Far East Journal of Theoretical Statistics
Volume 28, Number 2, 2009, Pages 133-155
Published Online: October 7, 2009
This paper is available online at http://www.pphmj.com

TESTING NORMALITY FOR LONGITUDINAL STUDIES WITH MISSING DATA

CHANGYONG FENG[*], HAIYAN SU[*], HONGYUE WANG[*], WAN TANG[*, †], QIN YU[*] and XIN M. TU[*, †]

[*]Department of Biostatistics and Computational Biology
University of Rochester
Rochester, NY 14642, U. S. A.

[†]Department of Psychiatry
University of Rochester
Rochester, NY 14642, U. S. A.
e-mail: feng@bst.rochester.edu

Abstract

Normal distribution underlies the premise for inference for many popular statistical models, especially for longitudinal data analysis such as the mixed-effects model. Validating the normality assumption is crucial for applications of such normal-based models to yield correct inferences. However, most of the current methods for testing the assumption of normality focus on univariate analysis. Although a few methods apply to multivariate outcomes, they do not address missing data. In this paper, we propose a new approach to test the normality assumption for longitudinal data analyses that simultaneously addresses missing data. We consider the two most popular missing data mechanisms in practical studies, the missing completely at random (MCAR) and the missing at random (MAR), and develop the corresponding methods to address them. We illustrate the approach with real as well as simulated data.

2000 Mathematics Subject Classification: 62H10, 47N30.

Keywords and phrases: inverse probability weighting, longitudinal data, missing completely at random, missing at random, monotone missing data pattern, *U*-statistics.

Received February 25, 2009

1. Introduction

Normal distribution is the most widely used assumption in theoretical and applied statistics. It forms the foundation for many classic multivariate tests as well as cutting-edge mixed-effects models for longitudinal and multi-level clustered data and structural equations model for inference of causal pathways (Diggle et al. [7]; Bollen [4]; Cole and Maxwell [5]). Applications of normal based models provide reasonably good estimates in many studies, however, severe departures from this assumption can lead to invalid inference and misleading conclusions. As Geary [13] points out, 'Normality is a myth; there never was, and will never be, a normal distribution.' His statement may be especially true for outcomes derived from most assessment instruments in mental health and psychosocial research. Since most instruments used in psychosocial research are based on item scores, they are intrinsically discrete. The treatment of such variables as though they were continuous is purely for analytic simplicity. For variables with a relatively large range such as the Hamilton rating scale for depression, such a treatment is convenient and sensible. However, because these variables are inherently discrete, they usually do not follow the normal distribution assumption. Thus, the use of parametric models is problematic and not justified in most applications. For this reason, the assumption of normality should not be taken for granted and must be checked routinely to ensure valid inference.

Over the past two decades, studies in biomedical and behavioral sciences have evolved from simple cross-sectional study designs to modern day longitudinal trials. As longitudinal study designs use subjects as their own controls, they provide a unique opportunity to study changes of outcomes of interest over time, causal effects and disease progression, in addition to providing more power for assessing treatment differences. In longitudinal data analysis, the generalized estimating equations (GEE) or weighted GEE (WGEE), and linear mixed-effects models (LMM) are among the most popular data analysis methods (Diggle et al. [7]). Since the estimation LMM is based on maximum likelihood method, it places very strong assumptions on the data and the model such as the observations from same individual over time have a multivariate normal distribution and the random effects is independent of the error term and have normal distributions. Thus, testing normality is crucial for valid inference and interpretation of results from such normal based models.

Missing data is one of the most important issues for longitudinal studies. Indeed,

this issue has been a major driving force behind the development of several ground-breaking statistical methods in the past two decades (e.g., Laird and Ware [18]; Liang and Zeger [21]; Zeger and Liang [42]; Wu and Carroll [40]; Lavori [19]; DeGruttola and Tu [6]; Follman and Wu [11]; Robins et al. [30]; Hogan and Laird [17]; Wulfsohn and Tsiatis [41]; Diggle et al. [7]). Methods of great complexity have been developed to address missing data, most such developments are centered around data analysis based on the assumed model (either parametric or semi-parametric model), but no methods have been developed to test normality with missing data in a longitudinal analysis setting.

In this paper, we develop a novel approach for testing normality for longitudinal study data that address the inherent missing data issue under the framework we recently developed for modelling second- and higher-order moments (Tu et al. [in press]). This approach provides a useful tool to validate the underlying assumption for normal-based models for longitudinal analysis and investigate causes of discrepancy arising when applying normal and non-normal based models.

The paper is organized as follows. In Section 2, we first briefly review existing methods for testing normality for univariate outcomes. In Section 3, we present our approach for longitudinal data. We first consider the complete data case and then discuss generalizations of the proposed approach to missing data under the two most popular missing data mechanisms in applications: the missing completely at random (MCAR) and missing at random (MAR) assumptions. In Section 4, we illustrate the proposed methods with real study data and investigate the size and power of the tests by Monte Carlo simulation. In Section 5, we discuss limitations of the work and directions of future research.

2. Tests for Normality for Univariate Outcome

Testing for normality is an age-old problem with many proposed solutions. For example, Mecklin and Mundfrom [27] reviewed dozens of procedures currently available for testing the normality assumption. They listed 106 references in their report and broke them down to four major categories: (1) methods based on graphical plots and correlation coefficient; (2) nonparametric goodness-of-fit methods; (3) methods based on skewness and kurtosis; and (4) methods based on empirical characterization. While this list is impressive, most methods are focused on univariate analysis. Since the 1970s, many attempts have been made to extend the univariate tests to a multivariate setting. However, none existing approach addresses

the issue of missing data, one of the most prominent and difficult problems in modern longitudinal data analysis.

In this section, we propose a new approach to address this fundamental limitation in existing approaches. In particular, we generalize the class of Lin-Mudholkar's *z*-tests of univariate normality to a longitudinal data setting (Lin and Mudholkar [22]; Mudholkar et al. [28]; Mudholkar et al. [29]). Although some of their tests also apply to multivariate outcomes, they fail to address missing data.

Following their work, our proposed test is premised on the following fundamental property of the normal distribution:

Proposition. *For an i.i.d.* (*independently, identically distributed*) *random sample, the sample mean and sample variance are independent if and only if the sample is from a normal population*.

Lukacs [24] later gave another justification using characterization function and extended the conclusion to multivariate outcomes.

