UNLV | **University Libraries**
University of Nevada, Las Vegas

Environmental & Occupational Health Faculty Publications

Environmental and Occupational Health

3-26-2020

Correlation Coefficients for a Study with Repeated Measures

Guogen Shan
University of Nevada, Las Vegas, guogen.shan@unlv.edu

Hua Zhang
Zhejiang Gongshang University

Tao Jiang
Zhejiang Gongshang University, jtao@263.net

Follow this and additional works at: https://digitalscholarship.unlv.edu/env_occ_health_fac_articles

Part of the Applied Mathematics Commons

Research Article

Correlation Coefficients for a Study with Repeated Measures

Guogen Shan [ID],[1] **Hua Zhang** [ID],[2] **and Tao Jiang** [ID][3]

[1]*Epidemiology and Biostatistics Program, 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]*School of Statistics and Mathematics, and School of Business, Zhejiang Gongshang University, Hangzhou, China*

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

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, ρ_{UW}, is the parameter of interest to quantify a relationship between them. Several methods have been proposed to calculate ρ_{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: $U_{ij} \perp U_{ij'}$ and $W_{ij} \perp W_{ij'}$. The longitudinal correlation ρ_{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 ρ_I. This is a naive approach that is easy to apply. Irimata and Li [21] found that ρ_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, ρ_{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 ρ_M.

These two correlations for repeated measures, ρ_I and ρ_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, ρ_{PS}.

2.4. Correlation Adjusting for the Visit Effect.
In the ρ_{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 ρ_{PVa} correlation. We use ρ_{PVb} 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: ρ_{PVa} is generally not equal to ρ_{PVb}.

2.5. Mixed-Effects Model.
Let $Y_i = (U_{i1}, W_{i1}, U_{i2}, W_{i2}, \ldots, U_{im_i}, W_{im_i})$ 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

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i, \tag{1}$$

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 ϵ_i follows a multivariate normal distribution $N(0, R_i)$. The detailed formula for D and R_i may be found in the article by Hamlett et al. [15]. The fixed effect is $\beta = (\beta_0, \beta_U, \beta_W)'$, where β_0 is the intercept, and β_U and β_W are the fixed effects of U and W, respectively. Correlation between U and W is computed as

$$\rho_{UW} = \text{Corr}(U, W), \qquad (2)$$

which is assumed to be independent of the visit.

Each subject has multiple visits, correlation within U is $\text{Corr}(U_{ij}, U_{ij'}) = \rho_U^{d(j-j')}$, and the correlation within W is $\text{Corr}(W_{ij}, W_{ij'}) = \rho_W^{d(j-j')}$, where $d(j - j') = 1$ for the CS structure and $d(j - j') = |j - j'|$ for the AR(1) structure. Since W_{ij} is correlated with both U_{ij} and $W_{ij'}$, therefore, U_{ij} and $W_{ij'}$ are correlated and their correlation is assumed to be $\delta \rho UW$, where δ is a factor which is generally less than 1. Let σ_U^2 and σ_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 conduct 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 $\rho_{UW} = 0.2, 0.5,$ and 0.8.

In the simulation studies for the AR(1) structure for the visit effect, the correlation within U is $\text{Corr}(U_{ij}, U_{ij'}) = \rho_U^{|j-j'|}$, with $\rho_U = 0.2, 0.5,$ and 0.8, and the correlation within W is $\text{Corr}(W_{ij}, W_{ij'}) = \rho_W^{|j-j'|}$, with $\rho_W = 0.2, 0.5,$ and 0.8. The factor δ in the correlation between U_{ij} and $W_{ij'}$ 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: $Z_i D Z_i'$ and R_i. We assume that a quarter of variance is from R_i and the remaining is from $Z_i D Z_i'$. 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 ρ_{UW} and the MSE when $\rho_{UW} = 0.2, \sigma_U^2 = 1,$ and $n = 60$ subjects. The MSE is defined as

$$\text{MSE} = \frac{1}{B} \sum_{b=1}^{B} (\widehat{\rho}_{UW}(b) - \rho_{UW})^2, \qquad (3)$$

where $\widehat{\rho}_{UW}(b)$ is the estimator of ρ_{UW} by using the b-th simulated data set. It can be seen that the correlations adjusting the visit effect, ρ_{PVa} and ρ_{PVb}, often underestimate the correlation, while the correlation adjusting the subject effect, ρ_{PS}, always overestimate the correlation. The remaining methods have correlations close to the nominal level. Although ρ_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 ρ_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 ρ_M is much large that leads to a large MSE. It is noted that ρ_{PVa} or ρ_{PVb} could have the lowest MSE in some cases, but their estimated correlations are generally much below the nominal level. For this reason, we exclude ρ_{PVa} and ρ_{PVb} 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, ρ_{CS}, is very similar to ρ_I under the independent assumption. The other mixed-effects model correlation ρ_{AR} has a similar correlation as ρ_{CS} and ρ_I. The MSE of ρ_{AR} is slightly smaller than the MSEs of ρ_{CS} and ρ_I when the correlations within U or W are small, and this trend is reversed when ρ_U and ρ_W are large. Similar results are observed when σ_U^2 is increased to 3.

