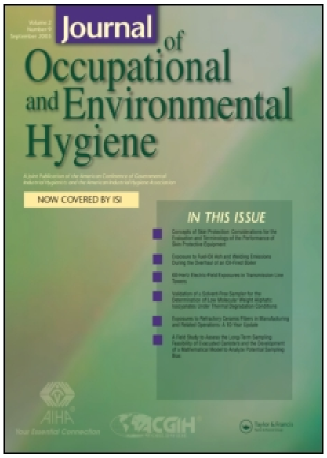

Journal of Occupational and Environmental Hygiene
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713657996

Statistical Methods for Establishing Equivalency of Several Sampling Devices

K. Krishnamoorthy [a]; Thomas Mathew [b]
[a] Department of Mathematics, University of Louisiana at Lafayette, Lafayette, Louisiana
[b] Department of Mathematics and Statistics, University of Maryland, Baltimore, Maryland

First Published on: 01 January 2008
To cite this Article: Krishnamoorthy, K. and Mathew, Thomas (2008) 'Statistical Methods for Establishing Equivalency of Several Sampling Devices', Journal of Occupational and Environmental Hygiene, 5:1, 15 - 21
To link to this article: DOI: 10.1080/15459620701755133
URL: http://dx.doi.org/10.1080/15459620701755133

Journal of Occupational and Environmental Hygiene, 5: 15–21
ISSN: 1545-9624 print / 1545-9632 online

DOI: 10.1080/15459620701755133

Statistical Methods for Establishing Equivalency of Several Sampling Devices

K. Krishnamoorthy[1] and Thomas Mathew[2]

[1]Department of Mathematics, University of Louisiana at Lafayette, Lafayette, Louisiana
[2]Department of Mathematics and Statistics, University of Maryland, Baltimore, Maryland

The problem of comparing several alternate sampling devices to the Occupational Safety and Health Administration (OSHA) standard, or the comparison of several sampling devices among themselves, is considered. A test based on the OSHA criterion that states that "90% of the readings of the sampling device should be within ±25% of the readings obtained by the standard" is developed. Type I error rates and powers of the test are studied using Monte Carlo simulation. The study indicates that the proposed test is quite satisfactory for applications. The method is illustrated using a simulated data set.

Keywords Intersection-Union test, power, Type I error

Address correspondence to: K. Krishnamoorthy, Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70504; e-mail: krishna@louisiana.edu.

INTRODUCTION

The comparison of an alternate sampling device to an Occupational Safety and Health Administration (OSHA) standard, or the comparison of two sampling devices or that of several sampling devices or methods, is a problem that is frequently encountered while developing cheaper or more efficient exposure assessment strategies. Applications of this type are numerous in the occupational hygiene literature. Some recent references include England et al.,(1) Rudzinski et al.,(2) O'Brien et al.,(3) Spicer and Gangloff,(4) Savely et al.,(5) and Wellons et al.(6)

Some of these authors have used the t-test to compare the means of the log-transformed data. Spicer and Gangloff(4) consider an application of Spearman's rank correlation for checking the agreement between microbial data from two comparison zones. They conclude that the nonparametric statistical analysis of bioaerosol data, as currently recommended for building assessment, has serious drawbacks; a very high Type II error (i.e., probability of not rejecting a false null hypothesis) has been noted for small samples.

First consider the problem of comparing just two monitoring methods or devices and ask the following question: Using sample data, how does one check if the two methods are equivalent, or how does one check if a new and cheaper (or easy to use) sampling device is equivalent to the OSHA standard? Consider the following options:

(i) Assuming lognormal distributions, compare the geometric means, i.e., means of the log-transformed data, using a student's t confidence interval or test
(ii) Assuming lognormal distributions, compare the arithmetic means using an appropriate procedure
(iii) Assuming lognormal distributions, develop a procedure for comparing the actual observations themselves instead of comparing their means.

Option *(i)* is simple to carry out, but it is clearly inappropriate because it merely compares the median of the original data (since comparing the means of the log-transformed data is equivalent to comparing the medians of the original data). Even if the comparison shows that the medians are close to each other, there is no guarantee that individual readings by the standard device and the alternative device are close. Option *(ii)* is an improvement as it does take into account the variability because the arithmetic mean of a lognormal distribution is a function of the mean and variance of the log-transformed data.