Consider an i.i.d. random sample X_i $(1 \le i \le n)$. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ and $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ be the sample mean and sample variance, respectively. By the proposition, testing for normality is equivalent to ascertaining the independence between $\overline{X}$ and *S*. Thus, the null H_0 that X_i are from a normal distribution is the same as the null of a zero correlation between $\overline{X}$ and *S*. Since only one pair $(\overline{X}, S)$ is available from the sample, we cannot apply standard methods and test the null by using say the Pearson correlation estimator. One approach to this problem is to construct a sample by using reassembling methods (e.g., Efron and Tibshirani [9, Chapter 10]). For example, in Lin and Mudholkar [22], they constructed a sample by Jackknifing the original data sample, i.e.,

$$U_{(i)} = \left[\frac{1}{n-1}\sum_{k\neq i} X_k^2 - \left(\sum_{k\neq i} X_k\right)^2\right]^{1/3},$$

$$\hat{\rho} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(U_{(i)} - \overline{U})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2 \sum_{i=1}^{n}(U_{(i)} - \overline{U})^2}}, \quad 1 \le i \le n. \tag{1}$$

In the above, the Pearson correlation $\hat{\rho}$ is computed based on the Jackknified-sample, $(X_i, U_{(i)})$ $(1 \le i \le n)$. By further applying Fisher's *z* transformation to *r*, $Z = \frac{1}{2}\log\frac{1+\hat{\rho}}{1-\hat{\rho}}$, it is shown by Lin and Mudholkar [22] that under the null hypothesis of normality the normalized *Z* has an approximate standard normal distribution,

$$\sqrt{\frac{n}{3}}Z = \sqrt{\frac{n}{3}}\frac{1}{2}\log\frac{1+\hat{\rho}}{1-\hat{\rho}} \sim_{\text{approx}} N(0, 1), \tag{2}$$

for large samples.

The *T* statistic and its approximate distribution above forms the basis for inference when using the transformed Jackknife-based Pearson correlation estimator *Z* to test the null of zero correlation between the sample mean and variance. Simulation studies have shown that this test statistic has good coverage (type I error) and power for small and moderate sample sizes (Lin and Mudholkar [22]).

3. Tests for Normality for Multivariate Outcome

In this section, we develop a test statistic for testing normality for longitudinal studies with missing data. Although tests for normality for multivariate outcomes are available, none addresses missing data. We start with the complete data case.

3.1. Complete data case

Consider a longitudinal study with *n* subjects and *m* assessments. Let $\mathbf{X}_i = (X_{i1}, X_{i2}, ..., X_{im})^\top$ be the response vector from the *i*th subject with X_{it} denoting the observation at the *t*th assessment $(1 \le j \le m)$. We assume that the i.i.d. sample $\mathbf{X}_i$ follows an *m*-variate multivariate normal distribution $N_m(\mu, \Sigma)$ with mean vector μ and variance matrix Σ $(1 \le i \le n)$. We assume that Σ full rank so that Σ is positive definite. Let $\Sigma^{1/2}$ be the symmetric square-root of Σ and $\mathbf{Y}_i = \Sigma^{-1/2}(\mathbf{X}_i - \mu) = (Y_{i1}, Y_{i2}, ..., Y_{im})^\top$. Then, $\text{Var}(\mathbf{Y}_i) = I_m$, the $m \times m$ identity matrix. Under the null hypothesis, $\mathbf{Y}_i \sim$ i.i.d. $N_m(\mathbf{0}, I_m)$. For the *normalized* $\mathbf{Y}_i$, the components Y_{it} of $\mathbf{Y}_i$ are now stochastically independent, which form the premise for our test statistic.

First, assume that μ and Σ are both known. Let $\tilde{\mathbf{Y}}_t = (Y_{1j}, Y_{2j}, ..., Y_{nj})^\top$ denote

the observations at time t $(1 \le t \le m)$. Then, by applying (1) and (2) to each $\tilde{\mathbf{Y}}_t$, we obtain the *z*-test statistic Z_t for each *t*. Further, Z_t are i.i.d. since $\tilde{\mathbf{Y}}_t$ are such random variables under the null. Thus, under the null hypothesis of multivariate normality, it follows from (2) that

$$T = \frac{n}{3}\sum_{t=1}^{m} Z_t^2 \sim_{\text{approx}} \chi_m^2, \tag{3}$$

where χ_m^2 denotes a chi-square distribution with degrees of freedom *m*.

In most applications, neither μ nor Σ is known. To compute the statistic *T* in (3), we must estimate μ and Σ. To this end, let $\hat{\mu}$ and $\hat{\Sigma}$ be the usual sample mean and variance based on the random vectors $\mathbf{X}_i$ $(1 \le i \le n)$. Then, we can apply (3) to $\mathbf{Y}_i = \hat{\Sigma}^{-1/2}(\mathbf{X}_i - \hat{\mu})$ and obtain a similar statistic *T*. Since both are consistent estimators of their respective parameter vector and matrix, we have:

$$\sqrt{n}(\hat{\mu} - \mu) = \mathbf{O}_p(1), \qquad \sqrt{n}(\hat{\Sigma} - \Sigma) = \mathbf{O}_p(1), \tag{4}$$

where $\mathbf{O}_p(\cdot)$ denotes the stochastic $\mathbf{O}(\cdot)$ for random vectors (matrices) (e.g., Serfling [33, Chap. 1]). It then follows from (4) that

$$\mathbf{Y}_i = \hat{\Sigma}^{-1/2}(\mathbf{X}_i - \hat{\mu}) = \Sigma^{-1/2}(\mathbf{X}_i - \mu) + \mathbf{o}_p(n^{-1/2}), \tag{5}$$

where $\mathbf{o}_p(\cdot)$ denotes the stochastic $\mathbf{o}(\cdot)$ for random vectors (e.g., Serfling [33, Chap. 1]). Thus, it follows from (5) that the revised statistic, *T* when computed based on the estimated $\hat{\mu}$ and $\hat{\Sigma}$, also follows an approximate χ_m^2 in (2).

3.2. Missing data case

In longitudinal studies, missing data are inevitable, even for well planned trials. In longitudinal cohort studies, subjects may simply quit the study or they may not show up at follow-up visits because of problems with transportation, weather, health conditions, relocations, etc. In clinical trial studies, missing data may also be the results of patients' deteriorated or improved health conditions due to treatment, treatment-related complications, treatment response. Some of the reasons such as patients' deteriorated or improved health conditions due to treatment are clearly treatment related, while others such as subject's relocation may not. Treatment-

related missing data will impact the properties of estimators such as bias and consistency when ignored and thus must be carefully studied and addressed.

