

Missing Data Methods for Partial Correlations

Gina M D'Angelo*, Jingqin Luo and Chengjie Xiong

Division of Biostatistics, Washington University School of Medicine, 660 S. Euclid Ave, St. Louis, MO 63110, USA

Abstract

In the dementia area it is often of interest to study relationships among regional brain measures; however, it is often necessary to adjust for covariates. Partial correlations are frequently used to correlate two variables while adjusting for other variables. Complete case analysis is typically the analysis of choice for partial correlations with missing data. However, complete case analysis will lead to biased and inefficient results when the data are missing at random. We have extended the partial correlation coefficient in the presence of missing data using the expectation-maximization (EM) algorithm, and compared it with a multiple imputation method and complete case analysis using simulation studies. The EM approach performed the best of all methods with multiple imputation performing almost as well. These methods were illustrated with regional imaging data from an Alzheimer's disease study.

Keywords: Partial correlation; Fisher-z transformation; Missing data; Missing at random; Expectation-maximization algorithm; Alzheimer's disease

Introduction

Recent advancements have fueled biomarker and imaging marker data collection in medical research studies. Associations between these markers need to be established first prior to understanding temporal relationships and assessments being made about predictions. In the early stages of medical research, data analysis is exploratory and often the direction of relationships between variables is unknown. A first step in medical research is to identify which markers are associated and correlation coefficients can assist in statistical assessments of this endeavor. For example, in the neuropsychological area and Alzheimer's disease area we often are interested in how various brain regions are structurally and functionally related. To provide a hint of the disease pathway, studying neurodegenerative diseases, particularly Alzheimer's disease, among those who are cognitively normal may shed light on the earlier stage of the disease. Obtaining some knowledge of how various regions are structurally related can point us to future research directions.

A common measure to assess whether imaging markers are related is the Pearson correlation coefficient; however it is often necessary to adjust for other variables such as demographic and other marker data to remove potential confounding effects. Partial correlations can be utilized to correlate two variables while adjusting for other variables. Often times, data are partially missing. Missing data approaches have mainly been devoted to regression models with minimal work done in the correlation area. No statistical methods have been developed to handle missing data in a partial correlation analysis. Our objective here is to develop statistical methods for partial correlations with missing data.

Missing data is such a common scenario in medical studies. Standard practice for correlations and partial correlations is to analyze only the data where all observations are fully observed. This is known as complete case analysis. The mean is typically of interest and has been addressed through estimating the regression coefficient from regression [1,2] or estimating the mean from a bivariate distribution [3], typically where the distribution is bivariate normal and the data are continuous. Here, we are interested in the second order statistics not the first order statistic.

Some work has been done with correlation estimation and missing

data [3,4], however no literature has been devoted to estimating the partial correlation with missing data. Minami and Shimizu [5] have proposed a maximum likelihood estimate and restricted maximum likelihood estimate for a correlation coefficient in a bivariate normal distribution. He and Nagaraja [6] proposed using estimation based on the concomitants of order statistics from the bivariate normal distribution. Their problem was specifically for a continuous variable and rank-based variable. In a related area, Truxillo [7] examined maximum likelihood estimation and multiple imputation to estimate the mean and covariance parameters where there is missing data.

In the presence of missing data, complete case analysis can lead to biased and inefficient results [1]. An Alzheimer's disease data set with missing imaging markers motivated us to extend the partial correlation coefficient using maximum likelihood estimation. We compare the expectation-maximization (EM) algorithm to complete case analysis and multiple imputation. We will limit our method to data that are missing at random where the missing data pattern is permitted to be nonmonotonic. Properties from these missing data methods will be compared with simulation studies. These missing data methods will be demonstrated using volumetric, diffusion tensor imaging (DTI), and Pittsburgh Compound-B (PIB) data from the Adult Children's Study conducted at the Washington University Knight Alzheimer's Disease Research Center.

Methods

Notation and methodology

We define our data to be three continuous variables (X, Y, Z) where we are interested in correlating X and Y adjusting for Z . Index $i=1,..,n$ indicates the i th subject where there are n subjects. We assume that $V=(X, Y, Z)$ has a multivariate normal distribution, i.e. $V \sim MVN(\mu, \Sigma)$.

***Corresponding author:** Gina D'Angelo, Division of Biostatistics, Washington University School of Medicine, 660 S. Euclid Ave, St. Louis, MO 63110, USA, Tel: +314-362-3758, Fax: +314-362-2693; E-mail: gina@wubios.wustl.edu

Received October 26, 2012; **Accepted** November 20, 2012; **Published** November 27, 2012

Citation: D'Angelo GM, Luo J, Xiong C (2012) Missing Data Methods for Partial Correlations. J Biom Biostat 3:155. doi:[10.4172/2155-6180.1000155](https://doi.org/10.4172/2155-6180.1000155)

Copyright: © 2012 D'Angelo GM, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

The likelihood is

$$L = \prod_{i=1}^n \frac{1}{(2\pi)^{D/2}} |\Sigma|^{-1/2} \exp \left[-\frac{1}{2} (\mathbf{v}_i - \mathbf{u})^T \Sigma^{-1} (\mathbf{v}_i - \mathbf{u}) \right], \quad (1)$$

where $D=3$ is the dimension of \mathbf{V} . The log-likelihood is

$$l = \sum_{i=1}^n \left[-(D/2) \log(2\pi) - (1/2) \log |\Sigma| - (1/2) (\mathbf{v}_i - \mathbf{u})^T \Sigma^{-1} (\mathbf{v}_i - \mathbf{u}) \right]. \quad (2)$$

The covariance matrix is

$$\Sigma = \begin{pmatrix} \sigma_x^2 & \sigma_x \sigma_y \rho_{xy} & \sigma_x \sigma_z \rho_{xz} \\ \sigma_x \sigma_y \rho_{xy} & \sigma_y^2 & \sigma_y \sigma_z \rho_{yz} \\ \sigma_x \sigma_z \rho_{xz} & \sigma_y \sigma_z \rho_{yz} & \sigma_z^2 \end{pmatrix}. \quad (3)$$

The partial correlations of x and y adjusting for z can be estimated by

$$\rho_{xy.z} = \frac{\rho_{xy} - \rho_{xz} \rho_{yz}}{\sqrt{1 - \rho_{xz}^2} \sqrt{1 - \rho_{yz}^2}}. \quad (4)$$

where the correlation between two variables x and y is

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}. \quad (5)$$

Since our data are multivariate normal the maximum likelihood estimate (MLE) of the mean is $\bar{\mathbf{v}} = \frac{1}{n} \sum_{i=1}^n \mathbf{v}_i$ and the MLE of the covariance matrix, Σ , is the cross product $\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\mathbf{v}_i - \bar{\mathbf{v}})^T (\mathbf{v}_i - \bar{\mathbf{v}})$. For our paper we assume the data are missing at random where the missingness depends on the data that are observed and not on the data that are missing. The missing indicator for x is r_1 , for y is r_2 and for z is r_3 , where $r_d=1$ indicates not missing and $r_d=0$ indicates missing for $d=1,2,3$. The missing data model is $p(r_1, r_2, r_3 | x, y, z) = p(r_1 | y, z, r_2, r_3) p(r_2 | x, z, r_3) p(r_3 | x, y)$.