However, this is still not adequate to judge the closeness of the individual measurements from the two devices. The preferred approach toward solving the problem is to follow Option (*iii*), assuming that lognormality holds. In fact, OSHA regulations allow the use of an alternate sampling device for exposure monitoring, provided equivalence to the standard device can be established following the criterion that *90% of the readings of the sampling device be within plus or minus 25% of the readings obtained by the standard device, with a 95% confidence level as demonstrated by a statistically valid protocol* (see OSHA regulation document 1910.1043). Note that this criterion requires that the actual observations, and not the means be compared. Krishnamoorthy and Mathew(7) provided a method for testing if an alternative sampling device is equivalent to an OSHA standard, following the above criterion. The necessary tables are also provided in their paper.

In this article, the authors adapted the approach in Krishnamoorthy and Mathew[7] to test the equivalence of several sampling devices simultaneously or to test if several sampling devices are equivalent to an OSHA standard. For these problems, the *intersection-union* principle (see Casella and Berger,[8] Section 8.2.3) was used to justify the use of the Krishnamoorthy and Mathew test, developed for comparing just two devices. Thus, it was concluded that several sampling devices are equivalent to an OSHA standard if each of them is equivalent to the OSHA standard, where the latter equivalence is concluded using the Krishnamoorthy and Mathew test.

The intersection-union test is thus simple to use in applications, and it does not require any new table values. Indeed, the same table values given in Krishnamoorthy and Mathew[7] can be used to carry out the intersection-union test. Note also that each individual test will be carried out at the same nominal significance level, say 5%. This guarantees that the overall Type I error probability will not exceed 5%. Thus, from a theoretical point of view, no new test procedure is developed in this article. The actual Type I error rate of the intersection-union test can be significantly below the nominal level, and consequently, the test can have low power. Another important point worth noting is that the actual Type I error probability and power will depend on the correlation among the measurements obtained by the several sampling devices. This is a crucial part in evaluating the Type I error rates and powers of the simultaneous test.

The article is organized as follows: The setup and notations are introduced in the next section. For the purpose of establishing the equivalence of two sampling devices, the Krishnamoorthy and Mathew[7] test has been briefly reviewed. The intersection-union test is then described to test if several sampling devices are equivalent to an OSHA standard or if several sampling devices are equivalent among themselves. Numerical results are reported regarding the Type I error probability and power of the proposed test. In particular, the numerical results clearly bring out the dependence of the Type I error probability and power on the underlying correlation. Finally, an example is presented using simulated data.

PRELIMINARIES

Let $(X_j, Y_{1j}, \ldots, Y_{kj})$, $j = 1, \ldots, n$, be readings taken by the standard device (X) and the alternative devices $(Y_i), i = 1, \ldots, k$. That is, X_j denotes the reading by the standard device on the jth sampling unit, and Y_{ij} denotes the reading by the ith alternative device on the jth sampling unit. Usually the readings are taken from the same unit or location, and so it is commonly assumed that for a given sampling unit j, $X_j, Y_{1j}, \ldots, Y_{kj}$ are dependent. But the authors make the realistic assumption that $(X_1, Y_{11}, \ldots, Y_{k1}), (X_2, Y_{12}, \ldots, Y_{k2}) \ldots (X_n, Y_{1n}, \ldots, Y_{kn})$ are independent. To develop a test for equivalency, the authors further assume a multivariate lognormal distribution for $(X, Y_1, \ldots, Y_k)$ or, equivalently, a multivariate normal distribution for $(\ln(X), \ln(Y_1), \ldots, \ln(Y_k))$. Specifically, the authors assume that the random vector $(\ln(X), \ln(Y_1), \ldots, \ln(Y_k))$ follows a $(k + 1)$ variate normal distribution with

$$\text{mean vector } \boldsymbol{\mu} = \begin{pmatrix} \mu_X \\ \mu_1 \\ \vdots \\ \mu_k \end{pmatrix} \text{ and covariance matrix}$$

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{xx} & \sigma_{x1} & \cdots & \sigma_{xk} \\ \sigma_{1x} & \sigma_{11} & \ldots & \sigma_{1k} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{kx} & \sigma_{k2} & \cdots & \sigma_{kk} \end{pmatrix}. \tag{1}$$

Let $D_i = \ln(Y_i) - \ln(X), i = 1, \ldots, k$ and let $\mathbf{D} = (D_1, \ldots, D_k)^T$, where T denotes the transpose of a matrix. To identify the distribution of$D$, the authors write

$$\mathbf{D} = \mathbf{A} \begin{pmatrix} \ln(X) \\ \ln(Y_1) \\ \vdots \\ \ln(Y_k) \end{pmatrix}_{(k+1)\times 1} \quad \text{with}$$

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{pmatrix}_{k\times(k+1)}. \tag{2}$$

