Correlation Coefficients for a Study with Repeated Measures

Guogen Shan 1, Hua Zhang 2, and Tao Jiang 3

1Epidemiology and Biostatistics Program, School of Public Health, University of Nevada Las Vegas, Las Vegas, NV 89154, USA
2School of Computer and Information Engineering, Zhejiang Gongshang University, Hangzhou, Zhejiang, China
3School of Statistics and Mathematics, and School of Business, Zhejiang Gongshang University, Hangzhou, China

Correspondence should be addressed to Tao Jiang; jtao@263.net

Received 3 January 2020; Revised 2 February 2020; Accepted 12 February 2020; Published 26 March 2020

Copyright © 2020 Guogen Shan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Repeated measures are increasingly collected in a study to investigate the trajectory of measures over time. One of the first research questions is to determine the correlation between two measures. The following five methods for correlation calculation are compared: (1) Pearson correlation; (2) correlation of subject means; (3) partial correlation for subject effect; (4) partial correlation for visit effect; and (5) a mixed model approach. Pearson correlation coefficient is traditionally used in a cross-sectional study. Pearson correlation is close to the correlations computed from mixed-effects models that consider the correlation structure, but Pearson correlation may not be theoretically appropriate in a repeated-measure study as it ignores the correlation of the outcomes from multiple visits within the same subject. We compare these methods with regard to the average of correlation and the mean squared error. In general, correlation under the mixed-effects model with the compound symmetric structure is recommended as its correlation is close to the nominal level with small mean square error.

1. Introduction

Repeated-measure designs are increasingly used in practice to evaluate the trajectory of measures. The Alzheimer’s Disease Neuroimaging Initiative (ADNI) study is a longitudinal study to investigate the progression of Alzheimer’s disease (AD) [1, 2]. This study evaluates the normal cognitive aging with the focus on mild cognitive impairment (MCI) and early AD. Brain structure and function are two research areas of interest in the ADNI study. As expected, brain structure volumes are often highly associated with results from cognitive tests [3–5]. In a longitudinal study, correlation for repeated measures should be calculated and reported. However, recent articles still only reported the Pearson correlation coefficient that ignores the correlation of outcomes from the same subject. For these reasons, it is important to compare the existing correlations for repeated measures and make recommendations for other researchers to use.

Bland and Altman [6, 7] discussed several approaches to compute correlations for repeated measures. They proposed calculating subject means to compute the Pearson correlation, where subject means eliminate the correlation of outcomes from the same subject. The second approach is to fit a linear regression model with one measure as the dependent variable and the other measure and the subject as the predictor variables. The second approach is similar to the one proposed by Christensen [8] who suggested computing correlation after adjusting for the subject effect [9–12]. In a repeated-measure study, the visit effect is the correlation within the subject. Lipsitz et al. [13] proposed computing partial correlation adjusting the visit effect. When data are correlated, mixed-effects models may be utilized to analyze data while controlling for these additional correlations. Lam et al. [14] were among the first to propose computing correlation between repeated measures under the compound symmetric (CS) correlation structure. Later, Hamlett et al. [15] developed programs to compute correlation under the CS structure by using the commercially available statistical software, SAS. In the work by Lam et al. [14], they also computed the correlation under the autoregressive correlation structure, AR(1). After that, Roy [16] developed SAS macros to compute correlation under the AR(1) structure and compared the correlations for repeated measures under
these two correlation structures with limited simulation studies.

The objective of this manuscript is to conduct extensive simulation studies to compare the existing correlations for repeated measures with regard to the average of correlation and the mean squared error (MSE) and identify the correlation method that has the best performance to be used in practice. In addition to the parameter of interest (correlation for repeated measures), there are several nuisance parameters in the variance-covariance matrix: variances, correlations within each outcome, and correlation between outcomes from different visits [17–20]. It is computationally intensive for these comparisons. We have to use supercomputers for simulation studies. However, it is computationally feasible to calculate correlations for an observed data set. We use one example from the ADNI study to illustrate the application of the considered methods to calculate correlation between hippocampal volumes and a neuropsychological assessment to evaluate verbal memory.

We organize this article as follows. In Section 2, we introduce the existing methods to calculate correlations for repeated measures. In Section 3, we conduct extensive Monte Carlo simulation studies to compare the performance of the considered correlations with regard to the average of correlation and the MSE. A real example from the ADNI study is then used to illustrate the application of these correlations. Lastly, we provide conclusions in Section 4 on computing correlation for repeated measures when heterogeneity of correlation is observed across visits.

