Discriminant Analysis for Repeated Measures Data: Effects of Mean and Covariance Misspecification on Bias and Error in Discriminant Function Coefficients

Tolulope T. Sajobi[1], Lisa M. Lix[1,2], Longhai Li[2], & William Laverty[2]

[1]School of Public Health, University of Saskatchewan, CANADA

[2]Department of Mathematics & Statistics, University of Saskatchewan, CANADA

Corresponding Author:

Lisa Lix, PhD P.Stat
School of Public Health
University of Saskatchewan
107 Wiggins Road
Saskatoon, SK CANADA S7N 5E5
Phone: 306-966-1617
Fax: 306-966-7920
e-mail: lisa.lix@usask.ca

Abstract Word Count: 47

Manuscript Word Count: 4387

Abstract

Discriminant analysis (DA) procedures based on parsimonious mean and/or covariance structures have recently been proposed for repeated measures (RM) data. This paper investigates bias and means square error of discriminant function coefficients (DFCs) of these DA procedures when mean and/or covariance structures are correctly specified and misspecified.

Key Words: multivariate; model misspecification; discriminant function coefficients; mean square error; bias

Discriminant Analysis for Repeated Measures Data: Effects of Mean and Covariance Misspecification on Bias and Error in Discriminant Function Coefficients

Introduction

Linear discriminant analysis (DA) is a multivariate procedure, originally proposed by Fisher (1936), for predicting group membership (predictive discriminant analysis; PDA) and/or describing group separation (descriptive discriminant analysis; DDA) (Huberty & Olejnik, 2006) on multiple variables. The classical linear PDA procedure has also been applied to repeated measures (RM) data (Feighner & Sverdlov, 2002; Levesque, Ducharme, Zarit, Lachance, & Giroux, 2008), in which study participants are measured on a single variable at two or more occasions. Classical linear DA will not result in an efficient classification rule in multivariate or RM data when there are a large number of variables or measurement occasions relative to sample size. In recent years, a number of PDA procedures for RM data have been proposed (Marshall & Baron, 2000; Roy & Khatree, 2005a, 2005b, 2007; Tomasko, Helms, & Snappin, 1999). Specifically, Roy and Khattree (2005a, 2005b) developed DA procedures based on parsimonious mean and covariance structures for both univariate (i.e., measurements on one outcome variable) and multivariate (i.e., measurements on two or more outcome variables) RM data to address the issue of classification efficiency when sample size is small. For univariate RM data, they proposed procedures based on a constant RM mean vector and either a compound symmetric (CS) or first-order autoregressive (AR-1) covariance. While these procedures can result in efficient classification rules in high-dimensional data (Roy & Khatree, 2007), they can also result in inflated misclassification error rates (MERs) when the mean and/or covariance structure is/are incorrectly specified.

Although these procedures were originally developed for PDA, the discriminant function coefficients (DFCs) that are produced can be used for DDA, that is, to quantify the relative importance of the measurement occasions for discriminating amongst groups (Thomas, 1992). In classical linear DA, it is known that bias and error variation of DFCs is influenced by a variety of characteristics of the data, including degree and pattern of separation between groups (i.e., group mean vectors), and magnitude of correlation among the outcome variables (Williams & Titus, 1998;

Williams, Titus, & Hines, 1991). However, to date, there has been little, if any research, about the effects of misspecifying the mean and/or covariance structure on DDA procedures for RM data.

The purpose of this study is to investigate the effects of RM mean and/or covariance misspecification on bias and error in DFCs of DDA procedures based on constant mean vectors and/or structured covariance matrices in univariate RM data. The manuscript is organized as follows: First, the investigated DA procedures are described. The results of a Monte Carlo study, which was conducted to investigate the effects of mean and covariance structure misspecification under a variety of data-analytic conditions, are presented. The manuscript concludes with some considerations about selecting a DDA procedure for RM data.

Estimation of DFCs in DA Procedures for RM Data

Throughout this manuscript, we focus on the case of $g = 2$ groups, although all procedures can also be generalized to $g > 2$. In general, the number of uncorrelated DFC vectors is equal to $g - 1$.

Let $\mathbf{y}_{ij}$ be the $p \times 1$ random vector of observed measurements for the ith study participant ($i = 1, \ldots, n_j$; $N = n_1 + n_2$) in the jth group ($j = 1, 2$). It is assumed that $\mathbf{y}_{ij} \sim \mathrm{N}_p(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, where $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ are the population mean vector and covariance for the jth group and are estimated by $\hat{\boldsymbol{\mu}}_j$ and $\hat{\boldsymbol{\Sigma}}_j$, respectively. The linear DFC vector is estimated by

$$\hat{\mathbf{a}} = \hat{\boldsymbol{\Sigma}}^{-1}(\hat{\boldsymbol{\mu}}_1 - \hat{\boldsymbol{\mu}}_2)\,. \tag{1}$$

For Fisher's (1936) linear DA procedure,

$$\hat{\boldsymbol{\Sigma}} = \frac{(n_1 - 1)\hat{\boldsymbol{\Sigma}}_1 + (n_2 - 1)\hat{\boldsymbol{\Sigma}}_2}{n_1 + n_2 - 2}, \tag{2}$$

and

$$\hat{\boldsymbol{\mu}}_j = \bar{\mathbf{y}}_j\,, \tag{3}$$

where $\bar{\mathbf{y}}_j = \dfrac{\sum_{i=1}^{n_j} \mathbf{y}_{ij}}{n_j}$. These quantities are estimated using the least-squares approach.

