

Article

Imputation-Based Variable Selection Method for Block-Wise Missing Data When Integrating Multiple Longitudinal Studies

Zhongzhe Ouyang, Lu Wang* and Alzheimer's Disease Neuroimaging Initiative[†]

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109, USA

* Correspondence: luwang@umich.edu

[†] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report.

A complete listing of ADNI investigators can be found at:

http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf
(accessed on 29 February 2024).

Abstract: When integrating data from multiple sources, a common challenge is block-wise missing. Most existing methods address this issue only in cross-sectional studies. In this paper, we propose a method for variable selection when combining datasets from multiple sources in longitudinal studies. To account for block-wise missing in covariates, we impute the missing values multiple times based on combinations of samples from different missing pattern and predictors from different data sources. We then use these imputed data to construct estimating equations, and aggregate the information across subjects and sources with the generalized method of moments. We employ the smoothly clipped absolute deviation penalty in variable selection and use the extended Bayesian Information Criterion criteria for tuning parameter selection. We establish the asymptotic properties of the proposed estimator, and demonstrate the superior performance of the proposed method through numerical experiments. Furthermore, we apply the proposed method in the Alzheimer's Disease Neuroimaging Initiative study to identify sensitive early-stage biomarkers of Alzheimer's Disease, which is crucial for early disease detection and personalized treatment.



Citation: Ouyang, Z.; Wang, L.; Alzheimer's Disease Neuroimaging Initiative. Imputation-Based Variable Selection Method for Block-Wise Missing Data When Integrating Multiple Longitudinal Studies. *Mathematics* **2024**, *12*, 951. <https://doi.org/10.3390/math12070951>

Academic Editor: Quanxin Zhu

Received: 12 March 2024

Revised: 22 March 2024

Accepted: 22 March 2024

Published: 23 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: multiple imputation; correlated data; data integration**MSC:** 62H99

1. Introduction

Multi-sources data are now attracting more attention in scientific research. A practical problem with multi-source data is block-wise missing. Our work is motivated by the existence of block-wise missingness in Alzheimer's Disease Neuroimaging Initiative (ADNI) data when investigating the biomarkers that are associated with Alzheimer's Disease (AD). In the ADNI study, healthy elderly subjects, as well as subjects with normal cognition (NC), mild cognitive impairment (MCI), or AD, were recruited to identify neuroimaging measures, cognitive measures and biomarkers that can effectively and timely detect cognitive and functional changes [1]. The ADNI data exhibit a block-wise missing structure along with the long duration of the study, and the high cost of certain measurements, etc. Besides the ADNI data, datasets with block-wise missing structure also exist across many other fields including environmental science, sociology, and economics. For example, a block-wise missing structure appears in human mortality data integrated from Italy and Switzerland [2] and in credit data collected from various institutions (Lan and Jiang [3]; Li et al. [4]).

Statistical analysis with missing covariates has been widely studied due to the prevalence of missing values in many datasets. Common methods for dealing with missing data include complete case analysis, maximum likelihood, inverse probability weighting,

and imputation. While complete case analysis is the easiest approach to implement, it has several drawbacks, such as potential bias in certain cases and a significant loss of information when the proportion of missingness is high. The maximum likelihood approach (e.g., Sabbe et al. [5]; Bondarenko and Raghunathan [6]; Audigier et al. [7]; von Hippel and Bartlett [8]) requires a specification on the distribution of variables, though this is unknown and unverifiable in practice. Inverse probability weighting (e.g., Chen et al. [9]; Creemers et al. [10]; Zubizarreta [11]; Hughes et al. [12]) heavily relies on the information from complete cases, which can be problematic when the fraction of completely observed subject is small.

Two big challenges with the above ADNI data are the high proportion of missingness and the large number of covariates, which make the complete case analysis and maximum likelihood approach inefficient. In addition to these two challenges, weighted methods cannot handle the problem in presence of multiple missing patterns. Compared to these methods with notable limitations, imputation methods are more appropriate for the ADNI data. Recently, multi-source data with block-wise missing, exemplified by the ADNI data, have drawn extensively attention in statistically research. Ref. [13] developed a classification framework, which was accomplished by three steps: feature selection, sample selection, and matrix completion. Ref. [2] proposed a dimension reduction method called generalized integrative principal component analysis (GIPCA). Under the assumption of identical type of distribution in the exponential family within each data source, GIPCA decomposed the overall effect into joint and individual effect across data sources. Ref. [14] imputed the missing data using a factor structure model, which considered the correlation between predictors and does not depend on missing mechanism. Ref. [15] developed a multiple block-wise imputation (MBI) approach by constructing estimating functions based on both complete and incomplete observation. Other related literature include those of [4,16,17].

However, these methods are not applicable to longitudinal studies. Using these methods on the ADNI data, they only select baseline measurement for each patient and simply delete the following measurements. Thus, these methods are inefficient for the ADNI data since they fail to take account of with-subject correlations. In this paper, we aim to develop a method for variable selection when integrating longitudinal data from multiple sources in the existing block-wise missing structure. We impute the block-wise missing data multiple times by using the information from both subjects with complete observation and subjects with missing values. We construct estimating equations based on imputed data and incorporate working correlation matrices to account for within-cluster correlation. With the generalized method of moment, we are capable of integrating data from multiple sources and identifying the relevant variables by introducing a penalty term.

This paper is organized as follows. Section 2 describes the setup and formalize the proposed method. In Section 3, we study the asymptotic properties of the proposed estimator. In Section 4, we develop an algorithm to implement the proposed method, followed by a simulation study conducted in Section 5 to evaluate the performance of the proposed method. In Section 6, we apply the proposed method to the ADNI study. Section 7 provides further discussions.

2. Methods

2.1. Setup

Suppose the dataset consists of n independent and identically distributed (*i.i.d.*) samples drawn from independent sources with disjoint covariates. Without loss of generality, we assume that the data are already sorted by missing patterns, and the total number of missing patterns is K with n_k samples in each pattern, where $\sum_{k=1}^K n_k = n$ and $k = 1, \dots, K$. Within each missing pattern, all subjects have the same missing structure and the covariates from any specific source are either fully observed or fully missing. Let $Y_{k,i} = (Y_{k,i1}, \dots, Y_{k,im_i})^T$ be the response vector for the i th subject in the k th pattern with m_i measurements. For ease of presentation, we assume that each sample has the same

number of measurements m . Furthermore, let $\mathbf{X}_{k,i} = (\mathbf{X}_{k,i1}, \dots, \mathbf{X}_{k,ip})$ be the corresponding covariate matrix for the i th subject in the k th pattern across all the measurements, where p is the number of covariates. We assume the underlying population-level model is as follows:

$$E(\mathbf{Y}_{k,i} | \mathbf{X}_{k,i}) = \mu(\mathbf{X}_{k,i}\boldsymbol{\beta}), \quad k = 1, \dots, K,$$

where $\mu(\cdot)$ is a known monotonic link function and $\boldsymbol{\beta}$ is a p -dimensional vector in the parameter space. Let $O(k)$ and $M(k)$ denote the index of observed covariates and missing covariates in the k th pattern, respectively. Define $R_i = 1$ if $\mathbf{X}_{k,i}$ is fully observed, otherwise 0. We assume the missing mechanism of $\mathbf{X}_{k,i}$ is missing completely at random [18].