Within our setting, one approach is to simply apply the approach in the complete data case discussed in the preceding section to the sub-sample of subjects with complete data, i.e., those with response vector $\mathbf{X}_i = (X_{i1}, X_{i2}, ..., X_{im})^\top$ observed at all m assessment times. Such a complete-data approach is at best inefficient. More importantly, it is likely to yield biased estimators if missing data does not satisfy the missing completely at random (MCAR) assumption (Rubin [31]). We first address efficiency under MCAR and then turn our attention to bias when MCAR fails.

As in the complete data case, we first discuss the case when both μ and Σ are known. Let $\mathbf{X}_i$ be i.i.d. m-variate random vectors with mean vector μ and variance matrix Σ $(1 \le i \le n)$. In the presence of missing data, some of the components of $\mathbf{X}_i$ are not observed. Let $\mathbf{X}_i^o$ denote the m_i-dimensional subvector consisting of the observed components, and μ_i^o and Σ_i^o denote, respectively, the m_i-dimensional subvector of μ and the $m_i \times m_i$ submatrix of Σ corresponding to $\mathbf{X}_i^o$. Then, under the null of normality, it follows from the properties of multivariate normal distribution that $\mathbf{X}_i^o$ are independently and normally distributed with mean μ_i^o and variance Σ_i^o (Seber [32, Chap. 1]). Thus, the normalized vector, $\mathbf{Y}_i^o = (\Sigma_i^o)^{-1/2}(\mathbf{X}_i^o - \mu_i^o)$, are independently distributed with an m_i-variate normal distribution with mean $\mathbf{0}$ and variance I_{m_i}.

For each t $(1 \le t \le m)$, let $\tilde{\mathbf{Y}}_t^o = (Y_{i_1 t}, Y_{i_2 t}, ..., Y_{i_{n_j} t})^\top$, where $i_1, i_2, ..., i_{n_t}$ index the n_t subjects with the response $Y_{i_k t}$ observed at time t $(1 \le k \le n_t)$. Then, as in the complete data case, by applying (1) to each $\tilde{\mathbf{Y}}_t^o$, we obtain the z-test statistic Z_t. Since under the null $\tilde{\mathbf{Y}}_t^o$ are independent, it follows that Z_t are also independent and

$$T = \sum_{t=1}^{m} \frac{n_j}{3} Z_t^2 \sim_{\text{approx}} \chi_m^2, \tag{6}$$

for large $n = \min\{n_t; 1 \le t \le m\}$.

When μ_i^o and Σ_i^o are unknown, we must replace them by consistent estimators. We consider estimation for the two most popular missing data mechanisms arising in applications next.

3.2.1. Missing completely at random (MCAR)

In statistics, we characterize the impact of missing data on model estimates through assumptions or missing data mechanisms. Such assumptions allow statisticians to ignore the multitude of reasons for missing data and focus on addressing their impact on estimation of model parameters. The missing completely at random assumption (MCAR) is used to define a class of missing data that do not affect model estimates when completely ignored. For example, in a treatment study, most treatment unrelated missing data such as patient's relocation and conflict of schedules all fall into this category. The MCAR corresponds to a lay person's notion of random missing, i.e., missing data are completely random with nothing to do with treatment conditions.

To define consistent estimators of μ and Σ to address missing data, define a vector of binary variables for indicating missing (or rather observed) data as follows:

$$r_{it} = \begin{cases} 1 & \text{if } X_{it} \text{ is observed} \\ 0 & \text{otherwise} \end{cases}, \quad \mathbf{r}_i = (r_{i1}, ..., r_{im})^\top, \quad R_i = \text{diag}(r_{it}), \quad 1 \le i \le n, \tag{7}$$

where X_{it} is the tth component of $\mathbf{X}_i$ and $\text{diag}(r_i)$ is an $m \times m$ diagonal matrix with r_{it} on the tth diagonal. For convenience, denote the baseline by $t = 1$ and assume no missing data at this time, i.e., $r_{i1} \equiv 1\ (1 \le i \le n)$.

Now, define our estimator of μ as follows:

$$\hat{\mu} = \left(\sum_{i=1}^{n} R_i \right)^{-1} \sum_{i=1}^{n} R_i \mathbf{X}_i. \tag{8}$$

Under MCAR, it follows from the weak law of large numbers (LLN) that

$$\hat{\mu} = \left(\sum_{i=1}^{n} R_i \right)^{-1} \sum_{i=1}^{n} R_i \mathbf{X}_i \to_p E^{-1}(R_i) E(R_i \mathbf{X}_i) = E^{-1}(R_i) E(R_i) E(\mathbf{X}_i) = E(\mathbf{X}_i).$$

Thus, (8) is a consistent estimator of μ.

The construction of a consistent estimator of Σ of $\mathbf{X}_i$ is more complex since unlike the mean μ, Σ involves second-order moments. To reduce algebraic complexity and effectively address the analytic difficulty, we utilize the theory of *U*-statistics to construct such an estimator (Hoeffding [16]; Serfling [33, Chap. 5]; Tu et al. [38]; Ma et al. [25]).

For $1 \le s \le t \le m$, consider the following *U*-statistic:

$$U_{st}(\mathbf{r}_i, \mathbf{X}_i; \mathbf{r}_j, \mathbf{X}_j) = \binom{n}{2}^{-1} \sum_{(i,\, j)\in C_2^n} h_{st}(\mathbf{X}_i, \mathbf{X}_j)$$

$$= \binom{n}{2}^{-1} \sum_{(i,\, j)\in C_2^n} \frac{1}{2} r_{is} r_{it} r_{js} r_{jt} (X_{is} - X_{js})(X_{it} - X_{jt}), \qquad (9)$$

where $C_2^n = \{(i, j) : 1 \le i, j \le n, i \ne j\}$ denotes the set of all distinct pairs (i, j) from the integer set $\{1, 2, ..., n\}$ and $h_{st}(\mathbf{X}_i, \mathbf{X}_j) = \frac{1}{2} r_{is} r_{it} r_{js} r_{jt} (X_{is} - X_{js})(X_{it} - X_{jt})$ is a *symmetric kernel* (with respect to *i* and *j*) of the *U*-statistic. Under MCAR,

$$\theta_{st} = E[h_{st}(\mathbf{X}_i, \mathbf{X}_j)] = E(r_{is} r_{it} r_{js} r_{jt}) E\left[\frac{1}{2}(X_{is} - X_{js})(X_{it} - X_{jt})\right]$$

$$= E^2(r_{is} r_{it}) \mathrm{Cov}(X_{is}, X_{it}) = E^2(r_{is} r_{it}) \sigma_{st}. \qquad (10)$$

It thus follows from the theory of *U*-statistics that U_{st} in (9) is an unbiased and consistent estimator of $\theta_{st} = E^2(r_{is} r_{it}) \sigma_{st}$. Then, in the absence of missing data, $r_{is} = r_{it} = 1$ for all i and t $(1 \le i \le n, 1 \le t \le m)$ such that U_{st} is a consistent estimator of the variance σ_{st}. In general, however, U_{st} is not an estimator of Σ since $E^2(r_{is} r_{it}) \ne 1$.

