HHS Public Access

Author manuscript

*Stat Med*. Author manuscript; available in PMC 2018 February 01.

Published in final edited form as:

*Stat Med*. 2012 September 28; 31(22): 2414–2427. doi:10.1002/sim.4437.

# Sufficient Dimension Reduction for Longitudinally Measured Predictors

**Ruth M. Pfeiffer**[a,*], **Liliana Forzani**[b], and **Efstathia Bura**[c]

[a]Biostatistics Branch, National Cancer Institute, Bethesda, MD 20892-7244

[b]Instituto de Matemática Aplicada del Litoral and Facultad de Ingenieŕýa Qúýmica, CONICET and UNL, Güemes 3450, (3000) Santa Fe, Argentina

[c]Department of Statistics, George Washington University, Washington, DC 20052

## Abstract

We propose a method to combine several predictors (markers) that are measured repeatedly over time into a composite marker score without assuming a model and only requiring a mild condition on the predictor distribution. Assuming that the first and second moments of the predictors can be decomposed into a time and a marker component via a Kronecker product structure, that accommodates the longitudinal nature of the predictors, we develop first moment sufficient dimension reduction techniques to replace the original markers with linear transformations that contain sufficient information for the regression of the predictors on the outcome. These linear combinations can then be combined into a score that has better predictive performance than the score built under a general model that ignores the longitudinal structure of the data. Our methods can be applied to either continuous or categorical outcome measures. In simulations we focus on binary outcomes and show that our method outperforms existing alternatives using the AUC, the area under the receiver-operator characteristics (ROC) curve, as a summary measure of the discriminatory ability of a single continuous diagnostic marker for binary disease outcomes.

### Keywords

Discrimination; AUC; Kronecker product; Sliced Inverse Regression; SIR

## 1. Introduction

Much research effort is devoted to searching for predictors, also called markers, which may aid in diagnosis of disease or physical impairment. Ideally one would obtain a single marker with very high specificity and sensitivity. However, such high performance markers are yet to be found for many diseases. Strategies for combining information from multiple diagnostic predictors are needed, since a combination may provide a better tool for diagnosis or screening applications than any single marker on its own. In earlier work [1], we proposed an approach to combine several biomarkers into a composite marker score without assuming

[*]Correspondence to: Ruth M. Pfeiffer, Biostatistics Branch, National Cancer Institute, Bethesda, MD 20892-7244. pfeiffer@mail.nih.gov, USA.

a model for the distribution of the predictors. Using sufficient dimension reduction (SDR) techniques, we replaced the original markers with a lower-dimensional version, obtained through linear transformations of markers that contain sufficient information for regression of the predictors on the outcome. We focused on a first moment method, Sliced Inverse Regression (SIR) [2], and also a second moment method, Sliced Average Variance Estimation [3].

In this paper we extend first moment based SDR methods to combining several longitudinally measured biomarkers into a composite marker score under a mild distributional assumption and by exploiting their longitudinal structure. We assume that the means and the second moments of the markers can be separated into a marker and a time specific component via a Kronecker product structure. This substantially reduces the complexity of the first moment based dimension reduction subspace and results in better predictive accuracy of the resulting score compared to standard first moment based SDR methods.

Following the presentation of background material (Section 2), we present the new results (Section 3) for first moment based SDR methods for longitudinally measured predictors and an algorithm that extends SIR (Section 4) to that setting. We carry out simulations to assess and compare the performance of the longitudinal extension of SIR (Section 5) and give a data example in Section 6 before concluding with a discussion in Section 7.

## 2. Background

We start with a brief overview of standard sufficient dimension reduction methods, and a brief review of results for the estimation of the first moment based dimension reduction subspace, before considering the setting of longitudinally measured markers.

### 2.1. Sufficient dimension reduction (SDR)

Suppose we are interested in inferring the relationship between a response variable $Y$ and a covariate vector $\mathbf{X} = (X_1, \ldots, X_p)^T \in \mathbb{R}^p$. When the number of covariates $p$ is large, it is very difficult to visualize how $Y$ changes as a function of the components of $\mathbf{X}$ which makes modeling challenging. Dimension reduction aims to reduce the complexity of the regression or classification problem prior to model fitting. In particular, SDR [4] aims to find a function of $\mathbf{X}$, $\mathbf{Q} : \mathbb{R}^p \to \mathbb{R}^d$ with $d \leq p$, such that $\mathbf{Q}(\mathbf{X})$ contains the same information as $\mathbf{X}$ about the response $Y$. This means that $F(Y|\mathbf{X} = \mathbf{x}) = F(Y|\mathbf{Q}(\mathbf{X}) = \mathbf{Q}(\mathbf{x}))$, where $F(\cdot|\cdot)$ is the conditional distribution function of $Y$ given the second argument. This means no information about $Y$ is lost when $\mathbf{X}$ is replaced by $\mathbf{Q}(\mathbf{X})$. For this reason, $\mathbf{Q}(\mathbf{X})$ is called a sufficient reduction for the regression of $Y$ on $\mathbf{X}$.

While in principle $\mathbf{Q}$ can be any function, only linear sufficient transformations $\mathbf{Q}(\mathbf{X}) = \boldsymbol{\eta}^T\mathbf{X} = (\boldsymbol{\eta}_1^T\mathbf{X}, \ldots, \boldsymbol{\eta}_d^T\mathbf{X})$ with $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_d) \in \mathbb{R}^{p\times d}$ have been used in SDR methodology so far (e.g., [2], [3], [4], [5], [6]). One important fact about linear sufficient reductions is that if the columns of a matrix $\boldsymbol{\gamma}$ span the vector space spanned by the columns of $\boldsymbol{\eta}$, then $\boldsymbol{\eta}^T\mathbf{X}$ and $\boldsymbol{\gamma}^T\mathbf{X}$ contain the same information about $Y$ and therefore inference needs

to focus on the subspace $\mathcal{S}$ = span($\boldsymbol{\eta}$) spanned by the columns of $\boldsymbol{\eta}$ and not on the matrix $\boldsymbol{\eta}$ itself. We call such a subspace a SDR subspace.

Under mild conditions the intersection of all SDR subspaces, called the central dimension reduction subspace and denoted by $\mathcal{S}_{Y|X}$ ([4], Section 6.4, p. 108–112), is also sufficient. In particular, $\mathcal{S}_{Y|X}$ exists when the support of the distribution of **X** is convex. The dimension $d$ = dim($\mathcal{S}_{Y|X}$), referred to as the structural dimension of the regression of $Y$ on **X**, can take on any value in the set {0, 1, …, $p$}. When $d < p$, the structural dimension of the regression is smaller than the number of predictors and the complexity of the regression is reduced.

If $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_d)$ is a basis for $\mathcal{S}_{Y|X}$, then the $d$ linear combinations $\eta^T\mathbf{X}=(\eta_1^T\mathbf{X},\ldots,\eta_d^T\mathbf{X})$ contain all the information in **X** about **Y**. Moreover, the number of these linear combinations of **X**, $d$, is the smallest number of linear combinations of **X** with this property. In this sense, $\eta^T\mathbf{X}=(\eta_1^T\mathbf{X},\ldots,\eta_d^T\mathbf{X})$ is a "minimal" transformation of the covariate vector **X** for the regression of $Y$ on it. Thus the goal of SDR methodology is to estimate $d$ and a set of basis vectors $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_d)$ such that span($\boldsymbol{\eta}$) $\subseteq \mathcal{S}_{Y|X}$. A detailed exposition of SDR methodology is provided in [4].

### 2.2. The first moment based SDR subspace

If **Q**(**X**) is a sufficient reduction for the forward regression $Y|\mathbf{X}$, then it is also a sufficient statistic for the inverse regression $\mathbf{X}|Y$ [5]. This result implies that we can use the inverse regression of **X** on $Y$ in order to find sufficient reductions for the forward regression of $Y$ on **X**. The advantage of inverse regression is that a complex, $p$-dimensional multiple forward regression of $Y$ on **X** that cannot easily be visualized is translated to $p$ univariate inverse regressions of each component of **X** on $Y$. Thus, a $p$-dimensional problem is mapped to $p$ one-dimensional ones so that the difficulty of inference in high dimensions is removed. Because of this, most SDR methods are based on inverse regression.

In general, the estimation of $\mathcal{S}_{Y|X}$ is based on finding a matrix $\boldsymbol{\Omega}$, called kernel, so that span($\boldsymbol{\Omega}$) $\subseteq \mathcal{S}_{Y|X}$. First moment methods such as Sliced Inverse Regression (SIR) [2], Principal Fitted Components (PFC) [6] and Parametric Inverse Regression (PIR) [7] use kernel matrices computed from first moments of $\mathbf{X}|Y$, whereas second moment methods, such as Sliced Average Variance Estimation (SAVE) [3], use second moment based kernels. In this paper we focus on first moment methods, and in particular on extensions of SIR to longitudinal settings.

The first moment based SDR (FMSDR) subspace is the span of the centered mean of the inverse regression of **X** on $Y$, E(**X**|$Y$) − E(**X**), scaled by the inverse of the marginal covariance of **X**, $\boldsymbol{\Sigma}$. That is,

$$\mathcal{S}_{FMSDR}=\text{span}(\Sigma^{-1}[\text{E}(\mathbf{X}|Y)-\text{E}(\mathbf{X})]) \quad (1)$$

This subspace is the estimation target of SIR with kernel matrix $\mathbf{\Omega}_{SIR} = \mathbf{\Sigma}^{-1}\text{cov}(\text{E}(\text{vec}(\mathbf{X}|Y)))$.

If the predictors $\mathbf{X}$ satisfy the *linearity condition* ([4], p. 188), then $\mathcal{S}_{FMSDR}$ is contained in the central subspace,

$$\mathcal{S}_{FMSDR} \subseteq \mathcal{S}_{Y|X}. \quad (2)$$

This condition requires that there exists a linear sufficient transformation $\boldsymbol{\eta}^T\mathbf{X}$ such that $\text{E}(\mathbf{X}|\boldsymbol{\eta}^T\mathbf{X}) - \text{E}(\mathbf{X})$ is a linear function of $\boldsymbol{\eta}^T\mathbf{X}$ [8]. An important feature of this condition is that it is required only for the marginal distribution of the predictors. The linearity condition holds when $\mathbf{X}$ has an elliptical distribution [9], for example a multivariate normal distribution, but normality is much stronger than needed. Coordinate-wise transformations or re-weighting of the predictors can help achieve ellipticity of $\mathbf{X}$ [10]. Moreover, the linearity condition holds approximately when $p$ is large [11].