2. Methods

For a repeated-measure study with \( n \) participants, each participant has several scheduled visits (\( m_i \) visits for the \( i \)-th subject). Suppose \( U \) and \( W \) are the two measures in a repeated-measure study and \( U_{ij} \) and \( W_{ij} \) are the outcomes of the \( i \)-th subject at the \( j \)-th visit, where \( i = 1, 2, \ldots, n \) and \( j = 1, 2, \ldots, m_i \). The correlation between \( U \) and \( W \), \( \rho_{UW} \), is the parameter of interest to quantify a relationship between them. Several methods have been proposed to calculate \( \rho_{UW} \), including independence models, partial correlation models, and mixed-effects models.

2.1. Independent Assumption. Bland and Altman [6, 7] were among the first to provide methods to compute longitudinal correlation coefficient. One of their approaches assumes the independence between outcomes from the same subject: \( U_{ij} \perp \perp U_{ij'} \) and \( W_{ij} \perp \perp W_{ij'} \). The longitudinal correlation \( \rho_{UW} \) is computed as the Pearson correlation by ignoring the correlation structure from repeated measures. This approach is referred to as the I approach, with the computed correlation as \( \rho_I \). This is a naive approach that is easy to apply. Irimata and Li [21] found that \( \rho_I \) for a pharmacokinetics data set is very close to other correlations computed from other complicated models.

2.2. Subject Means. As suggested by Bland and Altman [6], the correlation can be computed by using the averages at the subject level to eliminate the subject effect in repeated measures. This correlation is able to address the research question whether the average of one measure is related to the average of another. When correlation within each measure is large, \( \rho_{UW} \) at different visits should be similar to each other, and this average correlation model would have good performance. We refer to this correlation approach as the \( M \) approach with the notation of \( \rho_M \).

These two correlations for repeated measures, \( \rho_I \) and \( \rho_M \), are the Pearson correlation and can be computed by using many statistical software: such as the Proc corr procedure in SAS and the function cor or cor.test in R [22]. The next five correlations are computed from regression models (e.g., mixed-effects models), and we would like to suggest using SAS Proc mixed procedure for implementation. Detailed SAS programs are provided in the Appendix.

2.3. Correlation Adjusting for the Subject Effect. Christensen [8] proposed computing correlation for repeated measures by partialling out the subject effect. The subject effect can be removed from the two measures by fitting a multivariate regression model with both measures being the outcomes and the subject ID as the only covariate. The residuals are used to compute the final correlation, which is essentially a partial correlation method for repeated data. This correlation is referred to as the PS correlation that partials out the subject effect, \( \rho_{PS} \).

2.4. Correlation Adjusting for the Visit Effect. In the \( \rho_{PS} \) calculation, the correlation between the two measures is included in the multivariate model. In addition to that correlation, another correlation between measures at different visits may be considered. Lipsitz et al. [13] proposed computing partial correlation between outcome and one of the covariates by using this approach. When one of the two measures (e.g., measure \( U \)) is considered as the dependent variable, the other measure (\( W \)) is considered as the covariate. The correlation structure between visits is assumed to be compound symmetric. We refer this correlation as the \( \rho_{PVU} \) correlation. We use \( \rho_{PVW} \) for another correlation when \( W \) is considered as the dependent variable in the model. One of the properties for correlation is \( \rho_{UW} = \rho_{WU} \), but this property is not met here; \( \rho_{PVU} \) is generally not equal to \( \rho_{PVW} \).

2.5. Mixed-Effects Model. Let \( Y_i = (U_{i1}, W_{i1}, U_{i2}, W_{i2}, \ldots, U_{im}, W_{im}) \) be the outcomes from the \( i \)-th subject, with the vector length of \( 2m_i \). The complete data can be reorganized in a long format, with the columns subject ID, visit, mtype, and outcome, where mtype = “U” for the \( U \) measure and mtype = “W” for the \( W \) measure. The long format utilizes \( 2m_i \) rows for the outcomes from \( Y_i \). The linear mixed-effects model 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 and the random effect, respectively. The random effect \( b_i \) follows a multivariate normal distribution \( N(0, D) \), and the
measurement error $e_i$ follows a multivariate normal distribution $N(0, R)$. The detailed formula for $D$ and $R$ may be found in the article by Hamlett et al. [15]. The fixed effect is $\beta = (\beta_0, \beta_U, \beta_W)^\prime$, where $\beta_0$ is the intercept, and $\beta_U$ and $\beta_W$ are the fixed effects of $U$ and $W$, respectively. Correlation between $U$ and $W$ is computed as

$$\rho_{UW} = \text{Corr}(U, W),$$

which is assumed to be independent of the visit.