Now, let

$$w_{st} = \sum_{(i,\, j)\in C_2^n} r_{is} r_{it} r_{js} r_{jt},$$

$$\hat{\sigma}_{st} = w_{st}^{-1} U_{st} = \binom{n}{2}^{-1} \sum_{(i,\, j)\in C_2^n} w_{st}^{-1} h_{st}(\mathbf{X}_i, \mathbf{X}_j), \quad 1 \le s \le t \le m. \qquad (11)$$

Under MCAR, it follows from (10) and Slutsky's theorem (Serfling [33, Chap. 1]) that

$$\hat{\sigma}_{st} = w_{st}^{-1}U_{st} = \frac{1}{\binom{n}{2}^{-1}\sum_{(i,j)\in C_2^n} r_{is}r_{it}r_{js}r_{jt}}\binom{n}{2}^{-1}\sum_{(i,j)\in C_2^n} h_{st}(\mathbf{X}_i, \mathbf{X}_j)$$

$$\to_p \frac{E(h_{st}(\mathbf{X}_i, \mathbf{X}_j))}{E^2(r_{is}r_{it})} = \frac{E^2(r_{is}r_{it})\mathrm{Cov}(X_{is}, X_{it})}{E^2(r_{is}r_{it})} = \sigma_{st}. \tag{12}$$

Thus, $\hat{\sigma}_{st}$ defined in (12) is a consistent estimator of σ_{st} and $\hat{\Sigma} = [\hat{\sigma}_{st}]$ is a consistent estimator Σ.

3.2.2. Missing at random (MAR)

In many clinical trials, missing data are often associated with the treatment interventions under study. For example, a patient may quit the study if he/she feels that the study treatment has deteriorated his/her health conditions and any further treatment will only worsen the medical or psychological problems. Or, a patient may feel that he/she has completely responded to the treatment and does not see any additional benefit in continuing the treatment. In such cases, the missing data do not follow the MCAR model since they are predicted by treatment related responses. This class of missing data reasons are modeled by the MAR that posits that the occurrence of a missing response at an assessment time depends on the response history or observed pattern prior to the assessment point. Thus, MAR postulates a plausible and indeed applicable missing data condition that encompasses many treatment related missing data and constitutes a sensible statistical approach to address bias in such situations.

Under MAR, the missing data indicator $\mathbf{r}_i$ becomes dependent on the response vector $\mathbf{X}_i$. We must model this dependence in order to define consistent estimators of μ and Σ. To this end, let

$$\pi_{it} = \Pr[r_{it} = 1 | \mathbf{X}_i], \quad \pi_{ist} = \Pr[r_{is} = 1, r_{it} = 1 | \mathbf{X}_i],$$
$$1 \le s \le t \le m, \quad 1 \le i \le n. \tag{13}$$

We first assume that π_{it} and π_{ist} are known and derive consistent estimators of μ and Σ and then discuss how to model and estimate π_{it} and π_{ist} under MAR.

Let

$$\hat{\mu}_t = \left(\sum_{i=1}^{n} \frac{r_{it}}{\pi_{it}}\right)^{-1} \sum_{i=1}^{n} \frac{r_{it}}{\pi_{it}} X_{it},$$

$$\hat{\sigma}_{st} = \binom{n}{2}^{-1} \sum_{(i,\, j)\in C_2^n} h_{st}(\mathbf{X}_i,\, \mathbf{X}_j)$$

$$= \binom{n}{2}^{-1} \sum_{(i,\, j)\in C_2^n} \frac{1}{2} \frac{r_{is} r_{it} r_{js} r_{jt}}{\pi_{ist}\pi_{jst}} (X_{is} - X_{js})(X_{it} - X_{jt}),$$

$$1 \le s \le t \le m. \tag{14}$$

These *inverse probability weighted* (IPW) estimators generalize those defined in (8) and (11) to account for response-dependent missingness under MAR. We assume that π_{it} and π_{ist} are bounded away from 0. By applying LLN and Slutsky's theorem, it follows that

$$\hat{\mu}_t \to_p \left[E\left(\frac{r_{it}}{\pi_{it}}\right)\right]^{-1} E\left(\frac{r_{it}}{\pi_{it}} X_{it}\right). \tag{15}$$

By (13) and the iterated conditional expectation (Billingsley [3, Chapter 6]), we have

$$E\left(\frac{r_{it}}{\pi_{it}}\right) = E\left[E\left(\frac{r_{it}}{\pi_{it}} \mid \mathbf{X}_i\right)\right] = E[\pi_{it} E(r_{it} \mid \mathbf{X}_i)] = 1,$$

$$E\left(\frac{r_{it}}{\pi_{it}} X_{it}\right) = E\left[E\left(\frac{r_{it}}{\pi_{it}} X_{it} \mid \mathbf{X}_i\right)\right] = E[\pi_{it}^{-1} X_{it} E(r_{it} \mid \mathbf{X}_i)] = E(X_{it}) = \hat{\mu}_t. \tag{16}$$

Thus, it follows from (15) and (16) that $\hat{\mu}_t$ is a consistent estimator of μ_t $(1 \le t \le m)$.

