

Confidence Interval Estimation of the Correlation in the Presence of Non-Detects

Courtney E. McCracken¹, Stephen W. Looney²

¹Center for Research and Evaluation, Kaiser Permanente Georgia, Atlanta, USA

²Department of Population Health Sciences, Augusta University, Augusta, USA

Email: Courtney.E.McCracken@kp.org, slooney@augusta.edu

How to cite this paper: McCracken, C.E. and Looney, S.W. (2021) Confidence Interval Estimation of the Correlation in the Presence of Non-Detects. *Open Journal of Statistics*, 11, 463-475.

<https://doi.org/10.4236/ojs.2021.113029>

Received: March 17, 2021

Accepted: June 22, 2021

Published: June 25, 2021

Copyright © 2021 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This article deals with correlating two variables that have values that fall below the known limit of detection (LOD) of the measuring device; these values are known as non-detects (NDs). We use simulation to compare several methods for estimating the association between two such variables. The most commonly used method, simple substitution, consists of replacing each ND with some representative value such as LOD/2. Spearman's correlation, in which all NDs are assumed to be tied at some value just smaller than the LOD, is also used. We evaluate each method under several scenarios, including small to moderate sample size, moderate to large censoring proportions, extreme imbalance in censoring proportions, and non-bivariate normal (BVN) data. In this article, we focus on the coverage probability of 95% confidence intervals obtained using each method. Confidence intervals using a maximum likelihood approach based on the assumption of BVN data have acceptable performance under most scenarios, even with non-BVN data. Intervals based on Spearman's coefficient also perform well under many conditions. The methods are illustrated using real data taken from the biomarker literature.

Keywords

Confidence Interval, Coverage Probability, Left Censoring, Limit of Detection, Maximum Likelihood, Spearman Correlation

1. Introduction

In research studies involving clinical measurements such as biomarker concentrations, it is quite common to have specimens for which the concentration of the analyte is non-zero, but below the analytic limit of detection (LOD); that is, the measuring device used to determine the level of the analyte in the biological

specimen is unable to measure the concentration. For such specimens, all that we know is that the analyte is present and that the concentration is less than the LOD. Non-zero observations that are less than the LOD are commonly referred to as non-detects (NDs). For purposes of statistical analysis, non-detects as we have defined them are considered to be left-censored.

In the study by Amorim and Alvarez-Leite [1], NDs were of particular concern. The authors evaluated urinary *o*-cresol as a biomarker of toluene exposure by calculating the Pearson correlation coefficient (PCC) relating the level of *o*-cresol in the urine of workers exposed to toluene to the level of urinary hippuric acid in the same workers. Out of the 54 urine samples that Amorim and Alvarez-Leite analyzed, the *o*-cresol concentration was below its LOD (0.2 µg/ml) in 39 (72%); out of these 39 samples, the concentration of hippuric acid was below its LOD (0.1 mg/ml) in 4 (10%). Thus, there were only 15 samples for which the data were “complete” for both biomarkers. In the study by Atawodi *et al.* [2], NDs were also of great concern. These authors examined various hemoglobin adducts as biomarkers of tobacco smoke exposure by comparing the adduct levels of 18 current smokers with those of 52 “never smokers”. The hemoglobin adduct levels were below the LOD (9 fmol HPB/g Hb) in 7 (13%) of the 52 samples from the “never smokers”.

Perhaps the method that is most commonly used to deal with samples in which there are NDs is to remove the NDs and perform the statistical analysis using only the “complete data”. In the study by Lagorio *et al.* [3], the authors used this approach in their examination of trans, trans muconic acid (*t, t*-MA) as a biomarker for low-level benzene exposure. They calculated the Pearson correlations among *t, t*-MA concentrations in urine samples obtained from 10 Estonian shale oil workers; these concentrations were estimated using high-performance liquid chromatography (HPLC) following three different pre-analytical procedures (methanol dilution, filtration, ether extraction). Another method that is commonly used to analyze data sets in which NDs are present is to use “simple substitution”; in other words, a value is substituted in place of the NDs and then the “usual” statistical analysis is performed on the resulting “new” sample of data. The most commonly used values in this crude type of imputation include the LOD [1] [2], and LOD/2 [4]. We contend that the approaches that are commonly used to handle NDs have several shortcomings; the purpose of our study was to evaluate some of the commonly used methods, along with some that are not so common.

