Statistical Methods & Applications
https://doi.org/10.1007/s10260-020-00555-1

ORIGINAL PAPER

Bootstrap confidence intervals for correlation between continuous repeated measures

Guogen Shan[1] · Hua Zhang[2] · Jim Barbour[3] on behalf of for the Alzheimer's Disease Neuroimaging Initiative

Accepted: 15 December 2020

Abstract

Repeated measures designs are widely used in practice to increase power, reduce sample size, and increase efficiency in data collection. Correlation between repeated measurements is one of the first research questions that needs to be addressed in a repeated-measure study. In addition to an estimate for correlation, confidence interval should be computed and reported for statistical inference. The asymptotic interval based on the delta method is traditionally calculated due to its simplicity. However, this interval is often criticized for its unsatisfactory performance with regards to coverage and interval width. Bootstrap could be utilized to reduce the interval width, and the widely used bootstrap intervals include the percentile interval, the bias-corrected interval, and the bias-corrected with acceleration interval. Wilcox (Comput Stat Data Anal 22:89–98,1996) suggested a modified percentile interval with the interval levels adjusted by sample size to have the coverage probability close to the nominal level. For a study with repeated measures, more parameters in addition to sample size would affect the coverage probability. For these reasons, we propose modifying the percentiles in the percentile interval to guarantee the coverage probability based on simulation studies. We analyze the correlation between imaging volumes and memory scores from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study to illustrate the application of the considered intervals. The proposed interval is exact with the coverage probability guaranteed, and is recommended for use in practice.

Keywords Bootstrap confidence interval · Correction for repeated measures · Coverage probability · Longitudinal data · Proc mixed

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database(adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/orprovided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at:http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Extended author information available on the last page of the article

Published online: 21 January 2021

1 Introduction

Correlation between two measures has been studied for decades, and Pearson's correlation coefficient is traditionally used to quantify the degree to which two variables are related to each other (Pearson 1900; Casella and Berger 2002). The sign of Pearson's correlation shows the direction of relationship between two continuous measures. The range of Pearson's correlation is from −1 to 1. A value of zero indicates that there is no linear relationship between two measures.

In Pearson's correlation, each measurement is only recorded once for each participant. For a study with repeated measures, each measurement is collected multiple times with some correlation between these observations. The correlation between two measures is the parameter of interest in this article. When data from a repeated-measure study are analyzed by using the Pearson's method, the outcomes from the same patient are assumed to be independent from each other, which is not a reasonable assumption. Data from the same patient at different visits are correlated, not independent. Ignoring the dependency of measures at multiple visits from the same patient, the computed type I error rate could be much larger than the nominal level (Aarts et al. 2014; Bakdash and Marusich 2017). In addition, Pearson's correlation is often used to measure the correlation between two outcomes in a cross-sectional study, not a repeated-measure study. Therefore, it is not appropriate to use Pearson's correlation for repeated measures. In contemporary clinical trials, participants are often scheduled for multiple visits with one of the aims to study the trajectory of important outcomes.

When repeated measures are collected, one of the very first steps in data analysis is to estimate the correlation between these measures. Bland and Altman (1995a, b) developed a few methods to compute correlation for repeated measures by partialling out the visit effect. Lam et al. (1999) proposed using a linear mixed model to compute the longitudinal correlation based on maximum likelihood method using a special software. Later, Hamlett et al. (2004) implemented the mixed model using the commercially available software (SAS) to compute correlation under the assumption of a compound symmetric (CS) correlation structure. Roy (2006) computed correlation under the correlation matrix of autoregressive of order one. She found that the type of correlation matrix could affect the statistical inference (Crowder 1995).

In addition to estimates for correlation in the presence of repeated measures, it is also important to provide confidence intervals for statistical inference. Irimata et al. (2018) compared five methods to estimate correlation and their associated confidence intervals using a pharmacokinetics data with 18 subjects. They recommended the mixed model approach and the corresponding confidence interval. They derived the standard deviation of correlation under the CS correlation structure using the delta method to construct an asymptotic confidence interval. They also considered the bias-corrected (BC) bootstrap confidence interval. Bootstrap methods are computationally intensive as compared to the traditional statistical methods. But bootstrap is an efficient method to compute the empirical distribution of the parameter of interest that can then be utilized to construct

confidence intervals. The SAS macros to compute the asymptotic interval and the BC interval were attached to the article by Irimata et al. (2018) who only compared their performances using that pharmacokinetics data set. In addition to the BC interval, the percentile bootstrap interval and the bias-corrected with acceleration (BCa) bootstrap confidence interval are also popularly used in practice.

When sample size is small to medium (e.g., less than 100), the aforementioned confidence intervals may have unsatisfactory performance with regards to coverage probability. For this reason, Wilcox (2011) suggested adjusting the nominal level by the sample size in a regression setting (Wilcox 1996). We adopt his approach for correlation in the presence of repeated measures where more parameters (sample size, variance, covariance, and mean) could affect the coverage probability. We conduct extensive numerical studies to identify the modified percentile that guarantees the coverage probability and has the shortest width.

We organize this article as follows. In Sect. 2, we briefly introduce the mixed method for correlation for repeated measures and present the existing methods for confidence intervals. Then, we propose symmetric modified percentile intervals that guarantee the coverage probability. In Sect. 3, we conduct extensive simulation studies to compare the performance of the existing methods and the new method. We illustrate the application of these methods by using data from the Alzheimer's Disease Neuroimaging Initiative (ADNI) study in Sect. 4. Finally, we provide some comments in Sect. 5.

2 Methods

linear mixed models (LMMs) are recommended for correlation estimate in the presence of repeated measures (Irimata et al. 2018; Hamlett et al. 2004). LMMs are flexible and efficient to model the correlation structure for a study with repeated measures. In this article, we assume a CS covariance structure for repeated measurements as literature suggests (Lam et al. 1999; Hamlett et al. 2004). Suppose

$$Y_i = (Y_{i1}, Y_{i2}, \ldots, Y_{it_i}) = (U_{i1}, W_{i1}, U_{i2}, W_{i2}, \ldots, U_{it_i}, W_{it_i})$$

is a multivariate random vector with a list of variables at each visit for the $i-$th participant, where $Y_{ij} = (U_{ij}, W_{ij})$ is the random vector with two measures for the i-th participant at the j-th visit. Under the CS correlation structure, the correlation at each visit is the same. Then Y_{ij} is assumed to follow a bivariate normal distribution (Lam et al. 1999)

$$Y_{ij} \sim MVN\left(\begin{pmatrix}\mu_{U_j}\\ \mu_{W_j}\end{pmatrix}, \begin{pmatrix}\sigma_U^2 & \sigma_{UW}\\ \sigma_{UW} & \sigma_W^2\end{pmatrix}\right).$$

It follows that the correlation coefficient for repeated measures is

$$\rho = Corr(U_{ij}, W_{ij}) = \frac{\sigma_{UW}}{\sqrt{\sigma_U^2 \sigma_W^2}}, \tag{1}$$

where the correlation ρ is between -1 and 1.