Each subject has multiple visits, correlation within $U$ is

$$\text{Corr}(U_{ij}, U_{i'j'}) = \rho_U^{d(j-j')}$$

and the correlation within $W$ is

$$\text{Corr}(W_{ij}, W_{i'j'}) = \rho_W^{d(j-j')}$$

where $d(j - j') = 1$ for the CS structure and $d(j - j') = \mid j - j' \mid$ for the AR(1) structure. Since $W_{ij}$ is correlated with both $U_{ij}$ and $W_{i'j'}$, therefore, $U_{ij}$ and $W_{ij}$ are correlated and their correlation is assumed to be

$$\delta_{ij} U_{ij} W_{ij}$$

where $\delta$ is a factor which is generally less than 1. Let $\sigma_U^2$ and $\sigma_W^2$ be the variances of $U$ and $W$, respectively. These variances and covariances are used to derive the variance-covariance matrix under the CS structure (see Lam et al. [14] and Hamlett et al. [15]) and that under the AR(1) structure (see Lam et al. [14] and Roy [16]).

3. Results

We conducted simulation studies to compare the performance of the considered 7 methods for the correlation between repeated measures for a study with four visits. The mean values of $U$ and $W$ are assumed to be (2.0, 1.9, 1.7, 1.4) and (0.8, 0.7, 0.6, 0.5), with both measures decreasing as time goes. Such data are commonly available from cognitive tests on elderly population and other studies. The prespecified correlation for repeated measures is $\rho_{UW} = 0.2, 0.5, \text{and } 0.8$.

In the simulation studies for the AR(1) structure for the visit effect, the correlation within $U$ is

$$\text{Corr}(U_{ij}, U_{i'j'}) = \rho_U^{d(j-j')}$$

with $\rho_U = 0.2, 0.5, \text{and } 0.8$, and the correlation within $W$ is

$$\text{Corr}(W_{ij}, W_{i'j'}) = \rho_W^{d(j-j')}$$

with $\rho_W = 0.2, 0.5, \text{and } 0.8$. The factor $\delta$ in the correlation between $U$ and $W$ is assumed to be 0.6 in all simulations. The considered variances are $\sigma_U^2 = 1$ and $3$ and $\sigma_W^2 = 0.5$ and 1. The variance-covariance matrix can be separated into two parts: $Z_i D Z_i^\prime$ and $R_i$. We assume that a quarter of variance is from $R_i$ and the remaining is from $Z_i D Z_i^\prime$. This weight is needed in order to calculate the covariances. For each configuration, we simulate $B = 2,000$ data sets.

Under the AR(1) structure for the visit effect, Figure 1 presents the average of correlation $\hat{\rho}_{UW}$ and the MSE when $\rho_{UW} = 0.2$, $\sigma_U^2 = 1$, and $n = 60$ subjects. The MSE is defined as

$$\text{MSE} = \frac{1}{B} \sum_{b=1}^{B} (\hat{\rho}_{UW}(b) - \rho_{UW})^2,$$

where $\hat{\rho}_{UW}(b)$ is the estimator of $\rho_{UW}$ by using the $b$-th simulated data set. It can be seen that the correlations adjusting the visit effect, $\rho_{PV_a}$, and $\rho_{PV_b}$, often underestimate the correlation, while the correlation adjusting the subject effect, $\rho_{PS}$, always overestimate the correlation. The remaining methods have correlations close to the nominal level. Although $\rho_M$ is the best with the correlation around the nominal level, its MSE is much larger than the ones that have the correlations close to the nominal level. In the calculation of $\rho_M$, each subject only has one outcome for each measure, as compared to multiple outcomes in other correlation calculations. Due to the reduced number of outcomes, the variance of $\rho_M$ is much large that leads to a large MSE. It is noted that $\rho_{PV_a}$ or $\rho_{PV_b}$ could have the lowest MSE in some cases, but their estimated correlations are generally much below the nominal level. For this reason, we exclude $\rho_{PV_a}$ and $\rho_{PV_b}$ in the following simulation studies. When a study has the same number of visits for each subject, the estimated correlation by using the mixed-effects model with the CS structure, $\rho_{CS}$, is very similar to $\rho_I$ under the independent assumption. The other mixed-effects model correlation $\rho_{AR}$ has a similar correlation as $\rho_{CS}$ and $\rho_I$. The MSE of $\rho_{AR}$ is slightly smaller than the MSEs of $\rho_{CS}$ and $\rho_I$ when the correlations within $U$ or $W$ are small, and this trend is reversed when $\rho_U$ and $\rho_W$ are large. Similar results are observed when $\sigma_W^2$ is increased to 3.