Pearson correlation and Fisher-z

The main focus of the paper is to estimate the partial correlation. Two strategies are used to estimate the coefficient and its associated variance: 1) Pearson's correlation [8]; and 2) Fisher-z transformation [9-11]. It has been shown that the Pearson correlation coefficient has an approximate t-distribution with $(n-k-2)$ degrees of freedom [8] where $k=1$ is the number of variables partialled out where the standard error is

$$SE = \frac{\sqrt{1 - r^2}}{\sqrt{n - 3}}. \quad (6)$$

The second approach is using the Fisher-z. The Fisher-z approximation [9,10] is a common measure used for estimation and inference of correlations and partial correlations. RA Fisher discovered the correlation coefficient is not normally distributed [9,10]; and he suggested a function of the correlation coefficient that is approximately normally distributed. A property of the Fisher-z transformation is that its variance is a function of the sample size. When using the Fisher-z transformation the correlation is first transformed:

$$\tau = \frac{1}{2} \ln \left(\frac{1 + \rho_{xy.z}}{1 - \rho_{xy.z}} \right). \quad (7)$$

For the Fisher-z transformation, $\tau - \hat{\tau}$ has an approximately normal distribution of $N(0,1/(n-(p-q)-3))$ [8], where $p-q$ is the number of variables conditioned on. In this case we are only conditioning on 1 variable and the standard error is

$$SE = \sqrt{1/(n-1-3)}. \quad (8)$$

EM algorithm

Maximum likelihood estimation [1,12-14] is an approach to estimate parameters from the multivariate normal distribution in the presence of missing data. However, if the missingness pattern is nonmonotonic then the maximum likelihood estimate (MLE) is not tractable when factoring the likelihood. If the likelihood cannot be factored the parameter estimates will not be identifiable. An algorithm that can solve the MLE in general missing data problems is the expectation-maximization (EM) algorithm. Although the EM is a powerful tool to solve parameter estimates it can have convergence issues and can be slow.

The joint probability of \mathbf{V} is $p(\mathbf{v}; \boldsymbol{\varphi}) = p(\mathbf{v}_{obs} | \boldsymbol{\varphi}) p(\mathbf{v}_{mis} | \mathbf{v}_{obs}, \boldsymbol{\varphi})$ and the log-likelihood for complete data is of the form [1]

$$l(\boldsymbol{\varphi} | \mathbf{v}) = l(\boldsymbol{\varphi} | \mathbf{v}_{obs}) + \ln p(\mathbf{v}_{mis} | \mathbf{v}_{obs}, \boldsymbol{\varphi}) \quad (9)$$

where $\boldsymbol{\varphi} = (\boldsymbol{\mu}, \Sigma)$, $\mathbf{v} = (\mathbf{v}_{obs}, \mathbf{v}_{mis})$ and \mathbf{v}_{obs} and \mathbf{v}_{mis} denotes the observed and missing components of \mathbf{v} . For the EM algorithm, the E-step at the k th iteration is [12,13]

$$\mathbf{Q}(\boldsymbol{\varphi} | \boldsymbol{\varphi}^{(k)}) = E(l(\boldsymbol{\varphi}; \mathbf{v}) | \mathbf{v}_{obs}, \boldsymbol{\varphi}^{(k)}) \quad (10)$$

When missing data occurs at the i th observation, we take a sample $\mathbf{z}_{i1}, \dots, \mathbf{z}_{im_i}$ of size m_i for the i th observation from $P(\mathbf{v}_{mis,i} | \mathbf{v}_{obs,i}, \boldsymbol{\varphi}^{(k)})$ using the Gibbs sampler [13]. For continuous data, the E-step is [12,13].

$$\mathbf{Q}(\boldsymbol{\varphi} | \boldsymbol{\varphi}^{(k)}) = \sum_{i=1}^n \frac{1}{m_i} \sum_{j=1}^{m_i} l(\boldsymbol{\varphi} | \mathbf{z}_{ij}, \mathbf{v}_{obs,i}). \quad (11)$$

The M-step of the EM uses standard weighted methods, specifically the weighted mean and weighted covariance here, to estimate the MLE of the parameters at the $(k+1)$ iteration, $\boldsymbol{\varphi}^{(k+1)}$. The information matrix is of the form [12,13].

$$\mathbf{I}(\hat{\boldsymbol{\varphi}}) = -\ddot{\mathbf{Q}}(\hat{\boldsymbol{\varphi}}) - \sum_{i=1}^n \frac{1}{m_i} \sum_{j=1}^{m_i} \mathbf{S}_i(\hat{\boldsymbol{\varphi}} | \mathbf{z}_{ij}, \mathbf{v}_{obs,i}) \mathbf{S}_i(\hat{\boldsymbol{\varphi}} | \mathbf{z}_{ij}, \mathbf{v}_{obs,i})^T + \sum_{i=1}^n \dot{\mathbf{Q}}_i(\hat{\boldsymbol{\varphi}}) \dot{\mathbf{Q}}_i(\hat{\boldsymbol{\varphi}})^T \quad (12)$$

where $\hat{\boldsymbol{\varphi}}$ are estimates at convergence, $\mathbf{S}_i(\boldsymbol{\varphi} | \mathbf{z}_{ij}, \mathbf{v}_{obs,i}) = \frac{\partial l(\boldsymbol{\varphi} | \mathbf{z}_{ij}, \mathbf{v}_{obs,i})}{\partial \boldsymbol{\varphi}}$, $\dot{\mathbf{Q}}_i(\boldsymbol{\varphi} | \boldsymbol{\varphi}^{(k)}) = \frac{1}{m_i} \sum_{j=1}^{m_i} \frac{\partial l(\boldsymbol{\varphi} | \mathbf{z}_{ij}, \mathbf{v}_{obs,i})}{\partial \boldsymbol{\varphi}}$ and $\ddot{\mathbf{Q}}(\boldsymbol{\varphi} | \boldsymbol{\varphi}^{(k)}) = \sum_{i=1}^n \frac{1}{m_i} \sum_{j=1}^{m_i} \frac{\partial^2 l(\boldsymbol{\varphi} | \mathbf{z}_{ij}, \mathbf{v}_{obs,i})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^T}$.