When ρ_U is increased to 0.5 (the top plot in Figure 2), the averages of $\rho_I, \rho_{CS},$ and ρ_{AR} are generally above the nominal level, and the first two correlations are closer to the nominal level as compared to the third correlation ρ_{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 $\rho_U = 0.5, \sigma_U^2 = 1,$ and $n = 60$. Correlation ρ_{PS} does not perform well with the average correlations much below the nominal level in many configurations. We also found that ρ_M is likely to overestimate the correlation. It seems that ρ_M and ρ_{PS} have different trajectories as ρ_W increases. Both of these methods do not have satisfactory performance with regard to correlation under the CS structure, although ρ_{PS} has very good correlation estimates under the AR(1) structure. The other three correlations ($\rho_I, \rho_{CS},$ and ρ_{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 $\rho_U = \rho_W = 0.8$. 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 ρ_{AR} are still the best correlation estimates.

The aforementioned simulations 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 ρ_{CS} is slightly smaller than that of ρ_I, and their biggest difference occurs when both ρ_U and ρ_W are large. ρ_{AR} is more likely to overestimate the correlation. Although ρ_M has the correlation very close to the nominal level, it has the largest MSE as compared to other correlations. When variance is large, ρ_I and ρ_{CS} are the best correlations with the estimated correlations much closer to the
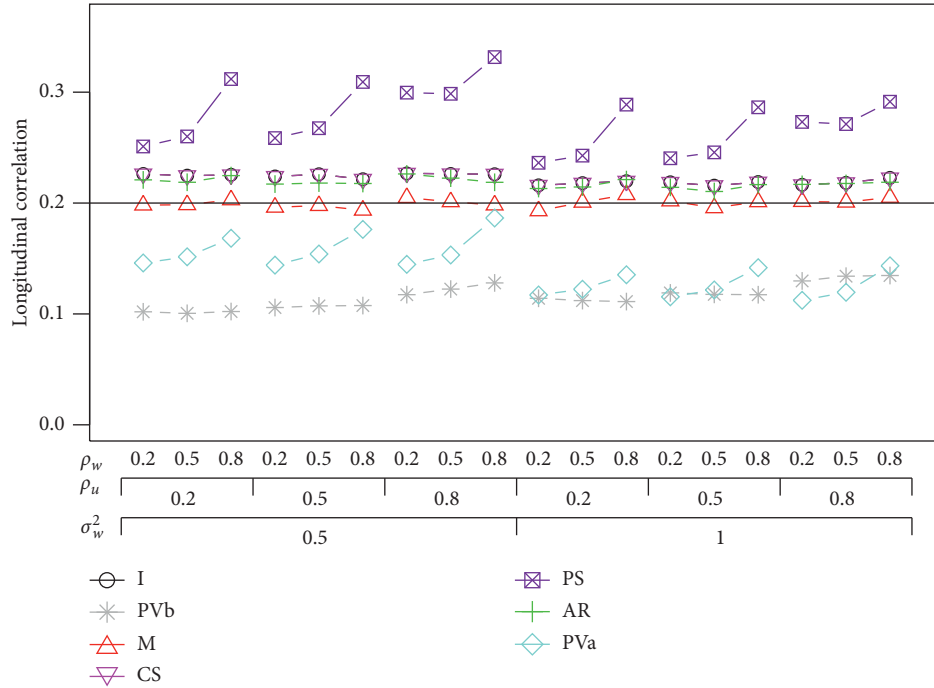
(a)
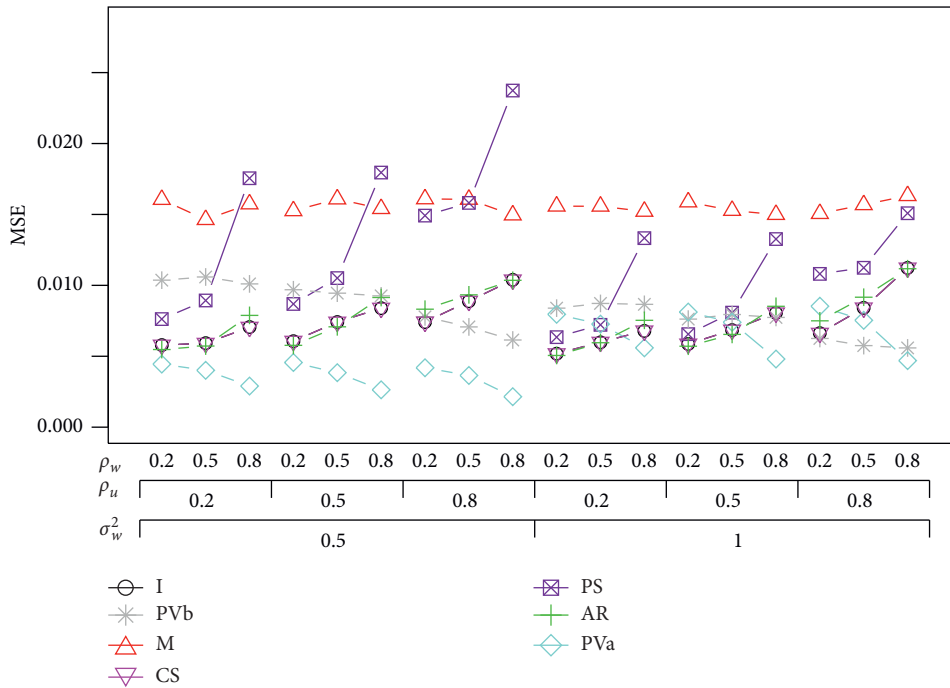
(b)