2.1 Correlation estimator from LMM

Lam et al. (1999) introduced the LMM to estimate correlation for repeated measures based on the maximum likelihood method by using a special software package. Later, Hamlett et al. (2004) utilized Proc Mixed in SAS instead of the special software by Lam et al. (1999) for correlation. A general LMM is presented as

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i,$$

where X_i and Z_i are the design matrices for the fixed effect β and the random effect b_i, respectively. Suppose b_i follows $N(0, G)$ and the random error is asymptotically distributed as $N(0, R)$. Then, we have

$$V_i = Var(Y_i) = Z_i G Z_i^{'} + R.$$

One may find detailed formulas for X_i, β, Z_i in the article by Hamlett et al. (2004). The variance covariance matrices for G and R are

$$G = \begin{pmatrix} \sigma_U^2\rho_U & \sigma_{UW}\delta \\ \sigma_{UW}\delta & \sigma_W^2\rho_W \end{pmatrix}, \quad R = I_{t_i} \otimes \begin{pmatrix} \sigma_U^2(1-\rho_U) & \sigma_{UW}(1-\delta) \\ \sigma_{UW}(1-\delta) & \sigma_W^2(1-\rho_W) \end{pmatrix}, \qquad (2)$$

where I_{t_i} is the identity matrix with the size of t_i, $\otimes$ is the Kronecker product of two matrices, $\rho_U = Corr(U_{ij}, U_{ij'})$ is the correlation within U, $\rho_W = Corr(W_{ij}, W_{ij'})$ is the correlation within W, and $\delta = Corr(U_{ij}, W_{ij'})/Corr(U_{ij}, W_{ij})$. The value of δ measures the degree of correlation between measures at different times. The quantities G and R can be estimated from the random and repeated statements in the Proc Mixed.

It follows that the elements in Eq. (1) can be expressed as

$$\sigma_U^2 = G_{11} + R_{11}, \sigma_W^2 = G_{22} + R_{22}, \text{ and } \sigma_{UW} = G_{12} + R_{12}.$$

Therefore, correlation for repeated measures, ρ, in Eq. (1) can be rewritten as a function of six parameters,

$$\rho = \phi(s) = \frac{G_{12} + R_{12}}{\sqrt{(G_{11} + R_{11})(G_{22} + R_{22})}},$$

where $s = (G_{11}, G_{22}, G_{12}, R_{11}, R_{22}, R_{12})$.

2.2 Asymptotic confidence interval

In order to calculate a confidence interval for ρ, one has to estimate its standard deviation. Asymptotic confidence intervals based on the limiting distribution of the correlation coefficient are traditionally used for statistical inference. The delta method utilizes the first order Taylor expansion to estimate its variance as

$$Var(\rho) = \left(\frac{\partial \phi}{\partial s}\right)' Var(s) \frac{\partial \phi}{\partial s},$$

where $\frac{\partial \phi}{\partial s} = (\frac{\partial \phi}{\partial G_{11}}, \frac{\partial \phi}{\partial G_{22}}, \frac{\partial \phi}{\partial G_{12}}, \frac{\partial \phi}{\partial R_{11}}, \frac{\partial \phi}{\partial R_{22}}, \frac{\partial \phi}{\partial R_{12}})'$ is the partial derivative vector, and $Var(s)$ is the variance-covariance matrix of s that can be estimated from the Asycov option in the Proc Mixed (Hamlett et al. 2004). The $(1-\alpha)\%$ confidence interval for ρ is then computed as

$$\left(\hat{\rho} - z_{1-\alpha/2}\sqrt{Var(\hat{\rho})}, \hat{\rho} + z_{1-\alpha/2}\sqrt{Var(\hat{\rho})}\right), \tag{3}$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution, e.g., $z_{1-\alpha/2} = 1.96$ when $\alpha = 0.05$. We refer this interval as the Asy interval. Irimata et al. (2018) developed a SAS macro (named MMCorr-NormalApprox) to compute the asymptotic confidence interval for the longitudinal correlation.

2.3 Bootstrap confidence intervals

Bootstrap confidence intervals are computationally intensive as compared to the traditional intervals. As more computational resources are available to statisticians and biostatisticians for data analysis, bootstrap intervals become feasible to be applied in practice to improve the performance of confidence intervals for correlation in the presence of repeated measures.

Suppose a study has n participants and each participant has m scheduled visits. In order to avoid breaking down the correlation within each participant, we bootstrap samples at the participant level. Suppose d samples are randomly selected from n participants with replacement. For each bootstrap sample set with the size of d, a LMM is used to fit the data, and ρ is computed from the fitted model. Suppose the bootstrapping procedure is repeated by B times. Then, the computed correlations from this bootstrap procedure are:

$$\Omega(\rho) = (\hat{\rho}_1, \hat{\rho}_2, \ldots, \hat{\rho}_B),$$

where $\Omega(\rho)$ is the sample space of all the ρ values. The sample space $\Omega(\rho)$ is used as the empirical distribution of ρ to compute confidence intervals. Suppose $(\hat{\rho}_{(1)}, \hat{\rho}_{(2)}, \ldots, \hat{\rho}_{(B)})$ is the ordered vector of $\Omega(\rho)$ from the smallest correlation to the largest value. The bootstrap confidence interval based on percentiles are defined as

$$(\hat{\rho}_{B\alpha_l}, \hat{\rho}_{B\alpha_u}),$$

where α_l and α_u are the lower and upper percentiles.

2.3.1 Percentile interval

The first bootstrap interval is the traditional percentile interval with

$$\alpha_l = \alpha/2 \quad \text{and} \quad \alpha_u = 1 - \alpha/2. \tag{4}$$

When $\alpha = 0.05$, α_l is 0.025 and α_u is 0.975. We refer this interval as the PCT interval.