Nonparametric methods have also been used to deal with samples in which NDs are present. In this approach, one treats all NDs as if they were tied at some value just below the LOD of the respective measuring device. For example, if one wished to correlate two analytes X and Y , at least one of which was undetectable in some specimens, one could use Spearman’s Rank Correlation Coefficient (denoted here by r_s). In this method, the original X and Y values are replaced by their respective midranks and the NDs are assigned the smallest midrank for

that variable. If one wished to compare the analyte levels between two groups and NDs were present in at least one of the two samples, one could compute the midranks after combining the data into a single sample; each of the NDs would then be assigned the smallest mid-rank. One could then use the Mann-Whitney-Wilcoxon (M-W-W) test or other nonparametric two-sample method based on these mid-ranks to compare the two groups in terms of the level of the analyte. In their evaluation of potential biomarkers of exposure to tobacco smoke, Atawodi *et al.* [2] used the M-W-W test to compare smokers and never smokers in terms of the level of a hemoglobin adduct; NDs were present in the sample of never smokers.

In a simulation study, Wang [5] found that none of the “standard” methods described above perform satisfactorily when correlating two measurements X and Y that are both subject to left-censoring, especially if X and Y are strongly positively correlated ($\rho \geq 0.5$). If the distribution of X and Y is bivariate normal (BVN), a preferable approach is to estimate the Pearson correlation between X and Y using maximum likelihood (ML) [6]. Multiple imputation could also be used if the appropriate missing data mechanism is present and other conditions are satisfied [7].

2. Methods

We performed a Monte Carlo simulation to compare 5 methods that can be used to obtain point and confidence interval (CI) estimates of the correlation between X and Y when both X and Y are left censored. These methods included several of the “standard methods” that have been used to analyze data in which NDs are present, as well as the ML method [6]. The methods compared were as follows:

(1) Simple Substitution: replace each ND by

- (a) LOD ;
- (b) $LOD/2$;
- (c) $LOD/\sqrt{2}$.

(2) Complex Substitution [8] [9]: Substitute $E(X_i | X_i < LOD_x)$ in place of each ND among the x -values and substitute $E(Y_i | Y_i < LOD_y)$ in place of each ND among the y -values. In other words, replace each ND for each variable by the conditional mean of that variable, given that it is known that the value is less than the LOD for that variable.

(3) Random Imputation from a Uniform Distribution: Substitute a randomly selected value from the interval $[0, LOD_x]$ in place of each ND among the x -values and substitute a randomly selected value from the interval $[0, LOD_y]$ in place of each ND among the y -values. The rationale for this method is that there may be nothing special about using the LOD or some fraction of it in place of the NDs; why not use any randomly generated number between 0 and the LOD.

(4) Maximum Likelihood [6].

(5) Spearman Correlation: All NDs among the x -values are treated as if they were tied at some value smaller than the smallest observed x -value; similarly, all

NDs among the y -values are treated as if they were tied at some value smaller than the smallest observed y -value. Assign midranks in the usual way and calculate r_s using these midranks.

For estimation methods (1)-(4) above, we used a 2nd-order Fisher z -transformation, which provides a more accurate estimate of the variance of $z(\hat{\rho})$, in the calculation of the 95% C.I. for ρ . The coverage probability of CIs based on this method has been shown to be closer to the nominal level than those based on the usual Fisher z -transformation, and the 2nd-order z -transformation poses no computational difficulties [10]. For estimation method (5), we evaluated both the jackknife and approximate bootstrap confidence interval (ABC) as methods for finding a 95% C.I. for the population value of the Spearman correlation. Defining the population value of the Spearman coefficient is controversial [11]; we followed Newton and Rudel [12] and defined the true value, ρ_s , to be the mean of the r_s values calculated from the Monte Carlo samples prior to applying the censoring schemes.

In the simulation study to compare the point estimation and confidence interval methods described above, we included various settings of several simulation parameters: 1) sample size ($n = 20, 30, 50, 75, 100, 200, 500$); 2) true correlation between X and Y prior to censoring ($\rho = -0.9, -0.6, -0.5, -0.25, 0.0, 0.1, 0.2, 0.25, 0.3, 0.4, 0.5, 0.6, 0.7, 0.75, 0.8$ and 0.9); 3) true bivariate distribution of X and Y (bivariate normal, bivariate gamma, bivariate beta); and 4) censoring proportions on X (p_1) and Y (p_2). We included 55 combinations of censoring percentages in the simulations, both balanced and unbalanced. Balanced combinations included $(p_1, p_2) = (0, 0), (10, 10), (20, 20), (25, 25), (30, 30), (40, 40), (50, 50), (60, 60), (70, 70), (75, 75), (80, 80),$ and $(90, 90)$. Unbalanced combinations included $(10, 0), (10, 5), (10, 50), (10, 75), (20, 50), (25, 75), (30, 75), (90, 45),$ and $(90, 0)$, along with 34 others. All together, we considered 18,480 different combinations of simulation parameter settings. We used a Monte Carlo simulation size (MCSS) of 5000.