The EM algorithm is based on the MLE and the distribution of $(\boldsymbol{\varphi} - \hat{\boldsymbol{\varphi}})$ is asymptotically normal where the mean is 0 and variance is \mathbf{I}^{-1} [1]. A consistent estimator of the variance is the sum of scores squared estimate, $\sum_i \dot{\mathbf{Q}}_i(\hat{\boldsymbol{\varphi}}) \dot{\mathbf{Q}}_i(\hat{\boldsymbol{\varphi}})^T$. This strategy will let us avoid further complications in deriving the second derivatives to estimate the variance.

To estimate the variance for the Pearson partial correlation we use the following details of the sum of scores squared approach and delta method. The delta method is used to estimate the variance of the Pearson's partial correlation:

$$\begin{bmatrix} \frac{\partial \rho_{xy.z}}{\partial \rho_{xy}} & \frac{\partial \rho_{xy.z}}{\partial \rho_{xz}} & \frac{\partial \rho_{xy.z}}{\partial \rho_{yz}} \end{bmatrix}^T \mathbf{I}^{-1} \begin{bmatrix} \frac{\partial \rho_{xy.z}}{\partial \rho_{xy}} & \frac{\partial \rho_{xy.z}}{\partial \rho_{xz}} & \frac{\partial \rho_{xy.z}}{\partial \rho_{yz}} \end{bmatrix}, \quad (13)$$

where the partial derivatives are

$$\frac{\partial \rho_{xy.z}}{\partial \rho_{xy}} = \frac{1}{\sqrt{1 - \rho_{xz}^2} \sqrt{1 - \rho_{yz}^2}}, \quad (14)$$

$$\frac{\partial \rho_{xy,z}}{\partial \rho_{xz}} = \frac{-\rho_{xz}\sqrt{1-\rho_{xz}^2}\sqrt{1-\rho_{yz}^2} + (\rho_{xy} - \rho_{xz}\rho_{yz})\rho_{xz}(1-\rho_{xz}^2)^{-1/2}\sqrt{1-\rho_{yz}^2}}{\left(\sqrt{1-\rho_{xz}^2}\sqrt{1-\rho_{yz}^2}\right)^2}, \quad (15)$$

and

$$\frac{\partial \rho_{xy,z}}{\partial \rho_{yz}} = \frac{-\rho_{xz}\sqrt{1-\rho_{xz}^2}\sqrt{1-\rho_{yz}^2} + (\rho_{xy} - \rho_{xz}\rho_{yz})\rho_{yz}(1-\rho_{yz}^2)^{-1/2}\sqrt{1-\rho_{xz}^2}}{\left(\sqrt{1-\rho_{xz}^2}\sqrt{1-\rho_{yz}^2}\right)^2}. \quad (16)$$

The information matrix is calculated with the second derivatives of the log-likelihood. However, here I is estimated by using the sum of scores squared. The sum of scores squared requires only the first derivatives of the log-likelihood. The first derivatives are

$$\frac{\partial l}{\partial \mathbf{u}} = \sum_{i=1}^n \frac{1}{m_i} \sum_{j=1}^{m_i} \Sigma^{-1} \left((\mathbf{z}_{ij}, \mathbf{v}_{obs,i}) - \mathbf{u} \right) \quad (17)$$

and

$$\frac{\partial l}{\partial \boldsymbol{\theta}} = \sum_{i=1}^n \frac{1}{m_i} \sum_{j=1}^{m_i} \left[-\left(\frac{1}{2} \right) \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \boldsymbol{\theta}} \right) - \left(\frac{1}{2} \right) \left((\mathbf{z}_{ij}, \mathbf{v}_{obs,i}) - \mathbf{u} \right)^T \left(-\Sigma^{-1} \frac{\partial \Sigma}{\partial \boldsymbol{\theta}} \Sigma^{-1} \right) \left((\mathbf{z}_{ij}, \mathbf{v}_{obs,i}) - \mathbf{u} \right) \right] \quad (18)$$

where $\boldsymbol{\theta} = (\sigma_x, \sigma_y, \sigma_z, \rho_{xy}, \rho_{xz}, \rho_{yz})$, $\frac{\partial}{\partial \boldsymbol{\theta}} \log |\Sigma| = \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \boldsymbol{\theta}} \right)$, and $\frac{\partial}{\partial \boldsymbol{\theta}} \Sigma^{-1} = -\Sigma^{-1} \frac{\partial \Sigma}{\partial \boldsymbol{\theta}} \Sigma^{-1}$ [8]. The scores are $\dot{\mathbf{Q}}_i(\boldsymbol{\phi} | \boldsymbol{\phi}^{(k)}) = \frac{1}{m_i} \sum_{j=1}^{m_i} \frac{\partial l(\boldsymbol{\phi} | \mathbf{z}_{ij}, \mathbf{v}_{obs,i})}{\partial \boldsymbol{\phi}}$. The sum of scores squared, $\sum_{i=1}^n \dot{\mathbf{Q}}_i(\boldsymbol{\phi}) \dot{\mathbf{Q}}_i(\boldsymbol{\phi})^T$, is a consistent estimate of the observed information [15,16].

A similar approach is employed for the Fisher-z transformation of the Pearson partial correlation. We use the delta method to estimate the variance of the Fisher-z transformation of the Pearson partial correlation:

$$\begin{bmatrix} \frac{\partial \tau}{\partial \rho_{xy}} & \frac{\partial \tau}{\partial \rho_{xz}} & \frac{\partial \tau}{\partial \rho_{yz}} \end{bmatrix}^T I^{-1} \begin{bmatrix} \frac{\partial \tau}{\partial \rho_{xy}} & \frac{\partial \tau}{\partial \rho_{xz}} & \frac{\partial \tau}{\partial \rho_{yz}} \end{bmatrix}, \quad (19)$$

where the partial derivatives are

$$\frac{\partial \tau}{\partial \rho_{xy}} = \frac{1}{2} \frac{\partial \rho_{xy,z}}{\partial \rho_{xy}} \left[\frac{1}{(1+\rho_{xy,z})} + \frac{1}{(1-\rho_{xy,z})} \right], \quad (20)$$

$$\frac{\partial \tau}{\partial \rho_{xz}} = \frac{1}{2} \frac{\partial \rho_{xy,z}}{\partial \rho_{xz}} \left[\frac{1}{(1+\rho_{xy,z})} + \frac{1}{(1-\rho_{xy,z})} \right], \quad (21)$$

and

$$\frac{\partial \tau}{\partial \rho_{yz}} = \frac{1}{2} \frac{\partial \rho_{xy,z}}{\partial \rho_{yz}} \left[\frac{1}{(1+\rho_{xy,z})} + \frac{1}{(1-\rho_{xy,z})} \right]. \quad (22)$$

As previously discussed the sum of scores squared, $\sum_{i=1}^n \dot{\mathbf{Q}}_i(\boldsymbol{\phi}) \dot{\mathbf{Q}}_i(\boldsymbol{\phi})^T$, is a consistent estimate of the observed information and is used.