2.3.2 BC and BCa intervals

Efron (1985) and Efron and Tibshirani (1994) proposed the bootstrap confidence interval by adjusting the lower and upper percentiles for both bias and skewness in the bootstrap distribution. The bias correction effect is estimated from the proportion of ρ_is being less than $\hat{\rho}$ estimated from the observed data:

$$\hat{z} = \Phi^{-1}\left(\frac{\sum_{i=1}^{B} I(\hat{\rho}_i < \hat{\rho})}{B}\right),$$

where $\Phi(.)$ is the cumulative distribution function of the standard normal distribution. The acceleration factor to adjust the skewness is calculated by using the jackknife approach. Let $\hat{\rho}(-i)$ be the estimate of ρ after removing the $i-$th participant from the observed data, $i = 1, 2, \ldots, n$. The average of these estimates is $\hat{\rho}_J = \sum_{i=1}^{n} \hat{\rho}(-i)/n$. The acceleration factor is computed as

$$\hat{a} = \frac{\sum_{i=1}^{n} (\hat{\rho}_J - \hat{\rho}(-i))^3}{6[\sum_{i=1}^{n} (\hat{\rho}_J - \hat{\rho}(-i))^2]^{3/2}}.$$

The second bootstrap interval is bias correlated with acceleration, with the lower and upper percentiles as

$$\alpha_l = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + Z_{\alpha/2}}{1 - \hat{a}(\hat{z}_0 + Z_{\alpha/2})}\right) \quad \text{and} \quad \alpha_u = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + Z_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 + Z_{1-\alpha/2})}\right). \tag{5}$$

This interval is well known as the BCa interval. The BCa interval becomes the BC interval when the acceleration factor is zero, $\hat{a} = 0$. Both the BCa interval and the BC interval have been applied to many important statistical problems (Efron and Tibshirani 1994; Shan et al. 2011).

2.3.3 MP interval

Wilcox (2011, 1996) proposed a modified percentile (MP) interval that makes the adjustments of the lower and upper percentiles for the confidence interval of the slope in a linear regression model. The MP percentiles were determined such that the confidence interval for the slope has the coverage close to the nominal level, but the MP interval does not guarantee the coverage. The suggested percentiles for a 95% two-sided confidence interval are

$$(\alpha_l, \alpha_u) = \begin{cases} (7/599, 593/599) & \text{if } n < 40; \\ (8/599, 592/599) & \text{if } 40 \le n < 80; \\ (11/599, 588/599) & \text{if } 80 \le n < 180; \\ (14/599, 585/599) & \text{if } 180 \le n < 250; \\ (15/599, 584/599) & \text{if } 250 \le n, \end{cases} \quad (6)$$

where $B = 599$ is chosen based on the finding from Hall (1986) that $(B+1)^{-1}$ is a multiple of $1-\alpha$. It is obvious that multiple B values could meet that condition, but simulations with $B = 599$ have comparable results with larger B values and much better than those with smaller B values (Wilcox 1996). It should be noted that $B = 599$ should be used in the MP intervals to make sure that these intervals are valid. Sample size is the only factor used to modify the percentiles in that approach. The MP interval is the same as the PCT interval when $n \ge 250$. The nominal level of the MP approach is much more than 95% when n is small. It should be noticed that the MP interval has asymmetric lower and upper percentiles when $n < 80$.

2.3.4 SMP interval

For a study with repeated measures, the actual coverage could be affected by multiple factors, including sample size, variance, covariance, mean, and the number of visits. The asymmetric intervals in the MP approach when $n < 80$, have only 1/599 percentile difference as compared to symmetric intervals. In general, symmetric intervals are preferable in practice. For these reasons, we propose using extensive simulation studies to identify the symmetric modified percentile (SMP) interval that guarantees the coverage probability, with the lower percentile and the upper percentile as

$$SMP\tau : (\alpha_l, \alpha_u) = \left(\frac{\tau}{599}, \frac{599-\tau}{599}\right),$$

where $\tau = 1, 2, \ldots, 20$. Coverage probability is the probability that the computed confidence intervals from simulated data sets contain the pre-specified ρ using the $SMP\tau$ lower and upper percentiles:

$$\sum_{i=1}^{M} \frac{I[\rho \in CI(i|\tau)]}{M},$$

where $CI(i|\tau) = (\hat{\rho}_{l-i|\tau}, \hat{\rho}_{u-i|\tau})$ is the confidence interval for ρ using the $i-$th simulated data set, M is the number of simulated data sets, and $I(.)$ is the index function.

In addition to coverage probability, width of a confidence interval is another criteria for comparing different intervals. The average width of these confidence intervals is calculated as

$$\sum_{i=1}^{M} \frac{\hat{\rho}_{u-i|\tau} - \hat{\rho}_{l-i|\tau}}{M}.$$

Since ρ is between -1 and 1, when the lower limit or the upper limit is beyond the range, we set the lower limit as -1 or the upper limit as 1. The bootstrap intervals always have the interval between -1 and 1 because these intervals are percentile intervals using the estimated ρ values which are bounded by -1 and 1.

The exact SMP interval is the one that has the shortest width among the SMP intervals whose coverages are above the nominal level. The *SMPj* interval contains $SMPj^{'}$ interval when $j<j^{'}$. Among all the SMP intervals have the coverage above the nominal level ($SMP1, SMP2, \ldots, SMPk$), the *SMPk* interval is the exact interval.

3 Simulation studies

We conduct extensive simulation studies to identify the SMP interval for correlation in the presence of repeated measures at the nominal level of 95%. The identified SMP interval is then compared with other intervals with regards to coverage probability and width of intervals.

In the simulations studies, we assume that each participant has $t_i = 4$ scheduled visits. The mean values are

$$Y_i = (U_{i1}, W_{i1}, U_{i2}, W_{i2}, U_{i3}, W_{i3}, U_{i4}, W_{i4}) = (2.0, 0.8, 1.9, 0.7, 1.7, 0.6, 1.4, 0.5),$$

with a decreasing trend over time for both measures, where U_{ij} and W_{ij} are the outcomes of U and W for the ith participant at the jth visit ($j = 1, 2, 3, 4$). The variance-covariance matrix has multiple nuisance parameters that need to be specified. The longitudinal correlation $\rho = 0.5$ is used in the following simulation studies. Other values in the covariance matrix in Equation (2) are set as: σ_U^2 and σ_W^2 from 0.1 to 3; ρ_U and ρ_W from 0.4 to 0.6; and $\delta = 0.4$. The covariance σ_{UW} in the variance-covariance matrix can be computed from the aforementioned values as: $\rho\sqrt{\sigma_U^2\sigma_W^2}$.

For each configuration, we simulate $M = 1,000$ data sets when $n = 30, 60$, and 100. An R function, *mvrnorm*, from the R package *MASS* is used to simulate these data sets. These $M = 1,000$ data sets are ready to compute the Asy interval. For bootstrap confidence intervals, we generate $B = 599$ bootstrap samples from each data set. The size of each bootstrap sample is set the same as that of the data set: $d = n$.