Figure 1 is an example illustrated what block-wise missing data look like, which consist of three sources with three missing patterns. Note that covariates in source 1 are completely observed in all three patterns, while covariates in source 2 are only observed in pattern 1 and 2, and covariates in source 3 are only observed in pattern 1 and 3. A similar structure also exists in the ADNI data. For example, variables in cerebrospinal fluid (CSF) are only measured in a subsample since CSF collection were mainly performed in phase II. Although complete cases analysis is feasible for ADNI data, it is inefficient especially the number of subjects with complete observation is limited. Thus, it is essential to leverage information from incomplete observation.

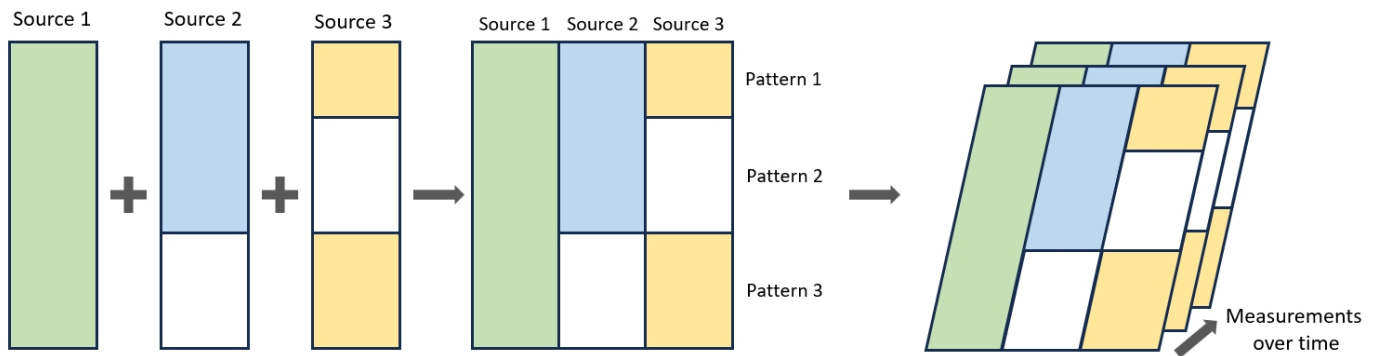


Figure 1. Example of block-wise missing data in longitudinal studies.

2.2. Proposed Method

One approach to utilizing incomplete data is by imputing missing values and performing statistical analysis based on the imputed dataset. Traditional methods impute missing values using information solely from complete cases. However, in scenarios involving block-wise missing data, the proportion of complete cases can be relatively small, resulting in unstable imputed values. To further illustrate how to incorporate information from subjects with partially observed values when imputing missing values, we continue to use the example given in Figure 1. Let $\mathbf{X}_{k,i(r)}$ be the r th imputed covariate vector for the i th subject from pattern k , $r = 1, \dots, R_k$. For instance, the missing values of $\mathbf{X}_{2,i}$, i.e., the covariates of source 3 in pattern 2, can be imputed using the information of all sources in pattern 1, which we denoted as $\mathbf{X}_{2,i(1)}$. Additionally, these can also be imputed based on the covariates in source 1 and source 3 for subjects from pattern 1 and pattern 3, which we denoted as $\mathbf{X}_{2,i(2)}$. Figure 2 illustrate how the above procedures work. When all the covariates are observed, $\mathbf{X}_{k,i(r)} = \mathbf{X}_{k,i}$. Similarly, we can define $\mu_{k,i(r)}(\boldsymbol{\beta})$ as the corresponding imputed conditional mean.

The intuition behind the proposed method stems from generalized estimating equations (GEE) and quadratic inference functions (QIF). Suppose $\mathbf{V}_{k,i}$ is the unknown true covariance matrix of $\mathbf{Y}_{k,i}$. Ref. [19] proposed that $\mathbf{V}_{k,i}$ can be estimated by $\mathbf{A}_{k,i}^{1/2} \mathbf{R}_{k,i}(\boldsymbol{\alpha}) \mathbf{A}_{k,i}^{1/2}$, where $\mathbf{A}_{k,i}$ is the diagonal matrix of the conditional variance of $\mathbf{Y}_{k,i}$ and $\mathbf{R}_{k,i}$ is a working correlation matrix that fully specified by a vector of parameter $\boldsymbol{\alpha}$. Ref. [20] proposed the QIF using the fact that the inverse of the correlation matrix $\mathbf{R}_{k,i}^{-1}$ can be approximated by $\sum_{j=1}^J a_{k,j} \mathbf{M}_j$, where $\mathbf{M}_1, \dots, \mathbf{M}_J$ are some basis matrices. For example, if we assume

the within-cluster correlation structure is exchangeable, $R_{k,i}^{-1}$ can be approximated by $a_1M_1 + a_2M_2$, where M_1 is the identity matrix and M_2 is a matrix with elements in the diagonal to be 0 and elements in the off-diagonal to be 1. The estimation of inverse of correlation matrix using linear combination has been intensively studied by [21]. The advantage of this linear approximation is that the parameter α can be treated as a nuisance parameter, leading to some improvement in computational efficiency. Then, the estimating function for the subject i in the k th pattern with the r th imputation is defined as:

$$\tilde{\mathbf{g}}_{k,i(r)}(\boldsymbol{\beta}) = \sum_{j=1}^J a_{k,j} \left\{ \frac{\partial \mu_{k,i}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{O(k)}} \right\}^T A_{k,i}^{-1/2} M_j A_{k,i}^{-1/2} \{Y_{k,i} - \mu_{k,i(r)}(\boldsymbol{\beta})\}.$$

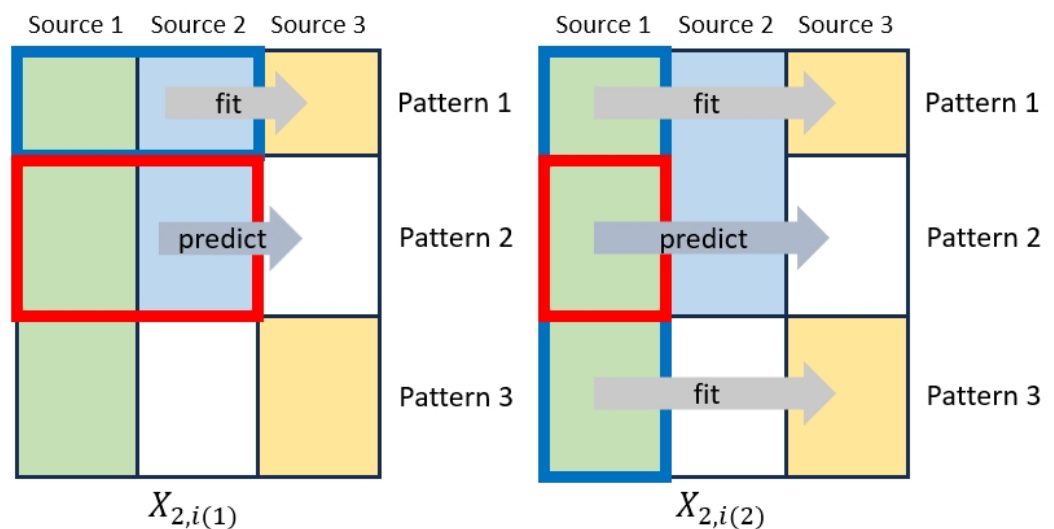


Figure 2. Two imputation approaches for missing covariates of source 3 in pattern 2. In the left figure, samples from pattern 1 and covariates in source 1 and source 2 are used to train the model, which is subsequently used to predict the missing covariates in pattern 2. Similarly, in the right figure, samples from pattern 1 and pattern 3 and covariates in source 1 are used to train the model.