FIGURE 1: Average correlation and the MSE for the 7 methods under the AR(1) correlation structure when $\rho_{UW} = 0.2$, $\sigma_U^2 = 1$, and $n = 60$.

nominal level as to the configurations with small variances. The mixed-effects model correlation ρ_{CS} performs slightly better than ρ_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.
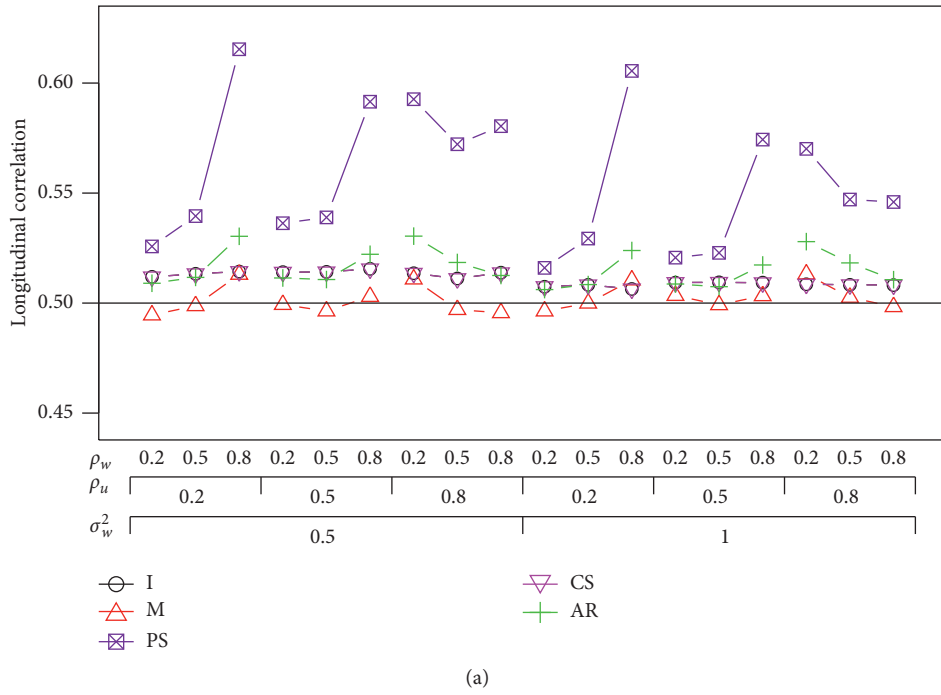
(a)
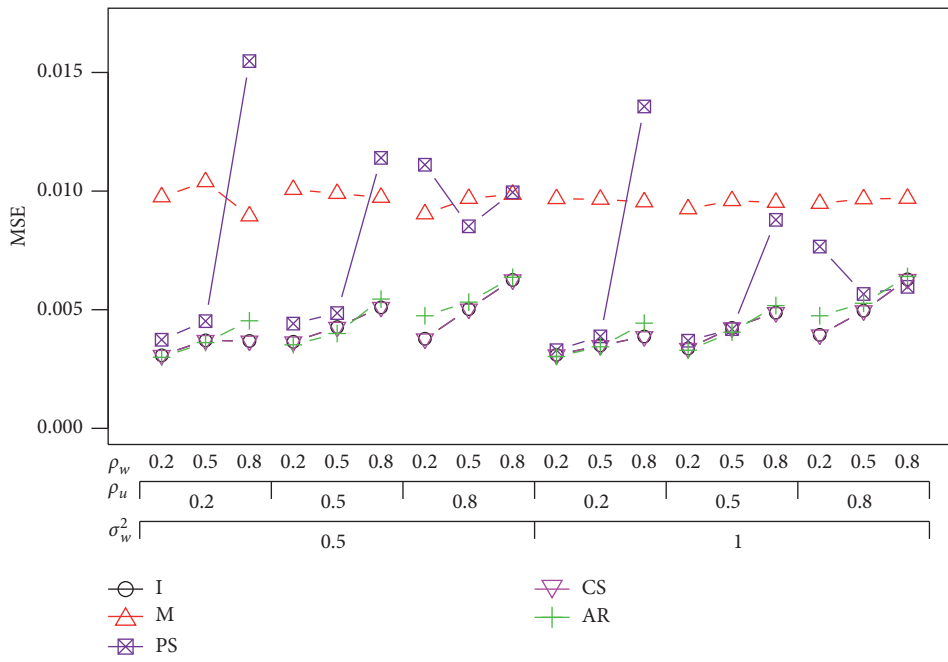
(b)
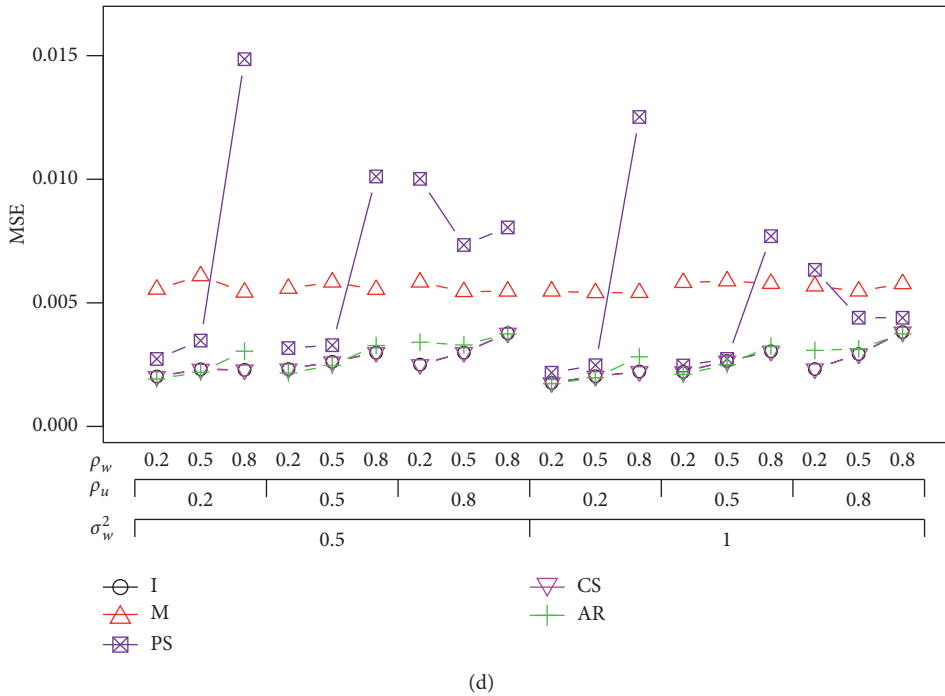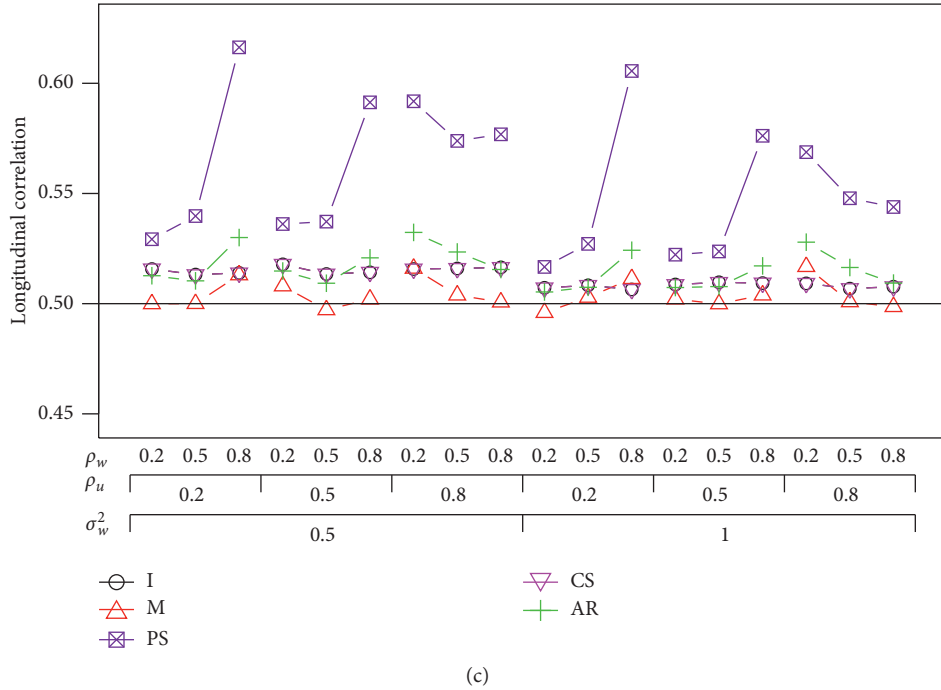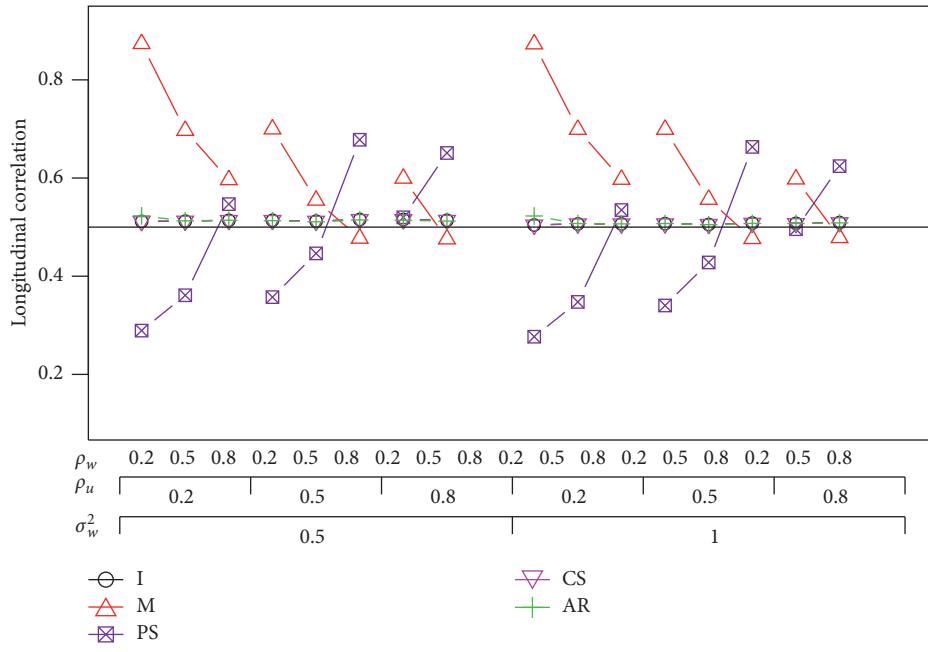
Figure 2: Continued.

