Supplementary File for the Manuscript entitled "Modeling and Prediction of Multiple Correlated Functional Outcomes"

A Computation for EM Algorithm

A.1 The E-Step

Let $E^*(\cdot)$ denote the expectation given the observed data $\{\widehat{R}_{ip}\}$. The goal of this step is to compute $E^*(Z_{ip})$, $E^*(Z_{ip}Z_{ip}^T)$ and $E^*(Z_{ip}, Z_{ip'})$. Suppose $Z_i|\widehat{R}_i \sim \mathrm{N}(m_i, V_i)$. That is,

$$f(Z_i|\widehat{R}_i) \propto \exp\left\{-(1/2)(Z_i - m_i)^T V_i^{-1}(Z_i - m_i)\right\}. \tag{1}$$

Since $f(Z_i|\widehat{R}_i) \propto f(Z_i, \widehat{R}_i)$ and

$$\begin{aligned} f(Z_i, \widehat{R}_i) &\propto \exp\left[-\frac{1}{2}\left\{\sum_p \frac{1}{\sigma_{\epsilon p}^2}\left(\widehat{R}_{ip} - B\Theta_p Z_{ip}\right)^T\left(\widehat{R}_{ip} - B\Theta_p Z_{ip}\right) + Z_i^T\Sigma^{-1}Z_i\right\}\right] \\ &\propto \exp\left\{-\frac{1}{2}\left(\sum_p \frac{1}{\sigma_{\epsilon p}^2} Z_{ip}^T\Theta_p^T B^T B\Theta_p Z_{ip} + Z_i^T\Sigma^{-1}Z_i\right) + \sum_p \frac{1}{\sigma_{\epsilon p}^2}\widehat{R}_{ip}^T B\Theta_p Z_{ip}\right\} \\ &= \exp\left\{-\frac{1}{2}Z_i^T(E^TE + \Sigma^{-1})Z_i + M_iZ_i\right\}, \end{aligned} \tag{2}$$

where

$$M_i = \left[\frac{1}{\sigma_{\epsilon 1}^2}\widehat{R}_{i1}^T B\Theta_1 \quad \frac{1}{\sigma_{\epsilon 2}^2}\widehat{R}_{i2}^T B\Theta_2 \quad \cdots \quad \frac{1}{\sigma_{\epsilon P}^2}\widehat{R}_{iP}^T B\Theta_P\right]_{1\times\sum_p K_p} \tag{3}$$

$$E = \begin{bmatrix} \frac{1}{\sigma_{\epsilon 1}}B\Theta_1 & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sigma_{\epsilon P}}B\Theta_P \end{bmatrix}_{mP\times\sum_p K_p}. \tag{4}$$

Comparing the two expressions, we have that

$$V_i = \left(E^TE + \Sigma^{-1}\right)^{-1} = V; \tag{5}$$

$$m_i = V^T M_i^T. \tag{6}$$

Suppose the matrix V and vectors m_i obtained from above take the forms of

$$V = \begin{pmatrix} V_{11} & \cdots & V_{1P} \\ \vdots & \ddots & \vdots \\ V_{P1} & \cdots & V_{PP} \end{pmatrix}, m_i = \begin{pmatrix} m_{i1} \\ \vdots \\ m_{iP} \end{pmatrix}. \tag{7}$$

The predictions for Z_{ip} and their moments required by the EM algorithm are

$$\begin{aligned} E^*(Z_{ip}) &= m_{ip}; \\ E^*(Z_{ip}Z_{ip}^T) &= m_{ip}m_{ip}^T + V_{pp}; \\ E^*(Z_{ip}Z_{ip'}^T) &= m_{ip}m_{ip'}^T + V_{pp'}. \end{aligned} \tag{8}$$

A.2 The M-Step

In the optimization, parameters are updated in order. The values of the parameters on the right hand side are plugged in using their current values. The details are as follows.