Multiple Imputation

Multiple imputation (MI) [17-21] is a popular missing data technique because it is included in many statistical packages. Essentially, multiple imputation is an approach that replaces the missing data with multiple simulated values. As the concept has been evolving since the 1970s, a number of researchers have shown the usefulness of multiple imputation and proven that statistical properties are improved in many settings [17-20]. The approach we consider is an imputation method developed by King et al. [17] that imputes the missing data with a bootstrapped expectation maximization sampling algorithm rather

than using the more traditional Markov-chain Monte Carlo-based imputation-posterior (IP) approach. The Amelia II library [17] was selected for the imputation schemes since it is faster than other existing software based on IP and leads to similar results. A disadvantage of imputation occurs when the imputation model is misspecified, resulting in estimates that will be biased and inefficient [22,23].

The basic idea of imputation is the missing variables are modeled jointly conditional on the fully observed data to provide a joint conditional probability for the posterior distribution. When a subject has only one variable missing only that missing value is filled in. When all 3 variables are missing none of the values for that subject is imputed.

The principal idea is to create M data sets of repeated imputations, $m=1,..,M$. We will refer to both the Pearson partial correlation and the Fisher-z transformation as coefficients. Based on these M imputed data sets, the coefficient estimates and their variances will be estimated for each data set: $(\hat{H}_1, \dots, \hat{H}_M)$ and (U_1, \dots, U_M) . The Pearson partial correlation calculated with equation (4) and its variance calculated with equation (6) are estimated for each completed data set. The Fisher-z transformation calculated with equation (7) and its variance calculated with equation (8) are estimated for each completed data set. Upon obtaining these M coefficient estimates and their variances, the equations listed directly below are used for the multiple imputation estimates of the coefficient and its variance. The average of the M coefficient estimates is [1,18]

$$\bar{H}_M = \frac{1}{M} \sum_{m=1}^M \hat{H}_m \quad (23)$$

The average of the M variances is [1,18]

$$\bar{U}_M = \frac{1}{M} \sum_{m=1}^M U_m \quad (24)$$

The between-variance is [1,18]

$$\mathcal{H}_M = \frac{1}{M} \sum_{m=1}^M (\hat{H}_m - \bar{H}_M)^T (\hat{H}_m - \bar{H}_M) / (M-1). \quad (25)$$

The total variance of the coefficients is [1,18]

$$T_M = \bar{U}_M + (1+M^{-1}) \mathcal{H}_M. \quad (26)$$

Inference for multiple imputation is based on a t-test with v degrees of freedom where the t-statistic $\frac{\bar{H}_M}{\sqrt{T_M}}$ is and $v = (M-1) \left(1 + \frac{1}{M+1} \frac{\bar{U}_M}{\mathcal{H}_M} \right)^2$ [1,18].

Simulation Study

We performed simulation studies to determine the finite sample properties of the various methods. The methods compared here are analysis on the full data (Full), complete case analysis (CC), the expectation-maximization algorithm (EM) and multiple imputation (MI). The full data is the generated data before deletion of missing values. First we estimated the Pearson partial correlation and its variance. Then we estimated the Fisher-z transformation and its variance. For simulation studies, we compared the mean of bias (bias), mean of the standard error (SE), square root of the mean squared error (MSE), relative efficiency (RE), and 95% coverage probabilities (95% Cov). Since all methods are not unbiased we used the MSE to calculate the relative efficiency where the MSE of each missing data method is compared to the MSE without missing data (Full). Each simulation study has 1000 replications.

We generated 3 variables to have a multivariate normal distribution, where $V \sim MVN(0, \Sigma)$, all the variances are 1 and the correlations, partial correlations, and Fisher-z transformation, $(\rho_{xy}, \rho_{xz}, \rho_{yz}, \rho_{xy.z}, \rho_{xz.z}, \rho_{yz.z}, \tau_{xy.z})$ considered are: (.105, .03, .21, .1, .1), (.35, .1, .15, .34, .36), (.64, .5, .4, .55, .63), (.9, .83, .7, .8, 1.1). We considered samples sizes of 50 (results in Supplementary Material Section), 200 (results in Supplementary Material Section), 500 and 1000 (results in Supplementary Material Section) and percentage of missingness of 20%, 35%, and 50%. The type of missing data mechanism is missing at random, where the missingness depends on the observed portion of the data and not on the unobserved portion of the data.

The missing data model we used to generate the missing data was $p(r_1, r_2, r_3 | x, y, z) = p(r_1 | y, z, r_2, r_3)p(r_2 | x, z, r_3)p(r_3 | x, y)$ where r_1 is the missing indicator for x , r_2 is the missing indicator for y , and r_3 is the missing indicator for z . The missing at random mechanism is specified by the models: $\text{logit}(p(r_1 | y, z, r_2, r_3)) = \beta_{0,1} + \beta_{1,1}y + \beta_{2,1}z + \beta_{3,1}r_2 + \beta_{4,1}r_3$, $\text{logit}(p(r_2 | x, z, r_3)) = \beta_{0,2} + \beta_{1,2}x + \beta_{2,2}z + \beta_{3,2}r_3$, and $\text{logit}(p(r_3 | x, y)) = \beta_{0,3} + \beta_{1,3}x + \beta_{2,3}y$. In these missing data models, all the regression coefficients except for the intercepts were fixed to be 1. Refer to Table 1 for the intercept values of the missing data models for each partial correlation.