Roy and Khatree (2005a) proposed a DA procedure based on constant RM mean vectors and CS covariance structure. With a CS structure, $\mathbf{\Sigma}$ has diagonal elements σ^2 and off-diagonal elements $\sigma^2\rho$. For constant RM mean vectors, $\hat{\boldsymbol{\mu}}_j = c_j \mathbf{1}_p$ and the maximum likelihood (ML) estimate of c_j is

$$\hat{c}_j = \frac{\mathbf{1}_p^{\mathrm{T}} \overline{\mathbf{y}}_j}{p}, \tag{4}$$

where $\mathbf{1}_p$ is a $p \times 1$ vector of ones, $^{\mathrm{T}}$ is the transpose operator, and $\overline{\mathbf{y}}_j$ is the sample mean vector for the jth group. The ML estimates of σ^2 and ρ can be obtained by simultaneously solving the following system of equations.

$$-Np(1-\rho)(1+(p-1)\rho)\sigma^2 + (1+(p-1)\rho)(a_1+a_2) - \rho(b_1+b_2) = 0, \tag{5}$$

and

$$-N(p-1)p(1+(p-1)\rho)(1-\rho)\rho\sigma^2 - (a_1+a_2)(1+(p-1)\rho)^2 + (b_1+b_2)(\rho^2(p-1)+1) = 0, \tag{6}$$

where $a_1 = \mathrm{tr}(\mathbf{W}_1)$, $a_2 = \mathrm{tr}(\mathbf{W}_2)$, $b_1 = \mathrm{tr}(\mathbf{JW}_1)$, $b_2 = \mathrm{tr}(\mathbf{JW}_2)$, $\mathbf{J} = \mathbf{1}_p \mathbf{1}_p^{\mathrm{T}}$,

$$\mathbf{W}_j = \sum_{i=1}^{n_j} (\mathbf{y}_{ij} - \overline{\mathbf{y}}_j)(\mathbf{y}_{ij} - \overline{\mathbf{y}}_j)^{\mathrm{T}}, \tag{7}$$

and tr is the trace operator. The DFCs are estimated by substituting the ML estimates of $\mathbf{\Sigma}$ and $\boldsymbol{\mu}_j$ in equation 1.

Roy and Khattree (2005a) proposed a DA procedure based on constant RM mean vectors and AR-1 covariance structure. With an AR-1 structure, Σ has diagonal elements σ^2 and off-diagonal elements $\sigma^2\rho^l$, where l is the number of lags between measurement occasions. Estimates of c_j, σ^2, and ρ are obtained by simultaneously solving

$$(p-2)\rho c_j - pc_j + pm_{j1} - (p-2)\rho m_{j2} = 0, \tag{8}$$

$$\begin{aligned} &Np\sigma^2(1-\rho^2) - (\beta_1\rho^2 - 2\gamma_1\rho + \alpha_1) + n_1c_1(\beta_2\rho^2 - 2\gamma_2\rho + \alpha_2) \\ &+ n_2c_2(\beta_3\rho^2 - 2\gamma_3\rho + \alpha_3) - (n_1c_1^2 + n_2c_2^2)((p-2)\rho^2 - 2(p-1)\rho + p) = 0, \end{aligned} \tag{9}$$

and

$$\begin{aligned}&N(p-1)\sigma^2\rho - N(p-1)\sigma^2\rho^3 - \{\rho(\alpha_1+\beta_1)-\gamma_1\rho^2-\gamma_1\}\\&+n_1c_1\{\rho(\alpha_2+\beta_2)-\gamma_2\rho^2-\gamma_2\}+n_2c_2\{\rho(\alpha_3+\beta_3)-\gamma_3\rho^2-\gamma_3\}\\&-(n_1c_1^2+n_2c_2^2)\{\rho(2p-2)-(p-1)\rho^2-(p-1)\}=0.\end{aligned} \tag{10}$$

The details of these equations are provided in the Appendix. The estimates of the DFCs are obtained by substituting the ML estimates of $\mathbf{\Sigma}$ and $\boldsymbol{\mu}_j$ in equation 1.

For DA procedure based on constant RM mean vectors and unstructured covariance, the ML estimate of $\boldsymbol{\mu}_j$ is as shown in equation 3 and $\mathbf{\Sigma}$ is estimated as

$$\hat{\mathbf{\Sigma}} = \frac{\sum_{j=1}^{2}\mathbf{W}_j}{N}, \tag{11}$$

where $\mathbf{W}_j$ is obtained from equation 7.

Methodology

The investigated procedures in the Monte Carlo study were: (a) DA procedure based on unstructured mean vectors and unstructured covariances (UN), (b) DA procedure based on constant mean vectors and unstructured covariances (STUN), (c) DA procedure based on constant mean vectors and CS covariances (STCS), and (d) DA based on constant mean vectors and AR-1 covariances (STAR).

The following conditions were manipulated in the study: (a) number of repeated measurements (p), (b) total sample size (N), (c) group sizes, (d) pattern and magnitude of correlation among the repeated measurements, and (e) RM mean vector configuration. The number of groups (g = 2) and the population distribution (normal) were fixed.