Furthermore, it is well known that the distribution of $\mathbf{D}$ is multivariate normal with

$$\text{mean vector } \boldsymbol{\mu}_D = \mathbf{A}\boldsymbol{\mu} = \begin{pmatrix} \mu_1 - \mu_x \\ \mu_2 - \mu_x \\ \vdots \\ \mu_k - \mu_x \end{pmatrix} = \begin{pmatrix} \mu_{d1} \\ \mu_{d2} \\ \vdots \\ \mu_{dk} \end{pmatrix}, \text{ say,} \tag{3}$$

and

$$\text{variance-covariance matrix } \boldsymbol{\Sigma}_D = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T = \begin{pmatrix} \eta_{11} & \eta_{12} & \cdots & \eta_{1k} \\ \eta_{21} & \eta_{22} & \cdots & \eta_{2k} \\ \vdots & \vdots & \vdots & \vdots \\ \eta_{k1} & \eta_{k2} & \cdots & \eta_{kk} \end{pmatrix}, \text{ say.} \tag{4}$$

Let θ_i denote the proportion of Y_i measurements that fall between $(1-\delta)X$ and $(1+\delta)X$. Since all the random variables are positive, the authors can express

$$\theta_i = P\left(1-\delta \le \frac{Y_i}{X} \le 1+\delta\right) = P\Bigg(\ln(1-\delta) \le \ln\left(\frac{X}{Y_i}\right) \le \ln(1+\delta)\Bigg) = P(\ln(1-\delta) \le D_i \le \ln(1+\delta)), \quad i = 1, \ldots, k. \tag{5}$$

To establish that all the alternative devices are equivalent to the standard device, the authors should check if the collected measurements provide evidence to conclude that $\theta_i > 1-p$, $i = 1, \ldots, k$, where p is a small value specified in advance, usually, 0.1, 0.05, or 0.01. If one wants to use the OSHA criterion to establish the equivalency of the devices, then one should use $\delta = 0.25$. The authors want to test the statistical hypotheses that

$$H_0 : \theta_i \le 1-p \text{ for some } i \text{ vs. } H_a : \theta_i > 1-p \text{ for all } i. \tag{6}$$

Also note that the hypotheses in Eq. 6 are set up so that rejection of the null hypothesis indicates that all alternate devices are equivalent to the standard device.

To develop a test for Eq. 6, the authors first define the necessary summary statistics. Let $x_1, \ldots, x_n$ be a sample of measurements obtained using the standard device and $y_{i1}, \ldots, y_{in}$ be measurements obtained using the ith alternative device, $i = 1, \ldots, k$. Define

$$u_{ij} = \ln(y_{ij}) - \ln(x_j), \quad j = 1, \ldots, n, \bar{u}_i = \frac{1}{n}\sum_{j=1}^{n} u_{ij} \text{ and}$$

$$s_{ui}^2 = \frac{1}{n-1}\sum_{j=1}^{n}(u_{ij}-\bar{u}_i)^2, \quad i = 1, \ldots, k. \tag{7}$$

Notice that the sample mean $\bar{u}_i$ is an estimate of the mean μ_{di} in (3) and the sample variance s_{ui}^2 is an estimate of the variance η_{ii} in (4).

A TEST FOR EQUIVALENCY OF A SINGLE ALTERNATIVE DEVICE TO THE STANDARD

For easy reference, Krishnamoorthy and Mathew's[7] test for establishing the equivalency of a single alternative device (say, the ith device) to the standard device is briefly outlined. In this case, the hypotheses of interest are

$$H_0 : \theta_i \le 1-p \text{ vs. } H_a : \theta_i > 1-p, \tag{8}$$

where $\theta_i = P(1-\delta \le Y_i/X \le 1+\delta) = P(\ln(1-\delta) \le D_i \le \ln(1+\delta))$. Krishnamoorthy and Mathew pointed out that the problem is somewhat easier to handle if a slightly different hypotheses is used compared with those in Eq. 8. Let $a = \ln(1-\delta)$ and $b = \ln(1+\delta)$. Noticing that D_j is normally distributed with mean μ_{di} and variance η_{ii}, it can be checked that $\theta_i > 1-p$ if

$$a < \mu_{di} - z_{1-p/2}\sqrt{\eta_{ii}} \text{ and } \mu_{di} + z_{1-p/2}\sqrt{\eta_{ii}} < b, \tag{9}$$

where z_c is the 100cth percentile of a standard normal distribution. This is because the conditions in Eq. 9 are equivalent to the condition that the tail probabilities $P(D_i \le a)$ and $P(D_i \ge b)$ are both less than or equal to $p/2$, which implies that $\theta_i = P(a < D_i < b) > 1-p$. Thus, instead of the hypotheses in Eq. 8, consider

$$H_{0i} : \mu_{di} - z_{1-p/2}\sqrt{\eta_{ii}} \le a \text{ or } b \le \mu_{di} + z_{1-p/2}\sqrt{\eta_{ii}} \quad \text{vs.} \quad H_{ai} : a < \mu_{di} - z_{1-p/2}\sqrt{\eta_{ii}} \text{ and } \mu_{di} + z_{1-p/2}\sqrt{\eta_{ii}} < b. \tag{10}$$