Each of the point estimates and corresponding 95% C.I. procedures described above were evaluated in terms of the following criteria: 1) bias (and absolute bias), 2) median absolute deviation, 3) confidence interval width, and 4) confidence interval coverage probability (CP). In this article, we present results from our comparisons of the CP of 95% CIs based on the different estimation methods.

3. Results

The maximum likelihood estimate (MLE) performed best overall in terms of all of the criteria that we considered, and it can be recommended for estimating ρ even when the assumption of BVN is violated. However, the ML method may not be able to produce a point estimate (due to failure of the optimization routine to converge) for extreme negative values of ρ , small sample sizes and/or extremely heavy or imbalanced censoring. This is more likely to occur when the

joint distribution of X and Y differs substantially from the BVN in terms of multivariate skewness and kurtosis. Our simulation study was designed so that the non-BVN distributions that we included represented substantial departures from the BVN in terms of Mardia's measures of multivariate skewness and kurtosis (denoted by $\beta_{1,p}$ and $\beta_{2,p}$, respectively). For the bivariate normal, $\beta_{1,2} = 0$ and $\beta_{2,2} = 8$. For the bivariate gamma we used in the simulation study, $\beta_{1,2} = 3.5$ and $\beta_{2,2} = 12$. For the bivariate beta we used, $\beta_{1,2} = 3$ and $\beta_{2,2} = 10$. The ML method performed quite well for most settings of the simulation parameters even when the simulated data were generated from these non-BVN distributions. This is somewhat surprising since the MLEs were derived under the assumption that X and Y followed a BVN distribution.

Tables 1-3 contains brief summaries of our simulated CP results for simple substitution based on LOD/2, complex substitution, random imputation, Spearman's r_s (jackknife interval), and Maximum Likelihood. We do not present the results for simple substitution based on LOD or $LOD/\sqrt{2}$. The results for $LOD/\sqrt{2}$ were comparable to, but generally inferior than, those based on LOD/2 for almost all simulation parameter settings. Simple substitution based on LOD was not competitive with LOD/2 in terms of coverage probability. We do not include results for the ABC-based CIs for the Spearman coefficient since this method required considerably more computation time than the jackknife and provided little or no improvement when compared to the jackknife intervals in terms of CP.

Table 1 summarizes the effects of censoring proportions on CP for only a subset of the censoring proportions that we examined: $\{(0, 0) (0.1, 0.7) (0.25, 0.25) (0.25, 0.75) (0.5, 0.5) (0.75, 0.375) (0.9, 0) (0.9, 0.9)\}$. The results for each censoring proportion in **Table 1** were obtained by calculating the median CP over all settings of the other simulation parameters (namely, the true value of ρ or ρ_s , and sample size). The results labelled "non-normal" in the bottom half of the table were obtained by averaging the CP results for the bivariate gamma with those for the bivariate beta. For example, based on BVN simulated data, the median CPs over all other simulation parameter settings for censoring proportions $p_1 = 0.1$ and $p_2 = 0.7$, were 94.8% for the ML method, 91.5% for the Spearman coefficient, 81.0% for complex substitution (CS), 77.0% for simple substitution (SS) with LOD/2, and 34.6% for random imputation (RI). Based on the non-BVN simulated data, the median CPs for the same censoring proportions for the ML, Spearman, CS, SS, and RI methods were 93.9%, 92.9%, 81.4%, 84.0%, and 11.6%, respectively.

We adopted the "liberal" guideline proposed by Bradley [13] for evaluating the robustness of a statistical test to aid us in determining if the CP of a CI based on a particular method differed in any meaningful way from the nominal 95% confidence level. According to the Bradley criterion, if the true significance level α differs from the nominal level by no more than $\alpha/2$, one can conclude that the test is robust. If the true significance level differs by more than $\alpha/2$ from the no-

minal level (either above or below), one can conclude that the test is not robust. In the present study, we applied the Bradley criterion as follows: if the estimated CP differed from the 0.95 nominal confidence level by no more than 0.025, the CP for the confidence interval method was deemed to be within acceptable limits. If the estimated CP differed by more than 0.025 from the nominal confidence level, the CP for that method was deemed to be unacceptable. Thus, for a 95% CI, the estimated CP had to be between 92.5% and 97.5% for a CI procedure to be classified as “acceptable.”