As in the discussion of MCAR, $\hat{\sigma}_{st} = \binom{n}{2}^{-1} \sum_{(i,\, j)\in C_2^n} h_{st}(\mathbf{X}_i,\, \mathbf{X}_j)$ is a *U*-statistic with a symmetric kernel $h_{st}(\mathbf{X}_i,\, \mathbf{X}_j)$. Further,

$$E(\hat{\sigma}_{st}) = E(h_{st}(\mathbf{X}_i,\, \mathbf{X}_j))$$

$$= E\left[E\left(\frac{1}{2} \frac{r_{is} r_{it} r_{js} r_{jt}}{\pi_{ist}\pi_{jst}} (X_{is} - X_{js})(X_{it} - X_{jt}) \mid \mathbf{X}_i,\, \mathbf{X}_j\right)\right]$$

$$= E\left[\frac{1}{2}\pi_{ist}^{-1}\pi_{jst}^{-1}(X_{is} - X_{js})(X_{it} - X_{jt})E(r_{is}r_{it}r_{js}r_{jt} \mid \mathbf{X}_i, \mathbf{X}_j)\right]$$

$$= E\left[\frac{1}{2}(X_{is} - X_{js})(X_{it} - X_{jt})\right] = \sigma_{st}.$$

It then follows from the theory of *U*-statistics that $\hat{\sigma}_{st}$ is an unbiased and consistent estimator of σ_{st}.

In most applications, π_{it} and π_{ist} are unknown. It is difficult to model π_{it} and π_{ist} as a function of $\mathbf{X}_i$ in general because of missing data responses (Little and Rubin [23]). As in the literature, we consider MAR, in which case π_{it} (π_{ist}) only depends on the observed part of $\mathbf{X}_i$. For nonparametric analysis, it is still quite complex to model the occurrence of missing data even under such an *ignorable missingness assumption* (Little and Rubin [23]). A common approach is to further impose the *monotone missing data pattern* (MMDP) assumption. MMDP eliminates a potentially large number of missing data patterns and makes it practical to model the occurrence of missing data under MAR (Robins et al. [30]).

Under MMDP, the response of the *i*th subject at t, X_{it}, is observed if and only if the subject's responses prior to time t, $X_{i1}, ..., X_{i(t-1)}$ are all observed. In terms of the missing data indicator vector, $\mathbf{r}_i$ follows the pattern that it has value 1 for the first m_i and 0 for the remaining $(m - m_i)$ components $(1 \le m_i \le m)$. Under MMDP, it is straightforward to model the binary missing data indicator r_{it} as a function of observed responses as well as other covariates using logistic regression. More specifically, let $\mathbf{X}_{it}^o = (X_{i1}, ..., X_{i(t-1)})^\top$ denote the subvector of observed responses of $\mathbf{X}_i^o$ up to and including time $t-1$. Then, under MMDP and MAR, it follows from (13) that

$$\begin{aligned}
&\pi_{it} = \Pr[r_{it} = 1 \mid \mathbf{X}_i] = \Pr[r_{it} = 1 \mid \mathbf{X}_{it}^o],\\
&\pi_{ist} = \Pr[r_{is} = 1, r_{it} = 1 \mid \mathbf{X}_i] = \Pr[r_{it} = 1 \mid \mathbf{X}_i] = \pi_{it},\\
&1 \le s \le t \le m, \quad 1 \le i \le n. \qquad (17)
\end{aligned}$$

Under MAR and MMDP assumption, MCAR can be viewed as a special case when π_{it} is functionally independent of $\mathbf{X}_{it}^o$.

To estimate π_{it} in (17), we can first model the one-step transition probability of the occurrence of missing data using a logistic regression model:

$$\text{logit}(p_{it}) = \text{logit}[\Pr(r_{it} = 1 | r_{i(t-1)} = 1, \mathbf{X}_{it}^{o})] = \beta_{0t} + \beta_t^{\top} \mathbf{X}_{it}^{o}, \quad 1 \le t \le m, \qquad (18)$$

where α_t and $\beta_t = (\beta_{1t}, ..., \beta_{(t-1)t})^{\top}$ are parameters of the model. By MMDP and the assumption of no missing data at $t = 1$, we have for $s \le t$:

$$\pi_{it} = \Pr[r_{it} = 1 | \mathbf{X}_{it}^{o}] = \Pr[r_{it} = 1 | r_{i(t-1)} = 1, \mathbf{X}_{it}^{o}] \Pr[r_{i(t-1)} = 1 | \mathbf{X}_{i(t-1)}^{o}] = \prod_{s=2}^{t} p_{is}.$$

In many applications, the one-step logistic model in (18) may only depend on the most recently observed response $X_{i(t-1)}$ and the predictor of the logistic model under this Markov condition is simply: $\beta_{0t} + \beta_{1t} X_{i(t-1)}$.

4. Applications

In this section, we illustrate the proposed approach with both real and simulated data. We first present applications of the tests to data from two studies in psychosocial research and then follow up with investigations of the performance of the tests with finite sample sizes. In all the examples, we set the statistical significance for inference at $\alpha = 0.05$. The approach is implemented in R and the software is available for download from our website:

http://www.urmc.rochester.edu/smd/biostat/service/index.html.

4.1. Testing normality for two real studies

Example 1. The study data for this example is from a PTSD (posttraumatic stress disorder) study conducted at the University of Pennsylvania Medical Center (Tu et al. [37]; Tu et al. [in press]). The goal of the study is to examine the longitudinal biological and psychological alterations in response to a cognitive behavioral therapy intervention that has been proven to induce a substantial reduction in PTSD symptoms in the majority of women victims of sexual and non-sexual assault. Ninety-five patients completed treatment, with scheduled assessments at pre- (baseline), post-treatment (10 weeks from baseline) and three follow-up times (in a six month interval).

For the two major response variables for examining symptom reductions in

PTSD and depression, the PTSD Symptom Scale (Foa et al. [10]) and the Beck Depression Inventory (BDI, Beck et al. [2]), all patients had complete data from baseline (pre-treatment) to the last follow-up assessment. As the linear mixed-effects model is the most popular in mental health and psychosocial research, it is important to assess the normal distribution assumption for these key outcomes before such models can be applied.

Since there is no missing data for the two key outcomes of interest, we applied the test for complete data discussed in Subsection 3.2 to the longitudinal data from each of the outcomes and obtained the chi-square test statistics, $T = 19.24$ and $T = 16.09,$ and associated p-values, 0.0017 and 0.0066, for the PTSD symptom scale and variable and BDI variables, respectively. The highly significant findings provide little support for normality for either of the outcomes and thus the linear mixed-effects model may not be an appropriate choice for examining the trajectories of the outcomes over time. Our conclusion agrees with the test results (p-values < 0.01) from the Henze-Zirkler and Mardia tests for normality based on skewness and kurtosis implemented in a SAS macro which can be downloaded from the SAS website (Henze and Zirkler [15]; Mardia [24]).