We first compute the actual coverage of each SMP interval: SMP1, SMP2, ..., and SMP20. Figure 1 presents the coverage probability of the proposed SMP intervals as a function of σ_U^2 from 0.6 to 3 when σ_W^2 is from 0.1 to 1 given $n = 60$, $\rho_U = \rho_W = 0.4$. The coverage probability of each interval is an increasing function of σ_U^2 when σ_W^2 is low, and this increasing trend is slowed down when σ_W^2 is higher. For each configuration, we identify the SMP*k* interval whose width is the shortest

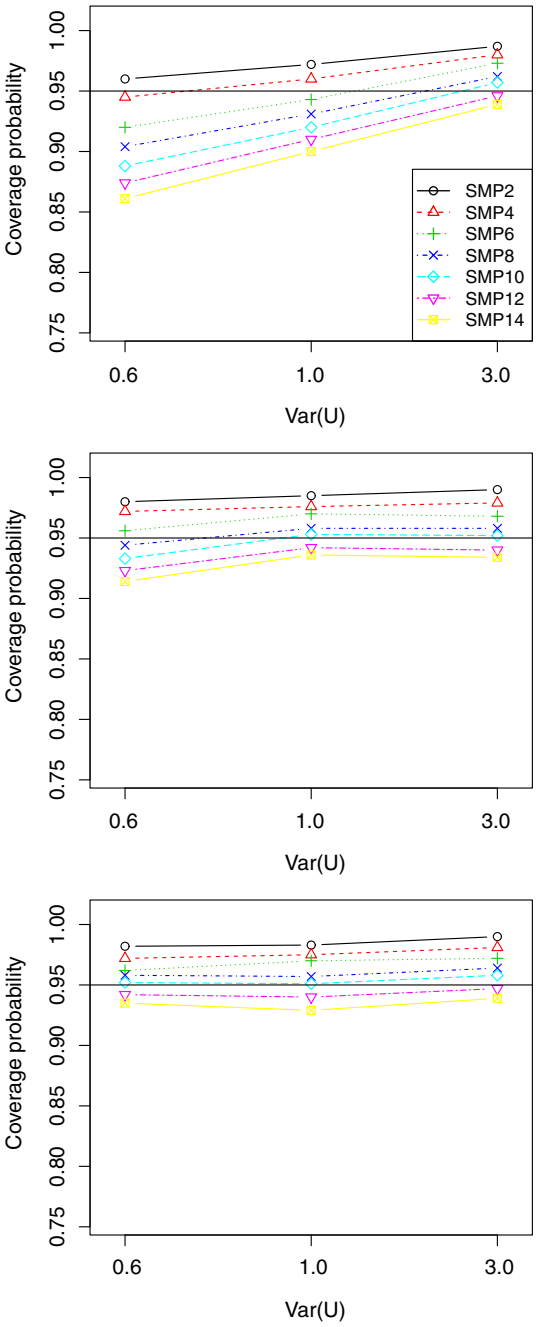

Fig. 1 Coverage probability of the proposed SMP intervals as a function of σ_U^2 from 0.6 to 3.0, given $\sigma_W^2 = 0.1$ (top), 0.6 (middle), and 1.0 (bottom), when $n = 60$, $\rho_U = \rho_W = 0.4$, and $\delta = 0.4$

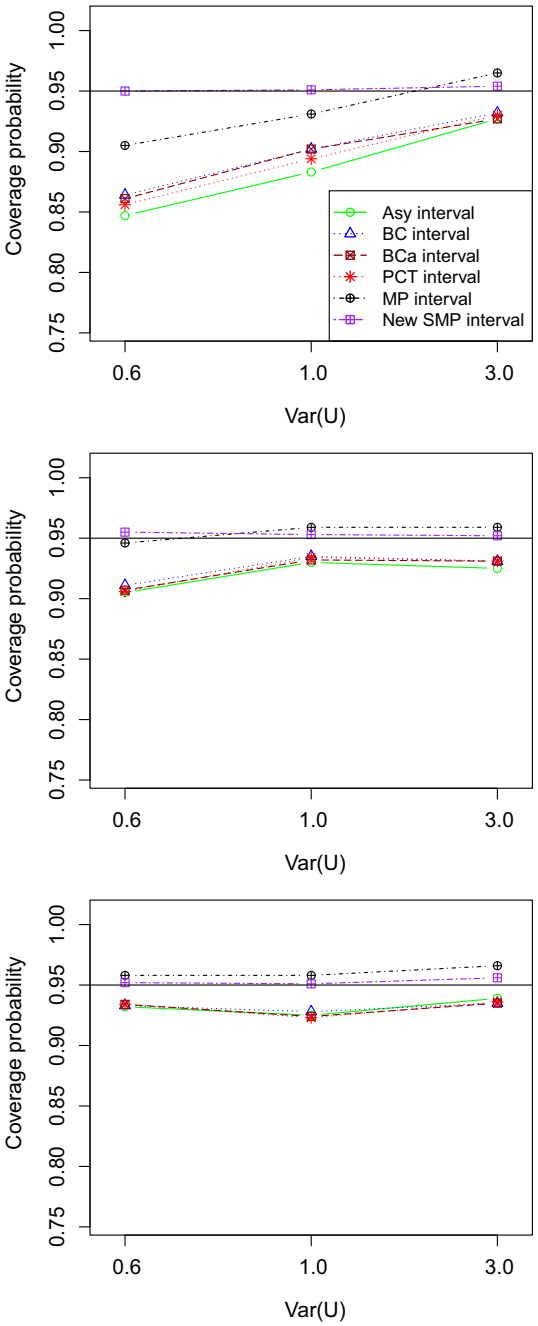

Fig. 2 Coverage comparison between the proposed SMP interval and other 5 intervals: coverage probability as a function of σ_U^2 from 0.6 to 3.0, given $\sigma_W^2 = 0.1$ (top), 0.6 (middle), and 1.0 (bottom), when $n = 60$, $\rho_U = \rho_W = 0.4$, and $\delta = 0.4$

Table 1 The optimal $SMP\tau$ interval for given σ_U^2 and σ_W^2, when $n = 60$, $\rho_U = \rho_W = 0.4$, and $\delta = 0.4$

σ_W^2	σ_U^2	τ
0.1	0.6	3
0.1	1.0	5
0.1	3.0	11
0.6	0.6	7
0.6	1.0	10
0.6	3.0	10
1.0	0.6	10
1.0	1.0	10
1.0	3.0	11

among the ones preserving the coverage. In the case that all SMP intervals do not guarantee the coverage probability, the SMP1 interval is used in the comparison.

After the exact SMP interval is identified, we compare the coverage probability of the SMP interval with the existing intervals in Fig. 2. We also present the τ values for the identified SMP interval for each configuration in Table 1. The coverage probability of the MP interval is close to the nominal level when σ_U^2 or σ_W^2 is relatively large. For these configurations, the confidence interval of the MP approach is much wider than that of the proposed SMP interval, see Fig. 3. Other existing intervals always have the coverage below the nominal level, which could be as low as 85% at the nominal level of 95%.