Step 1: Update the estimates for $\sigma_{\epsilon p}^2, p = 1, \ldots, P$. We want to minimize

$$E^* \left[\sum_{i=1}^{N} \left\{ m \log(\sigma_{\epsilon p}^2) + \frac{1}{\sigma_{\epsilon p}^2} \left(\widehat{R}_{ip} - B\Theta_p Z_{ip} \right)^T \left(\widehat{R}_{ip} - B\Theta_p Z_{ip} \right) \right\} \right].$$

The expectation of the quadratic form can be written as

$$\begin{aligned} & E^* \left\{ \left(\widehat{R}_{ip} - B\Theta_p Z_{ip} \right)^T \left(\widehat{R}_{ip} - B\Theta_p Z_{ip} \right) \right\} \\ = \; & \widehat{R}_{ip}^T \widehat{R}_{ip} - 2\widehat{R}_{ip}^T B\Theta_p E^*(Z_{ip}) + E^* \left(Z_{ip}^T \Theta_p^T B^T B\Theta_p Z_{ip} \right) \\ = \; & \widehat{R}_{ip}^T \widehat{R}_{ip} - 2\widehat{R}_{ip}^T B\Theta_p m_{ip} + \operatorname{tr} \left\{ (\Theta_p^T B^T B\Theta_p) E^*(Z_{ip}Z_{ip}^T) \right\} \\ = \; & \widehat{R}_{ip}^T \widehat{R}_{ip} - 2\widehat{R}_{ip}^T B\Theta_p m_{ip} + \operatorname{tr} \left\{ (\Theta_p^T B^T B\Theta_p)(V_{pp} + m_{ip}m_{ip}^T) \right\} \\ = \; & \widehat{R}_{ip}^T \widehat{R}_{ip} - 2\widehat{R}_{ip}^T B\Theta_p m_{ip} + \operatorname{tr}(m_{ip}^T \Theta_p^T B^T B\Theta_p m_{ip}) + \operatorname{tr}(B\Theta_p V_{pp} \Theta_p^T B^T) \\ = \; & (\widehat{R}_{ip} - B\Theta_p m_{ip})^T (\widehat{R}_{ip} - B\Theta_p m_{ip}) + \operatorname{tr}(B\Theta_p V_{pp} \Theta_p^T B^T). \end{aligned} \tag{9}$$

So, we want to minimize

$$Nm \log(\sigma_{\epsilon p}^2) + \frac{1}{\sigma_{\epsilon p}^2} \sum_{i=1}^{N} \left\{ (\widehat{R}_{ip} - B\Theta_p m_{ip})^T (\widehat{R}_{ip} - B\Theta_p m_{ip}) + \operatorname{tr}(B\Theta_p V_{pp} \Theta_p^T B^T) \right\}. \tag{10}$$

Hence, we update $\sigma_{\epsilon p}^2$ by

$$\widehat{\sigma}_{\epsilon p}^2 = \frac{1}{Nm} \sum_{i=1}^{N} \left\{ (\widehat{R}_{ip} - B\Theta_p m_{ip})^T (\widehat{R}_{ip} - B\Theta_p m_{ip}) + \operatorname{tr}(B\Theta_p V_{pp} \Theta_p^T B^T) \right\}, \tag{11}$$

for each $p = 1, \ldots, P$.

Step 2: Update Θ_p, by updating each column Θ_{pk} sequentially. For each k, we want to minimize

$$\sum_{i=1}^{N} E^* \left(\|\widehat{R}_{ip} - \sum_{l \neq k} B\Theta_{pl} Z_{ipl} - B\Theta_{pk} Z_{ipk}\|^2 \right) + \sigma_{\epsilon p}^2 \lambda_p \Theta_{pk}^T \int b''(t) b''(t)^T dt \Theta_{pk}. \tag{12}$$