We show in the Appendix that for $\mathbf{\Delta}_Y = \text{cov}(\mathbf{X}|Y)$ and $\mathbf{\Delta} = \text{E}(\mathbf{\Delta}_Y)$, the Principal Fitted Components (PFC) subspace defined by $\mathcal{S}_{PFC} = \text{span}\{\mathbf{\Delta}^{-1}[\text{E}(\mathbf{X}|Y) - \text{E}(\mathbf{X})]\}$ also equals the FMSDR subspace under the linearity condition. That is,

$$\mathcal{S}_{FMSDR} = \text{span}(\textstyle\sum^{-1}[\text{E}(\mathbf{X}|Y) - \text{E}(\mathbf{X})]) = \text{span}(\mathbf{\Delta}^{-1}[\text{E}(\mathbf{X}|Y) - \text{E}(\mathbf{X})]), \quad (3)$$

When the marginal distribution of the predictors $\mathbf{X}$ satisfies the linearity condition, we can see from (2) and (3) that one can scale the conditional data, $\mathbf{X}|Y$, using either the marginal covariance matrix $\mathbf{\Sigma}$ or the conditional $\mathbf{\Delta}$ to obtain $\mathcal{S}_{FMSDR}$, and

$$\mathcal{S}_{FMSDR} = \mathcal{S}_{PFC} = \mathcal{S}_{SIR} \subseteq \mathcal{S}_{Y|X}.$$

For $\mathcal{S}_{FMSDR}$ to recover all of $\mathcal{S}_{Y|X}$, and not just part of it, i.e. to obtain $\mathcal{S}_{FMSDR} = \mathcal{S}_{Y|X}$, $\text{cov}(\boldsymbol{\eta}^T\mathbf{X}|Y)$ must be positive definite for a linear sufficient transformation $\boldsymbol{\eta}^T\mathbf{X}$ [12]. This is true, for example, when the predictors $\mathbf{X}|Y$ have a multivariate normal distribution and the conditional covariance of $\mathbf{X}$ does not depend on $Y$.

Under the linearity condition any kernel matrix whose column space spans the same space as $\mathcal{S}_{FMSDR}$ can be used to estimate $\mathcal{S}_{Y|X}$ or a part of it. Different kernel matrices define different estimation methods, such as SIR, PIR and PFC. Here we focus on SIR and its extension to longitudinal data.

# 3. Longitudinal first moment based SDR subspace

## 3.1. Notation and assumptions

We now estimate the first moment based dimension reduction subspace $\mathcal{S}_{FMSDR}$ for longitudinal data. Throughout this section we assume that the linearity condition holds for $\mathbf{X}$.

In the longitudinal setting the $p \times 1$ predictor vector $\mathbf{X}$ is measured over $T$ time points. To be specific, let $Y_i \in \mathbb{R}$ be a one dimensional response variable for the $i$th individual in the study, with $p$-dimensional covariate vector $\mathbf{X}_{it} = (x_{i1t}, \ldots, x_{ipt})^T \in \mathbb{R}^p$ that is measured at time points $t = 1, \ldots, T_i$, for $i = 1, \ldots, n$, where $n$ is the total sample size. For notational simplicity we assume that the markers are measured at the same time points for all individuals and that all individuals have the same number of observations over time, i.e. $t_{ij} = t_j$, and $T_i = T$. The predictor vector over time can be represented as the $p \times T$-matrix

$$\mathbf{X}_i = (\mathbf{X}_{i1}, \ldots, \mathbf{X}_{iT}) = \begin{bmatrix} X_{i11} & \cdots & X_{i1T} \\ X_{i21} & \cdots & X_{i2T} \\ \vdots & \vdots & \vdots \\ X_{ip1} & \vdots & X_{ipT} \end{bmatrix}, \quad (4)$$

that corresponds to the $pT \times 1$ column vector $\mathrm{vec}(\mathbf{X}_i)$ comprised of the columns of $\mathbf{X}_i$ stacked one after the other.

If one ignores the time structure of the data, standard dimension reduction can be applied to find $\boldsymbol{\eta} \in \mathbb{R}^{pT \times d}$ such that $F(Y|\mathrm{vec}(\mathbf{X})) = F(Y|\boldsymbol{\eta}^T \mathrm{vec}(\mathbf{X}))$. However, the time structure is integral to the nature of longitudinal data and ignoring it could result in loss of accuracy in estimation for practically relevant sample sizes.

To accommodate the longitudinal structure of $\mathbf{X}|Y$, we assume the first two moments of $\mathbf{X}$ can be decomposed into a time and a marker component. In particular, we let both the mean and the covariance of the markers be Kronecker products of the two components to further improve classification or regression, as follows.

**Assumption 1—**The conditional mean of $\mathbf{X}$ given $Y$ has the following Kronecker product structure:

$$\mathrm{vec}(\mathrm{E}(\mathbf{X}|Y) - \mathrm{E}(\mathbf{X})) = (\boldsymbol{\alpha} \otimes \boldsymbol{\beta})\mathrm{vec}(\boldsymbol{\nu}_y), \quad (5)$$

for some $d \times s$ matrix $\boldsymbol{\nu}_y$ with $\mathrm{E}(\boldsymbol{\nu}_y) = 0$ and $\det(\mathrm{cov}(\mathrm{vec}(\boldsymbol{\nu}_y))) > 0$. The matrix $\boldsymbol{\beta} \in \mathbb{R}^{T \times s}$ captures the mean structure over time, and $\boldsymbol{\alpha} \in \mathbb{R}^{p \times d}$ captures the mean structure of the markers regardless of time.

A practically important case is that of binary outcomes, i.e. $Y = 0, 1$. For binary $Y$ the condition (5) is satisfied if, for example, $\text{vec}(E(\mathbf{X}|Y)) = \boldsymbol{\alpha}_Y \otimes \boldsymbol{\beta}$, which implies that the means of the markers change over time only by a multiplicative factor that affects all markers equally and is the same for the two groups defined by $Y$; that is, $\text{vec}(E(\mathbf{X}_t|Y)) = \beta_t \boldsymbol{\alpha}_Y$. Letting $p_Y = P(Y = y)$ and using that $E(\mathbf{X}) = p_0\boldsymbol{\alpha}_0 \otimes \boldsymbol{\beta} + (1 - p_0)\boldsymbol{\alpha}_1 \otimes \boldsymbol{\beta}$, we get that $\text{vec}(E(\mathbf{X}|Y = y) - E(\mathbf{X})) = (1 - p_y)(\boldsymbol{\alpha}_0 - \boldsymbol{\alpha}_1) \otimes \boldsymbol{\beta}$, which means that the first order moment condition is satisfied with $\nu_Y = (1 - p_y)$.

Moreover, we assume that the second moments of $\mathbf{X}$, either unconditional or conditional on $Y$, have a Kronecker product structure. In the next two sections we study the second moment assumptions and show that under this structure our target estimation subspace $\mathcal{S}_{FMSDR}$ is itself spanned by Kronecker products of the marker and time components.

Adopting this moment structure accommodates the longitudinal nature of the data by focusing on the predictor and time aspect separately and results in a substantial reduction in the number of parameters to estimate. Kronecker product structure of moments has been used previously in linear discriminant analysis procedures for repeated measurements normally distributed data, see e.g. [13], [14].

### 3.2. $\Sigma = \text{cov}(vec(\mathbf{X})) = \Sigma_1 \otimes \Sigma_2$

When the longitudinal data arise from a prospective cohort, it may be reasonable to assume that the $(Tp) \times (Tp)$ covariance matrix of $\text{vec}(\mathbf{X})$ can be written as $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2$, where $\boldsymbol{\Sigma}_1$ captures the the covariance structure of the $p$ markers and $\boldsymbol{\Sigma}_2$ models the temporal association. Thus, $\text{cov}(X_{it}, X_{js}) = \sigma^{(1)}_{ij}\sigma^{(2)}_{ts}$. In terms of correlation, this Kronecker product structure implies that

$$\text{cor}(X_{it}, X_{js}) = \frac{\sigma^{(1)}_{ij}}{\sqrt{\sigma^{(1)}_{ii}\sigma^{(1)}_{jj}}} \frac{\sigma^{(2)}_{ts}}{\sqrt{\sigma^{(2)}_{tt}\sigma^{(2)}_{ss}}}.$$

For the same marker measured at two different time points $\text{cov}(X_{it}, X_{is}) = \sigma^{(1)}_{ii}\sigma^{(2)}_{ts}$, and for two markers measured at the same time point $\text{cov}(X_{it}, X_{jt}) = \sigma^{(1)}_{ij}\sigma^{(2)}_{tt}$. That means that the correlation of two markers measured at the same time point does not depend on time, i.e. $\text{cor}(X_{it}, X_{jt}) = \sigma^{(1)}_{ij} / (\sigma^{(1)}_{ii}\sigma^{(1)}_{jj})^{-1/2}$ and the correlation of the same marker measured at two different time points is the same for all markers, $\text{cor}(X_{it}, X_{is}) = \sigma^{(2)}_{st} / (\sigma^{(2)}_{tt}\sigma^{(2)}_{ss})^{-1/2}$. As $\boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2 = (c\boldsymbol{\Sigma}_1) \otimes (c^{-1}\boldsymbol{\Sigma}_2)$, the matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ are determined up to a multiplicative constant and we assume without loss of generality that $\sigma^{(2)}_{TT} = 1$.

In Theorem 1 below we show that $\mathcal{S}_{FMSDR}$ has a Kronecker structure induced by the Kronecker structure of $\boldsymbol{\Sigma}$ and the conditional mean of $\mathbf{X}$. The proof is given in the Appendix.