The number of RMs was set at p = 3, 5, 7, and 9. Previous studies have considered values of p ranging from 3 to 10 (Roy & Khattree, 2005a; 2005b; Williams & Titus, 1988). Total sample sizes of N = 60, 90, and 120 were investigated, giving N/p ranging from 6.6 to 40.0.

Although previous simulation studies about DA procedures for RM data have primarily focused on equal group size conditions (Roy & Khattree, 2005a, 2005b), unequal group sizes have also been investigated for multivariate designs (Baron, 1991; He & Fung, 2000). The unequal group sizes selected for this study were (n_1, n_2) = (24, 36) for N = 60, (36, 54) for N = 90, and (48, 72) for N = 120. These were selected based on previous research (Baron, 1991; Lei & Koehly, 2003).

The standard error of DFCs is known to be influenced by the magnitude of correlation amongst the variables (Thomas & Zumbo, 1996). Six population correlation structures were investigated: (a) $\mathbf{Q}_1$: CS structure with parameter ρ = 0.3, (b) $\mathbf{Q}_2$: CS structure with ρ = 0.7, (c) $\mathbf{Q}_3$: AR-1 structure with ρ = 0.3, (d) $\mathbf{Q}_4$: AR-1 structure with ρ = 0.7, (e) $\mathbf{Q}_5$: unstructured with average correlation amongst the off-diagonal elements of 0.3, and (e) $\mathbf{Q}_6$: unstructured with average correlation amongst the off-diagonal elements of 0.7.

Pseudorandom observation vectors $\mathbf{y}_{ij}$ were generated from a multivariate normal distribution with mean $\boldsymbol{\mu}_j$ and correlation matrix $\mathbf{Q}_{mj} = \mathbf{Q}_m$ (m = 1 ,…, 6). A vector of standard normal deviates, $\mathbf{C}_{ij}$, was transformed to a vector of multivariate observations via $\mathbf{y}_{ij} = \boldsymbol{\mu}_j + \mathbf{L}\mathbf{C}_{ij}^{\mathrm{T}}$. Cholesky decomposition was used to obtain $\mathbf{L}$, an upper triangular matrix of dimension p satisfying the equality $\mathbf{L}^{\mathrm{T}}\mathbf{L} = \mathbf{Q}_{mj}$. Then $\mathbf{y}_{ij}$ was multiplied by $\mathbf{V}_j$, a diagonal matrix with elements σ_j to obtain multivariate observations with the desired variances and covariances, such that $\boldsymbol{\Sigma}_j = \mathbf{V}_j\mathbf{Q}_{mj}\mathbf{V}_j^{\mathrm{T}}$. We selected $\sigma_1^2 = \sigma_2^2$ =1 for all investigated conditions. The RANNOR function in SAS (SAS Institute Inc., 2008) was used to generate the standard normal deviates.

A variety of mean vector conditions have been investigated in previous research (Titus & Williams, 1988; Roy & Khattree, 2005a). In this study, three configurations for $\boldsymbol{\mu}_1$ were selected for each value of p (Table 1); for all conditions, $\boldsymbol{\mu}_2$ was the null vector. Configuration I had constant means for all RM occasions in both groups. Configurations II had non-constant RM mean with a quadratic, cubic, or polynomial pattern for the RM occasions in the first group and constant means in the second group. For Configuration III, a monotonic decreasing linear pattern was specified for the means in the first group and the means in the second group were constant.

Table 1. Configurations of $\boldsymbol{\mu}_1$ Investigated in the Simulation Study

p	I	II	III
3	(0.5, 0.5, 0.5)	(0.5, 1, 0.5)	(0.5, 0.25, 0)
5	(0.5, 0.5, 0.5, 0.5, 0.5)	(0.5, 1, 1.5, 1, 0.5)	(1, 0.75, 0.5, 0.25, 0)
7	(0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5)	(0.5, 1, 1.5, 2, 1.5, 1, 0.5)	(1.5, 1.25, 1, 0.75, 0.5, 0.25, 0)
9	(0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5)	(0.5, 1, 1.5, 2, 2.5, 2, 1.5, 1, 0.5)	(2, 1.75, 1.5, 1.25, 1, 0.75, 0.5, 0.25, 0)

Note: $\boldsymbol{\mu}_2$ was equal to the null vector for all conditions.

Overall 1493 combinations of simulation conditions were investigated with 5,000 replications for each combination. The study was conducted using SAS/IML software (SAS Institute Inc., 2008).