Similar to the first step, the expression is minimized when

$$\begin{aligned}\widehat{\Theta}_{pk} &= \left\{\sum_{i=1}^{N} E^*(Z_{ipk}^2)B^TB + \sigma_{\epsilon p}^2\lambda_p \int b''(t)b''(t)^T dt\right\}^{-1} \\ &\quad \times \sum_{i=1}^{N} B^T \left\{\widehat{R}_{ip}E^*(Z_{ipk}) - \sum_{l\neq k} B\Theta_{pl}E^*(Z_{ipk}Z_{ipl})\right\}, \end{aligned} \tag{13}$$

where $E^*(Z_{ipk}^2)$, $E^*(Z_{ipk})$ and $E^*(Z_{ipk}Z_{ipl})$ are the $(k,k)-, k-$ and $(k,l)-$ element of $m_{ip}m_{ip}^T + V_{pp}$, m_{ip} and $m_{ip}m_{ip'}^T + V_{pp'}$, respectively.

Step 3: Update Σ. We want to minimize

$$E^*\left\{\sum_{i=1}^{N}\left(\log|\Sigma| + Z_i^T\Sigma^{-1}Z_i\right)\right\} = \sum_{i=1}^{N}\left[\log|\Sigma| + \operatorname{tr}\left\{\Sigma^{-1}(V + m_i m_i^T)\right\}\right]. \tag{14}$$

Since $S = \sum_i \left(V + m_i m_i^T\right)$ is positive-definite, there is a unique positive-definite matrix $S^{1/2}$ such that $S^{1/2}S^{1/2} = S$. Let $W = S^{1/2}\Sigma^{-1}S^{1/2}$. Then we want to minimize

$$N\log|\Sigma| + \operatorname{tr}(S\Sigma^{-1}) = -N\log|W| + N\log|S| + \operatorname{tr}(W). \tag{15}$$

The positive-definite W can be diagonalized to get the diagonal elements $\tau_1, \ldots, \tau_K$, where $K = \sum_{p=1}^{P} K_p$. What we finally want to minimize is $-N\sum_{k=1}^{K}\log\tau_k + \sum_{k=1}^{K}\tau_k$. The solution is $\tau_k = N$ for all k. That is, $W = NI_K$ which implies that $\Sigma = S^{1/2}W^{-1}S^{1/2} = \frac{1}{N}S$. Hence, in this step, we update Σ by

$$\widehat{\Sigma} = \frac{1}{N}S = \frac{1}{N}\sum_i (V + m_i m_i^T). \tag{16}$$

Step 4: Orthogonalization and update of the covariance matrix. Suppose the matrix $\widehat{\Sigma}$ obtained in step 3 takes the form

$$\widehat{\Sigma} = \begin{pmatrix} \widehat{D}_1 & \widehat{C}_{12} & \cdots & \widehat{C}_{1P} \\ \widehat{C}_{21} & \widehat{D}_2 & \cdots & \widehat{C}_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{C}_{P1} & \widehat{C}_{P2} & \cdots & \widehat{D}_P \end{pmatrix}. \tag{17}$$

Let $\widehat{\Theta}_p\widehat{D}_p\widehat{\Theta}_p^T = Q_pS_pQ_p^T$ be the eigenvalue decomposition in which Q_p has orthogonal columns, and S_p is diagonal with diagonal elements in a decreasing order. We update $\widehat{\Theta}_p$ by Q_p and $\widehat{D}_p$ by S_p. Since the updating in this step corresponds to the transformation $Z_{ip} \leftarrow Q_p^T\widehat{\Theta}_pZ_{ip}$, we update $\widehat{C}_{pp'}$ by $Q_p^T\widehat{\Theta}_p\widehat{C}_{pp'}\widehat{\Theta}_{p'}^TQ_{p'}$. The sign of each column of $\widehat{\Theta}_p$ is changed as necessary so that the elements with the largest magnitude of each column are positive.