As expected, width of each SMP interval decreases as the sample size goes up from $n = 30$ to 100 in Fig. 4 when σ_U^2 from 0.6 to 3, $\sigma_W^2 = 0.1$, and $\rho_U = \rho_W = 0.4$. In Fig. 5, we observe that the BC interval, the BCa interval, and the PCT interval often have narrower intervals than the Asy interval. These four intervals have narrower intervals than the MP interval and the SMP interval, but these four intervals do not guarantee the coverage as seen in Fig. 6.

Figure 6 shows the coverage probability comparison between the exact SMP interval and other 5 intervals. As sample size increases, the actual coverage may go down when σ_W^2 is low. The actual coverage probabilities of the existing 5 intervals could be as low as 75% in some configurations. The proposed SMP interval always preserves the level of a confidence interval. When both variances are high, the Asy interval, the BC interval, the BCa interval and the PCT interval have their coverage probabilities closer to the nominal level although their coverage probabilities are still below the nominal level. The MP interval has conservative coverages when $\sigma_U^2 = 3$ where a wider interval is observed in Fig. 5 for the width comparison.

We observe similar results when $\sigma_W^2 = 0.6$ for the width and coverage probability comparison in Figs. 7 and 8. It can be seen that the MP approach generally have its coverage probability closer to the nominal level as compared to other existing intervals. The proposed new SMP interval guarantees the coverage, and the width of the SMP interval is shorter than that of the MP interval when variances are relatively large.

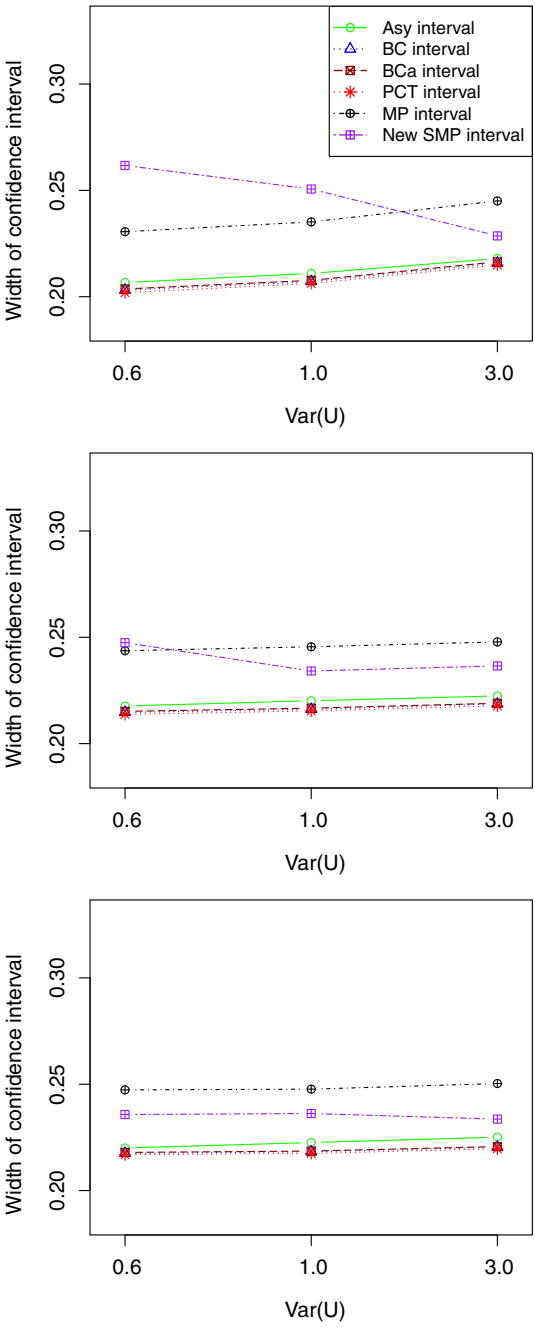

Fig. 3 Width comparison between the proposed SMP interval and other 5 intervals: interval width as a function of σ_U^2 from 0.6 to 3.0, given $\sigma_W^2 = 0.1$ (top), 0.6 (middle), and 1.0 (bottom), when $n = 60$, $\rho_U = \rho_W = 0.4$, and $\delta = 0.4$

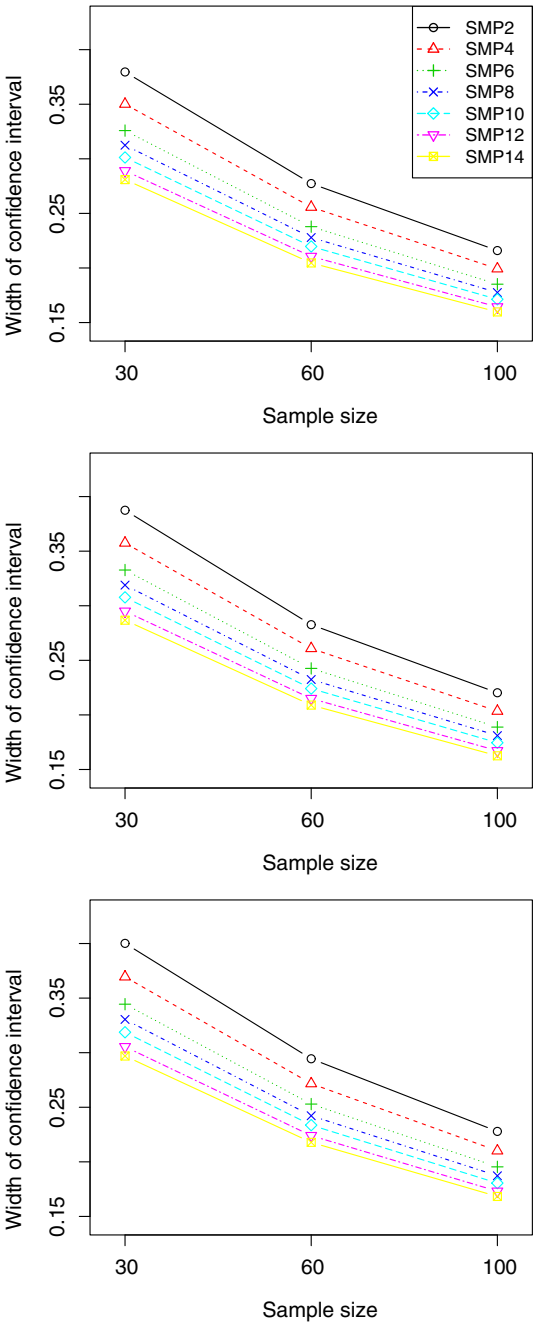

Fig. 4 Width of the proposed SMP intervals as a function of sample size n from 30 to 100, given $\sigma_U^2 = 0.6$ (top), 1.0 (middle), and 3.0 (bottom), when $\sigma_W^2 = 0.1$, $\rho_U = \rho_W = 0.4$, and $\delta = 0.4$