When $\rho_U$ is increased to 0.5 (the top plot in Figure 2), the averages of $\rho_I$, $\rho_{CS}$, and $\rho_{AR}$ are generally above the nominal level, and the first two correlations are closer to the nominal level as compared to the third correlation $\rho_{AR}$. We also present the correlation estimates when sample size $n$ is 100 in Figure 2. It can be seen that the MSEs become smaller as compared to the MSEs in the top plot (Figure 2) when sample size is 60.

Figure 3 shows the results when data sets are simulated under the CS structure given $\rho_I = 0.5$, $\sigma_U^2 = 1$, and $n = 60$. Correlation $\rho_{PS}$ does not perform well with the average correlations much below the nominal level in many configurations. We also found that $\rho_M$ is likely to overestimate the correlation. It seems that $\rho_M$ and $\rho_{PS}$ have different trajectories as $\rho_W$ increases. Both of these methods do not have satisfactory performance with regard to correlation under the CS structure, although $\rho_{PS}$ has very good correlation estimates under the AR(1) structure. The other three correlations ($\rho_I$, $\rho_{CS}$, and $\rho_{AR}$) have similar good performance with regard to correlation and the MSE. It should be noted that the variance-covariance matrix is not positively defined when $\rho_I = \rho_W = 0.8$. Therefore, data sets cannot be generated for that configuration. We also simulate data under the unstructured correlation structure and found that $\rho_I$, $\rho_{CS}$, and $\rho_{AR}$ are still the best correlation estimates.

The aforementioned simulation studies have data sets that each subject has the same number of visits. In practice, it is possible that the number of visits may not be exactly the same for all subjects. We assume the number of visits is either 2, 3, or 4. Each subject is randomly assigned to have 2, 3, or 4 visits with the same probability. We present the results with $n = 60$ in Figure 4 when variances are small ($\sigma_U^2 = 1$ and $\sigma_W^2 = 0.5$ and 1) and large ($\sigma_U^2 = 20$ and $\sigma_W^2 = 10$ and 30). The MSE of $\rho_{CS}$ is slightly smaller than that of $\rho_I$, and their biggest difference occurs when both $\rho_U$ and $\rho_W$ are large. $\rho_{AR}$ is more likely to overestimate the correlation. Although $\rho_M$ has the correlation very close to the nominal level, it has the largest MSE as compared to other correlations. When variance is large, $\rho_I$ and $\rho_{CS}$ are the best correlations with the estimated correlations much closer to the
nominal level as to the configurations with small variances. The mixed-effects model correlation \( \rho_{CS} \) performs slightly better than \( \rho_{I} \) with regard to the average of correlation and the MSE.

3.1. Example. We use one data set from the ADNI study to illustrate the application of the considered correlation methods, with 47 participants who had 5-year visits and completed imaging volumes and memory scores. Hippocampal volumes are found to be highly associated with the delayed recall scores from the Rey Auditory Verbal Learning Test (RAVLT delayed recall) [23]. The RAVLT delayed recall has the possible integer score from 0 to 15, which is often used to assess verbal memory. The higher the score is, the better the memory is.
Figure 2: Continued.
The computed correlations are presented in Table 1. Participants in this data set have the same number of visits. For this reason, $\rho_I$ is very similar to $\rho_{CS}$. $\rho_M$ is slightly larger than them, and $\rho_{AR}$ is smaller than them. Correlation adjusted by the subject effect $\rho_{PS}$ is much smaller than $\rho_{CS}$. Correlations adjusted by the visit effect highly depend on which variable is considered as the dependent variable in the linear regression model. When hippocampal volumes are used as the dependent variable, the estimated correlation is high (0.686), and it becomes too low (0.016) when RAVLT delayed recalls are considered as the dependent variable.