B Simulation Results

Figure S1 shows the estimates of mean, standard deviation and shape parameter functions from 100 simulated datasets. The black lines are the true values, and the gray lines are the estimates. The data were generated using a skew normal distribution as explained in the main paper. Even though there seems to be high variation in the estimates when $\alpha \approx 10$, the density of skew normal with α equal to 10 is not much different from when α is much higher. This means that we can still obtain a good estimate for the distribution even though the estimate for α in this range is not very accurate.

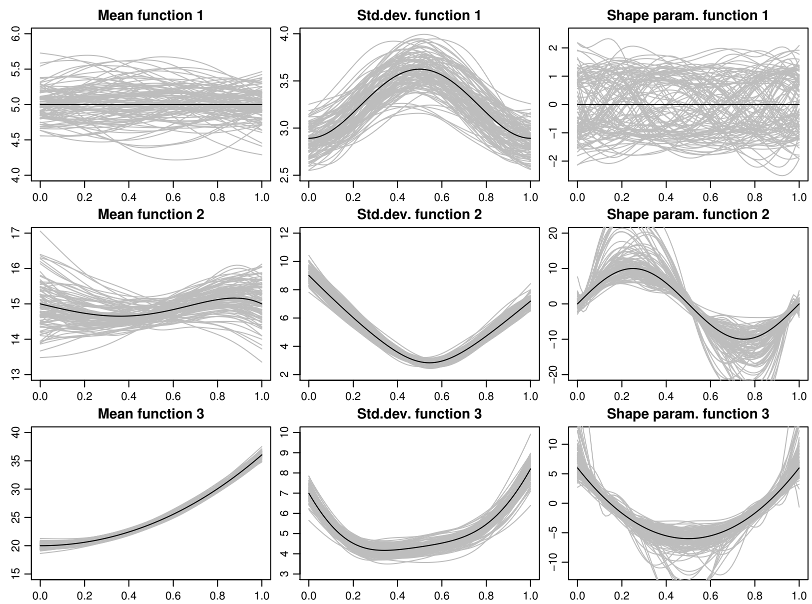

Figure S1: The estimates from 100 datasets of mean (first column), standard deviation (second column) and shape parameter (third column) functions for the simulated data. Black lines are true values.

Figure S2 shows the true and estimated correlations and cross correlations. The true correlations within an outcome are shown in rows 1 and the true cross correlations are shown in row 4. The estimated correlations within an outcome from 2 datasets are shown in rows 2 and 3. The estimated cross correlations from the same 2 datasets are shown in rows 5 and 6.The estimated variances of each latent process are shown in Figure S3. The values close to 1 indicate the closeness to our assumption that the latent processes have marginal variance of 1.

Table S1 shows the square root of the integrated mean square error (IMSE), integrated square bias (IBIAS) and integrated variance (IVAR) for the marginal parameter function and covariance parameters, as defined in the main paper. The contour plots of the pointwise square root of the mean square error for the covariance estimates are shown in Figure S4.

Table S1: Estimates of the square roots of IMSE, IBIAS and IVAR for the simulated data from 100 datasets

Parameter	$\sqrt{\text{IMSE}} \times 10^2$	$\sqrt{\text{IBIAS}} \times 10^2$	$\sqrt{\text{IVAR}} \times 10^2$
$\mu_1(t)$	23.66	2.51	23.52
$\mu_2(t)$	36.87	1.90	36.83
$\mu_3(t)$	36.66	3.94	35.84
$\sigma_1(t)$	16.72	2.20	16.58
$\sigma_2(t)$	29.24	1.78	29.19
$\sigma_2(t)$	28.57	3.36	28.37
$\alpha_1(t)$	98.00	6.93	97.76
$\alpha_2(t)$	210.7	72.42	197.9
$\alpha_2(t)$	105.6	16.71	104.2
$\text{cov}\,(R_1(t), R_1(t))$	4.25	0.43	4.23
$\text{cov}\,(R_2(t), R_2(t))$	4.58	0.57	4.55
$\text{cov}\,(R_3(t), R_3(t))$	4.52	0.73	4.46
$\text{cov}\,(R_1(t), R_2(t))$	6.30	0.64	6.27
$\text{cov}\,(R_1(t), R_3(t))$	6.49	0.98	6.42
$\text{cov}\,(R_2(t), R_3(t))$	6.36	0.66	6.32