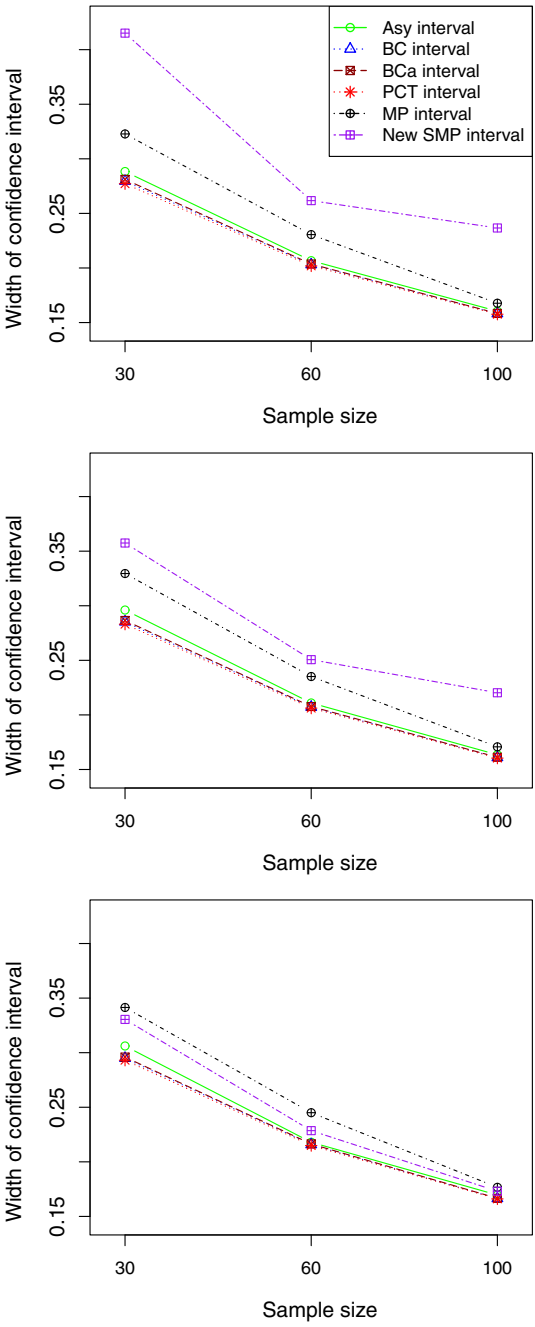

Fig. 5 Width comparison between the proposed SMP interval and other 5 intervals: interval width as a function of sample size n from 30 to 100, given $\sigma_U^2 = 0.6$ (top), 1.0 (middle), and 3.0 (bottom), when $\sigma_W^2 = 0.1$, $\rho_U = \rho_W = 0.4$, and $\delta = 0.4$

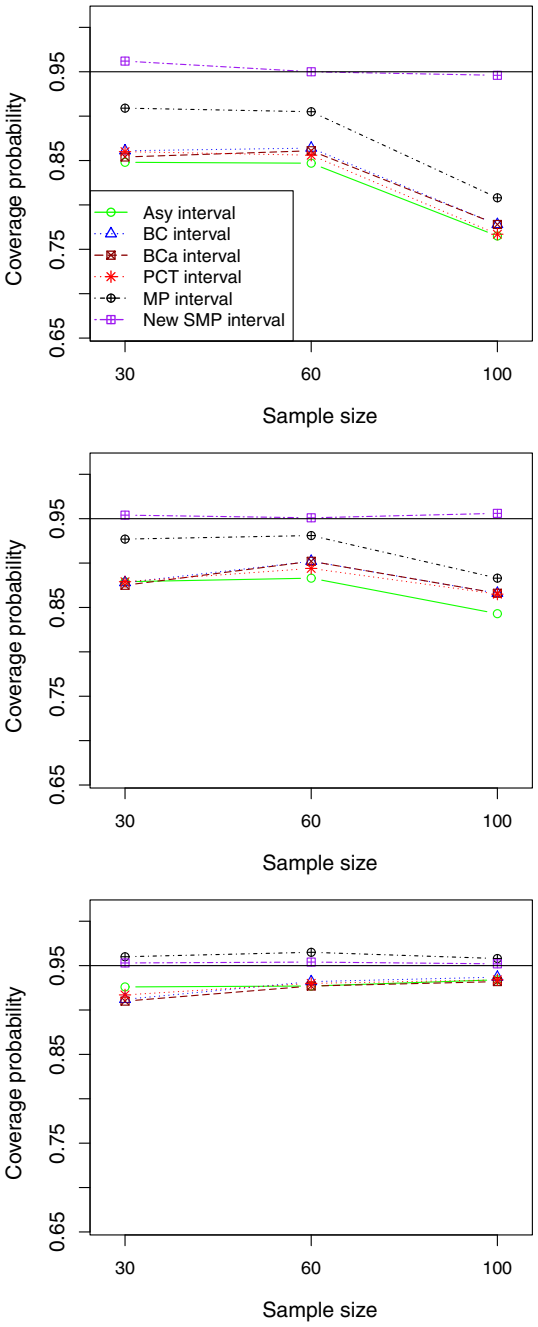

Fig. 6 Coverage comparison between the proposed SMP interval and other 5 intervals: coverage probability as a function of sample size n from 30 to 100, given $\sigma_U^2 = 0.6$ (top), 1.0 (middle), and 3.0 (bottom), when $\sigma_W^2 = 0.1$, $\rho_U = \rho_W = 0.4$, and $\delta = 0.4$