Example 2. Data for this second example is from a longitudinal study in examining whether clinically significant physical pain was associated with worse treatment outcomes of an interpersonal psychotherapy among 59 depressed women with childhood sexual abuse (Talbot et al. [35]). In the study, depression was assessed by the Structured Clinical Interview for DSM-IV (Spitzer et al. [34]), sexual abuse was assessed with a semi-structured interview (defined as any unwanted sexual contact, or any sexual contact with a family member 5 years or older than the patient, prior to the age of 18), and pain was assessed by two self-report items on the SF-36 measuring pain intensity and interference (Ware and Sherbourne [39]). Assessments were conducted at baseline, 10, 24, and 36 weeks, with about 30% missing data in the follow-up assessments.

Longitudinal models such as the mixed-effects model can be used to examine the relationship between pain and treatment outcomes. However, to ensure valid inference, we tested the normality for the three key depression outcomes: HAMD (Hamilton Rating scale for Depression, Hamilton [14]), BDI and TRMCS (level of physical health functioning based on SF-36). By applying the discussion in Subsection 3.2.2, we modeled the occurrence of missing data using the logistic model (18) with the predictor $\beta_0 + \beta_1 X_{i(t-1)}$ under the Markov assumption to

determine whether the missing mechanism is MCAR. For the three depression outcomes, the dependence of missingness on outcome was found for TRMCS; $\hat{\beta}_1 = 0.030$ (standard error = 0.017) and p-value = 0.08. We then applied the test statistic to TRMCS under MAR and HAMD and BDI under MCAR as discussed in Subsection 3.2.2. The test yielded a chi-square value of 11.13 with p-value $= 0.025$ for TRMCS, indicating that the normal assumption is questionable. The tests were not significant for the other two variables.

4.2. Simulation study

We conducted simulation studies to examine the empirical size and power of the proposed test with finite sample sizes under complete data, MCAR and MAR. For space consideration, we only report results for two sample sizes, $n = 50$ and 200. All simulations were performed with a Monte Carlo sample of 1000 using the R software (Free Software Foundation [12]).

4.2.1. Empirical size

To study the empirical size of the proposed test, we generated data from a 3 dimensional normal distribution $N_3(\mathbf{0}, \Sigma)$ with the mean vector $\mathbf{0}$ and variance $\Sigma = C(\rho)$, a compound symmetry correlation matrix with the correlation $\rho = 0.5$ (Diggle et al. [7]; Tu et al. [36]). The nominal size of the test was set at $\alpha = 0.05$. The empirical size is calculated as $\hat{\alpha} = \dfrac{\#\text{ of times } T > 7.81}{1000}$, where 7.81 is 95th percentile of the χ^2_3 distribution. The standard error of empirical size $\hat{\alpha}$ is $\sqrt{(0.05 \times 0.95)/1000} = 0.007$.

Shown in Table 1 are the estimated empirical size from the simulations for the complete and missing data under MCAR and MAR. For the complete data case, the empirical size is a bit underestimated when $n = 50$, but closer to the normal value for the large sample $n = 200$. For the missing data case, we assumed no missing data at baseline $t = 1$ and modeled the missing data mechanism according to MCAR or MAR using the methods discussed in Subsection 3.2.2. For MCAR, the missing probability $p = \Pr(r_{it})$ indicates the percentage of missing responses at time $t = 2$ and 3. Thus, when $p = 0.1$, there is 10% missing data at times 2 and 3. For MAR, we modeled and used the missing data indicator r_{it} as a function of the response at the previous time $t - 1$, $X_{i(t-1)}$, using a logistic regression with a predictor of the

form $\beta_0 + 0.5X_{i(t-1)}$, where $\beta_0 = \text{logit}(p)$ is determined by the missing data probability p in the absence of the predictor $X_{i(t-1)}$. Thus, for $p = 0.1$ for example, $\beta_0 = \text{logit}(0.1) = \log\left(\frac{0.1}{1-0.1}\right) = -2.2$.

Table 1. Empirical size for the proposed test for normality under complete data and missing data with MCAR and MAR as a function of sample size $n = 50$ and 200

n	Complete data	Missing data		
		Missing probability (p)	MCAR	MAR
50	0.03	0.1	0.03	0.04
		0.2	0.02	0.04
		0.3	0.04	0.04
200	0.04	0.1	0.05	0.04
		0.2	0.04	0.05
		0.3	0.04	0.05

4.2.2. Empirical power

To study empirical power, we considered two non-normal family distributions as alternatives for testing the null of multivariate normality: the multivariate t distributions and the two-component normal mixture distributions. We first briefly review these two classes of distributions.

1. Multivariate t

Let $\mathbf{U} \sim N_m(\mathbf{0}, \Sigma)$ and $V \sim \chi^2_\nu$ be independently distributed. Let $\mathbf{X} = \mu + (V/\nu)^{-\frac{1}{2}}\mathbf{U}$, where μ is an $m \times 1$ constant vector. Then, $\mathbf{X}$ follows an m-variate t-distribution with degrees of freedom ν. The probability density function of $\mathbf{X}$ is given by (Andersen [1]; Tu et al. [37]):

$$f_{MT}(\mathbf{x}) = \frac{\Gamma((\nu+m)/2)}{(\pi\nu)^{m/2}\Gamma(\nu/2)|\Sigma|^{1/2}}\left(1 + \frac{1}{\nu}(\mathbf{x}-\mu)^\top \Sigma^{-1}(\mathbf{x}-\mu)\right)^{-(\nu+m)/2}.$$

It is readily shown that $E(\mathbf{X}) = \mu$ and $\text{Var}(\mathbf{X}) = \frac{\nu}{\nu-2}\Sigma$ (Tu et al. [37]). Like the normal distribution, the multivariate t has a unique mode at μ. However, t has a

larger variance and a thicker tail than the normal distribution, although this difference diminishes as the degree of freedom *v* becomes larger.

2. Two-component multivariate normal mixture

The probability density function of a two-component normal mixture is given by:

$$f_{NM}(\mathbf{x}) = \pi\phi_m(\mu_1, \Sigma_1) + (1-\pi)\phi_m(\mu_2, \Sigma_2),$$

where $\phi_m(\mu, \Sigma)$ denotes the probability density function of $N_m(\mu, \Sigma)$. Unlike the multivariate *t*, the two-component normal mixture is generally not symmetric and bimodal with the two modes at μ_1 and μ_2. Thus, the multivariate *t* and the two-component normal mixture provides a symmetric and non-symmetric alternative to the normal distribution to evaluate the power of the proposed test.