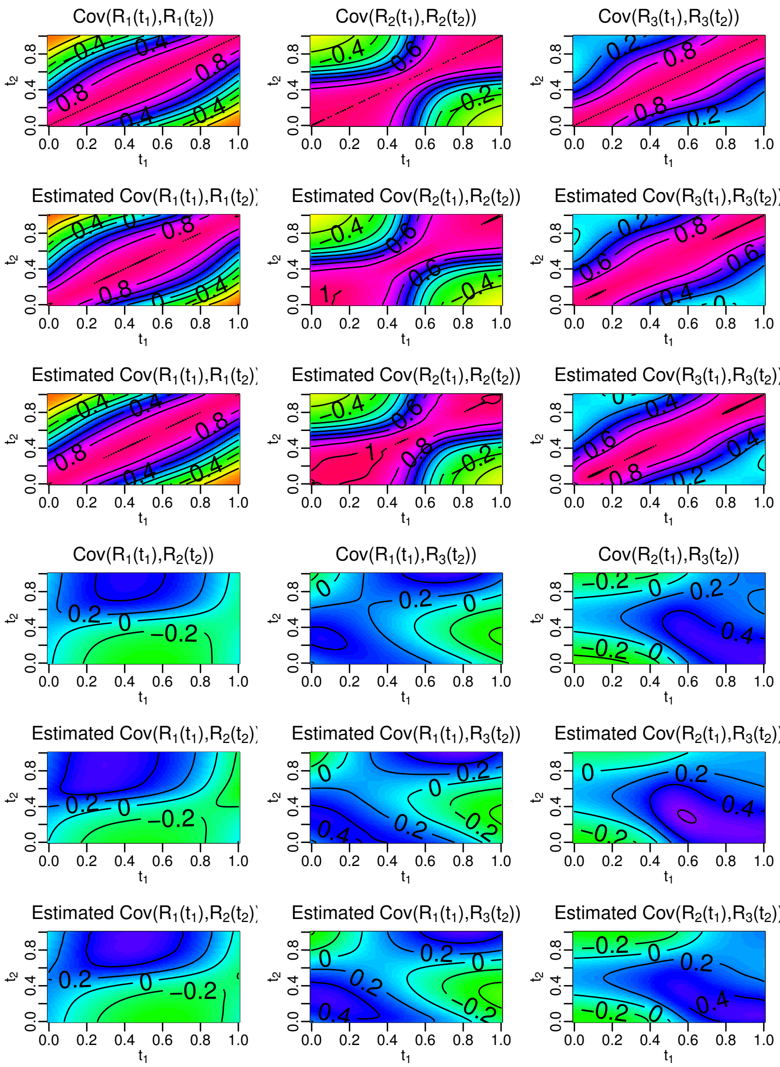

Figure S2: True and estimated covariances within each outcome and cross-covariances between outcomes. The true covariances and cross covariances are shown in rows 1 and 4, respectively. The estimated covariances within each outcome from 2 datasets are shown in rows 2 and 3. Rows 5 and 6 are the estimated cross-covariances from the same datasets. Thus, rows 2 and 3 should be compared with row 1 and rows 5 and 6 with row 4.