Here, we only take derivative with respect to $\boldsymbol{\beta}_{O(k)}$ to enhance numerical stability. Recall that $a_{k,j}$ is a linear coefficient that used to approximate the inverse of correlation matrix, and thus, it is the nuisance parameter. To avoid estimating these nuisance parameters, we define the extended score vector:

$$\mathbf{g}_{k,i(r)}(\boldsymbol{\beta}) = \begin{pmatrix} \left\{ \frac{\partial \mu_{k,i}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{O(k)}} \right\}^T A_{k,i}^{-1/2} M_1 A_{k,i}^{-1/2} \{Y_{k,i} - \mu_{k,i(r)}(\boldsymbol{\beta})\} \\ \vdots \\ \left\{ \frac{\partial \mu_{k,i}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_{O(k)}} \right\}^T A_{k,i}^{-1/2} M_J A_{k,i}^{-1/2} \{Y_{k,i} - \mu_{k,i(r)}(\boldsymbol{\beta})\} \end{pmatrix}.$$

Similarly, we obtain extended score vectors for all imputed covariate vectors and subjects. To integrate all score vectors, we aggregate the information by stacking them into a long vector:

$$\mathbf{g}(\boldsymbol{\beta}) = \begin{pmatrix} \mathbf{g}_1(\boldsymbol{\beta}) \\ \vdots \\ \mathbf{g}_K(\boldsymbol{\beta}) \end{pmatrix} = \begin{pmatrix} \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{g}_{1,i}(\boldsymbol{\beta}) \\ \vdots \\ \frac{1}{n_K} \sum_{i=1}^{n_K} \mathbf{g}_{K,i}(\boldsymbol{\beta}) \end{pmatrix},$$

where $\mathbf{g}_{k,i}(\boldsymbol{\beta}) = (\mathbf{g}_{k,i(1)}^T(\boldsymbol{\beta}), \dots, \mathbf{g}_{k,i(R_k)}^T(\boldsymbol{\beta}))^T$. Note that this might be an overdetermined system because the number of equations can exceed the number of parameters. To overcome

this difficulty, we adopt generalized method of moment [22] and add a penalty term. Therefore, the objective function becomes:

$$S(\beta) = \mathbf{g}(\beta)^T C(\beta)^{-1} \mathbf{g}(\beta) + \sum_{j=1}^p p_{\lambda_n}(|\beta_j|), \tag{1}$$

where:

$$C(\beta) = \text{diag} \left\{ \frac{1}{n_1} \sum_{i=1}^{n_1} \mathbf{g}_{1,i}(\beta) \mathbf{g}_{1,i}^T(\beta), \dots, \frac{1}{n_K} \sum_{i=1}^{n_K} \mathbf{g}_{K,i}(\beta) \mathbf{g}_{K,i}^T(\beta) \right\}$$

is a block-diagonal matrix under the assumption of independence among samples from different missing patterns and $p_{\lambda_n}(\cdot)$ is an arbitrary, investigator’s chosen, penalty function with a tuning parameter λ . Among many optional penalty functions, we adopt the non-convex smoothly clipped absolute deviation (SCAD) penalty [23]:

$$p_{\lambda_n}(|\beta|) = \lambda|\beta|I(|\beta| \leq \lambda) + \frac{2a\lambda|\beta| - \beta^2 - \lambda^2}{2(a-1)}I(\lambda < |\beta| \leq a\lambda) + \frac{\lambda^2(a+1)}{2}I(a\lambda < |\beta|)$$

for some $a > 2$, which possess desirable oracle property.

3. Asymptotic Properties

In this section, we investigate the asymptotic properties of the proposed estimator. In Section 3.1, we assume the sample size n is increasing while the number of parameters p is fixed, and demonstrate that the proposed estimator is \sqrt{n} -consistent and asymptotically normal. As sample size goes to infinity, the proposed method is capable of selecting out the relevant variables with probability goes to 1. We also show that the proposed estimator is asymptotically more efficient than single imputation method via incorporating information of samples with missing values. In Section 3.2, we suppose both the sample size n and the number of parameters p are increasing but n increases faster than p . We show that the consistency and sparsity still hold with diverging p . Without loss of generality, we assume $\hat{\beta}$ can be partitioned into two parts, i.e., $\hat{\beta} = (\hat{\beta}_{\mathcal{A}}^T, \hat{\beta}_{\mathcal{N}}^T)^T$, where $\hat{\beta}_{\mathcal{A}}$ corresponds to relevant variables with a non-zero true value, while $\hat{\beta}_{\mathcal{N}}$ consists of coefficients of irrelevant variables with a zero true value. For any function $g(\beta)$, we use $\dot{g}(\beta)$ to denote the first derivative of $g(\cdot)$ evaluated at β . We use similar notation for its other order derivatives.

3.1. Fixed Number of Parameters

To establish the asymptotic properties of the proposed estimator in the setting of increasing sample sizes and fixed number of parameter, we require the following regularity conditions:

- C.1 $E[X_{k,j}]^4 < \infty$ and $E[E[X_{k,j(r)}]]^4 < \infty$, for any $1 \leq k \leq K, 1 \leq j \leq p$, and $1 \leq r$, where the inner expectation is with respect to the imputed values.
- C.2 All the variance matrix $A_{k,i} \geq 0$ and $\|A_{k,i}\| < \infty$, for any $1 \leq k \leq K$ and $1 \leq i \leq n_k$.
- C.3 Let $\epsilon_{k,i} = A_{k,i}^{-1/2}(\mathbf{Y}_{k,i} - \mu_{k,i}(\beta_0))$. For any $1 \leq k \leq K$ and $1 \leq i \leq n_k$, $E(\epsilon_{k,i}) = 0$ and the fourth moment of $\epsilon_{k,i}$ exists.
- C.4 $\|\mu_{k,i}(\beta_0) - \mu_{k,i(r)}(\beta_0)\| = o_p(n_k^{-1/2})$, for any $1 \leq k \leq K$ and $1 \leq i \leq n_k$.
- C.5 The penalty function satisfied:
 - (a) $\liminf_{n \rightarrow \infty} \inf_{\beta_j \rightarrow 0^+} p'_{\lambda_n}(\beta_j) / \lambda_n > 0$;
 - (b) $\max_{j \in \mathcal{A}} \{p'_{\lambda_n}(\beta_{0j})\} = o_p(n^{-1/2})$;
 - (c) $\max_{j \in \mathcal{A}} \{p''_{\lambda_n}(\beta_{0j})\} = o_p(1)$.
- C.6 $\sqrt{n}\mathbf{g}(\beta_0) \xrightarrow{d} N(\mathbf{0}, \Sigma\Omega)$, where $\Sigma = \text{diag}\{\Sigma_1, \dots, \Sigma_K\}$ and $\Omega = \text{diag}\{\Omega_1, \dots, \Omega_K\}$, with $\Sigma_k = \text{cov}(\mathbf{g}_{k,i}(\beta_0))$ and Ω_k to be a diagonal matrix with n_k dimension and each element equals to $\lim_{n \rightarrow \infty} n/n_k$.