Results for $n=500$ are reported in Tables 2-5, where Table 2 contains results for $\rho_{xy.z}=.1$, Table 3 contains results for $\rho_{xy.z}=.34$, Table 4 contains results for $\rho_{xy.z}=.55$, and Table 5 contains results for $\rho_{xy.z}=.8$. Regardless of the amount of missing data and coefficient type, MI and EM both had no bias or very small bias, whereas the CC approach was biased. As the proportion of missingness increased so did the bias. As the correlation increased the Fisher-z had slightly more bias than the Pearson correlation approach. When the correlations were moderate

$(\rho_{xy}, \rho_{xz}, \rho_{yz}, \rho_{xy.z}, \rho_{xz.z}, \rho_{yz.z}, \tau_{xy.z})$	Intercept values
(.105, .03, .21, .1, .1)	
20% missing	$(\beta_{0,1} = 3.1, \beta_{0,2} = 2.4, \beta_{0,3} = 1.4)$
35% missing	$(\beta_{0,1} = 2.0, \beta_{0,2} = 1.6, \beta_{0,3} = 0.6)$
50% missing	$(\beta_{0,1} = 1.5, \beta_{0,2} = 0.5, \beta_{0,3} = 0)$
(.35, .1, .15, .34, .36)	
20% missing	$(\beta_{0,1} = 3.2, \beta_{0,2} = 2.4, \beta_{0,3} = 1.3)$
35% missing	$(\beta_{0,1} = 2.0, \beta_{0,2} = 1.7, \beta_{0,3} = 0.6)$
50% missing	$(\beta_{0,1} = 1.4, \beta_{0,2} = 0.6, \beta_{0,3} = 0)$
(.64, .5, .4, .55, .63),	
20% missing	$(\beta_{0,1} = 3.3, \beta_{0,2} = 2.5, \beta_{0,3} = 1.5)$
35% missing	$(\beta_{0,1} = 2.0, \beta_{0,2} = 1.7, \beta_{0,3} = 0.6)$
50% missing	$(\beta_{0,1} = 1.3, \beta_{0,2} = 0.5, \beta_{0,3} = 0)$
(.9, .83, .7, .8, 1.1)	
20% missing	$(\beta_{0,1} = 3.4, \beta_{0,2} = 2.4, \beta_{0,3} = 1.6)$
35% missing	$(\beta_{0,1} = 2.0, \beta_{0,2} = 1.7, \beta_{0,3} = 0.6)$
50% missing	$(\beta_{0,1} = 1.2, \beta_{0,2} = 0.4, \beta_{0,3} = 0)$

Table 1: Intercept values for the missing data models from simulation studies

		Pearson correlation				Fisher-z			
		Full	CC	EM	MI	Full	CC	EM	MI
20% missing									
E(Bias)		-0.001	-0.044	0.002	0.003	-0.001	-0.044	0.003	0.003
E(SE)		0.045	0.05	0.051	0.051	0.045	0.050	0.052	0.051
\sqrt{MSE}		0.044	0.066	0.049	0.05	0.045	0.066	0.05	0.051
RE		1	2.2	1.25	1.29	1	2.15	1.25	1.30
95% Cov		0.95	0.869	0.953	0.963	0.946	0.869	0.956	0.963
35% missing									
E(Bias)		0.002	-0.061	0.011	0.012	0.003	-0.061	0.012	0.012
E(SE)		0.045	0.056	0.056	0.056	0.045	0.056	0.057	0.057
\sqrt{MSE}		0.044	0.081	0.055	0.056	0.045	0.081	0.056	0.057
RE		1	3.4	1.56	1.64	1	3.32	1.59	1.66
95% Cov		0.953	0.826	0.948	0.966	0.951	0.826	0.952	0.961
50% missing									
E(Bias)		0	-0.081	0.021	0.021	0	-0.081	0.022	0.022
E(SE)		0.045	0.063	0.063	0.064	0.045	0.063	0.065	0.064
\sqrt{MSE}		0.044	0.104	0.067	0.069	0.045	0.104	0.068	0.071
RE		1	5.51	2.31	2.45	1	5.4	2.35	2.5
95% Cov		0.952	0.748	0.926	0.947	0.952	0.746	0.931	0.944

Note: full data (Full), complete case analysis (CC), the expectation-maximization algorithm (EM), multiple imputation (MI), standard error (SE), mean squared error (MSE), relative efficiency (RE), and 95% coverage probabilities (95% Cov)

Table 2: Summary statistics for coefficients from simulation study with partial correlation of .1 and sample size of 500.

($\rho_{xy.z}=.34, .55$; Tables 3 and 4), the standard errors were the largest for CC and often slightly smaller for the EM in comparison to MI. When the correlation was small ($\rho_{xy.z}=.1$; Table 2) the standard errors were similar across all methods and slightly larger for the EM and MI approach. However, when the correlation was large ($\rho_{xy.z}=.8$; Table 5) the standard errors were the same across methods, whereas for EM and MI they became smaller than CC as the percentage of missingness increased. However for the Fisher-z approach when the correlation was large ($\rho_{xy.z}=.8$; Table 5), the standard errors tended to be the largest for the EM and smallest for MI. The standard errors tended to increase as the proportion of missingness increased.

In all scenarios the EM and MI had similar MSE; however, the EM had a smaller MSE that was closer to the actual MSE. CC always had the largest MSE. The EM was the most efficient approach, followed by MI then CC. As the proportion of missingness increased all methods yielded less efficient estimates. The coverages for small and moderate correlations ($\rho_{xy.z} < .55$; Tables 2 and 3) yielded coverages that were close to the true nominal value for EM and MI and were too narrow for CC. This implies CC will always be too liberal. As the correlation increased ($\rho_{xy.z} \geq .55$ Tables 4 and 5) the correlation approach produced coverages that were too conservative, except when $\rho_{xy.z}=.55$ (Table 4) EM had coverages close to the true coverage. When $\rho_{xy.z}=.55$ (Table 4) the Fisher-z produced close to true coverages for EM and MI and was too narrow for CC. However, when the correlation was large ($\rho_{xy.z}=.8$, Table 5) the Fisher-z produced coverages that were slightly too narrow for MI and too wide for EM. The findings for the correlation coverages as the correlation increases are not surprising considering the sampling distribution becomes skewed as the correlation is further away from 0