The boldface values in **Table 1** indicate median CPs that were less than the lower acceptability criterion of 92.5%. Confidence intervals based on the ML method maintained an acceptable value of CP for all censoring proportions except (0.9, 0.9) with BVN data and (0, 0) and (0.9, 0.9) for non-BVN data. The CPs for CIs based on the Spearman coefficient were comparable to those for the ML-based CIs for many of the censoring proportions, but did not achieve the 92.5% acceptability criterion in several instances, especially for BVN simulated data. The complex substitution, simple substitution and random imputation-based CIs achieved the 92.5% level only when there was no censoring in the BVN simulated data.

Table 1. Comparison of median coverage probability of 5 C.I. methods, by censoring proportions.

Distribution	Censoring Proportions (p_1, p_2)	Method				
		Simple Substitution $LOD/2$	Complex Substitution	Random Imputation	Spearman's r_s (Jackknife Interval)	Maximum Likelihood
Normal	(0, 0)	94.8	94.8	94.8	93.8	94.8
Normal	(0.1, 0.7)	77.0	81.0	34.6	91.5	94.8
Normal	(0.25, 0.25)	85.8	92.2	77.8	94.1	94.8
Normal	(0.25, 0.75)	73.0	76.2	22.8	89.9	94.8
Normal	(0.5, 0.5)	78.9	84.3	43.6	92.5	94.7
Normal	(0.75, 0.375)	71.5	75.7	21.7	88.4	94.9
Normal	(0.9, 0)	21.9	26.5	0.9	33.8	94.9
Normal	(0.9, 0.9)	27.3	27.2	0.1	61.8	90.0
Non-Normal	(0, 0)	90.7	90.7	90.7	93.9	91.0
Non-Normal	(0.1, 0.7)	84.0	81.4	11.6	92.9	93.9
Non-Normal	(0.25, 0.25)	90.4	89.3	27.3	94.1	93.4
Non-Normal	(0.25, 0.75)	81.6	76.6	8.2	91.8	94.0
Non-Normal	(0.5, 0.5)	86.3	85.5	17.1	93.7	93.8
Non-Normal	(0.75, 0.375)	82.1	80.0	8.9	92.5	94.0
Non-Normal	(0.9, 0)	60.6	60.0	0.6	69.7	93.2
Non-Normal	(0.9, 0.9)	40.5	38.6	0.1	74.3	85.8

The effects of the true value of the correlation parameter (either Pearson's correlation or Spearman's coefficient) on the median CP of the CIs based on the various methods are illustrated in **Table 2**. As in **Table 1**, boldface values in **Table 2** identify median CPs that did not achieve the lower acceptability criterion of 92.5%. Confidence intervals based on the ML method achieved an acceptable CP for all values of ρ except -0.9 with BVN data; however, ML-based CIs did not perform as well with non-BVN data. Confidence intervals based on Spearman's coefficient generally performed as well as those based on the ML method for non-BVN simulated data, but failed to achieve the 92.5% acceptability criterion for several values of the true correlation when BVN simulated data were used. The complex substitution and simple substitution CIs achieved the 92.5% level for very few of the settings for the true value of ρ . The random imputation CIs achieved the 92.5% level only when $\rho = 0$.

The effects of sample size (n) on the median CP of the CIs based on the various methods are illustrated in **Table 3**. As in **Table 1** and **Table 2**, boldface values in **Table 3** identify median CPs that did not achieve the lower acceptability criterion of 92.5%. It is interesting to note that the ML-based CIs maintained an acceptable value of CP for all sample sizes except $n = 500$ for the non-BVN simulated data. Confidence intervals based on the Spearman coefficient performed

Table 2. Comparison of Median Coverage Probability of 5 C.I. Methods, by True Parameter Value.