**Theorem 1—**Suppose Assumption 1 holds and suppose $\text{cov}(\text{vec}(\mathbf{X})) = \boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2$. Then

$$\mathcal{S}_{FMSDR}=\text{span}(\sum\nolimits_1^{-1}\boldsymbol{\alpha}\otimes\sum\nolimits_2^{-1}\boldsymbol{\beta}) \quad (6)$$

It follows immediately that

$$\mathcal{S}_{SIR}=\text{span}(\boldsymbol{\Omega}_{SIR})=\text{span}(\sum^{-1}\text{cov}(\text{vec}(\text{E}(\mathbf{X}|Y))))=\text{span}(\sum\nolimits_1^{-1}\boldsymbol{\alpha}\otimes\sum\nolimits_2^{-1}\boldsymbol{\beta}).$$

### 3.3. $\Delta = \text{E}[\text{cov}(vec(\mathbf{X}|Y))] = \Delta_1 \otimes \Delta_2$

Sometimes assuming that the overall population covariance $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2$ may not be reasonable, for example, when the data arise from a retrospective case-control sample. A slightly less restrictive assumption is to impose a Kronecker product structure on the conditional covariances of the predictors given the response $Y$. That is, $\text{cov}(\text{vec}(\mathbf{X})|Y=y)=\boldsymbol{\Delta}_1^y\otimes\boldsymbol{\Delta}_2^y$. When either $\boldsymbol{\Delta}_1^y$ or $\boldsymbol{\Delta}_2^y$ do not depend on $Y$ this structure implies $\boldsymbol{\Delta} = \text{E}(\text{var}(\mathbf{X}|Y)) = \boldsymbol{\Delta}_1 \otimes \boldsymbol{\Delta}_2$, as well. If the markers are sampled over the same time points in the groups defined by $Y$, the assumption that the correlations of the markers over time $\boldsymbol{\Delta}_1$ do not depend on $Y$ is reasonable, even though correlations of the markers differ for the groups defined by $Y$. If $\boldsymbol{\Delta} = \text{E}(\text{cov}(\mathbf{X}|Y)) = \boldsymbol{\Delta}_1 \otimes \boldsymbol{\Delta}_2$, the following theorem, whose proof is analogous to that of Theorem 1, holds.

**Theorem 2—**Suppose Assumption 1 holds and suppose $\boldsymbol{\Delta} = \text{E}(\text{cov}(\mathbf{X}|Y)) = \boldsymbol{\Delta}_1 \otimes \boldsymbol{\Delta}_2$. Then

$$\mathcal{S}_{PFC}=\text{span}(\boldsymbol{\Delta}_1^{-1}\boldsymbol{\alpha}\otimes\boldsymbol{\Delta}_2^{-1}\boldsymbol{\beta}). \quad (7)$$

Theorems 1 and 2 imply that when the conditional first moment $\text{vec}(\text{E}(\mathbf{X}|Y) - \text{E}(\mathbf{X}))$ and either $\text{cov}(\text{vec}(\mathbf{X}))$ or $\text{E}(\text{cov}(\text{vec}(\mathbf{X})|Y))$ have a Kronecker product structure, then the structure of $\mathcal{S}_{FMSDR}$ is also a Kronecker product. In particular, when the assumptions of Theorem 1 are satisfied

$$\mathcal{S}_{FMSDR}=\mathcal{S}_{PFC}=\mathcal{S}_{SIR}=\text{span}(\sum\nolimits_1^{-1}\boldsymbol{\alpha}\otimes\sum\nolimits_2^{-1}\boldsymbol{\beta})\subseteq\mathcal{S}_{Y|X}. \quad (8)$$

If the conditions of Theorem 2 are satisfied,

$$\mathcal{S}_{FMSDR}=\mathcal{S}_{PFC}=\mathcal{S}_{SIR}=\text{span}(\boldsymbol{\Delta}_1^{-1}\boldsymbol{\alpha}\otimes\boldsymbol{\Delta}_2^{-1}\boldsymbol{\beta})\subseteq\mathcal{S}_{Y|X}. \quad (9)$$

Based on the results in (8) and (9), the estimation of $\mathcal{S}_{FMSDR}$ requires estimating many fewer parameters compared to the setting when no structure is imposed.

**Remark:** Under the additional assumption of multivariate normality of the markers, likelihood ratio tests can be used to assess if Assumptions 1 and 2 hold [15], [16].

## 4. The longitudinal SIR (LSIR) algorithm

We estimate $\mathcal{S}_{FMSDR}$ using the SIR kernel matrix $\mathbf{\Omega}_{SIR} = \mathbf{\Sigma}^{-1}\text{cov}(\text{E}(\mathbf{X}|\, Y))$ based on repeated measurements of $\mathbf{X}$. In Section 3 we showed that under the conditions of either Theorem 1 or 2 the SIR subspace has a Kronecker product structure, that is $\mathcal{S}_{FMSDR} = \text{span}(\mathbf{\Omega}_x) = \text{span}(\mathbf{\Omega}_{1x} \otimes \mathbf{\Omega}_{2x})$, where $\mathbf{\Omega}_{1x}$, a $p \times p$ matrix, captures the reduction with respect to the predictors and $\mathbf{\Omega}_{2x}$, a $T \times T$ matrix, captures the reduction of the markers with respect to time.

In the implementation of the algorithm, the predictors are standardized for numerical stability. Depending on whether the conditions of Theorem 1 or 2 are satisfied, the predictors can be standardized (centered and scaled) using $\mathbf{\Sigma}$ or $\mathbf{\Delta}$, respectively, for their scaling (see (8) and (9)). However, it can be shown by direct calculation, as in the proof of Corollary 3.4 in [6], that if one were to replace $\bar{\mathbf{\Sigma}}_1$ and $\bar{\mathbf{\Sigma}}_2$ by $\bar{\mathbf{\Delta}}_1$ and $\bar{\mathbf{\Delta}}_2$, respectively, in scaling the predictors, $\hat{\mathcal{S}}_{FMSDR}$(= $\hat{\mathcal{S}}_{SIR}$) would remain the same. Hence, the algorithm uses the simpler $\bar{\mathbf{\Sigma}}_1$ and $\bar{\mathbf{\Sigma}}_2$ to scale the data.

The standardized predictors $\mathbf{Z} = \mathbf{\Sigma}^{-1/2}(\mathbf{X} - \text{E}(\mathbf{X}))$ are used to compute the kernel matrix $\mathbf{\Omega}_Z = \text{cov}(\text{E}(\mathbf{Z}|\mathbf{Y}))$. The latter relates to the kernel matrix for the original $\mathbf{X}$ predictors through $\mathbf{\Omega}_X = \mathbf{\Sigma}^{1/2}\mathbf{\Omega}_Z\mathbf{\Sigma}^{-1/2}$.

Suppose that a sample $(\mathbf{X}_i,\ Y_i)$, $i = 1, \ldots, n$ is available, where $Y_i$ denotes the outcome variable and $\mathbf{X}_i$ the $(p \times T)$ matrix of predictors for individual $i$ given in (4).

- The $p$ vector of covariate values for person $i$ at time $t$ is $\mathbf{X}_{it} = (x_{i1t}, \ldots, x_{ipt})^T$, and $\bar{\mathbf{x}}_{\cdot t} = (\bar{x}_{1t}, \ldots, \bar{x}_{pt})^T$ denotes the $p$-vector of predictor means over all subjects at time $t$. Similarly, $\mathbf{X}_{ij} = (x_{ij1}, \ldots, x_{ijT})^T$ denotes the $T$ vector of the $j$th covariate values across all $T$ time points for individual $i$, and let $\bar{\mathbf{x}}_{j\cdot}$ be the $T$-vector of means of the $j$-th predictor, $j = 1, \ldots, p$. Let $n_t$ be the number of observations at time $t$ across all $p$ predictors. Similarly, $n_j$ is the number of observations of the $j$-th predictor across all time points.
- We normalize the predictor matrix $\mathbf{X}_i$ for the ith individual: $\text{vec}(\mathbf{Z}_i) = \bar{\mathbf{\Sigma}}^{-1/2}(\text{vec}(\mathbf{X}_i) - \text{vec}(\bar{\mathbf{x}}))$, where each entry of $\bar{\mathbf{x}}$ is the empirical mean for each predictor at each time point over the $i$ observations. We find $\bar{\mathbf{\Sigma}} = \bar{\mathbf{\Sigma}}_1 \otimes \bar{\mathbf{\Sigma}}_2$ by computing

$$\hat{\sum}_{1t} = \text{cov}(\mathbf{x}_{\cdot t}) = \frac{1}{n_t}\sum_{i=1}^{n_t}(\mathbf{x}_{it} - \bar{\mathbf{x}}_{\cdot t})(\mathbf{x}_{it} - \bar{\mathbf{x}}_{\cdot t})^T \quad (10)$$

$$\hat{\sum}_1 = \frac{1}{T}\sum_{t=1}^{T}\hat{\sum}_{1t}, \quad (11)$$

and

$$\hat{\sum}_{2j}=\mathrm{cov}(\mathbf{x}_{j\cdot})=\frac{1}{n_j}\sum_{i=1}^{n_j}(\mathbf{x}_{ij}-\overline{\mathbf{x}}_{j\cdot})(\mathbf{x}_{ij}-\overline{\mathbf{x}}_{j\cdot})^T \quad (12)$$

$$\hat{\sum}_{2}=\frac{1}{p}\sum_{j=1}^{p}\hat{\sum}_{2j}. \quad (13)$$

- As in standard SIR, we divide the support of $Y$ into $H$ slices. For continuous $Y$ that means that $Y$ is replaced by a discrete version $\tilde{Y}$ based on partitioning the observed range of $Y$ into $H$ fixed, non-overlapping slices. For categorical $Y$ the slices are the categories of $Y$. Let $\overline{\mathbf{z}}_{\cdot t}^{(h)}$ denote the $p$-vector of the standardized predictor means and $\overline{\mathbf{z}}_{j\cdot}^{(h)}$ be the $T$-vector of the means of the $j$-th standardized predictor, $j = 1, \ldots, p$, within slice $h$, $h = 1, \ldots, H$. The proportion of observations in slice $h$ at time $t$ across all $p$ predictors is $f_{\cdot t}^{(h)}$ and $f_{j\cdot}^{(h)}$ is the proportion of observations in slice $h$ of the $j$-th predictor across all time points. Let

$$\hat{\Omega}_{1t}=\sum_{h=1}^{H}f_{\cdot t}^{(h)}\overline{\mathbf{z}}_{\cdot t}^{(h)}\overline{\mathbf{z}}_{\cdot t}^{(h)T} \quad (14)$$

$$\hat{\Omega}_{1z}=\frac{1}{T}\sum_{t=1}^{T}\hat{\Omega}_{1t}, \quad (15)$$

and

$$\hat{\Omega}_{2j}=\sum_{h=1}^{H}f_{j\cdot}^{(h)}\overline{\mathbf{z}}_{j\cdot}^{(h)}\overline{\mathbf{z}}_{j\cdot}^{(h)T} \quad (16)$$

$$\hat{\Omega}_{2z}=\frac{1}{p}\sum_{j=1}^{p}\hat{\Omega}_{2j}. \quad (17)$$

Then,

$$\widehat{\mathrm{cov}}(\mathrm{E}(\boldsymbol{Z}|Y))=\hat{\Omega}_{LSIR}=\hat{\Omega}_{1z}\otimes\hat{\Omega}_{2z}. \quad (18)$$