It is easy to see that the H_0 in Eq. 10 holds if H_0 in Eq. 8 holds, and the H_a in Eq. 8 holds if H_a in Eq. 10 holds. In other words, whenever the H_0 in Eq. 10 is rejected at the level α, the H_0 in Eq. 8 will also be rejected at a level not exceeding α.

As $(\bar{u}_i, s_{ui}^2)$ is sample estimate of (μ_{di}, η_{ii}), a natural test is the one that rejects the null hypothesis in (10) whenever $a < \bar{u}_i - cs_{ui}$ and $\bar{u}_i + cs_{ui} < b$, where $\bar{u}_i$ and s_{ui} are as defined in Eq. 7, and the critical value c is to be determined so that the Type I error rate

$$P(a < \bar{u}_i - cs_{ui} \text{ and } \bar{u}_i + cs_{ui} < b | H_0) \le \alpha. \tag{11}$$

Krishnamoorthy and Mathew[7] have provided values of c for $\alpha = .01, .05$ and $.1$, $1-p = .9, .95$ and $.99$, and for various values of n ranging from 5 to 1000 and ∞. Thus, a test for establishing equivalency of the jth alternative device to the standard one, according to the OSHA criterion mentioned earlier, can be easily carried out by choosing $a = \ln(1-\delta) = \ln(.75) = -0.28768$ and $b = \ln(1+\delta) = \ln(1.25) = 0.22314$, and then checking if $a < \bar{u}_i - cs_{ui}$ and $\bar{u}_i + cs_{ui} < b$; if these inequalities hold, then it can be concluded that the jth device is equivalent to the standard device at the level of significance α.

A TEST FOR EQUIVALENCY OF SEVERAL ALTERNATIVE DEVICES TO THE STANDARD

The above test, along with the *intersection-union* principle, will now be used to develop a test for the hypotheses in Eq. 6. For details about the intersection-union principle, see Casella and Berger,[8] Section 8.2.3. Consider the hypotheses

$$H_0 : \mu_{di} - z_{1-p/2}\sqrt{\eta_{ii}} \le a \text{ or } b \le \mu_{di} + z_{1-p/2}\sqrt{\eta_{ii}} \text{ for some } i, \quad \text{vs.} \quad H_a : a < \mu_{di} - z_{1-p/2}\sqrt{\eta_{ii}} \text{ and } \mu_{di} + z_{1-p/2}\sqrt{\eta_{ii}} < b \text{ for all } i. \tag{12}$$

It is easy to check that the above H_a implies the H_a in Eq. 6, and so, whenever the H_0 in Eq. 12 is rejected, the null hypothesis in Eq. 6 is also rejected. So a test for Eq. 12 is developed using intersection-union principle. Notice that the null hypothesis in Eq. 12 is the union of H_{0i}'s in Eq. 10, and the alternative hypothesis in Eq. 12 is the intersection of H_{aj}'s in Eq. 10. That is, the hypotheses in Eq. 12 can be expressed as

$$H_0 : \bigcup_{i=1}^{k} H_{0i} \text{ vs. } H_a : \bigcap_{i=1}^{k} H_{ai},$$

where H_{0i} and H_{ai} are given in Eq. 10. Notice that the null hypothesis is true if one or more of the individual H_{0i}'s in Eq. 10 is true, and so it will be rejected only when all H_{0i}'s are false. Equivalently, the above H_0 will be rejected only when all the alternative devices are equivalent to the standard one. Thus, the intersection-union test rejects the null hypothesis if

$$a < \bar{u}_i - cs_{ui} \text{ and } \bar{u}_i + cs_{ui} < b, \quad i = 1, \ldots, k,$$

or equivalently, if all the intervals $\bar{u}_1 \pm cs_{u1}, \ldots, \bar{u}_k \pm cs_{uk}$ are contained in the interval $(a, b) = (\ln(1-\delta), \ln(1+\delta))$. In other words, if all these intervals lie within (a, b), then it can be concluded that all the alternate devices are equivalent to the standard one at the overall level α. The value of c is determined by Eq. 11, and it does not depend on the number of devices to be compared; as mentioned earlier, the values of c have been tabulated in Krishnamoorthy and Mathew,[7] and so the test is easy to apply.