(c)

(d)

FIGURE 2: Average correlation and the MSE under the AR(1) correlation structure with $n = 60$ (top) and $n = 100$ (bottom) when $\rho_{UW} = 0.5$ and $\sigma_U^2 = 1$.

The computed correlations are presented in Table 1. Participants in this data set have the same number of visits. For this reason, ρ_I is very similar to ρ_{CS}. ρ_M is slightly larger than them, and ρ_{AR} is smaller than them. Correlation adjusted by the subject effect ρ_{PS} is much smaller than ρ_{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 ρ_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.
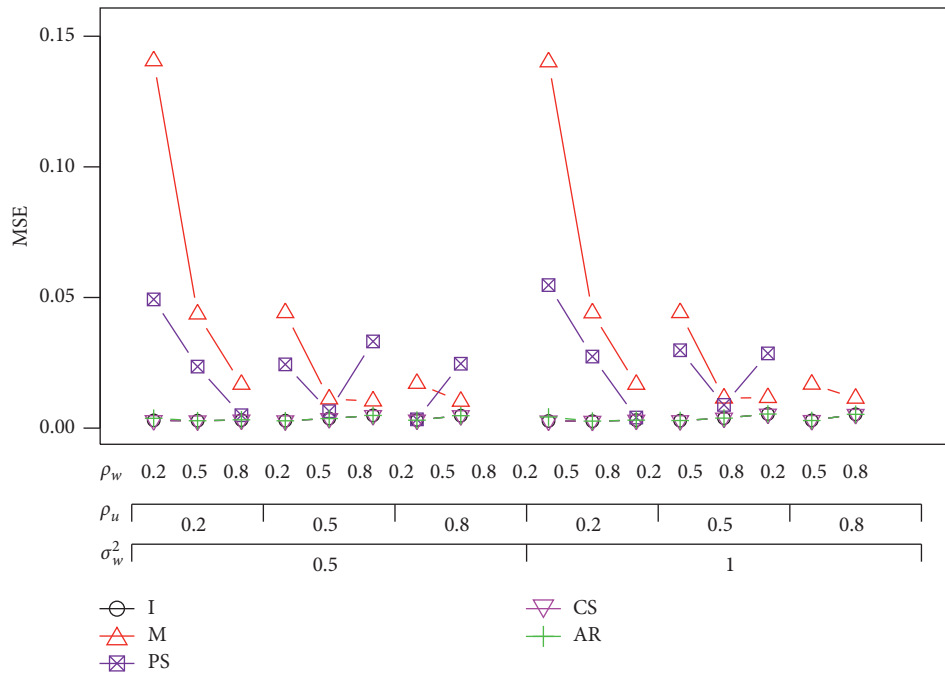
FIGURE 3: Average correlation and the MSE under the CS correlation structure when $\rho_{UW} = 0.5$, $\sigma_U^2 = 1$, and $n = 60$.

(a)