Two measures of performance were used to evaluate the DFCs: mean square error (MSE) and norm of the average bias (Crouxe & Dehon, 2001). The former is

$$b = \| \frac{1}{M} \sum_{k=1}^{M} (\hat{\mathbf{a}}_k - \mathbf{a}) \|, \tag{12}$$

and the latter is

$$e = \frac{1}{M} \sum_{k=1}^{M} \| \hat{\mathbf{a}}_k - \mathbf{a} \|^2 , \tag{13}$$

where $\mathbf{a}$ is the population vector of DFCs, $\|\mathbf{x}\|$ is the norm of $\mathbf{x}$ and M is the number of replications (i.e., M = 5000). Both measures takes values on the interval $[0, \infty)$ and the smaller the bias or error in the DFCs, the better. To adjust for the confounding effect of degree of separation between the two group means on bias and error, the MSE and bias in the DFCs were standardized using the distance between the two group mean vectors. Therefore,

$$b_{st} = \frac{b}{\| \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \|}, \tag{14}$$

and

$$e_{st} = \frac{e}{\| \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \|}. \tag{15}$$

Results

The average standardized MSE and bias values are summarized in Tables 2 to 5 for the four investigated values of p. As Table 2 reveals for p = 3, when the observations in both groups were sampled from populations with constant mean vectors (i.e., configuration I), MSE was smallest (and similar) for both the STCS and STAR DA procedures and largest for the UN procedure. When the data were sampled from a population with a non-constant mean configuration (i.e., configurations II or III), MSE and bias were smallest for either UN or STCS procedure and were substantially largest for STUN and STAR procedures. For example, under a CS covariance structure and when $\rho = 0.7$ and $p = 3$, UN and STAR procedures had the smallest and largest average MSE, respectively, when data were sampled from a population with mean configuration II, while UN and STUN procedures had the smallest and largest MSE respectively, when data were sampled from a population with mean configuration III.

For DA procedures based on constant mean vectors (i.e., STUN, STCS, and STAR), the average MSE decreased as the correlation among the RMs increased when the mean and covariance structure were correctly specified. This finding was observed regardless of the number of RMs. But when either the covariance or mean structure was misspecified, the average MSE increased as the correlation among the repeated measurements increased. For example, when $p = 3$ and under AR-1 population covariance structure, the average MSE for UN procedure was 0.35 and 0.64 when $\rho = 0.3$ and $\rho = 0.7$, respectively, while the average MSE of STAR procedure were 0.07 and 0.05 when $\rho = 0.3$ and $\rho = 0.7$, respectively, when data were sampled from a population with constant mean configuration (Table 2).

For DA procedures based on structured covariances, the average MSE and bias increased when the covariance structure was misspecified and the mean structures were correctly specified, regardless of the number of RMs. For example, under a AR-1 population covariance structure and when $\rho = 0.3$ and $p = 3$, the average MSE and bias of STCS procedure were 1.3 and 2.0 times the average MSE of STAR procedure, respectively, when the data were sampled from a population with mean configuration I. Similarly, the average MSE and bias of DA procedures based on structured covariances increased under a correctly specified population covariance but a misspecified mean structure. For example, when $p = 3$ and $\rho = 0.3$ under AR-1 population covariance structure, the average MSE and bias of the STAR procedure when the data were sampled from a population with mean configuration II were 6.4 and 7.0 times the average MSE and bias of STAR procedure under a constant mean configuration, respectively.

Table 2. Average standardized MSE and Bias by Covariance Structure, Magnitude of Correlation and Mean Configuration for p = 3

Covariance Structure	ρ	Mean Configuration	MSE				Bias			
			UN	STUN	STCS	STAR	UN	STUN	STCS	STAR
CS	0.3	I	0.34	0.11	**0.07**	0.09	0.08	0.08	**0.07**	0.15
		II	0.31	0.45	0.38	0.52	**0.09**	0..52	0.52	0.61
		III	0.52	0.64	0.61	0.63	**0.13**	0.98	0.98	0.98
	0.7	I	0.65	0.12	**0.05**	0.09	0.06	0.05	**0.05**	0.21
		II	0.65	1.89	1.81	2.38	**0.14**	1.20	1.20	1.38
		III	1.16	3.00	2.95	2.99	**0.25**	2.27	2.27	2.29
AR(1)	0.3	I	0.35	0.14	0.09	**0.07**	0.08	0.08	0.15	**0.08**
		II	0.30	0.56	0.33	0.44	**0.09**	0.59	0.47	0.56
		III	0.48	0.43	0.41	0.41	**0.11**	0.75	0.77	0.75
	0.7	I	0.64	0.13	0.08	**0.05**	0.06	0.06	0.22	**0.06**
		II	0.66	3.29	2.44	3.10	**0.16**	1.61	1.40	1.58
		III	1.01	1.11	1.06	1.06	**0.16**	1.34	1.36	1.34
UN	0.3	I	0.38	**0.13**	0.08	0.16	0.08	**0.08**	0.15	0.27
		II	**0.34**	0.33	0.41	0.53	**0.10**	0.42	0.54	0.60
		III	**0.61**	1.20	1.25	1.31	**0.18**	1.40	1.45	1.47
	0.7	I	0.67	**0.12**	0.05	0.12	0.06	**0.05**	0.08	0.27
		II	**0.66**	1.47	1.52	2.03	**0.13**	1.05	1.10	1.27
		III	**1.29**	4.34	4.41	4.48	**0.32**	2.77	2.81	2.83

Note: See Table 1 for a description of the mean configurations; CS = compound symmetric; AR-1 = first-order autoregressive; UN = unstructured; ρ = correlation parameter; UN = unstructured mean and covariance; STUN = structured mean and unstructured covariance; STCS = structured mean and CS covariance; STAR = structured mean and AR-1 covariance. Numbers in bold correspond to bias and error values of DA procedures for which the mean and covariance structures are correctly specified.