As in the investigation of normal coverage, we fixed the dimension at 3 and simulated data from a 3-variate multivariate *t* with $\mu = \mathbf{0}$, $\nu = 3$, $\Sigma = C(0.5)$. By using the hierarchal representation of the *t* distribution, we first generated independent normal vectors $\mathbf{U}_i$ from $N_3(\mathbf{0}, \Sigma)$ and independent variables V_i from χ^2_ν and then formed the desired *t* vectors by setting $\mathbf{X}_i = (V_i/\nu)^{-\frac{1}{2}}\mathbf{U}_i$ $(1 \le i \le n)$. For the normal mixture, we set $\mu_1 = \mu_2$, $\Sigma_1 = C(0.5)$ and considered two mixture distributions:

$$\text{Normal mixture } 1 : 0.8N_3(\mathbf{0}, \Sigma_1) + 0.2N_3(\mathbf{0}, 9\Sigma_1),$$

$$\text{Normal mixture } 2 : 0.9N_3(\mathbf{0}, \Sigma_1) + 0.1N_3(\mathbf{0}, 16\Sigma_1). \qquad (19)$$

With the different choice of Σ_2, the two normal mixtures in Subsection 4.2.2 provided different degrees of skewness in the data distributions. We generated each of the mixture normal samples defined in Subsection 4.2.2 also in a hierarchal fashion. We first simulated u_i from $U(0, 1)$, a uniform between 0 and 1 $(1 \le i \le n)$. Corresponding to each u_i, we then sampled $N_3(\mathbf{0}, \Sigma_1)$ if $u_i \le \pi$ and $N_3(\mathbf{0}, \Sigma_2)$ if otherwise. For each random sample, the test statistic *T* was calculated as discussed in Subsection 3.2.2, depending on the complete or missing data case. The empirical power was computed as $\frac{\#\text{ of times } T > 7.81}{1000}$.

Shown in Table 2 are the results of empirical power when the true distribution is the 3-variate *t*. For the complete data case, the empirical power is 0.56 when $n = 50$. The powers for the missing data case under MAR and MCAR range from 0.47 to 0.57 as the missing probability changes from 0.1 to 0.3. When $n = 200$, the empirical power is significantly increased to 0.85 for complete data. The powers for the missing data case under MAR and MCAR are comparable to those the one obtained for complete data.

Table 2. Empirical power for the multivariate *t* alternative under complete data and missing data with MCAR and MAR as a function of sample size $n = 50$ and 200

n	Complete data	Missing data		
		Missing probability (*p*)	MCAR	MAR
50	0.56	0.1	0.54	0.57
		0.2	0.47	0.54
		0.3	0.50	0.50
200	0.85	0.1	0.85	0.84
		0.2	0.83	0.84
		0.3	0.82	0.84

The empirical powers for the two normal mixtures are shown in Table 3. We can see that the test has less power for the first mixture distribution than for the second in both sample sizes. This is expected as the second normal mixture is more skewed than the first one. As in the *t* case, power significantly increased as the sample size became larger.

Table 3. Empirical size for the two normal mixture alternatives under complete data and missing data with MCAR and MAR as a function of sample size $n = 50$ and 200

n	Mixture	Complete data	Missing data		
			Missing probability (p)	MCAR	MAR
50	1	0.55	0.1	0.54	0.56
			0.2	0.54	0.55
			0.3	0.49	0.54
	2	0.72	0.1	0.70	0.71
			0.2	0.63	0.67
			0.3	0.62	0.70
200	1	0.63	0.1	0.64	0.75
			0.2	0.63	0.77
			0.3	0.65	0.77
	2	0.85	0.1	0.84	0.91
			0.2	0.83	0.89
			0.3	0.85	0.89

An interesting observation is that the power estimates for the missing data case are consistently larger under MAR than under MCAR for both multivariate t and normal mixtures. One plausible explanation is that although the increasing amount of missing data negatively affects power, the decreasing amount of observed data at the same leads to reduced information to ascertain the assumed analytic distributional structure of the data simulated, thus increasing the power to reject the null of normality. Also, unlike MCAR, missing data is not completely ignored under MAR and is accounted for by weighting the observed data using the inverse probability of the occurrence of missing data. Thus, it is not surprising that power is generally higher under MAR than MCAR.

5. Discussion

The normal distribution assumption underlies the premise for inference for many classic as well as modern statistical models. Although testing for normality is an age-old problem with many proposed solutions, modern longitudinal study designs involving multiple assessments over time have completely changed the

complexity of the underlying analytic issues. As a result, none of the existing methods can be applied to validate this key assumption which is the basis for a wide variety of normal-based mixed-effects models. Many published studies especially in psychosocial research use such models without even acknowledging this fundamental issue, yielding study findings that may misinform the public with misleading conclusions about treatment effects, disease etiology and progression. The proposed approach attempts to fill this important gap in the literature on longitudinal and clustered data analyses.

In developing the test, we generalized the inverse probability weighting (IPW) approach to address missing data under MAR. This approach is widely used in longitudinal data analysis using distribution-free models such as the WGEE (Robins et al. [30]). Our approach is novel in that we integrated IPW with the theory of *U*-statistics to address MAR when modeling the second-order moment based variance, rather than the mean response as in most applications of IPW.

The proposed approach is limited when applied to normal-normal based mixed-effects models, both the outcome and random effects following normal distributions, since it only addressed the normal assumption for the data distribution, not the distribution of the random effects. Thus, even if this assumption is valid, applications of such models may still yield incorrect results if the random effects fail to follow the normal distribution assumption. Current work is underway to address this latter issue.

Acknowledgements

This research is supported in part by NIH grants R01-DA012249 (Tang and Tu) and 1 UL1 RR024160-01 (Feng, Wang, Tang and Tu), and by a Merck Quantitative Sciences Fellowship grant provided by the Merck Company Foundation (Su). We are also grateful to Professor Mudholkar at the University of Rochester for many helpful discussions during the development of the paper.

References

[1] T. W. Anderson, An Introduction to Multivariate Statistical Analysis, 3rd ed., Wiley, Hoboken, NJ, 2003.

[2] A. T. Beck, C. H. Ward, M. Mendelson, J. Mock and J. Erbaugh, An inventory for measuring depression, Arch. Gen. Psychiatry 4(6) (1961), 561-571.

[3] P. Billingsley, Probability and Measure, 3rd ed., Wiley, New York, 1995.

[4] K. A. Bollen, Structural Equations with Latent Variables, Wiley, New York, 1989.