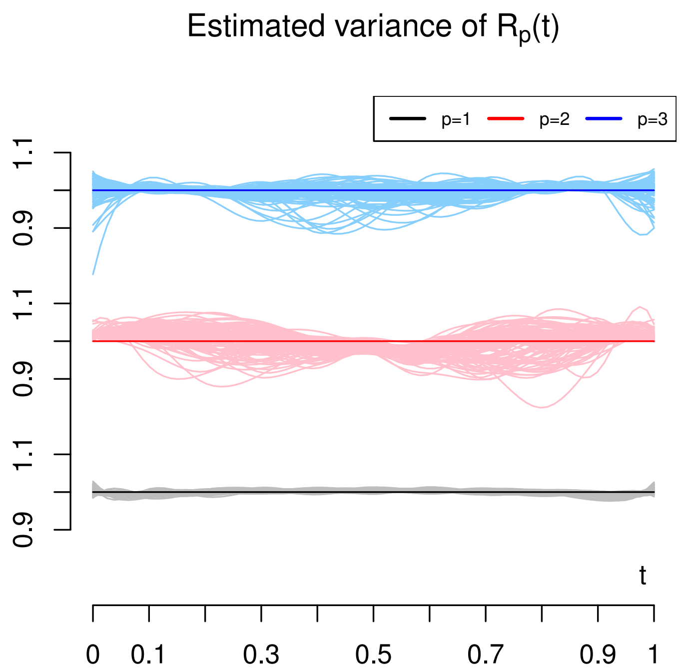

Figure S3: 100 estimates of the variance of the latent processes for the simulated data. Black, red and blue lines are the true values, which are 1.

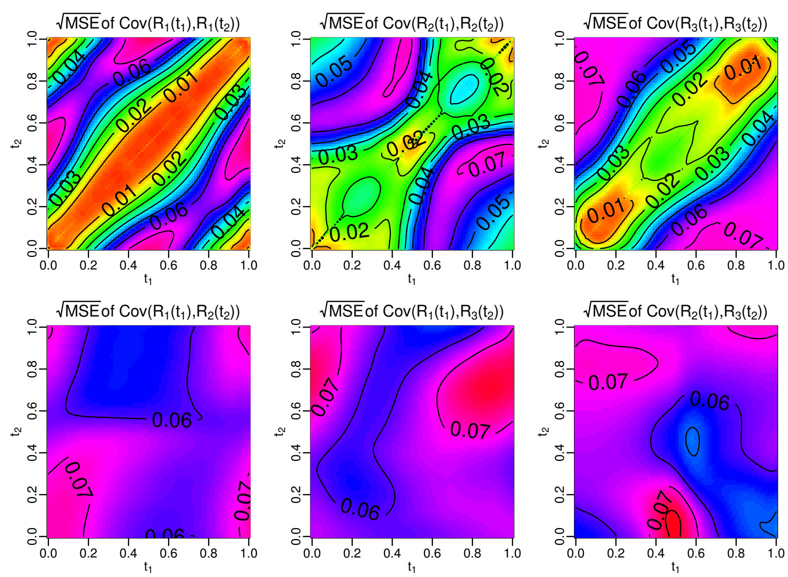

Figure S4: Pointwise square root of MSE for covariance within and between outcomes of the 100 estimates for the simulated data.

C Additional DTI Results

Figure S5 shows the estimates for the mean, variance and skewness for the three outcomes for both groups. The black lines are the estimates for the MS group and the red lines are the estimates for the healthy control group.

Figure S6 and Figure S7 display the estimated correlations within each outcome and the estimated cross correlations between outcomes for the healthy and MS groups.

Figure S8 shows the differences of the correlations within each outcome and between outcomes between the healthy and MS groups.