(b)
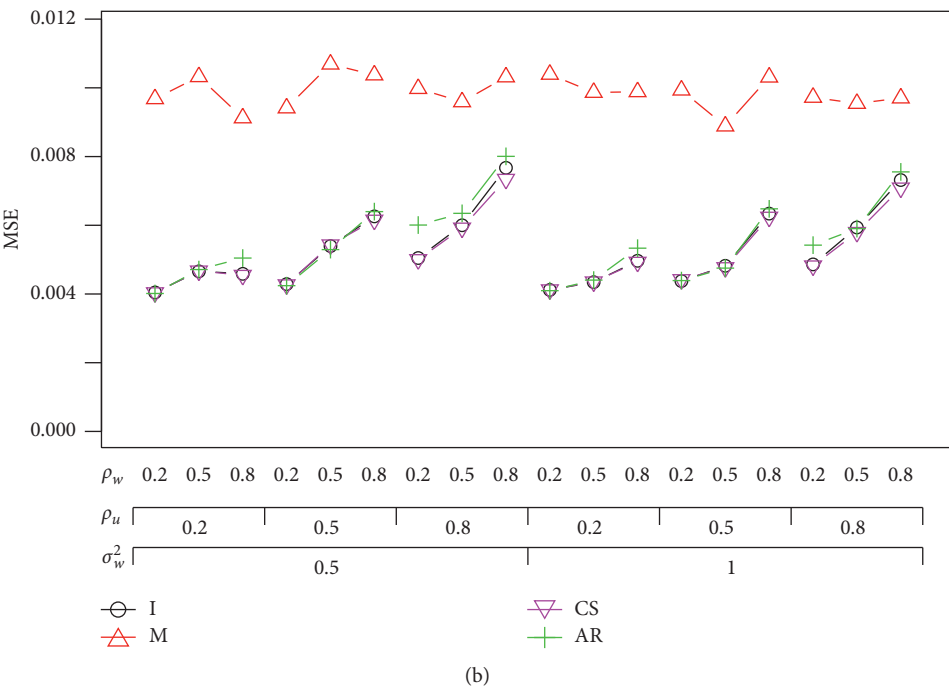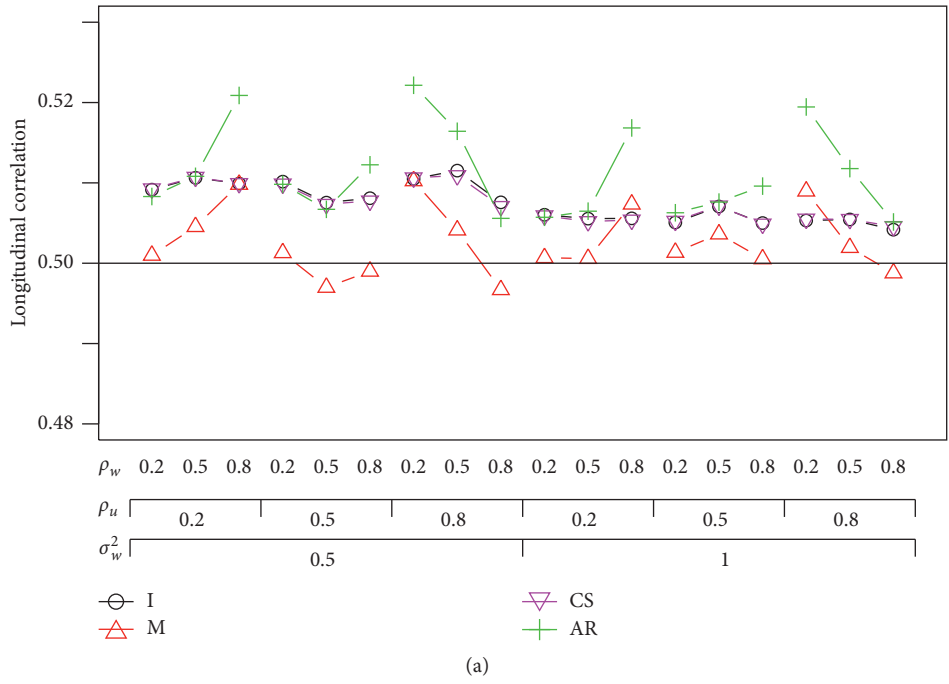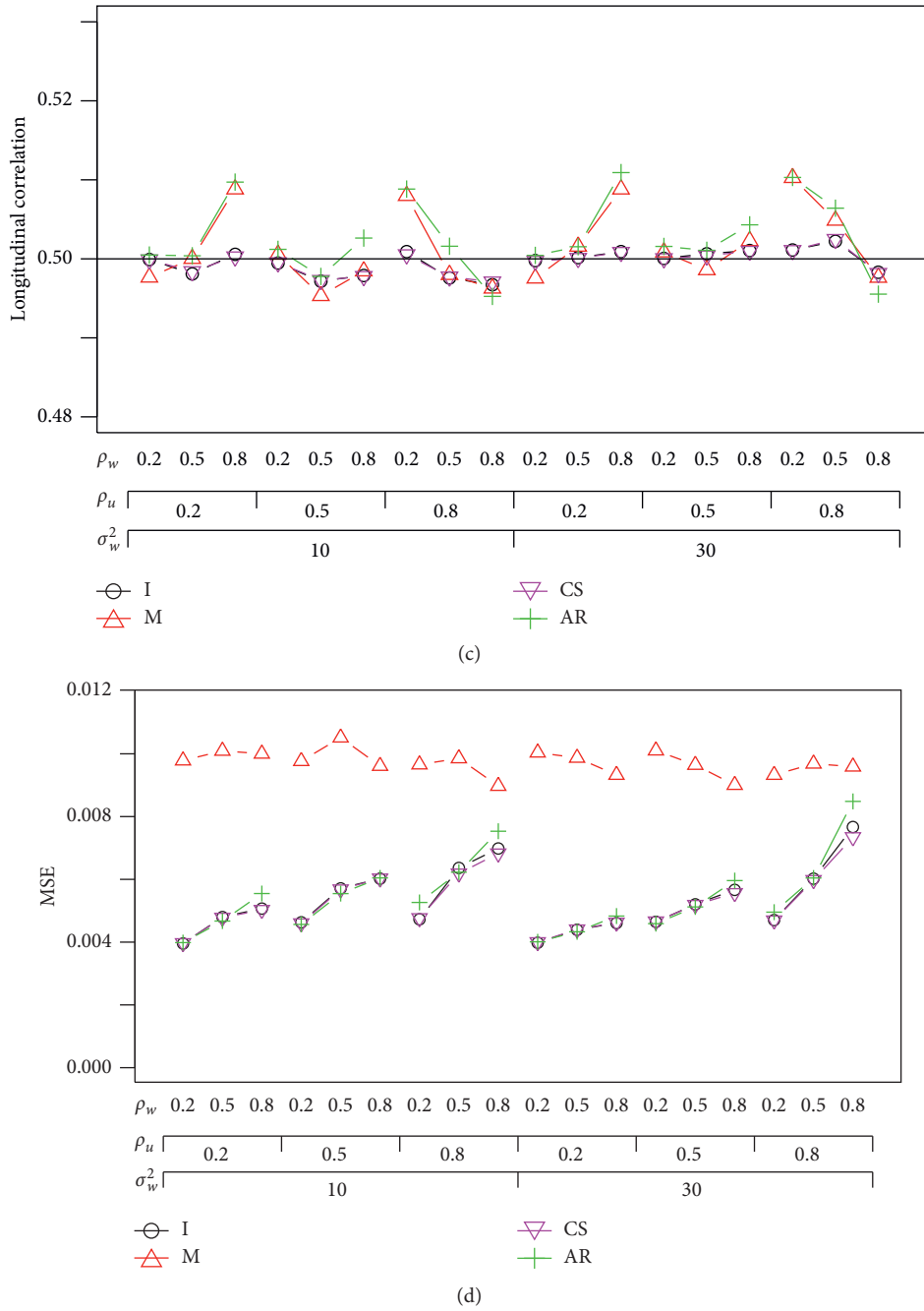
Figure 4: Continued.