C.1–C.3 are conditions that require the existence of the moment, which are easily satisfied. C.4 requires the imputed conditional mean converges to the true conditional mean in probability, which is satisfied as long as the imputed model is correctly specified and the missing mechanism is either missing completely at random. C.5 is a standard condition for SCAD penalty which is commonly used in variable selection method (Gao et al. [24]; Cho and Qu [25]; Tian et al. [26]). More specifically, (a) ensures the property of sparsity is satisfied, (b) and (c) ensure the property of consistency is satisfied, and (c) also guarantees that the objective function (1) is dominated by the first term. C.6 is used to establish the asymptotic normality of the proposed estimator.

Theorem 1. Under C.1–C.5, there exists a local minimizer $\hat{\beta}$ of $S(\beta)$ such that $\|\hat{\beta} - \beta_0\| = O_p(n^{-1/2})$.

Theorem 1 states the existence of a minimizer of the objective function and the minimizer will converge to the true coefficients at a rate of \sqrt{n} as the sample size increases. Next, we demonstrate that this estimator possesses the sparsity property and the estimator for the non-zero coefficient is asymptotically normal, as outlined in the following theorem.

Theorem 2. Under C.1–C.5, if $\lambda_n \rightarrow 0$ and there exist a sequence such that $\lambda_n \sqrt{n}/a_n \rightarrow \infty$ as $n \rightarrow \infty$, where $a_n = o_p(\sqrt{n})$, then the proposed estimator $\hat{\beta} = (\hat{\beta}_A^T, \hat{\beta}_N^T)^T$ satisfies the following properties:

1. (Sparsity) $P(\hat{\beta}_N = \mathbf{0}) \rightarrow 1$;
2. (Asymptotic Normality) Let $\mathbf{H} = E[\partial \mathbf{g}^T(\beta_0)/\partial \beta_A]$ and $\mathbf{V} = (\mathbf{H}\Sigma^{-1}\Omega^{-1}\mathbf{H}^T)^{-1}$ and if C.6 holds, then $\sqrt{n}(\hat{\beta}_A - \beta_{0,A}) \xrightarrow{d} N(\mathbf{0}, \mathbf{V})$.

The sparsity of the proposed estimator guarantees that the probability of selecting the true model approaches 1. We also obtained in Theorem 2 the asymptotic normality of $\hat{\beta}_A$, the estimator of coefficients for the relevant variables, which allows us to estimate its variance if \mathbf{H} and Σ are known. However, in practice, these are unknown to us. We can obtain the empirical variance covariance matrix of $\hat{\beta}_A$ by replacing \mathbf{H} with $\hat{\mathbf{H}}(\hat{\beta}) = \partial \mathbf{g}^T(\hat{\beta})/\partial \beta_A$ and replacing Σ with $\mathbf{C}(\hat{\beta})$, i.e., $\hat{\mathbf{V}} = (\hat{\mathbf{H}}\mathbf{C}^{-1}\Omega^{-1}\hat{\mathbf{H}}^T)^{-1}$. Next, we compare the empirical variance of the proposed estimator with the empirical variance of the single imputation approach.

Theorem 3. If a single imputation is used based on complete cases and denotes the asymptotic covariance matrix of β_A as $\tilde{\mathbf{V}}$, then under the conditions of Theorem 2, $\tilde{\mathbf{V}} - \mathbf{V}$ is positive semi-definite.

Theorem 3 claims that the proposed estimator is asymptotically more efficient than the single imputation approach, as it incorporates information from incomplete cases during imputation. The result of this Theorem is intuitive because the proposed method incorporates more samples into the imputation process.

3.2. Diverging Number of Parameters

In this subsection, we consider the setting where sample size n and number of coefficients p increase simultaneously. For certain properties to remain true, we require that n increases faster than p . We replace the notation p by p_n to indicate that the number of parameters also increases. We make the following assumptions:

D.1 For any i, j, k , $\dot{Q}_k(\beta_0) = o_p(p_n^{1/2}n^{-1/2})$ and:

$$\left\| \frac{\partial^2 Q_k(\beta_0)}{\partial \beta_i \partial \beta_j} - E \left\{ \frac{\partial^2 Q_k(\beta_0)}{\partial \beta_i \partial \beta_j} \right\} \right\| = O_p(n^{-1/2}).$$

D.2 There exist an open ball of β_0 and there exist a constant M such that each entries of $\ddot{Q}_k(\beta)$ is bounded by M , for any β in this open ball.

D.3 The penalty function satisfied:

- (a) $\liminf_{n \rightarrow \infty} \inf_{\beta_j \rightarrow 0^+} p'_{\lambda_n}(\beta_j) / \lambda_n > 0$;
- (b) $\max_{j \in \mathcal{A}} \{p'_{\lambda_n}(\beta_{0j})\} = o_p(p_n^{1/2} n^{-1/2})$;
- (c) $\max_{j \in \mathcal{A}} \{p''_{\lambda_n}(\beta_{0j})\} = o_p(p_n^{1/2})$.

D.1 and D.2 are analogous to C.1–C.4. D.3 is the modification of C.5 for diverging number of parameters.

Theorem 4. Under D.1–D.3, if $p_n = o(n^{1/4})$, there exists a local minimizer $\hat{\beta}$ of $S(\beta)$ such that $\|\hat{\beta} - \beta_0\| = O_p(p_n^{1/2} n^{-1/2})$.

From the result of Theorem 4, we find that the consistency still holds for the proposed estimator, even with a diverging number of parameters. Not surprisingly, the convergence rate is no longer \sqrt{n} , but $\sqrt{n/p_n}$. We also require that p_n does not increase faster than $n^{1/4}$ to ensure the model remains sparse. To be specific, the majority of the coefficients is zero.

Theorem 5. Under D.1–D.3, if $p_n = o(n^{1/4})$, $\lambda_n \rightarrow 0$, and $\lambda_n \sqrt{n/p_n} \rightarrow \infty$ as $n \rightarrow \infty$, then with probability tending to 1, the estimator $\hat{\beta} = (\hat{\beta}_{\mathcal{A}}^T, \hat{\beta}_{\mathcal{N}}^T)^T$ satisfies $P(\hat{\beta}_{\mathcal{N}} = \mathbf{0}) \rightarrow 1$.

Theorem 5 states the sparsity of the proposed estimator with a diverging number of parameters. This property guarantees that the proposed method can still select the true model with a probability approaching 1, even when the number of parameters is diverging.