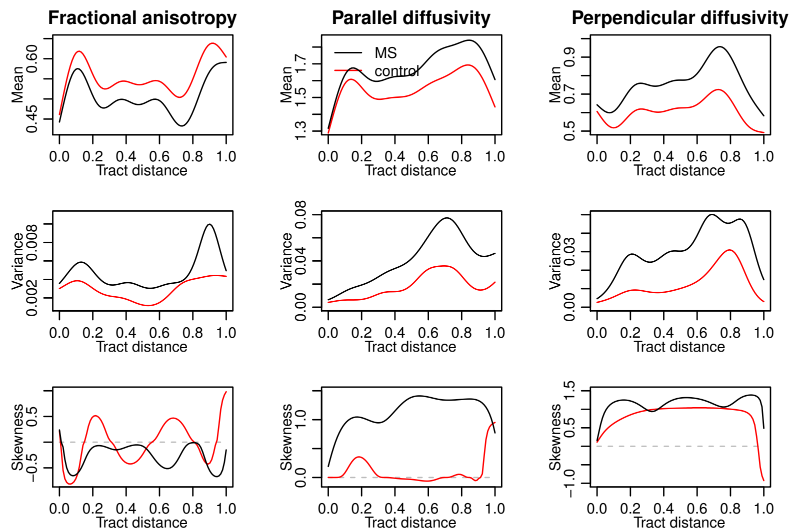

Figure S5: Estimates of the mean, variance and skewness for the three outcomes for the healthy control (red lines) and MS (black lines) groups. Gray dashed lines indicate zero skewness.

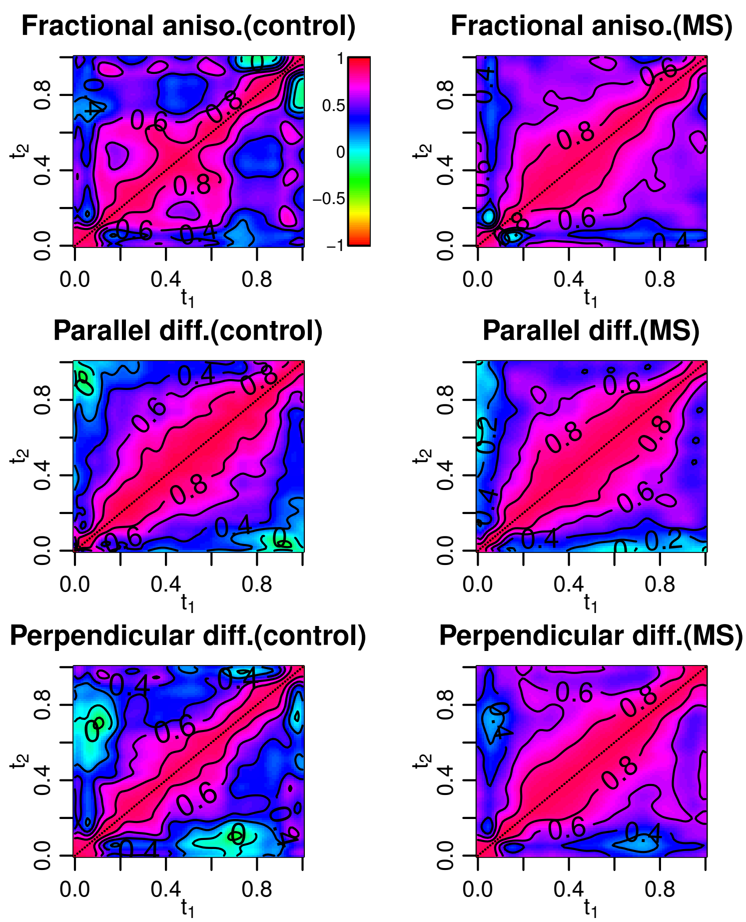

Figure S6: Estimated correlations within each outcome for the control group (left panel) and the MS group (right panel).