Distribution	True Value of ξ^*	Method				
		Simple Substitution <i>LOD</i> 2	Complex Substitution	Random Imputation	Spearman's r_s (Jackknife Interval)	Maximum Likelihood
Normal	-0.9	0.0	0.0	0.0	15.8	90.8
Normal	-0.5	62.0	82.5	45.5	91.6	94.6
Normal	-0.25	92.4	93.6	85.5	93.3	94.9
Normal	0.0	95.0	94.9	94.8	94.4	95.0
Normal	0.25	89.3	91.8	85.5	93.0	95.0
Normal	0.5	79.8	86.0	46.8	91.9	94.9
Normal	0.75	62.0	70.4	2.6	89.0	94.7
Normal	0.9	25.6	24.9	0.0	76.7	94.4
Non-Normal	-0.9	1.6	2.9	0.0	15.4	87.6
Non-Normal	-0.5	80.1	75.1	4.4	92.0	91.7
Non-Normal	-0.25	90.2	89.5	67.5	93.3	92.9
Non-Normal	0.0	94.9	95.0	94.9	94.2	89.3
Non-Normal	0.25	89.6	91.6	57.6	94.2	94.0
Non-Normal	0.5	86.3	85.8	11.3	93.4	93.5
Non-Normal	0.75	82.7	72.8	0.9	91.1	93.4
Non-Normal	0.9	38.5	31.7	0.6	79.8	91.4

$\xi^* = \rho$ for Pearson correlation, $\xi = \rho_s$ for Spearman's coefficient.

Table 3. Comparison of median coverage probability of 5 C.I. methods, by sample size.

Distribution	Sample Size	Method				
		Simple Substitution <i>LOD</i> /2	Complex Substitution	Random Imputation	Spearman's r_s (Jackknife Interval)	Maximum Likelihood
Normal	20	89.8	92.3	88.4	92.3	94.1
Normal	30	89.2	92.2	85.9	93.3	94.4
Normal	50	84.0	89.3	70.3	93.0	94.5
Normal	75	81.2	88.2	58.3	93.1	95.2
Normal	100	78.0	86.1	48.6	92.2	95.0
Normal	200	64.5	78.3	19.5	88.0	94.8
Normal	500	34.3	56.6	0.7	73.3	95.2
Non-Normal	20	89.6	91.8	67.5	92.4	93.3
Non-Normal	30	89.4	91.4	53.8	93.4	93.7
Non-Normal	50	88.0	88.6	31.6	93.7	93.8
Non-Normal	75	87.7	87.0	18.8	93.8	93.8
Non-Normal	100	87.2	85.1	11.2	93.9	93.7
Non-Normal	200	86.0	78.7	1.8	93.7	93.2
Non-Normal	500	82.0	59.7	0.0	90.6	91.7

almost as well as the ML method with the non-BVN data, but failed to maintain the 92.5% level for several sample sizes with the BVN data. The CIs based on the complex substitution, simple substitution and random imputation methods failed to achieve the 92.5% level for any of the sample sizes in **Table 3**.

4. Example

We used the data from the study by Amorim and Alvarez-Leite [1] described previously to illustrate the various point and confidence interval methods. The authors correlated urinary concentrations of *o*-cresol with urinary concentrations of hippuric acid as part of their evaluation of *o*-cresol as a biomarker for toluene exposure. Only 15 of the 54 subjects in their study had complete data on both *o*-cresol and hippuric acid. The Shapiro-Wilk test indicated that the bivariate normality assumption is untenable for these data, with $p < 0.0001$ for both the 15 *o*-cresol values and the 50 hippuric acid values.

Table 4 provides a summary of the results based on all of the estimation methods described in this article. The ML method for estimating ρ in the presence of non-detects yielded $\hat{\rho}_{ML} = 0.79$, with a 95% CI of (0.66, 0.87). Analyzing only the 15 cases with complete data yielded $r = 0.76$ with a 95% CI of (0.40, 0.92). Simple substitution with *LOD*/2, which was the method used by Amorim and Alvarez-Leite, yielded $\hat{\rho}_{LOD/2} = 0.79$, with a 95% CI of (0.65, 0.87). As can be seen in **Table 4**, the ML-based results differed very little from those based on the various substitution methods, with the exception of random imputation. However,

Table 4. Point and confidence interval estimates based on data from amorim and alvarez-leite [1].