SIZE AND POWER PROPERTIES OF THE INTERSECTION-UNION TEST

The intersection-union test is known to be a level α test; that is, the Type I error rate never exceeds the nominal level α. However, the test could be very conservative. As a consequence, it may not be very powerful. To appraise the Type I error rates and powers of the test, they are estimated using Monte Carlo simulation consisting of 100,000 runs. To estimate the maximum Type I error rate of the test, consider the situation of comparing two alternative devices to a standard. For this, the authors chose parameters $\mu_{di} = \mu_0 = (a+b)/2$ and $\sqrt{\eta_{ii}} = \sigma_0 = (b-a)/(2z_{1-p/2},$, $i = 1, 2$, which lie at the boundary of the parameter space underH_0 in Eq. 12; that is, when $(\mu_{di}, \sqrt{\eta_{ii}}) = (\mu_0, \sigma_0)$, they have $\mu_{di} - z_{1-p/2}\sqrt{\eta_{ii}} = a$ and $\mu_{di} + z_{1-p/2}\sqrt{\eta_{ii}} = b$, $i = 1, 2$. The values of a and b were chosen according to the OSHA criterion, and they are $a = \ln(1-\delta) = \ln(1-.25) = -0.287682$, $b = \ln(1+\delta) = \ln(1+.25) = 0.223144$, $p = 0.1$ and $\alpha = 0.05$. The Type I error rate also depends on the correlation between D_1 and D_2 (where D_1 and D_2 are defined below Eq. 1). To find the Type I errors, the values of (μ_{d1}, σ_{d1}) and (μ_{d2}, σ_{d2}) are chosen so that they are within or at the boundaries of the parameter space under H_0 in Eq. 12. The estimated Type I error rates are reported in Table I. The estimated Type I error rates are within the nominal level 0.05, as they should be. Also, it is observed that the Type I error rates are close to the nominal level for large samples and/or for large values of the correlation coefficient . This implies (as will be seen later) that the test will be more powerful if the correlation coefficient between D_1 and D_2 is large.

TABLE I. Type I Error Rates of the Intersection-Union Test

$(\mu_{d_1},\sigma_{d_1}), (\mu_{d_2}, \sigma_{d_2})$	ρ	$n = 10$ (2.375)^A	$n = 20$ (2.064)	$n = 50$ (1.873)
(μ_0, σ_0) , (μ_0, σ_0)	±0.820	0.020	0.022	0.023
(μ_0, σ_0) , (μ_0, σ_0)	±0.912	0.027	0.029	0.031
(μ_0, σ_0) , (μ_0, σ_0)	±0.995	0.047	0.048	0.048
(μ_0, σ_0) , $(\mu_0, .1)$	±0.323	0.029	0.045	0.050
(μ_0, σ_0) , $(\mu_0, .1)$	±0.500	0.037	0.047	0.049
(μ_0, σ_0) , $(\mu_0, .1)$	±0.998	0.050	0.050	0.050

Notes: $a = \ln(1-\delta) = \ln(.75)$, $b = \ln(1+\delta) = \ln(1.25)$;
$\mu_0 = (a+b)/2 = -0.032269$; $\sigma_0 = (b-a)/(2z_{.95}) = 0.15228$;
$p = 0.1$ and $\alpha = 0.05$; ρ = correlation between D_1 and D_2.
^{A}The numbers in parentheses are the critical values c satisfying Eq. 11 when $p = 0.1$ and $\alpha = 0.05$; these values are taken from Table I of Krishnamoorthy and Mathew.[7]

The authors also estimated the power of the test for some parameter values under H_a in Eq. 12 and for some values of ρ, and they are given in Table II. Notice that the values of (μ_{d1}, σ_{d1}) and (μ_{d2}, σ_{d2}) in Table II are chosen so that they are within the parameter space under H_a in Eq. 12. In Table II, for fixed parameters and ρ, the power increases as the sample size increases, which is a natural requirement of a test. Furthermore, if $D_1 = \ln(Y_1) - \ln(X)$ and $D_2 = \ln(Y_2) - \ln(X)$ are highly correlated, then the Type I error rates are very close to the nominal level (Table I), and as a result, the test is expected to be more powerful when the true parameters lie in the parameter space under H_a (Table II). In general, it is observed that, for a fixed value of $(\mu_{d1}, \sqrt{\eta_{11}}, \mu_{d2}, \sqrt{\eta_{22}})$, the power is increasing with increasing ρ. In the present problem, as the measurements are obtained using different devices from the same unit or location, one can expect that X, Y_1, and Y_2 are highly correlated, and so D_1 and D_2 are also expected to be strongly correlated. Thus, it is expected that