4. Implementation

Since directly minimizing the objective function is difficult, we incorporate an iterative procedure inspired by the implementation in [27], where they combined the minorization–maximization algorithm [28] with the Newton–Raphson algorithm. Given the current estimate of $\beta^{(t)}$ and tuning parameter λ_n , the objective function $S(\beta)$ can be locally approximated by (except a constant term):

$$Q(\beta^{(t)}) + \dot{Q}(\beta^{(t)})^T(\beta - \beta^{(t)}) + \frac{1}{2}(\beta - \beta^{(t)})^T \ddot{Q}(\beta^{(t)})^T(\beta - \beta^{(t)}) + \frac{1}{2} \beta^T D_{\lambda_n}(\beta^{(t)}) \beta, \tag{2}$$

where:

$$D_{\lambda_n}(\beta^{(t)}) = \text{diag} \left\{ \frac{p'_{\lambda_n}(|\beta_1^{(t)}|)}{\epsilon + |\beta_1^{(t)}|}, \dots, \frac{p'_{\lambda_n}(|\beta_p^{(t)}|)}{\epsilon + |\beta_p^{(t)}|} \right\}$$

and ϵ is a sufficiently small number (e.g., $\epsilon = 10^{-6}$). Thus, the search for estimator minimizing the objective function is equivalent to find an estimator that minimize (2). Notice that both $\ddot{Q}(\beta^{(t)})$ and $\ddot{Q}(\beta^{(t)})$ are unknown. Fortunately, from Lemma S2 in Supplementary Materials, $\dot{Q}(\beta^{(t)})$ can be approximated by:

$$M(\beta^{(t)}) = 2\dot{\mathbf{g}}^T(\beta^{(t)})C(\beta^{(t)})^{-1}\dot{\mathbf{g}}(\beta^{(t)}) \tag{3}$$

and $\ddot{Q}(\beta^{(t)})$ can be approximated by:

$$F(\beta^{(t)}) = 2\dot{\mathbf{g}}^T(\beta^{(t)})C(\beta^{(t)})^{-1}\dot{\mathbf{g}}(\beta^{(t)}). \tag{4}$$

Plugging (3) and (4) into (2) and applying the Newton–Raphson algorithm, we obtain the following formula to update $\beta^{(t+1)}$:

$$\beta^{(t+1)} = \beta^{(t)} - [F(\beta^{(t)}) + D_{\lambda_n}(\beta^{(t)})]^{-1} [M(\beta^{(t)}) + D_{\lambda_n}(\beta^{(t)})\beta^{(t)}].$$

We repeat the above procedure until $\|\beta^{(t+1)} - \beta^{(t)}\|$ is smaller than a pre-specified threshold or reach a pre-specified maximum number of iteration.

It is known that the sampling covariance matrix $C(\beta)$ may be singular in some cases [29]. To overcome the difficulty in computing the inverse of $C(\beta)$, we adopt the Moore–Penrose generalized inverse, which exists and is unique for any matrix.

In the implementation of the proposed method, we select tuning parameter λ_n with extended Bayesian Information Criterion (EBIC) criteria proposed by [30]:

$$EBIC_\gamma = n \log(RSS/n) + df_{\lambda_n} \{ \log(n) + 2\gamma \log(p) \}, \quad 0 \leq \gamma \leq 1,$$

where df_{λ_n} is the number of parameters of the model with tuning parameter λ_n and $RSS = \sum_{k=1}^K RSS_k$ is the residual sum of square of all the missing pattern with:

$$RSS_k = \frac{1}{R_k} \sum_{r=1}^{R_k} \sum_{i=1}^{n_k} \sum_{j=1}^m \{ y_{k,ij} - \mu_{k,ij(r)} \}^2.$$

5. Simulation

In this section, we implement a simulation study to compare the performance of the proposed method in variable selection against complete case analysis (CC) with SCAD penalty, single imputation (SI) with SCAD penalty, and the penalized generalized estimating Equation (PGEE). We use the same data structure as shown in Figure 1, where we have three missing patterns and three sources. The number of measurement is set to be three throughout this section. We replicate the simulation 100 times and use false positive rate (FPR) and false negative rate (FNR) to evaluate the performance of each method, which reflect the proportion of covariates that are irrelevant but falsely selected and the proportion of covariates that are relevant but fail to be selected, respectively. In the tuning parameter selection procedure, the parameter γ was set to 0.5 in EBIC. At the end of the iterative algorithm in Section 4, the estimated coefficient is considered as zero, if its absolute value is smaller than 0.01.

In the first setting, we simulate a dataset with a small proportion of complete cases, where $n_1 = 40$, $n_2 = 120$, $n_3 = 120$, and the missing rate is around 87%. The data with continuous outcome are generated from the model:

$$Y_{ij} = \mathbf{X}_{ij}^T \boldsymbol{\beta} + \varepsilon_i,$$

where $j = 1, \dots, 3$, $\mathbf{X}_{ij} = (x_{ij,1}, \dots, x_{ij,30})^T$ is a vector consisting of 30 covariates, and $\boldsymbol{\beta} = (1, 2, 0, \dots, 0, 1, 2, 0, \dots, 0, 1, 2, 0, \dots, 0)^T$. Here, each source consists of 10 covariates with the first two covariates having non-zero coefficients. $x_{ij,1}$ is a time-fixed covariate and we generate it from the standard normal distribution, whereas other covariates are time-varying covariates and follow multivariate normal distribution with mean zero and exchangeable covariance matrix with marginal variance 1 and correlation coefficient 0.5. We generate random error ε_i from the multivariate normal distribution with mean 0 and exchangeable covariance matrix with marginal variance 1 and correlation coefficient ρ . We always assume the true within-cluster correlation structure is known and considered ρ to be 0.3, 0.5, and 0.7 in each setting, which corresponded to mild, moderate, and strong within-cluster correlation. Let $\phi_i = 1 / (1 + \exp\{1 + x_{i1,1} + \dots + x_{i1,10}\})$. Then, n_1 , n_2 , and n_3 samples were sequentially drawn with probability proportional to the ϕ_i and assigned to the pattern 1, pattern 2, and pattern 3, respectively. Obviously, subjects with higher covariates value from source 1 at the baseline are more likely to be assigned to pattern 1,

followed by pattern 2 and then pattern 3. This data generating process implies a MAR mechanism for the missing covariates. The results of Table 1 summarize the performance of each method for three different ρ . All of these methods effectively control the FNR. However, FPR of the proposed method is lower than the other three methods. In other words, the proposed method is able to select most of relevant variables while controlling the error of selecting irrelevant variables. In addition, we notice that the proposed method is more capable of utilizing within-cluster correlation compared with PGEE since the proposed method performs better as the within-cluster correlation becomes stronger. This result demonstrates the superiority of the proposed method when the percentage of complete cases is small in the block-missing data.

Table 1. Simulation scenario 1 with continuous outcomes: comparing the proposed method, complete cases analysis, single imputation method, and PGEE in terms of false positive rate (FPR), false negative rate (FNR), FPR + FNR, and computation time in seconds ($n_1 = 40, n_2 = 100, n_3 = 100, p_1 = 10, p_2 = 10, p_3 = 10$).