		$(\rho_{xy}, \rho_{xz}, \rho_{yz}) = (.35, .1, .15)$, $\rho_{xy.z} = .34$, $\tau_{xy.z} = .36$, $n = 500$							
		Pearson correlation		Fisher-z					
		Full	CC	EM	MI	Full	CC	EM	MI
20% missing									
E(Bias)	0.001	-0.058	0.001	0.001	0.002	-0.064	0.001	0.002	
E(SE)	0.042	0.048	0.043	0.047	0.045	0.050	0.048	0.049	
\sqrt{MSE}	0.04	0.074	0.044	0.044	0.045	0.081	0.05	0.05	
RE	1	3.49	1.24	1.26	1	3.26	1.24	1.26	
95% Cov	0.96	0.809	0.933	0.963	0.952	0.768	0.938	0.954	
35% missing									
E(Bias)	0.002	-0.08	0.004	0.005	0.002	-0.087	0.005	0.006	
E(SE)	0.042	0.054	0.047	0.051	0.045	0.056	0.053	0.053	
\sqrt{MSE}	0.041	0.095	0.049	0.05	0.046	0.103	0.056	0.057	
RE	1	5.45	1.45	1.51	1	5.02	1.46	1.52	
95% Cov	0.954	0.716	0.928	0.958	0.942	0.679	0.931	0.947	
50% missing									
E(Bias)	0	-0.097	0.011	0.013	0.001	-0.105	0.014	0.015	
E(SE)	0.042	0.061	0.052	0.056	0.045	0.063	0.06	0.058	
\sqrt{MSE}	0.041	0.113	0.055	0.056	0.046	0.123	0.063	0.064	
RE	1	7.74	1.83	1.89	1	7.07	1.89	1.95	
95% Cov	0.96	0.651	0.927	0.96	0.949	0.621	0.935	0.941	

Note: full data (Full), complete case analysis (CC), the expectation-maximization algorithm (EM), multiple imputation (MI), standard error (SE), mean squared error (MSE), relative efficiency (RE), and 95% coverage probabilities (95% Cov)

Table 3: Summary statistics for coefficients from simulation study with partial correlation of .34 and sample size of 500.

[24]. However, with the Fisher-z transformation this skewness is greatly reduced. The very conservative coverages are reflecting this finding. It has been recommended by others [9-11,24] to use the Fisher-z with larger correlation values as we have shown.

We also considered various sample sizes (in supplement). Across all sample sizes, the CC approach was biased and both the EM and MI approaches had no bias or minimal bias. As the sample size increased, the standard error and MSE decreased and the CC estimates were less efficient and resulted in narrower coverages across all correlation values. We also examined the sum of scores squared approach to estimate the variance for all other methods (Full, CC, and MI) with the Pearson correlation. Based on this variance estimate, we found that the coverages were close to the true coverage with an occasional slightly narrower coverage for small to moderate correlations ($\rho_{xy.z} = .1, .34, .55$) and too wide for large correlations ($\rho_{xy.z} = .8$) (results not shown).

The CC method had the worst properties and is not recommended. EM and MI performed very similarly although the EM had a slightly smaller MSE and was slightly more efficient. When the correlation was large the EM produced more conservative coverages and MI was more liberal. We recommend the EM and if programming is a barrier then recommend MI. When the correlation is greater than .5 we recommend using the Fisher-z transformation since the coverages are closer to the nominal value.

Example

The Adult Children's Study is conducted at the Washington University Knight Alzheimer's Disease Research Center. The sample consists of cognitively normal subjects who all have a Clinical Dementia

Rating of 0 at baseline [25]. There was a decent amount of missing data among the imaging variables. Various imaging modalities are used to evaluate neurodegenerative diseases and are an important element of research to study Alzheimer's disease.

Magnetic resonance imaging has been a traditional marker to measure volumes, where whole brain and hippocampus are commonly utilized to track structural changes of the brain. Diffusion tensor imaging (DTI) is a newer structural imaging measure in Alzheimer's disease. DTI is a magnetic resonance imaging technique that measures water movement in the brain and can provide information about the structure of the white matter of the brain. DTI is represented in multiple measurements. Here we demonstrate radial diffusivity (RD) and fractional anisotropy (FA) from the corpus callosum genu region. Amyloid deposition is measured by Pittsburgh Compound B (PIB) positron emission tomography and the value is represented as the mean cortical binding potential. Although we are interested in studying a cognitively normal group their cerebrospinal fluid amyloid beta peptide 42 (CSF AB42) can vary and can be a confounder. CSF AB42 is a cerebral spinal fluid biomarker that is used in Alzheimer's disease research to distinguish those who have early-stage Alzheimer's disease. Since PIB and DTI are newer modalities we are interested in determining how they are related to whole brain volume. Therefore, we will correlate whole brain volume with two DTI measures and PIB while adjusting for CSF AB42 to remove the effects of CSF AB42.

The Adult Children's Study consists of 186 participants with a baseline measurement. Table 6 includes demographics of the Adult Children's Study participants. Of these 186 subjects, 36 (19%) are missing CSF AB42, 32 (17%) are missing whole brain volume, 34 (18%)

		$(\rho_{xy}, \rho_{xz}, \rho_{yz}) = (.64, .5, .4)$, $\rho_{xy.z} = .55$, $\tau_{xy.z} = .63$, $n = 500$							
		Pearson correlation		Fisher-z					
		Full	CC	EM	MI	Full	CC	EM	MI
20% missing									
E(Bias)	-0.002	-0.041	-0.005	-0.004	-0.002	-0.057	-0.005	-0.005	
E(SE)	0.037	0.043	0.035	0.041	0.045	0.050	0.051	0.048	
\sqrt{MSE}	0.033	0.056	0.036	0.036	0.047	0.077	0.051	0.052	
RE	1	2.91	1.19	1.21	1	2.63	1.18	1.20	
95% Cov	0.976	0.89	0.939	0.983	0.933	0.789	0.942	0.935	
35% missing									
E(Bias)	-0.001	-0.062	-0.009	-0.008	-0.001	-0.084	-0.012	-0.011	
E(SE)	0.037	0.048	0.039	0.044	0.045	0.056	0.055	0.051	
\sqrt{MSE}	0.032	0.075	0.04	0.04	0.046	0.101	0.057	0.057	
RE	1	5.49	1.55	1.57	1	4.79	1.51	1.5	
95% Cov	0.976	0.802	0.941	0.976	0.947	0.66	0.94	0.941	
50% missing									
E(Bias)	-0.002	-0.077	-0.013	-0.012	-0.002	-0.104	-0.017	-0.016	
E(SE)	0.037	0.056	0.043	0.049	0.045	0.064	0.062	0.055	
\sqrt{MSE}	0.032	0.092	0.046	0.047	0.045	0.122	0.064	0.066	
RE	1	8.41	2.1	2.17	1	7.2	2.0	2.07	
95% Cov	0.979	0.756	0.939	0.978	0.954	0.637	0.934	0.931	

Note: full data (Full), complete case analysis (CC), the expectation-maximization algorithm (EM), multiple imputation (MI), standard error (SE), mean squared error (MSE), relative efficiency (RE), and 95% coverage probabilities (95% Cov)

Table 4: Summary statistics for coefficients from simulation study with partial correlation of .55 and sample size of 500.