TABLE II. Powers of the Intersection-Union Test

$(\mu_{d_1},\sigma_{d_1}), (\mu_{d_2}, \sigma_{d_2})$	ρ	$n = 10$ (2.375)^A	Y (2.064)	$n = 50$ (1.873)
$(\mu_0,.1),(\mu_0,.1)$	±0.580	0.326	0.729	0.993
$(\mu_0,.1),(\mu_0,.1)$	±0.864	0.423	0.787	0.995
$(\mu_0,.1),(\mu_0,.1)$	±0.909	0.448	0.798	0.995
$(-.05,.1),(-.1,.1)$	±0.500	0.143	0.289	0.520
$(-.05,.1),(-.1,.1)$	±0.750	0.172	0.306	0.528
$(-.05,.1),(-.1,.1)$	±0.950	0.211	0.323	0.540

Notes: $a = \ln(1-\delta) = \ln(.75)$, $b = \ln(1+\delta) = \ln(1.25)$;
$\mu_0 = (a+b)/2 = -0.032269$; $\sigma_0 = (b-a)/(2z_{.95}) = 0.152280$;
ρ = correlation coefficient between D_1 and D_2.
^{A}The numbers in parentheses are the critical values c satisfying Eq. 11 when $p = 0.1$ and $\alpha = 0.05$; these values are taken from Table I of Krishnamoorthy and Mathew.[7]

the proposed test will be very satisfactory and powerful for practical applications.

Also note that if a given power is desired at a specified set of parameter values, numerical results similar to those in Table II can be used to determine the value of the sample size n. Other quantities being fixed, the required sample size is a decreasing function of $|\rho|$, as already noted.

UNION INTERSECTION TEST FOR COMPARING SEVERAL DEVICES

Suppose one wants to assess the equivalency of k devices. Let $y_{i1}, \ldots, y_{in}$ be a sample of measurements obtained using the ith device, $i = 1, \ldots, k$. Define

$$\theta_{ij} = P(a \leq \ln(Y_i) - \ln(Y_j) \leq b), \quad i > j, \quad j = 1, \ldots, k.$$

The hypotheses of interest are

$$H_0 : \theta_{ij} \leq 1 - p \text{ for some } i > j \text{ vs. } H_a : \theta_{ij} > 1 - p \text{ for all } i > j,\ j = 1, \ldots, k. \quad \textbf{(13)}$$

For this problem, define summary statistics as follows. Let $w_{ijl} = \ln(y_{il}) - \ln(y_{jl}),\ i > j, l = 1, \ldots, n$. Define

$$\bar{w}_{ij} = \frac{1}{n}\sum_{l=1}^{n} w_{ijl} \text{ and } s_{ij}^2 = \frac{1}{n-1}\sum_{l=1}^{n}(w_{ijl} - \bar{w}_{ij})^2, \quad i > j,\ j = 1, \ldots, k.$$

The null hypothesis in Eq. 13 will be rejected if all $k(k-1)/2$ intervals $\bar{w}_{ij} \pm cs_{ij},\ i > j,\ j = 1, \ldots, k$, are contained in the interval (a, b). Again, the value of c does not depend on k, and it is determined solely by the sample size n, p and α. Furthermore, note that the overall Type I error rate of the test will not exceed α.

AN ILLUSTRATIVE EXAMPLE

As mentioned in the Introduction, there are several papers addressing the problem of comparing several devices, and some of them reported only the means and variances of the measurements, obtained using the different devices. As these test procedures check the closeness of the measurements, individual measurements are needed to apply the test. Since the authors were unable to obtain any published data, for illustration purposes, generated a sample of 20 observations from a trivariate normal distribution (which can be regarded as log-transformed data from a trivariate lognormal distribution) were generated and are reported in Table III.

If dealing with a practical data set, then one has to check the assumption of normality. That is, one needs to verify that the data $(\ln(x), \ln(y_1), \ln(y_2))$ are from a trivariate normal distribution. Even though the data was generated from such a distribution, for the sake of illustration, one checks the assumption of normality using an approach given in Johnson and Wichern.[9, p.177] Toward this, let $X_i = (\ln(x_i), \ln(y_{1i}), \ln(y_{2i}))$, $i = 1, \ldots, 20$. The sample mean vector $\bar{\mathbf{X}}$ and the sample