	Method	FPR	FNR	FPR + FNR	Time
$\rho = 0.3$	Proposed	0.083	<0.001	0.083	2.38
	CC	0.204	<0.001	0.204	0.26
	SI	0.118	0.002	0.120	1.22
	PGEE	0.085	<0.001	0.085	0.62
$\rho = 0.5$	Proposed	0.093	0.007	0.100	2.42
	CC	0.205	<0.001	0.205	0.27
	SI	0.146	<0.001	0.146	1.29
	PGEE	0.126	0.007	0.133	0.65
$\rho = 0.7$	Proposed	0.110	<0.001	0.110	2.50
	CC	0.198	0.005	0.203	0.28
	SI	0.141	<0.001	0.141	1.33
	PGEE	0.132	0.017	0.149	0.67

In the second setting, we continue to investigate the proposed method’s performance with a continuous outcome, but we proportionally increase the sample size in each missing pattern to demonstrate the proposed method’s effectiveness in larger samples, where $n_1 = 120, n_2 = 300, n_3 = 300$. The results are described in Table 2. Unsurprisingly, the FPR and FNR of all the methods decreased compared with the first setting. We observe that the performance of the PGEE is very close to that of the single imputation method while the proposed method has a much lower FPR. In the meanwhile, complete cases analysis is still the worst option since the improvement is minor as the sample size increase, and even negligible when the within-cluster correlation is strong. Therefore, the proposed method is still able to maintain an appealing performance in the large sample size. The results from this setting further verify the efficiency gain of the proposed method in incorporating more information from the missing data compared to the single imputation.

Table 2. Simulation scenario 2 with continuous outcomes: comparing the proposed method, complete cases analysis, single imputation method, and PGEE in terms of false positive rate (FPR), false negative rate (FNR), FPR + FNR, and computation time in seconds ($n_1 = 120, n_2 = 300, n_3 = 300, p_1 = 10, p_2 = 10, p_3 = 10$).

	Method	FPR	FNR	FPR + FNR	Time
$\rho = 0.3$	Proposed	0.003	<0.001	0.003	4.31
	CC	0.101	<0.001	0.101	0.58
	SI	0.018	<0.001	0.018	2.55
	PGEE	0.010	<0.001	0.010	1.55
$\rho = 0.5$	Proposed	0.005	<0.001	0.005	4.37
	CC	0.135	<0.001	0.135	0.61
	SI	0.025	<0.001	0.025	2.55
	PGEE	0.023	<0.001	0.023	1.52

Table 2. Cont.

	Method	FPR	FNR	FPR + FNR	Time
$\rho = 0.7$	Proposed	0.015	<0.001	0.015	4.29
	CC	0.190	<0.001	0.190	0.54
	SI	0.049	<0.001	0.049	2.47
	PGEE	0.078	<0.001	0.078	1.37

In the third setting, we consider a correlated binary outcome with $n_1 = 120, n_2 = 300,$ and $n_3 = 300$. The data are generated from the model:

$$\log \frac{\pi_{ij}}{1 - \pi_{ij}} = \mathbf{X}_{ij}^T \boldsymbol{\beta} + \varepsilon_i,$$

where $j = 1, \dots, m, \mathbf{X}_{ij} = (x_{ij,1}, \dots, x_{ij,15})^T$ is a vector consisting of 15 covariates, and $\boldsymbol{\beta} = (1, 0, \dots, 0, -0.7, 0, \dots, 0, 0.5, 0, \dots, 0)^T$. Here, each source consists of five covariates, with the first covariate in each having non-zero coefficients. $x_{ij,1}$ is a time-fixed covariate and we generate it from the standard normal distribution, whereas other covariates are time-varying covariates and follow multivariate normal distribution with mean zero and exchangeable covariance matrix with marginal variance 1 and correlation coefficient 0.5. We generate random error ε_i from the multivariate normal distribution with mean 0 and exchangeable covariance matrix with marginal variance 1 and correlation coefficient 0.3. In this setting, $\phi_i = 1 / (1 + \exp\{1 + x_{i1,11} + \dots + x_{i1,15}\})$. The results are summarized in Table 3. Although the PGEE outperforms other methods in terms of FPR, its performance in FNR is poor. In contrast, the proposed method possesses a better balance between FPR and FNR. We still observed a better performance of the proposed method.

Table 3. Simulation scenario 3 with binary outcomes: comparison of the proposed method, complete cases analysis, single imputation method, and PGEE in terms of false positive rate (FPR), false negative rate (FNR), FPR + FNR, and computation time in seconds ($n_1 = 120, n_2 = 300, n_3 = 300, p_1 = 5, p_2 = 5, p_3 = 5, \rho = 0.3$).

Method	FPR	FNR	FPR + FNR	Time
Proposed	0.298	0.063	0.361	3.55
CC	0.334	0.218	0.552	0.32
SI	0.289	0.088	0.377	1.91
PGEE	0.071	0.537	0.608	0.74

6. Application

We apply our proposed method to the ADNI study. This study was launched in 2003 and has undertaken three different phases so far: ADNI 1, ADNI GO/2, and ADNI 3, which is designed to develop the effective treatment that can slow or stop the progression of AD. Our goal is to identify sensitive biomarkers of AD in the early stage from three data sources: magnetic resonance imaging (MRI), positron emission tomography (PET), and cerebrospinal fluid (CSF). We choose the mini-mental state examination (MMSE) [31] score as response variable, which has been widely used in the early diagnosis of AD [32]. The MRI data were analyzed by UCSF, who performed cortical reconstruction and volumetric segmentation with FreeSurfer. The processed MRI data primarily summarized average cortical thickness, standard deviation in cortical thickness, the volumes of cortical parcellations, the volumes of specific white matter parcellations, and the total surface area of the cortex [33]. The PET data were processed by UCB and quantities variables were obtained by standard uptake value ratio (SUVR) in amyloid florbetapir. The CSF data were acquired by ADNI Biomarker Core and Laboratory Medicine and Center for Neurodegenerative Diseases Research at UPENN. The block-wise missing emerged in this data. Less than half of patients lacked MRI measurements, few patients missed PET measurements, and only a small proportion

of patients had CSF measurements. One of the reasons for the block-wise missing data is that obtaining CSF measurements requires more invasive procedures (such as lumbar puncture), which are refused by the majority of patients. The goal of this analysis is to identify biomarkers that are highly predictive of MMSE.

We only use the ADNI GO/2 dataset and consider measurements at baseline, month 24, and month 48, since the majority of patients have records at these time points. We also notice that there exist some low-quality data, such as those missed baseline measurement or belonged to a missing pattern with few patients. For simplicity of analysis, we discard these low-quality data, which leads us to a study cohort of 669 patients. Among them, 280 patients missed the measurement at month 24 and 487 patients missed the measurements at month 48. There are 340 features in MRI data, 229 features in PET data, and 3 features in CSF data. These three datasets and MMSE data are joined by a unique identifier “visit code” provided by the ADNI study. In total, we have three missing patterns. Table 4 describes the missing pattern of this dataset. The number of patients with fully observed variables is 63, with a missing rate around 90.6%. From this extremely high proportion of missing data, we will see how the proposed method can substantially improve the prediction ability by incorporating the information of related samples with missing values. To assess the predictive performance of the proposed method, data are randomly split into a test data with a sample size 30 (roughly 5%) and the remaining data as training data, where the test data are drawn from the data with fully observed variables (missing pattern 1). This random split process was replicated 30 times. A variable is marked as a potential predictor of AD if its absolute coefficient value is greater than 0.01.