		$(\rho_{xy}, \rho_{xz}, \rho_{yz}) = (.9, .83, .7), \rho_{xy,z} = .8, \tau_{xy,z} = 1.1, n = 500$							
		Pearson correlation		Fisher-z					
		Full	CC	EM	MI	Full	CC	EM	MI
20% missing									
E(Bias)	0	-0.013	-0.002	-0.002	0.002	-0.032	-0.004	-0.004	
E(SE)	0.027	0.031	0.032	0.029	0.045	0.050	0.089	0.046	
\sqrt{MSE}	0.017	0.023	0.019	0.019	0.047	0.061	0.052	0.052	
RE	1	1.94	1.27	1.28	1	1.70	1.23	1.25	
95% Cov	0.997	0.991	0.998	0.998	0.941	0.89	0.999	0.915	
35% missing									
E(Bias)	0	-0.019	-0.004	-0.004	0.002	-0.048	-0.01	-0.008	
E(SE)	0.027	0.035	0.034	0.030	0.045	0.056	0.092	0.048	
\sqrt{MSE}	0.016	0.028	0.02	0.02	0.045	0.072	0.055	0.055	
RE	1	3.0	1.54	1.57	1	2.55	1.46	1.50	
95% Cov	0.999	0.995	0.999	0.998	0.948	0.858	1	0.911	
50% missing									
E(Bias)	0	-0.026	-0.01	-0.009	0.002	-0.065	-0.024	-0.021	
E(SE)	0.027	0.040	0.036	0.033	0.045	0.064	0.097	0.051	
\sqrt{MSE}	0.016	0.037	0.025	0.025	0.046	0.092	0.067	0.067	
RE	1	4.932	2.36	2.34	1	3.97	2.09	2.09	
95% Cov	0.999	0.984	0.997	0.995	0.955	0.821	0.997	0.886	

Note: full data (Full), complete case analysis (CC), the expectation-maximization algorithm (EM), multiple imputation (MI), standard error (SE), mean squared error (MSE), relative efficiency (RE), and 95% coverage probabilities (95% Cov)

Table 5: Summary statistics for coefficients from simulation study with partial correlation of .8 and sample size of 500.

	N	Mean (SD)
Age	186	62.0 (9.5)
CSF AB42	150	636.2 (215.9)
Education	171	16.0 (2.6)
Mini-mental state examination (MMSE)	171	29.3 (1)
Whole brain volume	154	0.80 (0.02)
Pittsburgh Compound B (PIB)	152	0.06 (0.16)
FA corpus callosum genu region	165	0.84 (0.09)
RD corpus callosum genu region	165	0.24 (0.15)

Note: cerebrospinal fluid amyloid beta peptide 42 (AB42), fractional anisotropy (FA), and radial diffusivity (RD)

Table 6: Demographics of Adult Children's Study.

are missing PIB, and 21 (11%) are missing FA corpus callosum genu region and RD corpus callosum genu region. For the corpus callosum genu regional analyses, 63 (34%) are missing at least 1 biomarker and 1 (1%) are missing all 3 biomarkers. With the PIB analysis, 54 (29%) are missing at least 1 biomarker and 18 (10%) are missing all 3 biomarkers. Our methods require data to be normally distributed. Whole brain volume, the RD corpus callosum genu region, and PIB were not normally distributed so we transformed each to be approximately normal. It was necessary to take: the cubic transformation of whole brain volume, the square root transformation of the RD corpus callosum genu region, and the log transformation of PIB.

Results for the partial correlations adjusted for CSF AB42 are reported in Table 7 of whole brain volume and the RD corpus callosum genu region, whole brain volume and the FA corpus callosum genu region, and whole brain volume and PIB. Overall, the coefficient values

were similar for the expectation-maximization algorithm (EM) and multiple imputation (MI) and differed from complete case analysis (CC). The magnitude of EM and MI was larger than CC for all analyses. This indicates that the correlation between whole brain volume and the corpus callosum genu region and between whole brain volume and PIB would appear larger if we used EM and MI. The EM standard error was the smallest except for whole brain volume and the FA corpus callosum genu region where it was the largest. These differences in magnitude and standard errors affect inference and it differed across methods. Correlation between whole brain volume and the RD corpus callosum genu region was statistically significant using the EM and MI approach and borderline with the CC approach. This is due to an increase in the correlation and decrease in standard error for both EM and MI. Also, the correlation between whole brain volume and the FA corpus callosum genu region was statistically significant using the MI approach, borderline with the EM approach, and not statistically significant with the CC approach. Once again, this is due to an increase in the correlation in EM and MI and decrease in standard error for MI. Inference did not differ across methods for the correlation between whole brain volume and PIB where the correlation was not statistically significant. This is due to such a small correlation between these imaging modalities.

In general, we found the correlation values using CC to probably be misleading. The inference differed across methods. Also, the standard errors tended to be smaller with the EM and MI approaches than with the CC approach. Based on these findings we suggest using the EM since the results were similar to what was found with the simulation studies. If programming is a barrier we recommend using MI.

Discussion

In preliminary studies it is necessary to establish correlations between variables of interest. Partial correlations are often used when there is a need to adjust for other variates. Quite frequently, variables are partially missing, and complete case methods can provide misleading results. We have demonstrated the need for methods to handle missing data when calculating partial correlations.