Table 3. Average standardized MSE and Bias by Covariance Structure, Magnitude of Correlation and Mean Configuration for p = 5

Covariance Structure	ρ	Mean Configuration	MSE				Bias			
			UN	STUN	STCS	STAR	UN	STUN	STCS	STAR
CS	0.3	I	0.56	0.14	**0.05**	0.13	0.06	0.06	**0.05**	0.21
		II	0.53	0.96	0.80	1.09	0.09	0.60	0.60	0.69
		III	0.63	1.21	1.13	1.16	0.12	0.89	0.89	0.91
	0.7	I	1.10	0.16	**0.02**	0.11	0.04	0.04	**0.03**	0.23
		II	1.35	4.40	4.19	5.20	0.18	1.39	1.39	1.54
		III	1.80	6.06	5.95	6.00	0.27	2.08	2.08	2.09
AR(1)	0.3	I	0.56	0.20	0.08	**0.05**	0.09	0.09	0.14	**0.07**
		II	0.46	0.76	0.37	0.48	0.09	0.48	0.38	0.45
		III	0.55	0.57	0.47	0.45	0.10	0.55	0.56	0.55
	0.7	I	1.06	0.21	0.08	**0.04**	0.05	0.05	0.22	**0.04**
		II	0.96	2.42	1.51	2.01	0.11	0.99	0.83	0.95
		III	1.08	0.86	0.76	0.72	0.10	0.72	0.74	0.72
UN	0.3	I	0.66	**0.20**	0.14	0.20	0.08	**0.08**	0.31	0.35
		II	**0.64**	2.26	1.33	1.67	**0.11**	0.96	0.77	0.86
		III	**0.75**	1.61	1.63	1.61	**0.15**	1.03	1.08	1.07
	0.7	I	1.15	**0.17**	0.03	0.10	0.04	**0.03**	0.07	0.22
		II	**1.40**	4.81	4.44	5.35	**0.18**	1.45	1.42	1.56
		III	**2.04**	7.57	7.66	7.76	**0.30**	2.33	2.36	2.37

Note: See Table 1 for a description of the mean configurations; CS = compound symmetric; AR-1 = first-order autoregressive; UN = unstructured; ρ = correlation parameter; UN = unstructured mean and covariance; STUN = structured mean and unstructured covariance; STCS = structured mean and CS covariance; STAR = structured mean and AR-1 covariance. Numbers in bold correspond to bias and error values of DA procedures for which the mean and covariance structures are correctly specified.

Table 4. Average standardized MSE and Bias by Covariance Structure, Magnitude of Correlation and Mean Configuration for p = 7

Covariance Structure	ρ	Mean Configuration	MSE				Bias			
			UN	STUN	STCS	STAR	UN	STUN	STCS	STAR
CS	0.3	I	0.78	0.19	**0.03**	0.17	0.06	0.06	**0.04**	0.27
		II	0.90	1.67	1.37	1.77	0.11	0.64	0.64	0.72
		III	0.97	1.96	1.81	1.83	0.15	0.86	0.86	0.87
	0.7	I	1.60	0.22	**0.02**	0.11	0.04	0.03	**0.02**	0.23
		II	2.72	7.68	7.31	8.56	0.22	1.48	1.48	1.60
		III	3.29	9.78	9.55	9.61	0.31	2.00	2.00	2.00
AR(1)	0.3	I	0.84	0.31	0.08	**0.04**	0.10	0.10	0.14	**0.07**
		II	0.87	1.16	0.43	0.58	0.10	0.44	0.34	0.41
		III	0.83	0.87	0.59	0.58	0.11	0.48	0.48	0.48
	0.7	I	1.56	0.31	0.08	**0.03**	0.06	0.06	0.20	**0.04**
		II	1.39	2.26	1.09	1.51	0.09	0.72	0.57	0.67
		III	1.42	0.96	0.70	0.70	0.09	0.51	0.54	0.51
UN	0.3	I	1.23	**0.33**	0.23	0.45	0.04	**0.04**	0.11	0.29
		II	**2.18**	4.70	7.21	7.64	0.24	1.51	1.55	1.68
		III	**2.54**	15.77	11.50	11.56	0.39	2.48	2.55	2.58
	0.7	I	1.73	**0.24**	0.03	0.15	0.05	**0.05**	0.05	0.34
		II	**2.94**	7.95	7.98	9.36	0.14	0.85	0.85	0.94
		III	**4.40**	14.93	15.59	15.84	0.19	1.20	1.20	1.22

Note: See Table 1 for a description of the mean configurations; CS = compound symmetric; AR-1 = first-order autoregressive; UN = unstructured; ρ = correlation parameter; UN = unstructured mean and covariance; STUN = structured mean and unstructured covariance; STCS = structured mean and CS covariance; STAR = structured mean and AR-1 covariance. Numbers in bold correspond to bias and error values of DA procedures for which the mean and covariance structures are correctly specified.