It was reported by Wang et al. [23] that the Pearson correlation $\rho_I$ between hippocampal volumes and RAVLT delayed recall scores is slightly above 0.4. They also provided the Pearson correlations for each group (AD, MCI, and control) which are all below the correlation using combined samples. The correlation within the control group is the lowest.
Figure 3: Average correlation and the MSE under the CS correlation structure when $\rho_{uw} = 0.5$, $\sigma_u^2 = 1$, and $n = 60$. 
Figure 4: Continued.
Figure 4: Average correlation and the MSE under the AR(1) correlation structure with a small variance $\sigma_u^2 = 1$ (top) and a large variance $\sigma_u^2 = 20$ (bottom) when $\rho_{UW} = 0.5$ and $n = 60$ for a study with unequal numbers of visits (2, 3, or 4 visits).

Table 1: Correlation between hippocampal volumes and RAVLT delayed recall scores using 47 participants with 5 visits from the ADNI study.

<table>
<thead>
<tr>
<th></th>
<th>$\rho_I$</th>
<th>$\rho_M$</th>
<th>$\rho_{PS}$</th>
<th>$\rho_{PV,a}$</th>
<th>$\rho_{PV,b}$</th>
<th>$\rho_{CS}$</th>
<th>$\rho_{AR}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left hippocampal and RAVLT delayed recall scores</td>
<td>0.421</td>
<td>0.468</td>
<td>0.151</td>
<td>0.016</td>
<td>0.686</td>
<td>0.421</td>
<td>0.392</td>
</tr>
<tr>
<td>Left hippocampal and RAVLT immediate recall scores</td>
<td>0.352</td>
<td>0.421</td>
<td>0.208</td>
<td>0.023</td>
<td>0.447</td>
<td>0.365</td>
<td>0.399</td>
</tr>
<tr>
<td>Right hippocampal and RAVLT delayed recall scores</td>
<td>0.361</td>
<td>0.398</td>
<td>0.149</td>
<td>0.014</td>
<td>0.652</td>
<td>0.361</td>
<td>0.327</td>
</tr>
<tr>
<td>Right hippocampal and RAVLT immediate recall scores</td>
<td>0.316</td>
<td>0.373</td>
<td>0.211</td>
<td>0.021</td>
<td>0.443</td>
<td>0.335</td>
<td>0.343</td>
</tr>
</tbody>
</table>
From Table 1, RAVLT delayed recall scores always have a larger correlation with left hippocampal volumes than the correlation with right hippocampal volumes for each correlation method. We also add RAVLT immediate recall scores to further illustrate the application of the considered methods. Its correlation with left hippocampal volumes is often larger than its correlation with right hippocampal volumes. The estimated $\rho_{CS}$ between hippocampal volumes and RAVLT delayed recalls is larger than that between hippocampal volumes and RAVLT immediate recalls.

4. Conclusions

From the simulation studies, $\rho_I$ under the independence assumption and $\rho_{CS}$ using the mixed-effects model with CS variance-covariance structure are shown to have similar correlation estimates when subjects have the same number of visits. But, $\rho_{CS}$ is appropriate as it models the data properly. The mixed-effects model correlation $\rho_{CS}$ is recommended for use as its correlation is close to the nominal level with small mean square error.

5. Discussions

Lam et al. [14] derived the detailed variance and covariance. The variances $\sigma_W^2$ and $\sigma_U^2$ and covariance $\sigma_{IU}^2$ are used to calculate $\rho_{IU}$. These variances and covariance estimates are not exactly the same from the independent model and the mixed-effects model with the CS structure: $\sigma_W^2 = 16.6846$ in the $\rho_I$ calculation and 16.6136 from the CS model. Because these estimated variances and covariance are very close between these two methods, the final estimated correlations are very similar. When a study has different number of follow-up for each participant, $\rho_I$ and $\rho_{CS}$ differ from each other [18, 24–26]. For a study with some possible outliers as seen in the data testing implementation of ADNI and/or provided 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.

Conflicts of Interest

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

Acknowledgments

Data collection and sharing for this project were 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 was 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 (http://www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study was coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data were disseminated by the Laboratory for Neuro Imaging at the University of Southern California. Shan’s research was partially supported by grants from the National Institute of General Medical Sciences from the National Institutes of Health; P20GM109025. Jiang’s work was supported by the National Natural Foundation of China under grant 11971433, and the First Class Discipline of Zhejiang –A (Zhejiang Gongshang University-Statistics).

Data Availability

The data used in preparation of this article were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (http://adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided 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.

Supplementary Materials

Supplement is the R code. SAS programs: correlation for repeated measures from 8 critically ill patients: pH and PaCO₂ [6]. (Supplementary Materials)
References