Method	n	Estimate	95% C.I.
Complete cases	15	0.76	(0.40, 0.92)
Substitute LOD	54	0.77	(0.63, 0.85)
Substitute $LOD/2$	54	0.79	(0.65, 0.87)
Substitute $LOD/\sqrt{2}$	54	0.78	(0.64, 0.86)
Random Imputation	54	0.63	(0.43, 0.76)
Complex Substitution	54	0.78	(0.65, 0.86)
Spearman	54	0.58	(0.34, 0.82)
Maximum Likelihood	54	0.79	(0.66, 0.87)

there is quite a discrepancy between the ML-based results and those based on the Spearman coefficient. The censoring proportions in this study were $p_1 = 0.07$ (4/54) for hippuric acid and $p_2 = 0.72$ (39/54) for *o*-cresol. The closest censoring proportions to these in **Table 1** are (0.1, 0.7). For non-BVN data (as is apparently the case for these data), only the CIs based on the MLE or the Spearman coefficient achieved acceptable CP for these censoring proportions (93.9% for ML and 92.9% for Spearman). The simulation results for non-BVN data in **Table 2** indicate that MLE-based CIs achieve acceptable CP (93.4%) when the true value of ρ is 0.75; this seems to be a reasonable assumption based on the ML-based point estimate in **Table 4** ($\hat{\rho}_{ML} = 0.79$). Similarly, Spearman-based CIs achieve acceptable CP (93.4%) when the true value of ρ_s is 0.5; this also appears to be a reasonable assumption based on the results in **Table 4** ($r_s = 0.58$). Finally, from the results for non-BVN data in **Table 3**, we see that CIs based on either the ML method (93.8%) or the Spearman coefficient (93.7%) achieve acceptable CP when n is 50, this is approximately so for the Amorim and Alvarez-Leite study ($n = 54$). Thus, based on our simulation results (summarized in **Tables 1-3**), we have no reason to doubt the validity of either the ML-based CI or the Spearman-based CI. Given the apparent departure from BVN for these data based on the Shapiro-Wilk test results, and the fact that the authors were evaluating *o*-cresol by examining its association with hippuric acid (not necessarily its linear association), we recommend that the results for Spearman's coefficient be used: $\hat{\rho}_s = 0.58$ with a 95% CI of (0.34, 0.82). Thus, the association between *o*-cresol and hippuric acid appears to be quite a bit weaker than that claimed by Amorim and Alvarez-Leite ($r = 0.777$).

The use of Spearman's coefficient as the measure of association for these data is consistent with Amorim and Alvarez-Leite's use of the nonparametric Kruskal-Wallis test to compare the level of urinary *o*-cresol across the three groups of toluene-exposed subjects in their study: workers in shoe factories, painting sectors of metal industries, and printing shops. If the primary goal of the authors had been to examine the linear association between urinary hippuric acid and

urinary *o*-cresol, we would recommend that the ML-based estimates be used instead of the Spearman-based estimates. Even though the BVN assumption appears to be violated for the data in this study, we feel it would still be safe to use the PCC to estimate the degree of linear relationship since our simulation results show that the ML-based method is preferable to any of the other methods of estimating the PCC when LODs are present in both variables, regardless of the bivariate distribution of X and Y .

R code for computing each of the point estimates and corresponding confidence intervals described in this article is available from the second author.

5. Discussion

In this article, we compared 5 methods that can be used to obtain a confidence interval for the correlation between two variates X and Y , both of which are left censored. Other authors have recently proposed that alternative methods be considered; for example, Weaver *et al.* [14] considered a Bayesian approach. In this article, we restrict our attention to estimators based on the frequentist approach. Some authors have considered estimation of the elements of a covariance matrix of dimension $p \times p$ under left censoring [15] [16]. The estimation problem considered in the present article corresponds to the case $p = 2$. Jones *et al.* [15] did consider $p = 2$; however, their focus was on the bias of the point estimate of ρ and they did not consider confidence interval estimation as in the present article. Pesonen *et al.* [16] considered $p = 3$ and $p = 10$, so their results are not applicable to the present article. Other recent publications have considered the estimation of Lin's concordance correlation coefficient (CCC) in the presence of left censoring [17] [18]. Domthong [17] proposed a new class of bivariate survival functions and examined their usefulness in estimating the CCC. Lapidus *et al.* [18] examined the use of multiple imputation to estimate the CCC. The PCC considered in the present article is a special case of the CCC, so our future work will focus on adapting these methods to the estimation of the PCC.