Table 5. Average standardized MSE and Bias by Covariance Structure, Magnitude of Correlation and Mean Configuration for p = 9

Covariance Structure	ρ	Mean Configuration	MSE				Bias			
			UN	STUN	STCS	STAR	UN	STUN	STCS	STAR
CS	0.3	I	1.33	0.31	**0.03**	0.25	0.07	0.07	**0.03**	0.33
		II	1.54	2.56	2.04	2.51	0.13	0.66	0.66	0.74
		III	1.64	2.88	2.53	2.59	0.16	0.84	0.84	0.85
	0.7	I	2.18	0.29	**0.01**	0.11	0.03	0.03	**0.02**	0.22
		II	5.14	11.58	10.97	12.40	0.29	1.54	1.54	1.64
		III	6.12	14.07	13.66	13.72	0.37	1.96	1.96	1.96
AR(1)	0.3	I	1.19	0.47	0.07	**0.04**	0.12	0.12	0.13	**0.07**
		II	0.98	1.41	0.51	0.75	0.09	0.41	0.31	0.40
		III	1.40	1.38	0.74	0.78	0.13	0.44	0.44	0.47
	0.7	I	2.17	0.46	0.07	**0.02**	0.07	0.07	0.19	**0.03**
		II	2.05	2.51	0.86	1.22	0.09	0.58	0.43	0.51
		III	2.03	1.27	0.69	0.70	0.10	0.41	0.43	0.41
UN	0.3	I	1.95	**0.47**	0.09	0.33	0.08	**0.07**	0.22	0.41
		II	**4.73**	10.85	12.28	12.84	0.32	1.46	1.63	1.67
		III	**6.85**	35.01	30.47	30.74	0.43	2.40	2.26	2.27
	0.7	I	2.86	**0.37**	0.01	0.12	0.04	**0.03**	0.06	0.23
		II	**8.52**	24.32	23.45	25.40	0.43	2.26	2.25	2.35
		III	**10.07**	32.21	31.44	32.00	0.56	2.98	2.97	2.99

Note: See Table 1 for a description of the mean configurations; CS = compound symmetric; AR-1 = first-order autoregressive; UN = unstructured; ρ = correlation parameter; UN = unstructured mean and covariance; STUN = structured mean and unstructured covariance; STCS = structured mean and CS covariance; STAR = structured mean and AR-1 covariance; Numbers in bold correspond to bias and error values of DA procedures for which the mean and covariance structures are correctly specified.

For STUN procedure, the average bias increased when the mean and covariance structures were misspecified, but STCS procedure had the smallest MSE when the data were sampled from a population with a constant mean configuration, regardless of the number of RM. For example, when $p = 7$, under an unstructured population covariance structure and when $\rho = 0.3$ and $p = 7$, the average MSE and bias of STUN procedure were 0.70 and 2.75 times the average MSE and bias of STCS procedures, respectively, when the data were sampled from a population with a constant mean configuration (Table 4).

Moreover, for each DA procedure, the average MSE and bias due to misspecification of the covariance structure increased as the magnitude of correlation and number of RMs increased. For example, when $p = 5$ and under a CS population covariance structure, the average MSEs of STAR procedure were 2.6 and 5.5 times the average MSE of STCS procedure for $\rho = 0.3$ and $\rho = 0.7$, respectively, when data were sampled from a population with a constant mean configuration (Table 3). The corresponding bias values for STAR procedure were 4.2 and 10.7 times the bias of STCS procedure when $\rho = 0.3$ and $\rho = 0.7$, respectively. Similarly, when $p = 9$, the average MSEs of STCS procedure were 8,3 and 11.0 times the average MSE of STAR for $\rho = 0.3$ and $\rho = 0.7$, respectively, while the corresponding average bias values were 11.0 times the average bias of STCS procedure when $\rho = 0.3$ and $\rho = 0.7$ (Table 5).

Finally, analyses revealed that the average MSE for each of the DA procedures decreased as the sample size increased. For example, the average MSEs of UN procedure were 7.82, 3.77, and 2.50 when $N = 60$, 90, and 120 respectively. In contrast, the average bias for each DA procedure remained largely unchanged as the sample size increased, regardless of the mean configuration and number of RM. For example, the overall average bias of STAR procedure were 2.12, 2.10 and 2.10 when $N = 60$, 90, and 120, respectively.

Conclusions

This manuscript investigated the effects of RM mean and/or covariance structure misspecification on bias and error in DFCs for DA procedures based on parsimonious mean and/or covariance structures. As expected, the bias and error in the DFCs of the investigated procedures increased when the RM mean and/or covariance structures were misspecified. The average bias and error variation due to misspecification of the RM mean structure was greater than the average bias and error variation due to RM covariance structure misspecification for all of the investigated procedures. While DA procedures based on parsimonious RM mean and covariance structures had negligible bias when the mean and covariances are correctly specified, UN DA procedure had the smallest bias when the data were sampled from a population with non-constant mean configuration.

Based on the study findings, we recommend adopting a DA procedure based on unstructured mean vectors and covariance matrices when the researcher has prior knowledge to suggest that the mean longitudinal profile for each group will change across the repeated measures occasions. If the mean longitudinal profile in each group is not expected to increase or decrease across the measurement occasions, either the STCS or STAR procedure are recommended because they require estimation of the fewer number of parameters, although any of the procedures can be expected to perform well in terms of both bias and error variation.