- We separately compute the singular value decompositions (SVDs) $\hat{\boldsymbol{\Omega}}_{1z}=\hat{\mathbf{U}}_1\hat{\mathbf{D}}_1\hat{\mathbf{R}}_1^T$ and $\hat{\boldsymbol{\Omega}}_{2z}=\hat{\mathbf{U}}_2\hat{\mathbf{D}}_2\hat{\mathbf{R}}_2^T$, to write $\hat{\boldsymbol{\Omega}}_{LSIR}=\hat{\boldsymbol{\Omega}}_{1z}\otimes\hat{\boldsymbol{\Omega}}_{2z}=(\hat{\mathbf{U}}_1\otimes\hat{\mathbf{U}}_2)\,(\hat{\mathbf{D}}_1\otimes\hat{\mathbf{D}}_2)\,(\hat{\mathbf{R}}_1^T\otimes\hat{\mathbf{R}}_2^T)$.
- To test for dimension, we estimate the rank of $\bar{\boldsymbol{\Omega}}_{1z}$ and $\bar{\boldsymbol{\Omega}}_{2z}$ separately using the weighted chi-square test for dimension in [17]. The dimension of $\bar{\boldsymbol{\Omega}}_{LSIR}$ is estimated by the product of the two estimated ranks since rank($\bar{\boldsymbol{\Omega}}_{LSIR}$) = rank($\bar{\boldsymbol{\Omega}}_{1z}$) × rank($\bar{\boldsymbol{\Omega}}_{2z}$).
- Letting rank($\bar{\boldsymbol{\Omega}}_{iz}$) = $d_i$, an estimate of the central subspace is given by

$$\hat{\mathscr{S}}_{FMSDR}=\hat{\mathscr{S}}_{SIR}-\mathrm{span}(\hat{\sum}_1^{-1}(\hat{U}_{11},\ldots,\hat{U}_{1d_1})\otimes\hat{\sum}_2^{-1}(U_{21},\ldots,\hat{U}_{2d_2})).$$

- To test for contributions of specific combinations of time points or predictors, we apply the test statistic proposed in Theorem 2 [18] separately to $\bar{\boldsymbol{\Omega}}_{1z}$ and $\bar{\boldsymbol{\Omega}}_{2z}$. That is, we use a Wald type test to test hypotheses of the form $H_0$ : $\mathbf{C}$vec($\mathbf{U}_{d_i}$) = $\mathbf{0}$ vs $H_1$ : $\mathbf{C}$vec($\mathbf{U}_{d_i}$) ≠ $\mathbf{0}$ where $\mathbf{U}_{d_i}$ = ($U_{i1}$, …, $U_{id_i}$), $i$ = 1, 2 for a prespecified $r \times kd_i$ matrix $\mathbf{C}$ of zeroes and ones, where $k = p$ for testing contributions of specific marker combinations and $k = T$ for specific time points. The rank of $\mathbf{C}$, $r$, equals the number of the elements of $\mathbf{U}_{d_i}$ set to zero. The test for marker contribution requires the computation of the asymptotic distribution of $\bar{\boldsymbol{\Omega}}_z$, which is derived in the Appendix.

We focus on the longitudinal SIR algorithm that naturally accommodates categorical responses as the slices are by default the categories. For continuous responses, alternatively one can use the longitudinal version of either PIR [7] or PFC [6], where the inverse regressions are fitted parametrically and the estimation is expected to be more accurate.

## 5. Simulations

In the simulations we focus on binary outcomes, $Y$ = 0, 1. We compare LSIR to standard SIR applied to the data ignoring their longitudinal structure, i.e. SIR is applied to $\mathbf{X}$ treated as the $pT \times 1$ vector, vec($\mathbf{X}$). For binary $Y$, SIR and LSIR estimate at most a single direction in $S_{Y|\mathbf{X}}$, that is $\mathcal{S}_{FMSDR}$ is given by a vector. The projection onto the space spanned by the SIR or LSIR kernel matrix can thus directly be used as a scalar diagnostic score. For categorical or continuous $Y$, if the dimension of the subspace is estimated to be larger than one, the projections need to be further combined to obtain a scalar score, for example by using the procedure in [1].

We quantify the discriminatory performance of the SIR and LSIR diagnostic scores with respect to the AUC, the area under the receiver-operator characteristics (ROC) curve. The ROC curve plots sensitivity against (1-specificity) (true vs. false positivity) for all thresholds that can be used to define ”test positive” (see also [19], page 67). Two diagnostic tests can be compared by calculating the difference between the areas under their two ROC curves (AUC), with the larger area corresponding to the ”better” test. The AUC values vary between 0.5 and 1, where 1 corresponds to perfect discriminatory accuracy of the test and

0.5 to no discriminatory ability. The AUC can also be expressed as the probability that the scalar diagnostic score for a randomly selected case $S_1$ exceeds that for a randomly selected control $S_0$, i.e. $AUC = P(S_1 > S_0)$.

We generated samples of $p = 5$ and $p = 10$ markers $\mathbf{X}^T = (X_1, \ldots, X_p)$ measured over $T = 5$ and $T = 10$ times points from two normal populations with equal covariance matrices, $(\mathbf{X}|Y = i) \sim MVN(\boldsymbol{\alpha}_i \otimes \boldsymbol{\beta}, \boldsymbol{\Delta})$, $i = 0, 1$. Each $\boldsymbol{\alpha}_i$, $i = 0, 1$ is a vector of length $p$ and $\boldsymbol{\beta}$ is a vector of length $T$. We let $\boldsymbol{\alpha}_1 = p^{-1/2}(1, \ldots, 1)$, $\boldsymbol{\alpha}_0$ has entries $\alpha_0(k) = 0$ for $k = 1, 3, 5$ and $\alpha_0(k) = p^{-1/2}$ for $k \neq 1, 3, 5$, and the entries of $\boldsymbol{\beta}$ are $\beta(k) = (T - k + 1)^{-1}$. These values were chosen to reflect practically relevant AUC values. The choice of the $\boldsymbol{\beta}$ coefficients leads to larger differences in the group means for later time points, that is measurements more proximal to $Y$ contribute more to discriminating the two groups. We let the number of observations in the $Y = 0$ and $Y = 1$ groups be equal, $N = N_0 = N_1$. To obtain unbiased estimates of the AUC, we used an independently generated sample of 500 cases and 500 controls.

For reference, we also computed the AUC value using 100, 000 cases and 100, 000 controls for the linear marker combination under the assumption that $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are known exactly. We report this value as the ”true AUC.”

The overall covariance matrix of the predictors $\boldsymbol{\Sigma}$ has the Kronecker structure, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_1 \otimes \boldsymbol{\Sigma}_2$. We assumed an AR(1) structure for $\boldsymbol{\Sigma}_1$, that is $\mathrm{cor}(X_{ij}, X_{ik}) = \rho_p^{|k-j|}$, and similarly for $\boldsymbol{\Sigma}_2$, $\mathrm{cor}(X_{ij}, X_{kj}) = \rho_T^{|k-j|}$, for various choices of $\rho_p$ and $\rho_T$. The group specific covariance matrix $\boldsymbol{\Delta}$ were computed using

$$\begin{aligned}\sum &= \mathrm{cov}(\mathrm{vec}(\mathbf{X}))E(\mathrm{cov}(\mathrm{vec}(\mathbf{X})|Y)) + \mathrm{cov}(E(\mathrm{vec}(\mathbf{X})|Y)) \\ &= \boldsymbol{\Delta} + E\{E(\mathrm{vec}(\mathbf{X})|Y)E(\mathrm{vec}(\mathbf{X})^T|Y)\} - E(\mathrm{vec}(\mathbf{X}))E(\mathrm{vec}(\mathbf{X})^T).\end{aligned}$$

For a model with $p = 5$ markers measured over $T = 10$ time points and the same mean and covariance matrices for both groups, $(\mathbf{X}|Y = i) \sim MVN(0, \boldsymbol{\Delta})$, $i = 0, 1$, the AUC estimates (with standard errors) were 0.52 (0.01) for the truth, 0.51 (0.01) for LSIR and 0.51 (0.01) for SIR for $N = 100$ cases and controls. Both methods thus performed well when there was no difference in the distribution of the predictors between the two groups.

Table 1 presents means in 100 simulations for various values of $p$, $T$ and correlations $\rho_T$ and $\rho_p$. For $p = T = 5$with $\rho_p = \rho_T = 0.3$, the AUC values (with empirical standard errors in parentheses) for LSIR and SIR were 0.728(0.021) and 0.699(0.023) for $N = 100$ cases and controls, and 0.748(0.013) and 0.741(0.014) for $N = 500$ cases and controls. For the same setting with $p = T = 10$ the differences were more pronounced, with AUC values of LSIR and SIR 0.643(0.034) and 0.577(0.029) for $N = 100$ cases and controls and 0.699(0.015) and 0.661(0.017) for $N = 500$ cases and controls.