[5] D. A. Cole and S. E. Maxwell, Testing mediational models with longitudinal data: questions and tips in the use of structural equation modeling, J. Abnormal Psychology 112(4) (2003), 558-577.

[6] V. DeGruttola and X. M. Tu, Modeling the relationship between disease progression and survival time, Biometrics 50 (1994), 1003-1015.

[7] P. J. Diggle, P. Heagerty, K. Y. Liang and S. L. Zeger, Analysis of Longitudinal Data, 2nd ed., Oxford Univ. Press, New York, 2002.

[8] A. Donner and N. Klar, Cluster randomization trials in epidemiology: theory and application, J. Statist. Plann. Inference 42(1-2) (1994), 37-56.

[9] B. Efron and R. J. Tibshirani, An Introduction to the Bootstrap, Chapman and Hall, London, 1998.

[10] E. B. Foa, A. Ehlers, D. M. Clark, D. F. Tolin and S. M. Orsillo, The posttraumatic cognitions inventory (PTCI): development and validation, Psychol. Assess. 11(3) (1999), 303-314.

[11] D. Follmann and M. Wu, An approximate generalized linear model with random effects for informative missing data, Biometrics 51(1) (1995), 151-168.

[12] Free Software Foundation, Inc., 59 Temple Place - Suite 330, Boston, MA 02111, USA.

[13] R. C. Geary, The distribution of Student’s ratio for non-normal samples, Suppl. J. Roy. Statist. Soc. 3 (1947), 178-184.

[14] M. Hamilton, A rating scale for depression, J. Neurol. Neurosurg. Psychiatry 23(1) (1960), 56-62.

[15] N. Henze and B. Zirkler, A class of invariant consistent tests for multivariate normality, Comm. Statist. Theory Methods 19(10) (1990), 3595-3617.

[16] W. Hoeffding, A class of statistics with asymptotically normal distribution, Ann. Math. Statistics 19 (1948), 293-325.

[17] J. W. Hogan and N. M. Laird, Mixture models for the joint distribution of repeated measures and event times, mixture models for the joint distribution of repeated measures and event times, Stat. Med. 16(3) (1997), 239-257.

[18] N. Laird and J. Ware, Random-effects models for longitudinal data, Biometrics 38 (1982), 963-974.

[19] P. W. Lavori, Clinical trials in psychiatry: should protocol deviation censor patient data?, Neuropsychopharmacology 6(1) (1992), 39-48.

[20] P. Lavori, E. Laska and E. Ulenhuth, Statistical issues for the clinical evaluation of psychotropic drugs, Clinical Evaluation of Psychotropic Drugs: Principles and Guidelines, R. Prien and D. Robinson, eds., pp. 139-160, Raven Press, New York, 1994.

[21] K. Y. Liang and S. L. Zeger, Longitudinal data analysis using generalized linear models, Biometrika 73(1) (1986), 13-22.

[22] C. C. Lin and G. S. Mudholkar, A simple test for normality against asymmetric alternatives, Biometrika 67(2) (1980), 455-461.

[23] R. J. A. Little and D. B. Rubin, Statistical Analysis with Missing Data, Wiley, New York, 1987.

[24] E. Lukacs, A characterization of the normal distribution, Ann. Math. Statistics 13 (1942), 91-93.

[25] Y. Ma, W. Tang, C. Feng and X. M. Tu, Inference for kappas for longitudinal study data: applications to sexual health research, Biometrics 64(3) (2008), 781-789.

[26] K. V. Mardia, Measures of multivariate skewness and kurtosis with applications, Biometrika 57(3) (1970), 519-530.

[27] C. J. Mecklin and D. J. Mundfrom, An appraisal and bibliography of tests for multivariate normality, Inter. Stat. Rev. 72(1) (2004), 123-138.

[28] G. S. Mudholkar, M. McDermott and D. K. Srivastava, A test of *p*-variate normality, Biometrika 79(4) (1992), 850-854.

[29] G. S. Mudholkar, D. K. Srivastava and C. T. Lin, Some *p*-variate adaptations of the Shapiro-Wilk test of normality, Comm. Statist. Theory Methods 24(4) (1995), 953-985.

[30] J. M. Robins, A. Rotnitzky and L. P. Zhao, Analysis of semiparametric regression models for repeated outcomes in the presence of missing data, J. Amer. Statist. Assoc. 90(429) (1995), 106-121.

[31] D. B. Rubin, Inference and missing data, Biometrika 63(3) (1976), 581-592.

[32] G. A. F. Seber, Multivariate Observations, Wiley, New York, 1984.

[33] R. Serfling, Approximation Theorems of Mathematical Statistics, Wiley, 1980.

[34] R. L. Spitzer, M. Gibbon and J. B. W. Williams, Structured Clinical Interview for Axis I DSM-IV Disorders, Biometrics Research Department, New York State Psychiatric Institute, 1994.

[35] N. L. Talbot, Y. Conwell, M. W. O'Hara, S. Stuart, E. A. Ward, S. A. Gamble, A. Watts and X. Tu, Interpersonal psychotherapy for depressed women with sexual abuse histories: a pilot study in a community mental health center, J. Nervous and Mental Disease 193(12) (2005), 847-850.

[36] X. M. Tu, J. Kowalski, J. Zhang, K. G. Lynch and P. Crits-Christoph, Power analyses for longitudinal trials and other clustered designs, Stat. Med. 23(18) (2004), 2799-2815.

[37] X. M. Tu, J. Kowalski, P. Crits-Christoph and R. Gallop, Power analyses for correlations from clustered study designs, Stat. Med. 25(15) (2006), 2587-2606.

[38] X. M. Tu, C. Feng, J. Kowalski, W. Tang, H. Wang, C. Wan and Y. Ma, Correlation analysis for longitudinal data: applications to HIV and psychosocial research, Stat. Med. 26(22) (2007), 4116-4138.

[39] J. E. Ware, Jr. and C. D. Sherbourne, The MOS 36-item short-form health survey (SF-36): I. Conceptual framework and item selection, Medical Care 30(6) (1992), 473-483.

[40] M. C. Wu and R. J. Carroll, Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process, Biometrics 44(1) (1988), 175-188.

[41] M. S. Wulfsohn and A. A. Tsiatis, A joint model for survival and longitudinal data measured with error, Biometrics 53(1) (1997), 330-339.

[42] S. L. Zeger and K.-Y. Liang, Longitudinal data analysis for discrete and continuous outcomes, Biometrics 42(1) (1986), 121-130.