TABLE III. Simulated Data

Unit	ln(x)	ln(y_1)	ln (y_2)	u_1	u_2	ln(y_1) − ln(y_2)	Q_j
1	1.430	1.449	1.460	−0.019	−0.030	−0.011	0.231
2	4.461	4.395	4.583	0.066	−0.122	−0.188	0.272
3	2.462	2.539	2.441	−0.077	0.021	0.098	0.431
4	2.985	2.929	2.925	0.056	0.060	0.004	0.906
5	4.296	4.055	4.324	0.241	−0.028	−0.269	1.736
6	1.428	1.177	1.494	0.251	−0.066	−0.317	2.043
7	4.787	4.486	4.643	0.301	0.144	−0.157	2.143
8	5.275	5.296	5.271	−0.021	0.004	0.025	2.384
9	3.691	3.966	3.782	−0.275	−0.091	0.184	2.464
10	3.471	3.523	3.495	−0.052	−0.024	0.028	2.625
11	1.200	1.100	1.179	0.100	0.021	−0.079	2.686
12	4.498	4.566	4.414	−0.068	0.084	0.152	2.946
13	3.537	3.500	3.528	0.037	0.009	−0.028	2.949
14	2.298	2.540	2.268	−0.242	0.030	0.272	3.485
15	3.778	4.119	3.739	−0.341	0.039	0.380	3.544
16	1.229	1.184	1.271	0.045	−0.042	−0.087	3.789
17	1.920	2.027	1.829	−0.107	0.091	0.198	4.013
18	1.704	1.793	1.648	−0.089	0.056	0.145	4.816
19	3.858	4.199	3.923	−0.341	−0.065	0.276	4.888
20	2.944	2.929	3.050	0.015	−0.106	−0.121	8.650

Notes: $\ln(x)$ = log-transformed measurements by the standard device, $\ln(y_i)$ = log-transformed measurements by the alternative device i, i = 1,2, $u_i = \ln(y_i) - \ln(x)$; Q_j's are observed quantiles in Eq. 14.

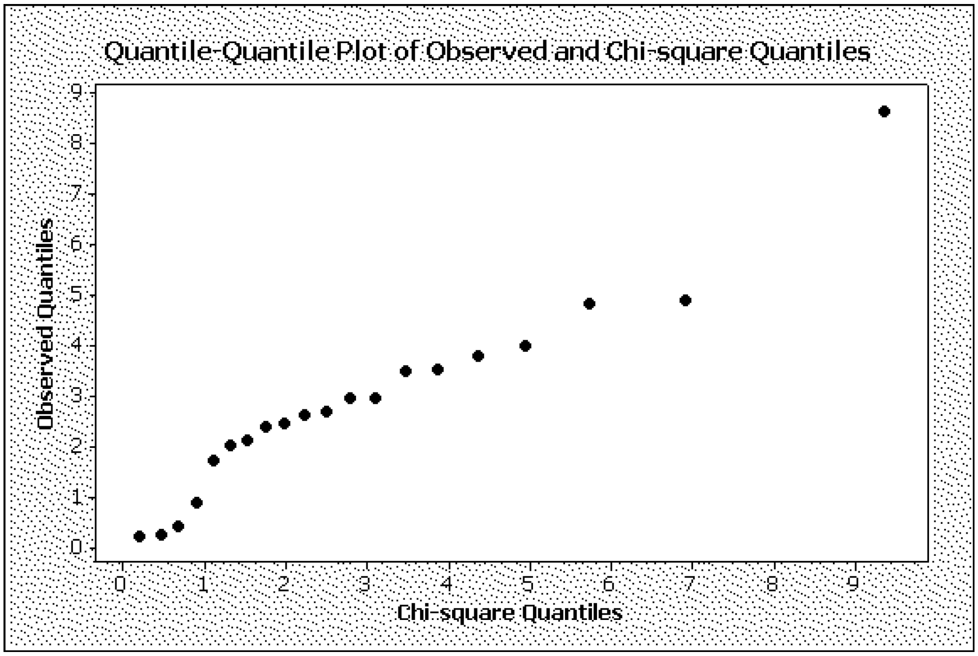

FIGURE 1. Quantile-quantile plot of simulated data

covariance matrix $\mathbf{S}$ of the X_i's were computed as

$$\bar{\mathbf{X}} = \begin{pmatrix} 3.0626 \\ 3.0886 \\ 3.0634 \end{pmatrix} \text{ and } \mathbf{S} = \begin{pmatrix} 1.6643 & 1.6699 & 1.6605 \\ 1.6699 & 1.7082 & 1.6672 \\ 1.6605 & 1.6672 & 1.6615 \end{pmatrix}.$$

Let

$$Q_i = (X_i - \bar{\mathbf{X}})^T \mathbf{S}^{-1}(X_i - \bar{\mathbf{X}}), \quad i = 1, \ldots, 20. \qquad \textbf{(14)}$$

To verify the assumption of multivariate normality, one needs to check that Q_i's fit a chi-square distribution with degrees of freedom 3. The computed values of Q_i's are given in Table III, and the quantile-quantile plot of the Q_i's and the χ^2 distribution with degrees of freedom 3 is given in Figure 1. As the plot shows a linear pattern close to the 45° line, the normality assumption is tenable.