Based on the paired $t$-test, LSIR had significantly higher AUC values than SIR, even for $N = 500$ cases and $N = 500$ controls. For example, for $N = 500$ cases and controls, for $p = T = 5$ with $\rho_p = \rho_T = 0.3$ the paired $t$-test $p$-value was 0.0002 and for $p = T = 5$ with $\rho_p = 0$, $\rho_T = 0.4$ the paired $t$-test $p$-values was 0.0001.

Results were very similar when the markers were uncorrelated, that is $\rho_T = \rho_p = 0$. In this setting the AUC was higher than for the correlated case, around 0.79 for the settings studied. SIR always resulted in lower AUC values for smaller sample sizes than LSIR. For $N = 100$, the difference was around 5%, a similar difference to the correlated case.

We also assessed the robustness of LSIR to violations of the moment assumptions. First, we violated the assumption of the Kronecker structure of the mean, but let $\mathbf{\Sigma}$ have a Kronecker structure. The mean vector for the controls was equal to zero, and the mean vector for cases had values $(1/3pT, \ldots, pT/3pT)$. For $p = 5$ and $T = 6$ with $\rho_t = 0.4$ and $\rho_p = 0.3$ the true AUC value was 0.73, and the AUCs (with empirical standard errors in parentheses) for LSIR and SIR were 0.64(0.03) and 0.62(0.03) respectively, for $N = 100$ cases and controls and 0.68(0.02) and 0.67(0.02) for $N = 500$ cases and controls. For $N = 1000$ cases and controls LSIR and SIR both had $AUC = 0.68$ with standard error 0.02. We then assumed that the means had a Kronecker structure, with the same choices of parameters as for the models in Table 1, but the true covariance matrix $\mathbf{\Sigma}$ does not have a Kronecker structure. Instead $\mathbf{\Sigma}$ had an AR(1) structure with $\rho = 0.3$. For $p = T = 5$ the AUC values (with empirical standard errors in parentheses) for LSIR and SIR were 0.718(0.019) and 0.726(0.021) for $N = 100$ cases and controls, and 0.767(0.015) and 0.770(0.016) for $N = 500$ cases and controls. For the same setting with $p = 5$ and $T = 10$ the AUC values of LSIR and SIR were 0.769(0.016) and 0.781(0.015) for $N = 100$ cases and controls and 0.699(0.015) and 0.661(0.017) for $N = 500$ cases and controls. When the population means were zero and $(1/3pT, \ldots, pT/3pT)$ for controls and cases respectively, and the assumption on $\Sigma$ was violated so that instead of a Kronecker product $\Sigma$ had an $AR(1)$ structure with $\rho = 0.4$, the true AUC value was 0.75, and the AUCs for LSIR and SIR were 0.681(0.030) and 0.645(0.024) respectively for $N = 100$ cases and controls and 0.70(0.02) and 0.69(0.02) respectively for $N = 500$ cases and controls. Thus the LSIR procedure appears to be robust to violations of the assumption on the second moments.

*R*-code for the LSIR algorithm, testing of dimension and testing of marker contributions is available from the first author upon request. *R*-code that implements standard SIR is available in the package *dr* at http://cran.rproject.org/.

## 6. Data Analysis

We used data from the Vorarlberg Health Monitoring and Promotion Program [20], one of the world's largest ongoing population-based risk factor surveillance programs, to illustrate LSIR and compare it to SIR. The aim was to classify men into two groups; those who developed cancer at the end of follow up and those who did not, using three serum biomarkers. The markers were the log transformed values of uric acid ($UA$), blood glucose ($GLUC$) and total serum cholesterol ($CHOL$). We used data from 100 male cancer cases who each had marker measurements at four time points prior to diagnosis, and selected a sample of 100 controls who also had measurements at four time points and at the end of follow up had not developed cancer. Thus the $p = 3$ predictors $\mathbf{X}_t = (UA_t, GLUC_t, CHOL_t)^T$, $t = 1, \ldots, 4$, were measured at $T = 4$ time points, which means that the dimension of this classification problem is $3 \times 4 = 12$. The histograms of the log-transformed markers in the cases and controls at time point $t = 1$, given in Figure 1, appeared reasonably symmetric.

In cases, the means of the markers for the four time points were $\bar{\mathbf{X}}_{t1}$ = (4.45, 5.38, 3.48), $\bar{\mathbf{X}}_{t2}$ = (4.47, 5.41, 3.50), $\bar{\mathbf{X}}_{t3}$ = (4.50, 5.40, 3.43), $\bar{\mathbf{X}}_{t4}$ = (4.54, 5.43, 3.46) and in controls $\bar{\mathbf{X}}_{t1}$ = (4.47, 5.43, 3.64), $\bar{\mathbf{X}}_{t2}$ = (4.45, 5.39, 3.55), $\bar{\mathbf{X}}_{t3}$ = (4.49, 5.42, 3.60), $\bar{\mathbf{X}}_{t4}$ = (4.52, 5.38, 3.56). To check the assumption of Kronecker structure of the means for binary outcomes $Y$, it suffices to assess if the means of the markers in the control and case groups vary only by a constant that may depend on time. Visual inspection of the ratios of the means for the different time points did not suggest violations of Assumption 1.

The correlations of the three markers across the four time points were similar, with

$$R_{UA}=\begin{bmatrix} 1.00 & 0.60 & 0.56 & 0.43 \\ 0.60 & 1.00 & 0.61 & 0.51 \\ 0.56 & 0.61 & 1.00 & 0.64 \\ 0.43 & 0.51 & 0.64 & 1.00 \end{bmatrix}, R_{GLUC}=\begin{bmatrix} 1.00 & 0.70 & 0.69 & 0.58 \\ 0.70 & 1.00 & 0.68 & 0.69 \\ 0.69 & 0.68 & 1.00 & 0.68 \\ 0.58 & 0.69 & 0.68 & 1.00 \end{bmatrix}, R_{CHOL}=\begin{bmatrix} 1.00 & 0.83 & 0.77 & 0.74 \\ 0.83 & 1.00 & 0.80 & 0.80 \\ 0.77 & 0.80 & 1.00 & 0.84 \\ 0.74 & 0.80 & 0.84 & 1.00 \end{bmatrix}.$$

Similarly, the correlations of the markers with each other for the four time points were similar,

$$\begin{aligned} R_{t1}&=\begin{bmatrix} 1.00 & 0.04 & 0.23 \\ 0.04 & 1.00 & 0.25 \\ 0.23 & 0.25 & 1.00 \end{bmatrix}, R_{t2} \\ &=\begin{bmatrix} 1.00 & 0.05 & 0.12 \\ 0.05 & 1.00 & 0.10 \\ 0.12 & 0.10 & 1.00 \end{bmatrix}, R_{t3} \\ &=\begin{bmatrix} 1.00 & 0.16 & 0.11 \\ 0.16 & 1.00 & 0.11 \\ 0.11 & 0.11 & 1.00 \end{bmatrix}, R_{t4} \\ &=\begin{bmatrix} 1.00 & 0.13 & 0.02 \\ 0.13 & 1.00 & 0.17 \\ 0.02 & 0.17 & 1.00 \end{bmatrix}. \end{aligned}$$

We used a likelihood ratio test based on assuming a multivariate normal distribution for the markers, to test the hypothesis of an unstructured covariance matrix against the alternative of a Kronecker product structure of the covariance matrix of the data [13]. Under the null hypothesis of a Kronecker structure, $-2LogLikelihood = -1709.9$, and under the alternative hypothesis, $-2LogLikelihood = -1781.60$, corresponding to a $p$-value 0.19 based on a chi-square distribution with $Tp(Tp+1)/2 - T(T+1)/2 - p(p+1)/2 = 62$ degrees of freedom. Thus, there was no strong evidence in the data against the null hypothesis of Kronecker product covariance structure.

The test for dimension estimated the dimension to be one for standard SIR ($p$-value=0.008 for testing dimension 0 vs 1), and also 1 for LSIR for the time and marker parts separately.

For LSIR, the coefficients for the projection corresponding to time component were $\boldsymbol{\beta}$ = (0.46, −0.31, 0.51, −0.66), and the coefficients for the marker component $\boldsymbol{\alpha}$ = ($UA, GLUC, CHOL$) = (0.12, 0.99, 0.04). The resulting combined vector of coefficients for the projection was ($UA_1$, $GLUC_1$, $CHOL_1$,…, $UA_4$, $GLUC_4$, $CHOL_4$) = (−0.05, −0.45, −0.02, 0.04, 0.30, 0.01, −0.06, −0.50, −0.02, 0.08, 0.66, 0.03). When we tested the contribution of individual markers using LSIR, there was evidence based on the Wald test that $GLUC$ significantly contributed to $\boldsymbol{\alpha}$ ($p$-value= 0.04), while $UA$ and $CHOL$ did not ($p$-value= 0.73 and $p$-value= 0.54, respectively). Applying the test to the time component showed that $t_4$ contributed significantly to $\boldsymbol{\beta}$ ($p$ = 0.05).

Based on standard SIR, the coefficients for the projection were ($UA_1$, $GLUC_1$, $CHOL_1$, …, $UA_4$, $GLUC_4$, $CHOL_4$) = (−0.16, −0.52, −0.02, 0.21, 0.25, 0.10, −0.01, −0.35, −0.12, 0.06, 0.67, −0.02). When we tested for significance of the contributions of individual markers at given time points, glucose was significant at the first time point ($p$-value= 0.01)s and the fourth time point ($p$-value< 0.01). However, the contribution of glucose at all four time points jointly was not found to be significantly associated with outcome.