To reduce the effect of mean and/or covariance structure misspecification on bias and error in the DFCs, preliminary tests of model fit could be undertaken before adopting a DDA procedure for RM data. Graphical exploration of the data, likelihood ratio tests, or penalized log-likelihood measures like the Akaike information criterion have all been proposed to guide the specification of mean and covariance structures (Fitzmaurice, Laird, & Ware, 2004)

The limitations of this study should be noted. We focused on normally distributed data. The impact of mean and/or covariance misspecification on bias and error in the DFCs when data are sampled from non-normal distribution has not been investigated. While mild departures from multivariate non-normality are known to have little effect on classification accuracy of classical DA procedure (Ashikaga & Chang, 1981), classification accuracy can be severely affected under large departures (Lachenbruch, Sneeringer, & Revo, 1973; Baron, 1991; McLachlan, 1992). Inferences about DFCs of the linear DA procedures may also be affected by the degree of departure from the assumption of multivariate normality (McLachlan, 1992). The DA procedures considered in manuscript also focused only on complete data, an assumption which may not be satisfied in RM studies, which are often characterized by missing observations and unbalanced measurements occasions (Fairclough et al, 1998). In the simulation study, the RM variances were assumed to be constant across variables and groups. Linear DA procedures rest on the assumption of covariance homogeneity (Huberty & Olejnik, 2006). Departures from this assumption may result in reduced classification accuracy (Solberg, 1988). DFCs have been shown to be relatively robust to violation of this assumption when the data are normally distributed (Owen & Chmielewski, 1985), but it is not known if this robustness will continue to be evident when the covariance and/or mean vector is misspecified.

A number of opportunities for future research exist in the development of DDA procedures for RM data. Although several studies have examined the effects of population distribution on classification accuracy, there is limited investigation of the effects of population distribution and other data characteristics on bias and error in DFCs. Existing studies in this area have only focused on the effects of sample size, number of outcome variables, and mean configuration on bias and variation in DFCs when data were sampled from normally distributed data (Williams & Titus, 1991; Owen & Chmielewski, 1985). This study investigated DA procedures based on constant mean vectors and/or structured covariances. However, the assumption of a constant repeated measures group mean structure may not be tenable when the interest is in the assessment of the relative importance of measurement occasions that discriminate between groups. DA procedures based on non-constant mean vectors and CS or AR-1 covariance structures can be further investigated. These procedures which assume non-constant mean configurations and parsimonious structures will be useful for assessing the relative importance of information collected at each measurement occasions in univariate repeated measures studies.

In summary, although the adoption of a DA procedure based on a parsimonious mean and/or covariance structure can reduce the number of parameters to estimate, which is beneficial when sample size is small (Roy & Khattree, 2005a), this study shows that bias and error variation in the DFCs can be large, particularly when there is misspecification of the RM mean structure. A researcher's choice of a DA procedure for RM data is dependent, in part, on the trade-off between parsimony in parameter estimation and bias and/or error in the DFCs.

Acknowledgements

This research was supported by a Canadian Institutes of Health Research (CIHR) Vanier Graduate Scholarship to the first author, and a CIHR New Investigator Award to the second author.

References

Ashikaga, T., & Chang, P. C. (1981). Robustness of Fisher's linear discriminant function under two-component mixed-normal models. *Journal of the American Statistical Association, 76;* 375 – 676.

Baron, A. E. (1991). Misclassification among methods used for multiple group discrimination – The effects of distributional properties. *Statistics in Medicine*, *10*, 757 – 766.

Beaumont, J. L., Lix, L. M., Yost, K. J., & Hahn, E. A. (2006). Application of robust statistical methods for sensitivity analysis of health-related quality of life outcomes. *Quality of Life Research, 15*; 349 – 356.

Croux, C., & Dehon, C. (2001). Robust linear discriminant analysis using S-estimators. *Canadian Journal of Statistics, 29;* 473 – 493.

Fairclough, D. L., Peterson, H. F., Cella, D., Bonomi, P. (1998). Comparison of several model-based methods for analysing incomplete quality of life data in cancer clinical trials. *Statistics in Medicine, 17;* 781 – 796.

Feighner, J. P., & Sverdlov, L. (2002). The use of discriminant analysis to separate a study population by treatment subgroups in a clinical trial with a new pentapeptide antidepressant. *Journal of Applied Research, 2*; 17 – 18.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, *7*, 179 – 188.

Fitzmaurice, G., Laird, N. M., & Ware, J. H. (2004). *Applied Longitudinal Analysis*. New Jersey: Wiley.

He, X., & Fung, W. K. (2000). High breakdown estimation for multiple populations with applications to discriminant analysis. *Journal of Multivariate Analysis, 72*; 151 – 162.

Huberty, C. J, & Wisenbaker, J. M. (1992). Variable importance in multivariate group comparisons. *Journal of Educational Statistics, 17;* 75 – 91.

Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and Discriminant Analysis*. New York: Wiley.

Lachenbruch, P. A., Sneeringer, C., & Revo, L. T. (1973). Robustness of the linear and quadratic discriminant function to certain types of non-normality. *Communications in Statistics, 1;* 39 – 57.

Lei, P., & Koehly, L. M. (2003). Linear discriminant analysis versus logistic regression: a comparison of classification errors in the two-group case. *Journal of Experimental Education, 72*; 25 – 49.

Levesque, L., Ducharme, F., Zarit, S. H., Lachance, L., & Giroux, F. (2008). Predicting longitudinal patterns of psychological distress in older husband caregivers: further analysis of existing data. *Aging Mental Health*, 12, 333 – 343.

Marshall, G., & Baron, A. E. (2000). Linear discriminant models for unbalanced longitudinal data. *Statistics in Medicine, 19*; 1969 – 1981.