(c)

(d)

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.

	ρ_I	ρ_M	ρ_{PS}	ρ_{PVa}	ρ_{PVb}	ρ_{CS}	ρ_{AR}
Left hippocampal and RAVLT delayed recall scores	0.421	0.468	0.151	0.016	0.686	0.421	0.392
Left hippocampal and RAVLT immediate recall scores	0.352	0.421	0.208	0.023	0.447	0.365	0.399
Right hippocampal and RAVLT delayed recall scores	0.361	0.398	0.149	0.014	0.652	0.361	0.327
Right hippocampal and RAVLT immediate recall scores	0.316	0.373	0.211	0.021	0.443	0.335	0.343

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 ρ_{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, ρ_I under the independence assumption and ρ_{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, ρ_{CS} is appropriate as it models the data properly. The mixed-effects model correlation ρ_{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 σ_U^2 and σ_W^2 and covariance σ_{UW} are used to calculate ρ_{UW}. 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 ρ_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, ρ_I and ρ_{CS} differ from each other [18, 24–26]. For a study with some possible outliers as seen in the data testing association between pH and $PaCO_2$ [6], their difference is substantial. We provide the SAS programs by using that example in the Appendix.

When CS or AR(1) correlation structure for the visit effect is applied in the mixed-effects models [10, 25, 27], the computed correlation is the same at different visits. In the observation of the heterogeneity of correlations at different visits, the unstructured correlation may be considered for the visit effect. Under the heterogeneity assumption, correlation can be computed at each visit from a mixed-effects model [28–30]. This brings some challenges to explain the results, such as the overall correlation and the trend of correlation. Alternatively, when a study has a monotonic relationship between correlation and visit, one may include an additional predictor: visit, in the statistical model, to calculate a monotonic correlation for repeated measures.

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.pd.

Conflicts of Interest

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

Supplementary Materials

Supplement is the R code. SAS programs: correlation for repeated measures from 8 critically ill patients: pH and $PaCO_2$ [6]. (*Supplementary Materials*)

References

[1] E. Bayram, N. Kaplan, G. Shan, and J. Z. K. Caldwell, "The longitudinal associations between cognition, mood and striatal dopaminergic binding in Parkinson's disease," *Aging, Neuropsychology, and Cognition*, pp. 1–14, 2019.

[2] C. Bernick, H. Zetterberg, G. Shan, S. Banks, and K. Blennow, "Longitudinal performance of plasma neurofilament light and tau in professional fighters: the professional fighters brain health study," *Journal of Neurotrauma*, vol. 35, no. 20, pp. 2351–2356, 2018.

[3] G. Shan, "More efficient unconditional tests for exchangeable binary data with equal cluster sizes," *Statistics & Probability Letters*, vol. 83, no. 2, pp. 644–649, 2013.

[4] G. Shan and C. Ma, "Exact methods for testing the equality of proportions for binary clustered data from otolaryngologic studies," *Statistics in Biopharmaceutical Research*, vol. 6, no. 1, pp. 115–122, 2014.

[5] G. Shan and G. E. Wilding, "Powerful exact unconditional tests for agreement between two raters with binary endpoints," *PLoS One*, vol. 9, no. 5, Article ID e97386, 2014.

[6] J. M. Bland and D. G. Altman, "Calculating correlation coefficients with repeated observations: part 1–correlation within subjects," *BMJ*, vol. 310, no. 6977, p. 446, 1995.