We also fitted a logistic regression model that included all predictors to the data. For binary outcomes, logistic regression is equivalent to standard SIR and yields the exact same projection. The corresponding log-odds ratios were ($UA_1$, $GLUC_1$, $CHOL_1$, …, $UA_4$, $GLUC_4$, $CHOL_4$) = (−1.01, −3.21, −0.12, 1.18, 1.47, 0.62, −0.02, −2.18, −0.69, 0.32, 4.33, −0.14). Based on the logistic model, only glucose measured at the first and fourth time point was significantly associated with outcome ($p$-value< 0.01).

As we did not have an independent test set, we used 2-fold Monte Carlo cross validation to obtain unbiased estimates for the AUC. That is, we randomly assigned each of the observations in the case and control group to one of two partitions of equal size. Subsequently we used one of the partitions as the training set to build the LSIR and SIR predictors, and evaluated the performance of our predictor on the remaining partition, labeled the test set. This procedure was repeated one thousand times. For each of the test sets, the AUC was computed and then averaged. The dimension of both SIR and LSIR subspaces was estimated to be one. The mean AUC value for the SIR predictor was 0.60(0.04) while it was 0.63(0.04) for the LSIR score. The 95% confidence interval around the difference between the AUC values for LSIR and SIR was (−0.04, −0.02), indicating that LSIR had better discriminatory performance than SIR. The difference in AUC values for this example agrees with what we saw in the simulations for a small number of predictors.

In summary, LSIR and SIR provided similar results. Based on LSIR we identified glucose as a significant predictor and found that the fourth time point was important. This is plausible, as measurements taken at the fourth time point were most proximal to cancer diagnosis. SIR also sfound the coefficient for glucose measured at time one and four was significant. The importance of glucose at the first time point is somewhat more difficult to explain, and could be a chance finding.

## 7. Discussion

In this paper we show that under the assumption that the first two moments of repeatedly measured predictors have a Kronecker product structure, which reflects the longitudinal structure of the predictors, the first moment based dimension reduction subspace also has a Kronecker product structure. We propose an algorithm that utilizes this structure to implement a longitudinal version of SIR, which we call LSIR. This substantially reduces the complexity of the estimation of the first moment based dimension reduction subspace. The reduction in estimation burden leads to a noticeable improvement in the discriminatory ability of a score obtained by projecting the data onto that subspace for practically relevant sample sizes. For binary outcomes, the improvement in the area under the curve for LSIR was approximately 5% higher than a score computed from standard SIR, that ignored the repeated measurement structure. This seemingly small difference constitutes a substantial improvement in discriminatory power [21]. To see this consider the simple case of a single continuous predictor that in both cases and controls arises from a normal population that differ only in their means, that is $X_i \sim N(\mu_i, \sigma^2)$, $i = 0, 1$. The corresponding AUC is given by $AUC = \Phi((\mu_1 - \mu_0)/\sigma^2)$. The odds ratio (OR) associated with $X$ in a logistic model based on a case control sample is given by $\exp((\mu_1 - \mu_0)/\sigma^2)$. When $\nu = (\mu_1 - \mu_0)/\sigma^2 = 0.55$, we obtain $AUC = 0.71$. In order to improve the AUC by 5%, we need to increase $\nu$ to $\nu = 0.70$, or equivalently, we need to increase the OR associated with $X$ from 1.73 to 2.01, corresponding to an 18% increase in the OR.

An important step before any data analysis is to test the assumption about the Kronecker structure of the moments. Under the additional assumption of multivariate normally distributed markers, likelihood ratio tests can be employed following [16]. Nevertheless, we found in simulations that even under violations of the Kronecker product structure, fitting a more parsimonious covariance matrix did not result in a noticeable or even statistically significant loss of discriminatory power.

Moment structures similar to the ones we assumed in this paper appeared for the first time in growth curve models (see [22]) that are used to analyze longitudinal and repeated measures data. These models were used in discrimination and classification of multivariate repeated measures data, for example, by [13] and [14]. While all these approaches study variations of such structures on the mean and covariance of $\mathbf{X}|Y$, they also require normally distributed data. In contrast, we do not require normality or a specific distribution and our approach is computationally tractable even for large numbers of markers. In addition, the LSIR algorithm can easily be adapted to handle missing data and a variable number of observations for each individual.

One of the reviewers of this paper alerted us to the fact that a very similar approach to our method, currently unpublished, was proposed as part of a Ph.D. thesis [23] independently. In particular, Theorem 1 is a common result in our paper and [23] (see chapter 8), but our other methodological results, such as Theorem 2 and testing for dimension and marker contribution when the moments have a Kronecker product structure, are new and appear only in our paper, to the best of our knowledge.

SDR was first introduced into the analysis of longitudinal data by Li and Yin [24] who applied Li et al.'s [25] dimension reduction method to settings where both the outcome and the predictors are measured repeatedly over time. Time is considered as a categorical covariate, dimension reduction subspaces are estimated separately for each time point, and then combined into a single matrix. However, this approach ignores the correlation across time, an integral feature of longitudinal data.

In summary, we present a new SDR approach for longitudinally measured predictors. By exploiting the structure of moments of the longitudinal predictors we obtain more accurate estimates of the dimension reduction subspace and hence more accurate marker projections for classification. A limitation of the proposed work is that first moment methods for binary outcomes can detect at most one dimension in the central dimension reduction subspace and thus may miss important information of the predictors. In future work we plan to extend this approach to second moment based methods.

## Acknowledgments

We thank the associate editor and the reviewers for helpful comments and suggestions.

## References

1. Pfeiffer RM, Bura E. A model free approach to combining biomarkers. Biometrical Journal. 2008; 50:558–570. [PubMed: 18663762]
2. Li KC. Sliced inverse regression for dimension reduction (with discussion). Journal of the American Statistical Association. 1991; 86:316–342.
3. Cook RD, Weisberg S. Discussion of "Sliced inverse regression for dimension reduction. Journal of the American Statistical Association. 1991; 86:328–332.
4. Cook, RD. Regression Graphics: Ideas for studying regressions through graphics. New York: Wiley; 1998.
5. Cook RD. Fisher lecture: Dimension reduction in regression. Statistical Science. 2007; 22:1–26.
6. Cook RD, Forzani L. Principal Fitted Components for Dimension Reduction in Regression. Statistical Science. 2008; 23:485–501.
7. Bura E, Cook RD. Estimating the structural dimension of regressions via parametric inverse regression. Journal of the Royal Statistical Society: Series B (Statistical Methodology). 2001; 63:393–410.
8. Li KC, Duan N. Regression analysis under link violations. Annals of Statistics. 1989; 17:1009–1052.
9. Eaton, ML. Multivariate Statistics. A Vector Space Approach. New York: John Wiley & Sons, Inc; 1983.
10. Cook RD, Nachtsheim CJ. Reweighting to achieve ellipically contoured covariates in regression. Journal of the American Statistical Association. 1994; 89:592–599.
11. Hall P, Li KC. On almost linearity of low-dimensional projections from high-dimensional data. Annals of Statistics. 1993; 21 :867–889.
12. Shao Y, Cook RD, Weisberg S. Marginal tests with sliced average variance estimation. Biometrika. 2007; 94:285–296.
13. Roy A, Khattree R. Discrimination and classification with repeated measures data under different covariance structures. Communications in Statistics– Simulation and Computation. 2005; 34:167–178.
14. Srivastava MS, Von Rosen T, Von Rosen D. Estimation and testing in general multivariate linear models with Kronecker product covariance structure. Sankhya, Ser A. 2009; 71:137–163.

15. Roy A, Khattree R. Tests for mean and covariance structures relevant in repeated measures based discriminant analysis. Journal of Applied Statistical Science. 2003; 12:91–104.
16. Roy A, Khattre R. On implementation of a test for Kronecker product covariance structure for multivariate repeated measures data in. Statistical Methodology. 2005; 2:297–306.
17. Bura E, Yang J. Dimension estimation in dufficient dimension reduction: a unifying approach. Journal of Multivariate Analysis. 2011; 102:130–142.
18. Bura E, Pfeiffer RM. On the distribution of the left singular vectors of a random matrix and its applications. Statsitics & Probability Letters. 2008; 15:2275–2280.
19. Pepe, MS. Oxford Statistical Science Series. Oxford University Press; 2003. The Statistical Evaluation of Medical Tests for Classification and Prediction.
20. Ulmer H, Kelleher C, Diem G, Concin H. Long-term tracking of cardiovascular risk factors among men and women in a large population-based health system: the Vorarlberg Health Monitoring & Promotion Programme. European Heart Journal. 2003; 24:1004–1013. [PubMed: 12788300]
21. Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. American Journal of Epidemiology. 2004; 159:882–890. [PubMed: 15105181]
22. Potthoff RF, Roy SN. A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika. 1964; 51:313–326.
23. Kim, MK. PhD dissertation. Pennsylvania State University; 2010. On dimension folding of matrix or array valued statstitical objects.
24. Li LX, Yin XR. Longitudinal data analysis using sufficient dimension reduction method. Computational Statistics & Data Analysis. 2009; 53:4106–4115.
25. Li B, Cook RD, Chiaromonte F. Dimension reduction for the conditional mean in regressions with categorical predictors. Annals of Statistics. 2003; 31:1636–1668.
26. Henderson HV, Searle SR. On deriving the inverse of a sum of matrices. SIAM Review. 1981; 23:53–60.