Our simulation results showed that when the simulated data were from a BVN distribution, the ML-based CIs had median CP above 92.5% under all conditions that we considered except when $p_1 = p_2 = 0.9$ or $\rho = -0.9$. Furthermore, the ML-based CIs were superior to all other CI methods in terms of median CP under all simulation scenarios using BVN data. Interestingly, for non-BVN simulated data, the ML-based CIs were still superior to those based on other estimation methods for almost all scenarios that we considered. Spearman-based CIs performed acceptably as long as $|\rho_s|$ was small or moderate, the sample size was not too large (*i.e.*, less than 500), and the censoring proportions for X and Y were not too large and there was little or no imbalance. The Spearman-based CIs generally performed better for non-BVN data than for BVN data.

The complex substitution method was proposed by Lynn [8] for use when only one of the variables is subject to NDs and McCracken [9] extended the method to the situation in which both variables are subject to NDs. The CPs of CIs

based on this method were acceptable for some settings of the simulation parameters. However, complex substitution-based CIs were typically inferior to those based on either the MLE or the Spearman coefficient or both. The distribution of the simulated data (either BVN or non-BVN) had only a minimal impact on the performance of complex substitution-based CIs. The CP of these CIs was generally superior to that of CIs based on simple substitution, but they were comparable for many of the settings of the simulation parameters that we considered. Confidence intervals based on the random imputation method were greatly inferior to the CIs based on all of the other methods for almost all settings of the simulation parameters, and rarely achieved the 92.5% acceptability criterion for CP.

The simple substitution-based CIs generally did better in terms of CP when the data were not BVN. However, these intervals did not yield acceptable CP except in the presence of no censoring (**Table 1**) or when the population correlation was zero (**Table 2**). Handelsman and Ly [19] recommended simple substitution with $LOD/\sqrt{2}$ when estimating bivariate correlations for serum steroid measurements; however, their study only considered situations in which either X or Y but not both are left censored.

For several of the simulation scenarios we included, the ML method failed to produce a point estimate of ρ (and hence a confidence interval) due to the failure of the optimization routine to converge. In case this happens with a real data set, we recommend using a Spearman-based CI if one wishes only to measure the strength of association between X and Y , and not to measure the strength of linear association between X and Y . If estimation of ρ (as a measure of linear association) is the primary aim of the analysis, and the MLE cannot be obtained due to lack of convergence, we recommend that complex substitution be used to estimate ρ . However, for some combinations of the simulation settings that we considered, complex substitution-based CIs performed extremely poorly in terms of CP. Under these conditions, the CI based on the CS method should be considered as only a rough approximation.