We extended the expectation-maximization (EM) algorithm for the partial correlation and compared it to multiple imputation and complete case analysis when all variables are missing at random. Both

	Pearson correlation			Fisher-z		
	CC	EM	MI	CC	EM	MI
All adjusted for AB42						
WBV and RD Genu (n)	123	186	185	123	186	185
Coef	-0.162	-0.239	-0.261	-0.164	-0.243	-0.279
SE	0.090	0.076	0.083	0.092	0.080	0.085
p-value	0.074	0.002	0.007	0.074	0.002	0.006
WBV and FA Genu (n)	123	186	185	123	186	185
Coef	0.131	0.170	0.202	0.132	0.172	0.184
SE	0.090	0.098	0.088	0.092	0.101	0.084
p-value	0.15	0.085	0.039	0.15	0.09	0.046
WBV and PIB (n)	132	186	168	132	186	168
Coef	0.022	0.028	0.040	0.022	0.028	0.047
SE	0.088	0.083	0.083	0.088	0.074	0.083
p-value	0.80	0.74	0.63	0.80	0.74	0.57

Note: full data (Full), complete case analysis (CC), the expectation-maximization algorithm (EM), multiple imputation (MI), cerebrospinal fluid amyloid beta peptide 42 (AB42), whole brain volume (WBV), corpus callosum genu region (Genu),radial diffusivity (RD), fractional anisotropy (FA), and Pittsburgh Compound B (PIB)

Table 7: Correlations of whole brain volume and other imaging data adjusted for CSF AB42.

the Pearson correlation coefficient and Fisher-z transformation were considered for all approaches. We have demonstrated that complete case analysis has poor performance and should not be used. We showed that of all methods the EM had the best statistical properties. Multiple imputation performed almost as well as EM. Multiple imputation is recommended when there is a limitation with statistical programming. There can be a computational cost with the EM which could be a consideration when selecting a missing data method. For example, multiple imputation took about 7 seconds and the EM took around 36 seconds for our example data.

In this manuscript we considered the partial Pearson correlation coefficient adjusting for a single covariate, as this is a common request in the clinical world. A limitation of this manuscript is that we targeted normally distributed data for our methods and made an assumption that our data comes from a trivariate normal distribution. Our methodology depends on the data coming from a trivariate normal distribution since the Pearson correlation is directly derived from it. However, for data that are non-normally distributed a transformation such as the Box-Cox or the ladder of powers [26] can be used for a normality approximation. In addition, our method can be extended to the Spearman correlation when the data are not normal. We demonstrated non-normally distributed data in the real application. Future work will involve addressing multiple covariates and categorical data which is quite intensive and will require changing our methodology and assumptions made.

Another limitation of this manuscript is that the second derivative was not used to calculate the information matrix for the EM. This may improve variance estimation of the EM. A suggestion for our future work is to use the bootstrap to estimate the variances for the EM. The nonparametric bootstrap method does not depend on distributional assumptions and provides an empirical estimate of the distribution and its variance. The disadvantage of the bootstrap is the computational time. At this time, we are investigating parallel processing to speed up the computational time for the bootstrap.

Based on our findings we recommend using the EM to estimate partial correlations, and can use multiple imputation as an alternative in the event programming is a consideration. Also, we recommend using the Fisher-z transformation when the correlation is larger than .5. The authors intend on developing a R package for the code; meanwhile, code can be requested from the corresponding author.

Acknowledgment

The project described was supported by National Institute On Aging (NIA) grant K25 AG035062 for Gina D'Angelo and supported by NIA grant R01 AG029672 and R01 AG034119 for Chengjie Xiong. This study was also partly supported by the NIA grant P50 AG005681, P01 AG003991, P01 AG26276, and U01 AG032438 for Chengjie Xiong. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIA or the National Institutes of Health.

References

1. Little RJA, Rubin DB (2002) *Statistical Analysis with Missing Data*, second ed. Wiley, New Jersey.
2. Molenberghs G, Kenward MG (2007) *Missing Data in Clinical Studies*. Wiley, West Sussex, England.
3. Dahiya RC, Korwar RM (1980) Maximum likelihood estimates for a bivariate normal distribution with missing data. *Ann Stat* 8: 687-692.
4. Roth PL (1994) Missing data: a conceptual review for applied psychologists. *Personnel Psychology* 47: 537-560.
5. Minami M, Shimizu K (1998) Estimation for a common correlation coefficient in bivariate normal distributions with missing observations. *Biometrics* 54: 1136-1146.
6. He Q, Nagaraja HN (2009) Correlation estimation using concomitants of order statistics from bivariate normal samples. *Commun Stat Theory Methods* 38: 2003-2015.
7. Truxillo C (2005) Maximum likelihood parameter estimation with incomplete data. In *SUGI 30 Proceedings*, Philadelphia.
8. Anderson TW (1984) *An Introduction to Multivariate Statistical Analysis*. (2ndedn), Wiley, New York.
9. Fisher R (1921) On the 'probable error' of a coefficient of correlation deduced from a small sample. *Metron* 1: 3-32.
10. Fisher R (1924) The distribution of the partial correlation coefficient. *Metron* 3: 329-332.
11. Hotelling H (1953) New light on the correlation coefficient and its transforms. *Journal of Royal Statistical Society. Series B (Methodological)* 15: 193-232.
12. Ibrahim JG, Chen MH, Lipsitz SR (1999) Monte carlo EM for missing covariates in parametric regression models. *Biometrics* 55: 591-596.
13. Ibrahim JG, Chen MH, Lipsitz SR, Herring AH (2005). Missing-data methods for generalized linear models: a comparative review. *J Am Stat Assoc* 100: 332-346.
14. McLachlan GJ, Krishnan T (2008) *The EM Algorithm and Extensions*. second ed. Wiley, New Jersey.
15. Louis TA (1982) Finding the observed information matrix when using the EM algorithm. *Journal of Royal Statistical Society. Series B(Methodological)* 44: 226-233.
16. Horton NJ, Laird NM (2001) Maximum likelihood analysis of logistic regression models with incomplete covariate data and auxiliary information. *Biometrics* 57: 34-42.
17. King G, Honaker J, Joseph A, Scheve K (2001) Analyzing incomplete political science data: an alternative algorithm for multiple imputation. *Am Polit Sci Rev* 95: 49-69.
18. Rubin DB (1987) *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York.
19. Schafer JL (1997) *Analysis of Incomplete Multivariate Data*. Chapman and Hall/ CRC, Boca Raton, Florida.
20. Barnard J, Meng XL (1999) Applications of multiple imputation in medical studies: from AIDS to NHANES. *Stat Methods Med Res* 8: 17-36.
21. Kenward MG, Carpenter J (2007) Multiple imputation: current perspectives. *Stat Methods Med Res* 16: 199-218.
22. Horton NJ, Kleinman KP (2007) Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. *Am Stat* 61: 79-90.
23. D'Angelo GM, Kamboh MI, Feingold E (2010) A likelihood-based approach for missing genotype data. *Hum Hered* 69: 171-183.
24. Kendall M, Stuart A (1977) *The Advanced Theory of Statistics*. (4thedn), Macmillan, New York.
25. Morris JC (1993) The clinical dementia rating (CDR): Current version and scoring rules. *Neurology* 43: 2412-2414.
26. Tukey JW (1977). *Exploratory Data Analysis*. Addison Wesley, Reading, MA.