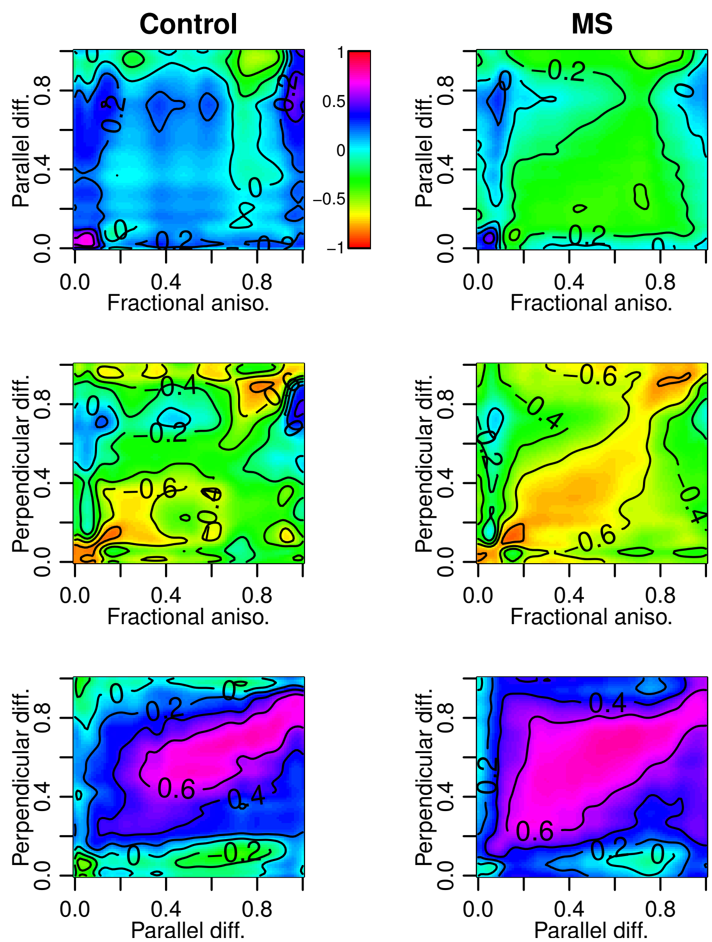

Figure S7: Estimated cross-correlations between different outcomes for the control group (left panel) and the MS group (right panel).

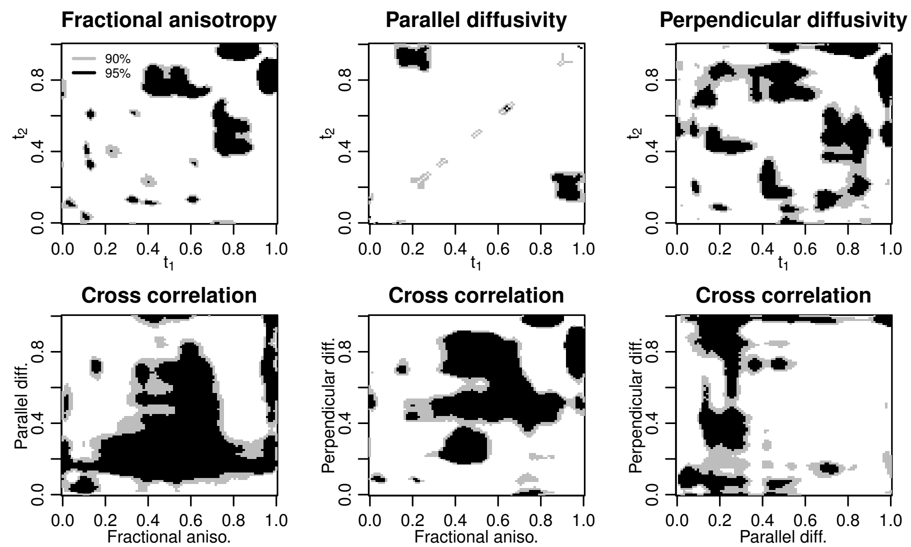

Figure S8: The differences of the correlations within each outcome (top panel) and between outcomes (bottom panel) between the healthy and MS groups. The gray and black regions are the areas at which the differences are statistically significant at 90% and 95% levels, respectively. The computation is based on bootstrap pointwise confidence intervals using 1000 samples.