## 8. Appendix A

## Proof of (3)

By the linearity condition using (2) we have

$$\sum^{-1}(\mathrm{E}(\mathbf{X}|Y)-\mathrm{E}(\mathbf{X}))=\boldsymbol{\eta}\boldsymbol{\nu}_Y \quad (19)$$

with $\boldsymbol{\eta}\in \mathcal{S}_{Y|X}$ and $\mathrm{cov}(\boldsymbol{\nu}_Y) > 0$ and $\boldsymbol{\eta}^T\boldsymbol{\eta}$ the identity matrix. Then $\mathrm{E}(\mathbf{X}|Y) - \mathrm{E}(\mathbf{X}) = \boldsymbol{\Sigma}\boldsymbol{\eta}\boldsymbol{\nu}_Y$, $\mathrm{span}(\boldsymbol{\eta}) = \boldsymbol{\Sigma}^{-1}\mathrm{span}((\mathrm{E}(\mathbf{X}|Y) - \mathrm{E}(\mathbf{X}))$ and

$$\sum=\mathrm{E}(\mathrm{cov}(\mathbf{X}|Y))+\mathrm{cov}(\mathrm{E}(\mathbf{X}|Y))=\boldsymbol{\Delta}+\mathrm{cov}(\mathrm{E}(\mathbf{X}|Y))=\boldsymbol{\Delta}+\sum\boldsymbol{\eta}\mathrm{cov}(\boldsymbol{\nu}_Y)\boldsymbol{\eta}^T\sum.$$

Using that $(\mathbf{A} + \mathbf{BCD})^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{C}^{-1} + \mathbf{DA}^{-1}\mathbf{B})^{-1}\mathbf{DA}^{-1}$ [26] we obtain

$$\boldsymbol{\Delta}^{-1}=\sum^{-1}+\boldsymbol{\eta}\left(\mathrm{cov}(\boldsymbol{\nu}_Y)^{-1}+\boldsymbol{\eta}^T\sum^{-1}\boldsymbol{\eta}\right)^{-1}\boldsymbol{\eta}^T$$

and therefore from (19),

$$\begin{aligned}&\boldsymbol{\Delta}^{-1}\text{span}\left(\text{E}(\mathbf{X}|Y)\right.\\&\quad\left.-\text{E}(\mathbf{X})\right)\\&=\sum^{-1}\text{span}\left(\text{E}(\mathbf{X}|Y)-\text{E}(\mathbf{X})\right)\\&+\text{span}\left(\boldsymbol{\eta}\left(\text{cov}(\boldsymbol{\nu})^{-1}+\boldsymbol{\eta}^T\sum^{-1}\boldsymbol{\eta}\right)^{-1}\boldsymbol{\nu}_Y\right)\\&=\sum^{-1}\text{span}\left(\text{E}(\mathbf{X}|Y)-\text{E}(\mathbf{X})\right)\end{aligned}$$

since span($\boldsymbol{\eta}$) = $\boldsymbol{\Sigma}^{-1}$span((E($\mathbf{X}$| $Y$) − E($\mathbf{X}$)).

## Proof of Theorem 1

From the definition of the FMSDR subspace we have

$$\mathscr{S}_{FMSDR}=\sum^{-1}\text{span}\left(\text{vec}(\text{E}(\mathbf{X}|Y)-\text{E}(\mathbf{X}))\right)=(\sum_1^{-1}\otimes\sum_2^{-1})\text{span}(\text{cov}(\text{vec}(\text{E}(\mathbf{X}|Y))))$$
$$=(\sum_1^{-1}\otimes\sum_2^{-1})\text{span}(\boldsymbol{\alpha}\otimes\boldsymbol{\beta})\text{cov}(\boldsymbol{\nu}_y)\,(\boldsymbol{\alpha}^T\otimes\boldsymbol{\beta}^T)=(\sum_1^{-1}\otimes\sum_2^{-1})\text{span}(\boldsymbol{\alpha}\otimes\boldsymbol{\beta})=\text{span}(\sum_1^{-1}\boldsymbol{\alpha}\otimes\sum_2^{-1}\boldsymbol{\beta}).$$

The second equality follows from Prop. 2.6 in [9], p. 75.

## Asymptotic distribution of $\bar{\boldsymbol{\Omega}}_{LSIR}$

First, notice that if vec(E($\mathbf{X}$| $Y$) − E($\mathbf{X}$)) = ($\boldsymbol{\alpha}\otimes\boldsymbol{\beta}$)vec($\boldsymbol{\nu}_y$), the same moment condition holds for the standardized predictors $\mathbf{Z}$,

$$\begin{aligned}\text{vec}(\text{E}(\boldsymbol{Z}|Y))&=\sum^{-1/2}(\boldsymbol{\alpha}\otimes\boldsymbol{\beta})\text{vec}(\boldsymbol{\nu}_y)\\&=(\sum\nolimits_1\otimes\sum\nolimits_2)^{-1/2}(\boldsymbol{\alpha}\otimes\boldsymbol{\beta})\text{vec}(\boldsymbol{\nu}_y)\\&=[(\sum\nolimits_1^{-1/2}\boldsymbol{\alpha})\otimes(\sum\nolimits_2^{-1/2}\boldsymbol{\beta})]\text{vec}(\boldsymbol{\nu}_y)\\&=(\boldsymbol{\alpha}_z\otimes\boldsymbol{\beta}_z)\text{vec}(\boldsymbol{\nu}_y).\end{aligned}$$

where $\boldsymbol{\alpha}_z=\sum_1^{-1/2}\boldsymbol{\alpha}$ and $\beta_z=\sum_2^{-1/2}\beta$. We derive the asymptotic distribution of $\bar{\boldsymbol{\Omega}}_{1t}$ given by (14). The proof for $\bar{\boldsymbol{\Omega}}_{2t}$ is analogous. Recall that $\overline{\mathbf{z}}_{\cdot t}^{(h)}$ denotes the $p$-vector of the standardized predictor means within slice $h$, $h$ = 1, …, $H$. The proportion of observations in slice $h$ at time $t$ across all $p$ predictors is $f_{\cdot t}^{(h)}$. Let $\hat{\boldsymbol{Z}}_t=(\overline{\boldsymbol{Z}}_{\cdot t}^{(1)}\sqrt{f_t^{(1)}},\ldots,\overline{\boldsymbol{Z}}_{\cdot t}^{(H)}\sqrt{f_t^{(H)}})$ to obtain $\hat{\Omega}_{1t}=\hat{\boldsymbol{Z}}_t\hat{\boldsymbol{Z}}_t^T$. Since $\bar{\boldsymbol{\Omega}}_{1t}$ is a sample covariance matrix, for each $t$ = 1, …, $T$,

$$n^{1/2}\text{vec}(\hat{\boldsymbol{\Omega}}_{1t}-\boldsymbol{\Omega}_{1t})\rightarrow N_{pd}(\mathbf{0},\mathbf{V}_{1t}) \quad (20)$$

where $\boldsymbol{\Omega}_{1t} = \text{cov}(\text{E}(\mathbf{Z}_t|Y)) = (\boldsymbol{\alpha}_z \otimes \boldsymbol{\beta}_{z,t})\text{cov}(\boldsymbol{\nu}_y)(\boldsymbol{\alpha}_z \otimes \boldsymbol{\beta}_{z,t})^T$ with $\boldsymbol{\beta}_{z,t}$ the $t$-th column of the matrix $\boldsymbol{\beta}_z$, and $\mathbf{V}_{1t}=\sum_h p_t^{(h)}\mathbf{V}_{1t}^{(h)}$ is the $p^2 \times p^2$ matrix with $\mathbf{V}_{1t}^{(h)}=\text{var}[\text{vec}(\boldsymbol{Z}_{\cdot t}^{(h)}-E(\boldsymbol{Z}_{\cdot t}^{(h)}|Y=h))\,(\boldsymbol{Z}_{\cdot t}^{(h)}-E(\boldsymbol{Z}_{\cdot t}^{(h)}|Y=h))^T|Y=h]$ and $p_t^{(h)}=P(Y \text{ falls in slice } h)$.

Since $\text{cov}(\boldsymbol{\nu}_y)$ is positive definite, $\text{cov}(\boldsymbol{\nu}_y) = \mathbf{U}\mathbf{U}^T$. Thus,

$$\boldsymbol{\Omega}_1=\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\Omega}_{1t}=\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{\alpha}_z\otimes\boldsymbol{\beta}_{z,t})\mathbf{U}\mathbf{U}^T(\boldsymbol{\alpha}_z\otimes\boldsymbol{\beta}_{z,t})^T=\frac{1}{T}\sum_{t=1}^{T}(\boldsymbol{\alpha}_z\mathbf{U}\boldsymbol{\beta}_{z,t}^T\boldsymbol{\beta}_{z,t}\mathbf{U}^T\boldsymbol{\alpha}_z^T)=\boldsymbol{\alpha}_z\mathbf{U}(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{\beta}_{z,t}^T\boldsymbol{\beta}_{z,t})\mathbf{U}^T\boldsymbol{\alpha}_z^T.$$

Using the eigendecomposition of the $s \times s$ matrix $\sum_{t=1}^{T}\boldsymbol{\beta}_{z,t}^T\boldsymbol{\beta}_{z,t}/T=\mathbf{V}\mathbf{D}\mathbf{V}^T$, $\boldsymbol{\Omega}_1 = \boldsymbol{\alpha}_z\mathbf{U}\mathbf{V}\mathbf{D}\mathbf{V}^T\mathbf{U}^T\boldsymbol{\alpha}^T = (\boldsymbol{\alpha}_z \otimes (\mathbf{V}\mathbf{D}^{1/2}))\text{cov}(\boldsymbol{\nu}_y)(\boldsymbol{\alpha}_z \otimes (\mathbf{V}\mathbf{D}^{1/2}))^T$. Therefore from (20) we obtain

$$n^{1/2}\text{vec}(\frac{1}{T}\sum_{t=1}^{T}\hat{\boldsymbol{\Omega}}_{1t}-\boldsymbol{\Omega}_1) \to N_{pd}(\mathbf{0}, \mathbf{V}_1)$$

with $\mathbf{V}_1=\frac{1}{T^2}\sum_{t=1}^{T}\mathbf{V}_{1t}+\frac{1}{T^2}\sum_t\sum_{s,s\neq t}\text{cov}(\text{vec}(\hat{\boldsymbol{\Omega}}_{1t}), \text{vec}(\hat{\boldsymbol{\Omega}}_{1s}))$. Now,

$$\begin{aligned}\text{cov}(\text{vec}(\hat{\boldsymbol{\Omega}}_{1t}), \text{vec}(\hat{\boldsymbol{\Omega}}_{1s}))&=\text{cov}(\text{vec}(\sum_{h=1}^{H} f_{\cdot t}^{(h)}\overline{\mathbf{z}}_{\cdot t}^{(h)}\overline{\mathbf{z}}_{\cdot t}^{(h)T}), \text{vec}(\sum_{h=1}^{H} f_{\cdot s}^{(h)}\overline{\mathbf{z}}_{\cdot s}^{(h)}\overline{\mathbf{z}}_{\cdot s}^{(h)T}))\\&=\sum_{h=1}^{H} f_{\cdot t}^{(h)} f_{\cdot s}^{(h)}\text{cov}(\text{vec}(\overline{\mathbf{z}}_{\cdot t}^{(h)}\overline{\mathbf{z}}_{\cdot t}^{(h)T}), \text{vec}(\overline{\mathbf{z}}_{\cdot s}^{(h)}\overline{\mathbf{z}}_{\cdot s}^{(h)T}))\end{aligned}$$