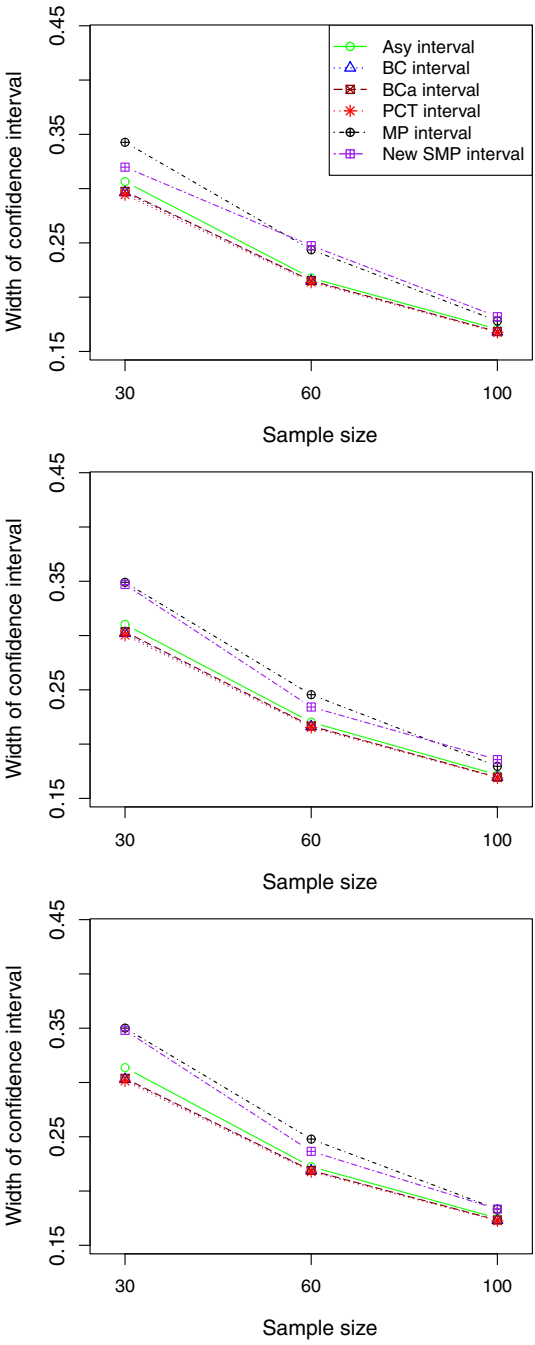

Fig. 7 Width comparison between the proposed SMP interval and other 5 intervals: interval width as a function of sample size n from 30 to 100, given $\sigma_U^2 = 0.6$ (top), 1.0 (middle), and 3.0 (bottom), when $\sigma_W^2 = 0.6$, $\rho_U = \rho_W = 0.4$, and $\delta = 0.4$

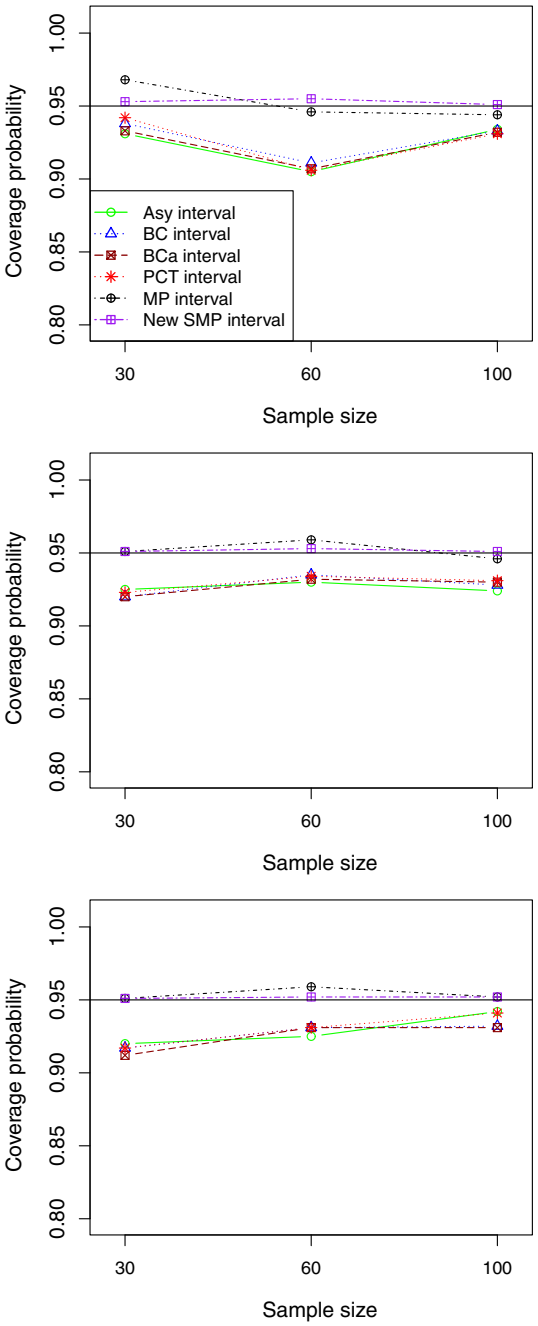

Fig. 8 Coverage comparison between the proposed SMP interval and other 5 intervals: coverage probability as a function of sample size n from 30 to 100, given $\sigma_U^2 = 0.6$ (top), 1.0 (middle), and 3.0 (bottom), when $\sigma_W^2 = 0.6$, $\rho_U = \rho_W = 0.4$, and $\delta = 0.4$

4 Example

The ADNI study is a longitudinal study with one of the goals to improve prevalence and treatment for Alzheimer's Disease (Cummings 2018; Shan et al. 2018; Shan 2013). We use data of 47 participants who have completed 5 year visits and have measurements for imaging volumes and memory scores.

By fitting a LMM (Hamlett et al. 2004), we have the estimated correlation of 0.421 between repeated hippocampal volumes and repeated scores on Rey Auditory Verbal Learning Test (RAVLT) delayed recall. We use the following four steps to compute exact SMP interval for ρ.

Step 1 Compute the variance–covariance matrix between repeated hippocampal volumes and repeated RAVLT delayed recall scores, and their mean vectors.

Step 2 We utilize the variance–covariance matrix and the mean values from Step 1 to simulate $M = 1000$ data sets with $n = 47$ in each data set.

Step 3 For each data set, we calculate the 95% SMP intervals using $B = 599$ bootstrap samples. The $M = 1,000$ data sets are simulated using an R function *mvrnorm*, and the bootstrapping samples are simulated from SAS procedures.

Step 4 For the proposed intervals SMP1 to SMP20, the SMP9 interval is identified as the optimal interval with the coverage probability above 95%. Then, the SMP9 interval is reported as the confidence interval for ρ.

The SMP9 interval and the existing intervals are presented in Table 2. The Asy interval has the shortest width, followed by the BC interval, the PCT interval, the BCa interval, the SMP9 interval, and the MP interval. The lower limit of the MP interval is the smallest among all these intervals. The lower limits of these intervals are all above 0, indicating that RAVLT delayed recall scores and hippocampal volumes are significantly correlated. If the null correlation is set as 0.15, the PM interval and the SMP interval fail to reject the null hypothesis while other intervals reject the null hypothesis. The MP interval is asymmetric when sample size is below 80. In this example, the lower limit of the MP interval is from the SMP8 interval, while the upper limit of the MP interval is from the SMP7 interval. For that reason, the MP interval is wider than the SMP interval in this example.