Table 5 summarizes the average number of biomarkers selected by each method, along with the most frequently selected biomarkers. We also report the post-model-selection p -value. Our method successfully identifies biomarkers that align with findings reported in existing Alzheimer’s Disease research literature. In comparison to PGEE, the other three methods consistently select amyloid- β as a biomarker of AD, whose accumulation in cells is an early event of AD [34]. Phosphorylated tau, another widely accepted biomarker, has been validated by multiple large-scale, multi-center studies [35]. Studies found that neurons in AD patients are more likely to loss the superior temporal sulcus [36]. Two distinct normalization methods of summary measures for the standardized uptake value ratio (SUVR) of the florbetapir tracer, in the composite reference region and the whole cerebellum reference region, may potentially serving as AD biomarkers [37]. Besides these biomarkers, the proposed method additionally identifies several well-established and potential biomarkers. The size of the region of interest (ROI) in the left and right hemisphere precuneus area of the cortex, as well as cortical volume of left precuneus, summarize the health status of precuneus, which may be atrophy in the early stage of AD. The size and volume of the ROI in the left and right inferior lateral ventricle reflect disease progression (Bartos et al. [38]; Song et al. [39]). White matter changes in cerebral or subcortical areas can appear in other neurological conditions and normal aging, their connections with AD potentially make them useful biomarkers for distinguishing AD from normality, especially when considered along with other biomarkers in future investigations. While the surface area of the left caudal middle frontal and the cortical volume of the right caudal anterior cingulate are both associated with AD, more research is required to further explore these associations.

Table 4. Data composition and missing pattern for the subset of ADNI data; “O” denotes the observed data and “-” denotes the missing data.

Missing Pattern	MRI (340)	PET (229)	CSF (3)	Number of Patients
1	O	O	O	63
2	O	O	-	384
3	-	O	-	222

Table 5. Comparison of the mean of the number of selected biomarkers (MNSB) whose absolute value of coefficient is greater than 0.01 based on 30 replications in application to ADNI data. Time is the computation time in seconds.

Method	MNSB	Top Selected Biomarkers	Time
Proposed	16	ABETA, PTAU, ST30SV *, ST15SA, ST89SV, ST151SV, ST52CV *, ST73CV, SUMMARYSUVR COMPOSITE REFNORM *, SUMMARYSUVR WHOLECEREBNORM *, CTX LH PRECUNEUS VOLUME, CTX RH PRECUNEUS SUVR, LEFT INF LAT VENT VOLUME, RIGHT INF LAT VENT VOLUME, CTX LH SUPERIORTEMPORAL SUVR *, LEFT CEREBRAL WHITE MATTER VOLUME	1550
CC	3	ABETA, TAU *, SUMMARYSUVR COMPOSITE REFNORM *	280
SI	9	ABETA *, TAU *, PTAU *, ST1SV, ST4SV *, ST52CV, SUMMARYSUVR COMPOSITE REFNORM, SUMMARYSUVR WHOLECEREBNORM, CC MID ANTERIOR VOLUME	1216
PGEE	1	ST52TA *	18

* Post-model-selection *p*-value < 0.05.

7. Discussion

It is well known that variable selection is a challenge for model robustness, estimator stability and efficiency, as well as precise predictability. However, another non-negligible issue when integrating longitudinal studies is missingness in the covariate, especially in block-wise missing data. Specifically, with block-wise missing data, the percentage of complete observations is relatively small while traditional statistical methods heavily rely on information of complete cases. In this paper, we develop new methods to extend the MBI approach in a longitudinal study under the setting of block-wise missing data. Under certain regularity conditions, the desirable properties, consistency, sparsity, and asymptotic normality still hold. In addition, the proposed method demonstrates superior efficiency compared to the single imputation approach. It is worth noting that dropout missing data are also very common in longitudinal studies, which typically cause bias in many cases. In future work, it will be of great interest to develop methods to handle dropout missingness and incorporate inverse probability weighting in the proposed method.

One limitation of this paper is that we assume a homogeneous missing pattern across measurements within a single patient. Although this assumption may be restrictive in real data analysis, it is not hard to fulfill in multi-source data.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/math12070951/s1>.

Author Contributions: Investigation, L.W. and Z.O.; methodology, Z.O.; software, Z.O.; formal analysis, Z.O.; data support, Alzheimer’s Disease Neuroimaging Initiative. All authors have read and agreed to the published version of the manuscript.

Funding: This research is partially supported by the NIH Grants P50-DA-054039-02, P30-ES-017885-10-A1, R01-ES-033515-01, and CDC Grant R01-CE-003497-01.

Data Availability Statement: Data are publicly available (<https://adni.loni.usc.edu>, accessed on 29 February 2024).

Acknowledgments: Data collection and sharing for this project was funded by the Alzheimer’s Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug

Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org), accessed on 29 February 2024. The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Mueller, S.G.; Weiner, M.W.; Thal, L.J.; Petersen, R.C.; Jack, C.; Jagust, W.; Trojanowski, J.Q.; Toga, A.W.; Beckett, L. The Alzheimer’s disease neuroimaging initiative. *Neuroimaging Clin.* **2005**, *15*, 869–877. [[CrossRef](#)] [[PubMed](#)]
- Zhu, H.; Li, G.; Lock, E.F. Generalized integrative principal component analysis for multi-type data with block-wise missing structure. *Biostatistics* **2020**, *21*, 302–318. [[CrossRef](#)]
- Lan, Q.; Jiang, S. A method of credit evaluation modeling based on block-wise missing data. *Appl. Intell.* **2021**, *51*, 6859–6880. [[CrossRef](#)]
- Li, J.; Zhang, Q.; Chen, S.; Fang, K. Weighted multiple blockwise imputation method for high-dimensional regression with blockwise missing data. *J. Stat. Comput. Simul.* **2023**, *93*, 459–474. [[CrossRef](#)]
- Sabbe, N.; Thas, O.; Ottoy, J.P. EMLasso: Logistic lasso with missing data. *Stat. Med.* **2013**, *32*, 3143–3157. [[CrossRef](#)]
- Bondarenko, I.; Raghunathan, T. Graphical and numerical diagnostic tools to assess suitability of multiple imputations and imputation models. *Stat. Med.* **2016**, *35*, 3007–3020. [[CrossRef](#)]
- Audigier, V.; Husson, F.; Josse, J. Multiple imputation for continuous variables using a Bayesian principal component analysis. *J. Stat. Comput. Simul.* **2016**, *86*, 2140–2156. [[CrossRef](#)]
- von Hippel, P.T.; Bartlett, J.W. Maximum likelihood multiple imputation: Faster imputations and consistent standard errors without posterior draws. *Stat. Sci.* **2021**, *36*, 400–420. [[CrossRef](#)]
- Chen, B.; Yi, G.Y.; Cook, R.J. Weighted generalized estimating functions for longitudinal response and covariate data that are missing at random. *J. Am. Stat. Assoc.* **2010**, *105*, 336–353. [[CrossRef](#)]
- Creemers, A.; Aerts, M.; Hens, N.; Molenberghs, G. A nonparametric approach to weighted estimating equations for regression analysis with missing covariates. *Comput. Stat. Data Anal.* **2012**, *56*, 100–113. [[CrossRef](#)]
- Zubizarreta, J.R. Stable weights that balance covariates for estimation with incomplete outcome data. *J. Am. Stat. Assoc.* **2015**, *110*, 910–922. [[CrossRef](#)]
- Hughes, R.A.; Heron, J.; Sterne, J.A.; Tilling, K. Accounting for missing data in statistical analyses: Multiple imputation is not always the answer. *Int. J. Epidemiol.* **2019**, *48*, 1294–1304. [[CrossRef](#)] [[PubMed](#)]
- Thung, K.H.; Wee, C.Y.; Yap, P.T.; Shen, D.; Initiative, A.D.N.; Alzheimer’s Disease Neuroimaging Initiative. Neurodegenerative disease diagnosis using incomplete multi-modality data via matrix shrinkage and completion. *NeuroImage* **2014**, *91*, 386–400. [[CrossRef](#)] [[PubMed](#)]
- Zhang, Y.; Tang, N.; Qu, A. Imputed factor regression for high-dimensional block-wise missing data. *Stat. Sin.* **2020**, *30*, 631–651. [[CrossRef](#)]
- Xue, F.; Qu, A. Integrating multisource block-wise missing data in model selection. *J. Am. Stat. Assoc.* **2021**, *116*, 1914–1927. [[CrossRef](#)]
- Yu, G.; Li, Q.; Shen, D.; Liu, Y. Optimal sparse linear prediction for block-missing multi-modality data without imputation. *J. Am. Stat. Assoc.* **2020**, *115*, 1406–1419. [[CrossRef](#)] [[PubMed](#)]
- He, Y.; Feng, Y.; Song, X. Variable selection for high-dimensional generalized linear model with block-missing data. *Scand. J. Stat.* **2023**, *50*, 1279–1297. [[CrossRef](#)]
- Rubin, D.B. Inference and missing data. *Biometrika* **1976**, *63*, 581–592. [[CrossRef](#)]
- Liang, K.Y.; Zeger, S.L. Longitudinal data analysis using generalized linear models. *Biometrika* **1986**, *73*, 13–22. [[CrossRef](#)]
- Qu, A.; Lindsay, B.G.; Li, B. Improving generalised estimating equations using quadratic inference functions. *Biometrika* **2000**, *87*, 823–836. [[CrossRef](#)]
- Anderson, T.W. Asymptotically Efficient Estimation of Covariance Matrices with Linear Structure *Ann. Stat.* **1973**, *1*, 135–141. Available online: <http://www.jstor.org/stable/2958164> (accessed on 29 February 2024).
- Hansen, L.P. Large sample properties of generalized method of moments estimators. *Econom. J. Econom. Soc.* **1982**, pp. 1029–1054. [[CrossRef](#)]

23. Fan, J.; Li, R. Variable selection via nonconcave penalized likelihood and its oracle properties. *J. Am. Stat. Assoc.* **2001**, *96*, 1348–1360. [[CrossRef](#)]
24. Gao, X.; Pu, D.Q.; Wu, Y.; Xu, H. Tuning parameter selection for penalized likelihood estimation of Gaussian graphical model. *Stat. Sin.* **2012**, *22*, 1123–1146.
25. Cho, H.; Qu, A. Model selection for correlated data with diverging number of parameters. *Stat. Sin.* **2013**, *23*, 901–927. [[CrossRef](#)]
26. Tian, R.; Xue, L.; Liu, C. Penalized quadratic inference functions for semiparametric varying coefficient partially linear models with longitudinal data. *J. Multivar. Anal.* **2014**, *132*, 94–110. [[CrossRef](#)]
27. Johnson, B.A.; Lin, D.; Zeng, D. Penalized estimating functions and variable selection in semiparametric regression models. *J. Am. Stat. Assoc.* **2008**, *103*, 672–680. [[CrossRef](#)]
28. Hunter, D.R.; Li, R. Variable selection using MM algorithms. *Ann. Stat.* **2005**, *33*, 1617. [[CrossRef](#)]
29. Hu, Y.; Song, P.X.K. Sample size determination for quadratic inference functions in longitudinal design with dichotomous outcomes. *Stat. Med.* **2012**, *31*, 787–800. [[CrossRef](#)] [[PubMed](#)]
30. Wang, T.; Zhu, L. Consistent tuning parameter selection in high dimensional sparse linear regression. *J. Multivar. Anal.* **2011**, *102*, 1141–1151. [[CrossRef](#)]
31. Folstein, M.F.; Folstein, S.E.; McHugh, P.R. “Mini-mental state”: A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **1975**, *12*, 189–198. [[CrossRef](#)] [[PubMed](#)]
32. Mendiondo, M.S.; Ashford, J.W.; Kryscio, R.J.; Schmitt, F.A. Modelling mini mental state examination changes in Alzheimer’s disease. *Stat. Med.* **2000**, *19*, 1607–1616. [[CrossRef](#)]
33. Yuan, L.; Wang, Y.; Thompson, P.M.; Narayan, V.A.; Ye, J.; Alzheimer’s Disease Neuroimaging Initiative. Multi-source feature learning for joint analysis of incomplete multiple heterogeneous neuroimaging data. *NeuroImage* **2012**, *61*, 622–632. [[CrossRef](#)] [[PubMed](#)]
34. LaFerla, F.M.; Green, K.N.; Oddo, S. Intracellular amyloid- β in Alzheimer’s disease. *Nat. Rev. Neurosci.* **2007**, *8*, 499–509. [[CrossRef](#)] [[PubMed](#)]
35. Hampel, H.; Blennow, K.; Shaw, L.M.; Hoessler, Y.C.; Zetterberg, H.; Trojanowski, J.Q. Total and phosphorylated tau protein as biological markers of Alzheimer’s disease. *Exp. Gerontol.* **2010**, *45*, 30–40. [[CrossRef](#)] [[PubMed](#)]
36. Gómez-Isla, T.; Hollister, R.; West, H.; Mui, S.; Growdon, J.H.; Petersen, R.C.; Parisi, J.E.; Hyman, B.T. Neuronal loss correlates with but exceeds neurofibrillary tangles in Alzheimer’s disease. *Ann. Neurol. Off. J. Am. Neurol. Assoc. Child Neurol. Soc.* **1997**, *41*, 17–24. [[CrossRef](#)] [[PubMed](#)]
37. Wong, D.F.; Rosenberg, P.B.; Zhou, Y.; Kumar, A.; Raymont, V.; Ravert, H.T.; Dannals, R.F.; Nandi, A.; Brašić, J.R.; Ye, W.; et al. In vivo imaging of amyloid deposition in Alzheimer disease using the radioligand 18F-AV-45 (flobetapir F 18). *J. Nucl. Med.* **2010**, *51*, 913–920. [[CrossRef](#)]
38. Bartos, A.; Gregus, D.; Ibrahim, I.; Tintëra, J. Brain volumes and their ratios in Alzheimer s disease on magnetic resonance imaging segmented using Freesurfer 6.0. *Psychiatry Res. Neuroimaging* **2019**, *287*, 70–74. [[CrossRef](#)]
39. Song, M.; Jung, H.; Lee, S.; Kim, D.; Ahn, M. Diagnostic classification and biomarker identification of Alzheimer’s disease with random forest algorithm. *Brain Sci.* **2021**, *11*, 453. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.