McLachlan, G. J. (1992). *Discriminant Analysis and Statistical Pattern Recognition*. New York: Wiley.

Owen, J. G., Chmielewski, M. A. (1985). On canonical variates analysis and the construction of confidence ellipses in systematic studies. *Systematic Zoology, 34;* 366 – 374.

Roy, A., & Khattree, R. (2005a). Discrimination and classification with repeated measures data under different covariance structures. *Communications in Statistics – Simulation and Computation, 34*; 167 – 178.

Roy, A., & Khattree, R. (2005b). On discrimination and classification with multivariate repeated measures data. *Journal of Statistical Planning and Inference, 134;* 462 – 485.

Roy, A., & Khattree, R. (2007). Classification of multivariate repeated measures data with temporal autocorrelation. *Advances in Data Analysis and Classification, 1*; 175 - 199.

SAS Institute Inc. (2008). *SAS/IML user's guide, version 9.2.* Cary, NC: SAS Institute, Inc.

Solberg, H. E., 1988. Discriminant analysis. *Critical Reviews in Clinical Laboratory Sciences, 9*; 209 – 242.

Thomas, D. R. (1992). Interpreting discriminant functions: a data analytic approach. *Multivariate Behavioral Research, 27*; 335 – 362.

Thomas, D. R., & Zumbo, B. D. (1996). Using a measure of variable importance to investigate the standardization of discriminant coefficients. *Journal of Educational and Behavioral*

Statistics, 21; 110 – 130.

Tomasko, L., Helms, R. W., & Snappin, S. M. (1999). A discriminant analysis extension to mixed models. *Statistics in Medicine, 18*; 1249 – 1260.

Williams, B. K., & Titus, K. (1988). Assessment of sampling stability in ecological applications of discriminant analysis. *Ecology, 69;* 1275 – 1285.

Williams, B. K., Titus, K., Hines, J. E. (1991). Stability and bias of classification rates in biological applications of discriminant analysis. *The Journal of Wildlife Management, 54;* 331 – 341.

Appendix: ML Estimates for the AR-1 Covariance Structure

As described earlier in the manuscript, more details about ML estimation of the coefficients of STAR procedure is provided here. In equation 8,

$$m_{j1} = \frac{\mathbf{1}_p^{\mathrm{T}}\bar{\mathbf{y}}_j}{p}, \tag{A-1}$$

$$m_{j2} = \frac{\mathbf{1}_p^{\mathrm{T}}\bar{\mathbf{y}}_j - \bar{\mathbf{y}}_{j1} - \bar{\mathbf{y}}_{jp}}{(p-2)}, \tag{A-2}$$

and $\bar{\mathbf{y}}_{j1}$ and $\bar{\mathbf{y}}_{jp}$, are respectively, the first and *p*th elements of the vector $\bar{\mathbf{y}}_j$. In equations 9 and 10, $\beta_1 = \mathrm{tr}(\mathbf{W}_0) - \mathbf{W}_{0,11} - \mathbf{W}_{0,pp}$, $\beta_2 = \alpha_1 - \mathbf{W}_{5,11} - \mathbf{W}_{5,pp}$, and $\beta_3 = \alpha_3 - \mathbf{W}_{6,11} - \mathbf{W}_{6,pp}$. Furthermore, $\alpha_1 = \mathrm{tr}(\mathbf{W}_0 + \mathbf{W}_1 + \mathbf{W}_2 + \mathbf{W}_3 + \mathbf{W}_4)$, $\alpha_2 = \mathrm{tr}(\mathbf{W}_5)$, and $\alpha_3 = tr(\mathbf{W}_6)$; $\mathbf{W}_0 = \mathbf{W} + \mathbf{W}_3 + \mathbf{W}_4$. Also,

$$\gamma_1 = \sum_{k=2}^{p} \mathbf{W}_{0,k-1k}, \tag{A-3}$$

$$\gamma_2 = \sum_{k=2}^{p} \mathbf{W}_{5,k-1k}, \tag{A-4}$$

and

$$\gamma_3 = \sum_{k=2}^{p} \mathbf{W}_{6,k-1k} \tag{A-5}$$

where $\mathrm{Wu}_{,k-1k}$ is the (k-1,k)th element of $\mathbf{W}_{\mathrm{u}}$ (u = 0,…,6) and k = 1,…,p.

In these equations,

$$\mathbf{W} = \sum_{j=1}^{2}\sum_{i=1}^{n_j}(\mathbf{y}_{ij} - \bar{\mathbf{y}}_j)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_j)^{\mathrm{T}} \tag{A-6}$$

$\mathbf{W}_3 = \bar{\mathbf{y}}_1\bar{\mathbf{y}}_1^{\mathrm{T}}$, $\mathbf{W}_4 = \bar{\mathbf{y}}_2\bar{\mathbf{y}}_2^{\mathrm{T}}$, $\mathbf{W}_5 = \mathbf{1}_p^{\mathrm{T}}\bar{\mathbf{y}}_1 + \bar{\mathbf{y}}_1\mathbf{1}_p^{\mathrm{T}}$, and $\mathbf{W}_6 = \mathbf{1}_p^{\mathrm{T}}\bar{\mathbf{y}}_1 + \bar{\mathbf{y}}_1\mathbf{1}_p^{\mathrm{T}}$.