[7] J. M. Bland and D. G. Altman, "Statistics notes: calculating correlation coefficients with repeated observations: part 2—correlation between subjects," *BMJ*, vol. 310, no. 6980, p. 633, 1995.

[8] R. Christensen, *Plane Answers to Complex Questions: The Theory of Linear Models*, Springer, New York, NY, USA, 1996.

[9] S. Litwin, B. Stanley, and E. A. Ross, "Two-sample binary phase 2 trials with low type I error and low sample size," *Statistics in Medicine*, vol. 36, no. 9, 2017.

[10] G. Shan, *Exact Statistical Inference for Categorical Data*, Academic Press, San Diego, CA, USA, 1st edition, 2015.

[11] G. Shan, G. E. Wilding, A. D. Hutson, and S. Gerstenberger, "Optimal adaptive two-stage designs for early phase II clinical trials," *Statistics in Medicine*, vol. 35, no. 8, pp. 1257–1266, 2016.

[12] G. Shan, H. Zhang, T. Jiang, H. Peterson, D. Young, and C. Ma, "Exact p values for Simon's two-stage designs in clinical trials," *Statistics in Biosciences*, vol. 8, no. 2, pp. 351–357, 2016.

[13] S. R. Lipsitz, T. Leong, J. Ibrahim, and S. Lipshultz, "A partial correlation coefficient and coefficient of determination for multivariate normal repeated measures data," *Journal of the Royal Statistical Society: Series D*, vol. 50, no. 1, pp. 87–95, 2001.

[14] M. Lam, K. A. Webb, and Donnell, "Correlation between two variables in repeated measures," in *Proceedings of the American Statistical Association Biometrics Section*, pp. 213–218, Alexandria, VA, USA, 1999.

[15] A. Hamlett, L. Ryan, and R. Wolfinger, "On the use of PROC MIXED to estimate correlation in the presence of repeated measures," *Proc Statistics and Data Analysis*, pp. 129–198, 2004.

[16] A. Roy, "Estimating correlation coefficient between two variables with repeated observations using mixed effects model," *Biometrical Journal*, vol. 48, no. 2, pp. 286–301, 2006.

[17] G. Shan, "A better confidence interval for the sensitivity at a fixed level of specificity for diagnostic tests with continuous endpoints," *Statistical Methods in Medical Research*, vol. 26, no. 1, pp. 268–279, 2017.

[18] G. Shan, "Exact confidence limits for the probability of response in two-stage designs," *Statistics*, vol. 52, no. 5, pp. 1086–1095, 2018.

[19] G. Shan and J. J. Chen, "Optimal inference for Simon's two-stage design with over or under enrollment at the second stage," *Communications in Statistics—Simulation and Computation*, vol. 47, no. 4, pp. 1157–1167, 2018.

[20] G. Shan and S. Gerstenberger, "Fisher's exact approach for post hoc analysis of a chi-squared test," *PLoS One*, vol. 12, no. 12, Article ID e0188709, 2017.

[21] K. Irimata, W. Paul, and X. Li, *Estimation of Correlation Coefficient in Data with Repeated Measures*, SAS Global, Washington, DC, USA, 2018.

[22] G. Shan and W. Wang, "ExactCIdiff: an R package for computing exact confidence intervals for the difference of two proportions," *The R Journal*, vol. 5, no. 2, pp. 62–71, 2013.

[23] X. Wang, W. Zhou, T. Ye, X. Lin, and J. Zhang, "Alzheimer's disease neuroimaging initiative. The relationship between hippocampal volumes and delayed recall is modified by APOE $\epsilon 4$ in mild cognitive impairment," *Frontiers in Aging Neuroscience*, vol. 11, p. 36, 2019.

[24] G. Shan, "Exact confidence limits for the response rate in two-stage designs with over- or under-enrollment in the second stage," *Statistical Methods in Medical Research*, vol. 27, no. 4, pp. 1045–1055, 2018.

[25] G. Shan, S. Banks, J. B. Miller et al., "Statistical advances in clinical trials and clinical research," *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, vol. 4, no. 1, pp. 366–371, 2018.

[26] G. Shan, H. Zhang, and T. Jiang, "Adaptive two-stage optimal designs for Phase II clinical studies that allow early futility stopping," *Sequential Analysis*, vol. 38. , 2019 In press.

[27] G. Shan and A. Hutson, *Tests for Comparing Two Ordered Multinomials*, Wiley StatsRef: Statistics Reference Online, Hoboken, NJ, USA, 2018.

[28] G. Shan, C. Dodge-Francis, and G. E. Wilding, "Exact unconditional tests for dichotomous data when comparing multiple treatments with a single control," *Therapeutic Innovation & Regulatory Science*, vol. 54, pp. 411–417, 2020.

[29] G. Shan and W. Wang, "Exact one-sided confidence limits for Cohen's kappa as a measurement of agreement," *Statistical Methods in Medical Research*, vol. 26, no. 2, pp. 615–632, 2017.

[30] G. Shan, H. Zhang, and T. Jiang, "Minimax and admissible adaptive two-stage designs in phase II clinical trials," *BMC Medical Research Methodology*, vol. 16, no. 1, p. 90, 2016.