The test for establishing equivalency of the y_1 and y_2 devices to the standard device x is illustrated below. The necessary summary statistics were computed as

$$\bar{u}_1 = 0.0260, s_{u1} = 0.181, \bar{u}_2 = 0.0008 \text{ and } s_{u2} = 0.070.$$

For $n = 20$, $p = 0.10$, and $\alpha = 0.05$, the critical value c is 2.064; this value can be obtained from Table I of Krishnamoorthy and Mathew.[7] Using these values, the authors computed

$$\bar{u}_1 \pm cs_{u1} = (-0.34758, 0.39958) \text{ and}$$
$$\bar{u}_2 \pm cs_{u2} = (-0.14348, 0.14548).$$

Comparing these results with $a = \ln(0.75) = -0.287682$ and $b = \ln(1.25) = 0.223144$, it can be seen that only the second alternative device satisfies $a < \bar{u}_2 - cs_{u2}$ and $\bar{u}_2 + cs_{u2} < b$, and so it cannot be concluded that both alternative devices agree adequately with the standard device. It is highly probable, however, that the second device agrees adequately with the standard device.

To compare all three devices among themselves (i.e., assume that device x is also one of the devices to be compared with the other two), the mean and standard deviation of the difference $w_{12} = \ln(y_1) - \ln(y_2)$ needs to be computed. The mean as $\bar{w}_{12} = 0.025$ and the standard deviation as $s_{12} = 0.188$. were computed. The interval to be compared with $(a, b) = (-0.28768, 0.22314)$ is $\bar{w}_{12} \pm (2.064)s_{12} =$ (-0.36303, 0.39898). As this latter interval is not contained in (a, b), it cannot be concluded that alternative devices 1 and 2 are equivalent.

ACKNOWLEDGMENTS

This research was supported by a grant from the National Institute of Occupational Safety and Health (NIOSH).

REFERENCES

1. **England, E., R. Key-Schwartz, J. Lesage, G. Carlton, R. Streicher, and R. Song:** Comparison of sampling methods for monomer and

polyisocyanates of 1,6-hexamethylene disocyanate during spray fishing operations. *Appl. Occup. Environ. Hyg. 15*:472–478 (2000).
2. **Rudzinski, W.E., J. Yin, E. England, G. Carlton, R. Key-Schwartz, and J. Lesage:** A comparison of solid sampler methods for the determination of hexamethylene-based isocyanates in spray-painting operations. *Am. Ind. Hyg. Assoc. J. 62*:246–250 (2001).
3. **O'Brien, D.M., G.M. Piacitelli, W.K. Sieber, R.T. Hughes, and J.D. Catalano:** An evaluation of short-term exposures to metalworking fluids in small machine shops. *Am. Ind. Hyg. Assoc. J. 62*:342–348 (2001).
4. **Spicer, R.C., and H.J. Gangloff:** Limitations in application of Spearman's rank correlation to bioaerosol sampling data. *Am. Ind. Hyg. Assoc. J. 61*:362–366 (2000).
5. **Savely, S.M., R.J. Emery, and T.H. Connor:** A comparison of methods for determining lead content in drinking water: A portable anodic stripping voltammetry instrument method versus the standard EPA 239.2 method.*Am. Ind. Hyg. Assoc. J. 61*:557–562 (2000).
6. **Wellons, S.L., E.G. Trawick, M.F. Stowers, S.L.P. Jordan, and T.L. Wass:** Laboratory and hospital evaluation of four personal monitoring methods for glutaraldehyde in ambient air. *Am. Ind. Hyg. Assoc. J. 59*:96–103 (1998).
7. **Krishnamoorthy, K., and T. Mathew:** Statistical methods for establishing equivalency of a sampling device to the OSHA standard. *Am. Ind. Hyg. Assoc. J. 63*:567–571 (2002).
8. **Casella, G., and R.L. Berger:** *Statistical Inference*. Pacific Grove, Calif.: Duxbury.
9. **Johnson, R.A., and D.W. Wichern:** *Applied Multivariate Statistical Analysis*. Saddle River, N.J.: Prentice Hall.