For $t = 1, \ldots, T$, let $\tilde{\boldsymbol{Z}}_t^{(h)}=\overline{\mathbf{z}}_{\cdot t}^{(h)}\overline{\mathbf{z}}_{\cdot t}^{(h)T}$, with elements $[\tilde{\boldsymbol{Z}}_t^{(h)}]_{kl}=(f_{\cdot t}^{(h)})^2\sum_{i=1}^{n_t}Z_{ikt}^{(h)}\sum_{i=1}^{n_t}Z_{ilt}^{(h)}$. The $p^2$ diagonal elements of $\text{cov}(\text{vec}(\tilde{\boldsymbol{Z}}_t^{(h)}), \text{vec}(\tilde{\boldsymbol{Z}}_s^{(h)}))$ are given by $\text{var}([\tilde{\boldsymbol{Z}}_t^{(h)}]_{kl})$ and $\text{var}([\tilde{\boldsymbol{Z}}_s^{(h)}]_{qr})$, and the off diagonal elements are $\text{cov}(\text{vec}([\tilde{\boldsymbol{Z}}_t^{(h)}]_{kl}), \text{vec}([\tilde{\boldsymbol{Z}}_s^{(h)}]_{qr}))$, where $k, l, q, r = 1, \ldots, p$.

The asymptotic behavior of $\hat{\boldsymbol{\Omega}}_2$ can be derived in a similar manner, by using $\mathbf{Z}^T$ instead of $\mathbf{Z}$ and $\text{vec}(\text{E}(\boldsymbol{Z}^T|Y))=(\boldsymbol{\beta}_z \otimes \boldsymbol{\alpha}_z)\text{vec}(\boldsymbol{\nu}_y^T)$ in the above calculations, leading to

$$n^{1/2}\text{vec}(\frac{1}{p}\sum_{j=1}^{p}\hat{\boldsymbol{\Omega}}_{2j}-\boldsymbol{\Omega}_2) \to N_{pd}(\mathbf{0}, \mathbf{V}_2),$$

where $\boldsymbol{\Omega}_2=\boldsymbol{\beta}_z\mathbf{U}(\frac{1}{p}\sum_{j=1}^{p}\boldsymbol{\alpha}_{z,j}^T\boldsymbol{\alpha}_{z,j})\mathbf{U}^T\boldsymbol{\beta}_z^T$. From the expressions for $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ it can be seen that $\text{span}(\boldsymbol{\Omega}_1 \otimes \boldsymbol{\Omega}_2) = \text{span}(\boldsymbol{\alpha}_z \otimes \boldsymbol{\beta}_z)$.

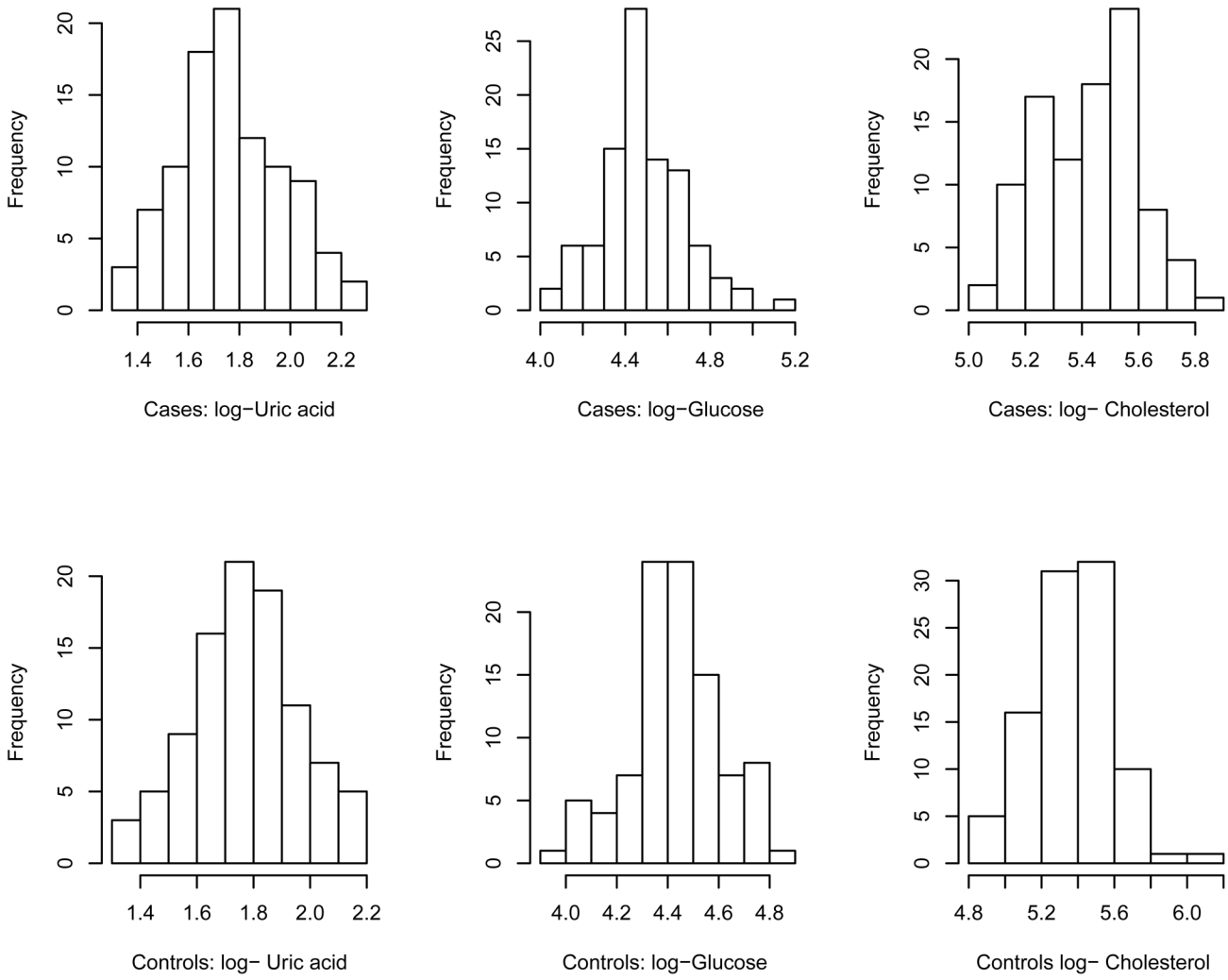


**Figure 1.**
Distribution of the three biomarkers after log-transformation in cases (top panel) and controls (bottom panel).

**Table 1**

Mean AUC values over 100 simulations when $\boldsymbol{\Sigma}$ has a Kronecker structure and the correlations over time and markers follow an AR(1) structure for $N$ cases and $N$ controls. Standard errors are given below the mean estimates. The size of the test set for the AUC computation was 500 cases and 500 controls.

| $p$, $T$, $\rho_p$, $\rho_T$ | AUC | method | $N$ = 100 | $N$ = 200 | $N$ = 300 | $N$ = 400 | $N$ = 500 |
|---|---|---|---|---|---|---|---|
| 5,5,0.3,0.3 | 0.743 | *LSIR* | 0.728 | 0.743 | 0.746 | 0.747 | 0.748 |
| | | | 0.021 | 0.015 | 0.016 | 0.015 | 0.013 |
| | | *SIR* | 0.699 | 0.728 | 0.735 | 0.737 | 0.741 |
| | | | 0.023 | 0.016 | 0.018 | 0.017 | 0.014 |
| 5,10,0.3,0.3 | 0.758 | *LSIR* | 0.743 | 0.765 | 0.772 | 0.775 | 0.776 |
| | | | 0.023 | 0.017 | 0.015 | 0.016 | 0.015 |
| | | *SIR* | 0.694 | 0.735 | 0.750 | 0.758 | 0.761 |
| | | | 0.024 | 0.023 | 0.016 | 0.018 | 0.017 |
| 10,10,0.3,0.3 | 0.691 | *LSIR* | 0.643 | 0.677 | 0.690 | 0.695 | 0.699 |
| | | | 0.034 | 0.022 | 0.019 | 0.017 | 0.015 |
| | | *SIR* | 0.577 | 0.619 | 0.641 | 0.654 | 0.661 |
| | | | 0.029 | 0.020 | 0.020 | 0.021 | 0.017 |
| 5,5,0,0.4 | 0.724 | *LSIR* | 0.698 | 0.715 | 0.721 | 0.725 | 0.725 |
| | | | 0.026 | 0.019 | 0.016 | 0.015 | 0.017 |
| | | *SIR* | 0.670 | 0.699 | 0.708 | 0.715 | 0.716 |
| | | | 0.025 | 0.019 | 0.018 | 0.016 | 0.018 |
| 5,10,0,0.4 | 0.749 | *LSIR* | 0.747 | 0.766 | 0.772 | 0.777 | 0.778 |
| | | | 0.029 | 0.017 | 0.015 | 0.014 | 0.015 |
| | | *SIR* | 0.700 | 0.737 | 0.750 | 0.760 | 0.765 |
| | | | 0.025 | 0.019 | 0.016 | 0.014 | 0.015 |
| 10,10,0,0.4 | 0.698 | *LSIR* | 0.618 | 0.661 | 0.673 | 0.679 | 0.683 |
| | | | 0.039 | 0.022 | 0.019 | 0.018 | 0.019 |
| | | *SIR* | 0.568 | 0.607 | 0.626 | 0.637 | 0.649 |
| | | | 0.021 | 0.024 | 0.021 | 0.020 | 0.018 |