Li *et al.* [20] also examined the use of the ML method to estimate a bivariate correlation in the presence of left censoring of both X and Y . They provided R code for implementing this method and thoroughly evaluated its performance in a simulation study in which they used several of the same parameter settings as in our simulation study [9]. Their results for the CP of 95% CIs based on the ML method under assumptions of both BVN and non-BVN data are comparable to ours. However, Li *et al.* did not provide a comparison of ML-based CIs with those based on other methods, as we have done here. The results we have presented in **Tables 1-3** enable the analyst to select an estimation method that is likely to give acceptable results for the CP depending on the degree of censoring, the true value of ρ , and the sample size, as we illustrated in the Example.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Amorin, L. and Alvarez-Leite, E. (1997) Determination of *o*-cresol by Gas Chromatography and Comparison with Hippuric Acid Levels in Urine Samples of Individuals Exposed to Toluene. *Journal of Toxicology Environmental Health*, **50**, 401-408. <https://doi.org/10.1080/009841097160438>
- [2] Atawodi, S.E., Lea, S., Nyberg, F., Mukeria, A., Constantinescu, V., Ahrens, W., *et al.* (1998) 4-Hydroxyl-1-(3-pyridyl)-1-Butanone-Hemoglobin Adducts as Biomarkers of Exposure to Tobacco Smoke: Validation of a Method to be Used in Multicenter Studies. *Cancer Epidemiology Biomarkers and Prevention*, **7**, 817-821.
- [3] Lagorio, S., Crebelli, R., Ricciarello, R., Conti, L., Iavarone, I., Zona, A., Ghittori, S. and Carere, A. (1998) Methodological Issues in Biomonitoring of Low Level Exposure to Benzene. *Occupational Medicine*, **8**, 497-504. <https://doi.org/10.1093/occmed/48.8.497>
- [4] Cook, D.G., Whincup, P.H., Papacosta, O., Strachan, D.P., Jarvis, M.J. and Bryant, A. (1993) Relation of Passive Smoking as Assessed by Salivary Cotinine Concentration and Questionnaire to Spirometric Indices in Children. *Thorax*, **48**, 14-20. <https://doi.org/10.1136/thx.48.1.14>
- [5] Wang, H. (2006) Correlation Analysis for Left-Censored Biomarker Data with Known Detection Limits. Unpublished Master's Thesis, Louisiana State University Health Sciences Center, School of Public Health, Biostatistics Program, New Orleans, Louisiana.
- [6] Lyles, R.H., Williams, J.K. and Chuachoowong, R. (2001) Correlating Two Viral Load Assays with Known Detection Limits. *Biometrics*, **57**, 1238-1244. <https://doi.org/10.1111/j.0006-341X.2001.01238.x>
- [7] Scheuren, F. (2005) Multiple Imputation: How It Began and Continues. *The American Statistician*, **59**, 315-319. <https://doi.org/10.1198/000313005X74016>
- [8] Lynn, H. (2001) Maximum Likelihood Inference for Left-Censored HIV RNA Data. *Statistics in Medicine*, **20**, 33-45. [https://doi.org/10.1002/1097-0258\(20010115\)20:1%3C33::AID-SIM640%3E3.0.CO;2-O](https://doi.org/10.1002/1097-0258(20010115)20:1%3C33::AID-SIM640%3E3.0.CO;2-O)
- [9] McCracken, C.E. (2013) Correlation Coefficient Inference for Left-Censored Biomarker Data with Known Detection Limits. Unpublished Ph.D. Dissertation, Augusta University, Department of Biostatistics, Augusta, Georgia.
- [10] Li, L., Wang, W. and Chan, I. (2004) Correlation Coefficient Inference on Censored Bioassay Data. *Journal of Biopharmaceutical Statistics*, **15**, 501-512. <https://doi.org/10.1081/BIP-200056552>
- [11] Gibbons, J. and Chakraborti, S. (2003) Nonparametric Statistical Inference. 4th Edition, Marcel Dekker Inc., New York.
- [12] Newton, E. and Rudel, R. (2007) Estimating Correlation with Multiply Censored Data Arising from the Adjustment of Singly Censored Data. *Environmental Science and Technology*, **41**, 221-228. <https://doi.org/10.1021/es0608444>
- [13] Bradley, J.V. (1978) Robustness? *British Journal of Mathematical and Statistical Psychology*, **31**, 144-152. <https://doi.org/10.1111/j.2044-8317.1978.tb00581.x>
- [14] Weaver, B.P., Kaufeld, K. and Warr, R. (2020) Estimating Correlations with Censored Data. *Quality Engineering*, **32**, 521-527. <https://doi.org/10.1080/08982112.2019.1698744>
- [15] Jones, M.P., Perry, S.S. and Thorne, P.S. (2015) Maximum Pairwise Pseudo-Likelihood Estimation of the Covariance Matrix from Left Censored Data. *Jour-*

-
- nal of Agricultural, Biological, and Environmental Statistics*, **20**, 83-99.
<https://doi.org/10.1007/s13253-014-0185-y>
- [16] Pesonen M., Pesonen, H. and Nevalainen, J. (2015) Covariance Matrix Estimation for Left-Censored Data, *Computational Statistics and Data Analysis*, **92**, 13-25.
<https://doi.org/10.1016/j.csda.2015.06.005>
- [17] Domthong, U. (2014) A New Class of Bivariate Weibull Distribution to Accommodate the Concordance Correlation Coefficient for Left-Censored Data. Unpublished Ph.D. Dissertation, Pennsylvania State University, Department of Public Health Sciences, Hershey, Pennsylvania.
- [18] Lapidus, N., Chevret, S. and Resche-Rigon, M. (2014) Assessing Assay Agreement Estimation for Multiple Left-Censored Data: A Multiple Imputation Approach. *Statistics in Medicine*, **33**, 5298-5309. <https://doi.org/10.1002/sim.6319>
- [19] Handelsman, D.J. and Ly, L.P. (2019) An Accurate Substitution Method to Minimize Left Censoring Bias in Serum Steroid Measurements. *Endocrinology*, **160**, 2395-2400.
<https://doi.org/10.1210/en.2019-00340>
- [20] Li, Y., Gillespie, B.W., Shedden, K. and Gillespie, J.A. (2018) Profile Likelihood Estimation of the Correlation Coefficient in the Presence of Left, Right or Interval Censoring and Missing Data. *The R Journal*, **10**, 159-179.
<https://doi.org/10.32614/RJ-2018-040>