Table 2 Confidence interval and its width for the correlation between hippocampal volumes and scores on Rey Auditory Verbal Learning Test (RAVLT) delayed recall, by using 47 participants from the ADNI study

Asy CI		BC CI		BCa CI		PCT CI		MP CI		SMP9 CI	
Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper	Lower	Upper
0.222	0.619	0.196	0.601	0.178	0.596	0.190	0.599	0.143	0.631	0.148	0.616
Width of confidence interval											
0.397		0.405		0.418		0.409		0.488		0.468	

The correlation is estimated as 0.421

As pointed out by one of the reviewers, the true value of ρ and the variance-covariance matrix are unknown. We use the real data to estimate them by fitting a linear mixed effects model in Step 1, and then use these estimated values in Step 2 to simulate bootstrapping samples to compute confidence intervals. Theoretically, one should use the true values to simulate data. Since they are unknown, we used the estimated values to replace the true values in the simulations, which may introduce some variations to the coverage probability of confidence intervals.

5 Discussion

In this article, we propose the SMP interval for correlation in the presence of repeated measures, compared with the traditionally used asymptotic interval, the PCT interval, the BC interval, the BCa interval, and the MP interval. The developed software programs are available upon request. In the calculation of these intervals, the correlation for repeated measures ρ is estimated from a LMM under the assumption of a CS covariance structure. This assumption leads to a common ρ across different visits. In the case that correlations are not the same at different visits, suppose ρ_j is the correlation for the j-th visit. When a trend of correlation can be assumed, one may be interested in testing $\rho_1 \leq \rho_2 \leq \ldots, \leq \rho_{t_i}$. Then, the statements in the SAS Proc mixed should be modified in order to compute the correlation at each visit.

The confidence interval width could be affected by multiple factors. Increasing sample sizes would decrease the standard error of the longitudinal correlation. Then, the confidence interval width becomes shorter. There are several other parameters affecting the coverage property of confidence intervals for correlation as presented in simulation studies, including covariance (Shan and Ma 2014). When the variances are relatively large as compared to the mean values, confidence intervals become wide enough to have a high coverage probability. Research on limit theory for the coverage probability of correlation would be an interesting methodology topic to explore.

In a controlled study, the number of visits for each participant is often the same (Bernick et al. 2018). For a study with the number of visits being substantially different from one participant to another, the missing mechanism has to be investigated before data analysis. When the follow up time is the same within each center in a multi-center clinical trial, the bootstrap samples may be selected from each center for confidence interval calculation. The correlation studied in this article is the one without controlling for other variables. When other measures are correlated with the two considered measures in calculating the longitudinal correlation, a partial longitudinal correlation (Shan et al. 2020) is then the parameter of interest to provide a reliable correlation after removing the effects from other correlated measures.

When large volumes of data are involved, it can take an inordinately long period of time to process the data. Often, the same set of computations must be performed over and over again on different portions of the data. The computations in this paper

were performed using SAS. One of the lesser-known capabilities of SAS is parallel processing which can significantly reduce computational time.

Acknowledgements The authors are very grateful to Editor, Associate Editor and two referees for their insightful comments that help improve the manuscript. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

References

Aarts E, Verhage M, Veenvliet JV, Dolan CV, Van Der Sluis S (2014) A solution to dependency: using multilevel analysis to accommodate nested data. Nat Neurosci 17(4):491–496

Bakdash JZ, Marusich LR (2017) Repeated measures correlation. Front Psychol 8(MAR):456

Bernick C, Zetterberg H, Shan G, Banks S, Blennow K (2018) Longitudinal performance of plasma neurofilament light and tau in professional fighters: the professional fighters brain health study. J Neurotrauma 35(20):2351–2356

Bland JM, Altman DG (1995) Calculating correlation coefficients with repeated observations: part 1—correlation within subjects. BMJ (Clin Res Ed) 310(6977):446

Bland JM, Altman DG (1995) Calculating correlation coefficients with repeated observations: part 2—correlation between subjects. BMJ 310(6980):446

Casella G, Berger RL (2002) Statistical inference, 2nd edn. Thomson Learning, Belmont, CA

Crowder M (1995) On the use of a working correlation matrix in using generalised linear models for repeated measures. Biometrika 82(2):407–410

Cummings J (2018) Lessons learned from Alzheimer disease: clinical trials with negative outcomes. Clin Transl Sci 11(2):147–152

Efron B (1985) Bootstrap confidence intervals for a class of parametric problems. Biometrika 72(1):45–58

Efron B, Tibshirani RJ (1994) An introduction to the bootstrap, softcover edition edn. Monographs on statistics and applied probability. Chapman and Hall/CRC, London

Hall P (1986) On the number of bootstrap simulations required to construct a confidence interval. Ann Stat 14(4):1453–1462

Hamlett A, Ryan L, Wolfinger R (2004) On the use of PROC MIXED to estimate correlation in the presence of repeated measures. In: Proceedings of statistics and data analysis, pp 129–198

Irimata K, Wakim P, Li X (2018) Estimation of correlation coefficient in data with repeated measures. In: SAS paper, p 2424

Lam M, Webb KA, Donnell DE(1999) Correlation between two variables in repeated measures. In: Proceedings-American statistical association biometrics section, pp 213–218

Pearson K (1900) On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philos Mag Ser 5 50(302):157–175

Roy A (2006) Estimating correlation coefficient between two variables with repeated observations using mixed effects model. Biom J Biom Z 48(2):286–301

Shan G (2013) A note on exact conditional and unconditional tests for Hardy–Weinberg equilibrium. Hum Hered 76(1):10–17

Shan G, Banks S, Miller JB, Ritter A, Bernick C, Lombardo J, Cummings JL (2018) Statistical advances in clinical trials and clinical research. Alzheimer's Dement Transl Res Clin Interv 4:366–371

Shan G, Bayram E, Caldwell JZK, Miller JB, Shen JJ, Gerstenberger S (2020) Partial correlation coefficient for a study with repeated measurements. Stat Biopharm Res. https://doi.org/10.1080/19466315.2020.1784780

Shan G, Ma C (2014) Exact methods for testing the equality of proportions for binary clustered data from otolaryngologic studies. Stat Biopharm Res 6(1):115–122

Shan G, Vexler A, Wilding GE, Hutson AD (2011) Simple and exact empirical likelihood ratio tests for normality based on moment relations. Commun Stat Simul Comput 40(1):129–146

Wilcox Rand (2011) Modern statistics for the social and behavioral sciences: a practical introduction. CRC Press, Boca Raton

Wilcox RR (1996) Confidence intervals for the slope of a regression line when the error term has nonconstant variance. Comput Stat Data Anal 22(1):89–98

Affiliations

Guogen Shan[1] · Hua Zhang[2] · Jim Barbour[3] on behalf of for the Alzheimer's Disease Neuroimaging Initiative

✉ Guogen Shan
guogen.shan@unlv.edu

[1] Department of Epidemiology and Biostatistics, School of Public Health, University of Nevada, Las Vegas, Las Vegas, NV 89154, USA

[2] School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou, Zhejiang, China

[3] Experian Information Solutions, Inc., Costa Mesa, CA 92626, USA