Appendix A
Parameter Estimation in EM

In the E (Expectation) step, we calculate the responsibilities $w_{p,v,b}^{(k)} = Q_p(q_{v,b} = k)$ for $p \in P_{v,b}^{tr}$ using the current values of the parameters in iteration j:

$$[C_{p,v,b}^{(1)}]^{(j)}$$
$$= [\pi_{v,b}^{(1)}]^{(j)}[D_{p,v,b}^{(1)}]^{(j)}\mathcal{N}(x_{p,v,b}|x_{p,-v,b}[\beta_{v,b}^{(1)}]^{(j)}, [{\sigma_{v,b}^{(1)}}^2]^{(j)})$$
$$[C_{p,v,b}^{(2)}]^{(j)}$$
$$= [\pi_{v,b}^{(2)}]^{(j)}[D_{p,v,b}^{(2)}]^{(j)}\mathcal{N}(x_{p,v,b}|x_{p,v,-b}[\beta_{v,b}^{(2)}]^{(j)}, [{\sigma_{v,b}^{(2)}}^2]^{(j)})$$
$$[C_{p,v,b}^{(3)}]^{(j)}$$
$$= [\pi_{v,b}^{(3)}]^{(j)}[D_{p,v,b}^{(3)}]^{(j)}\mathcal{N}(x_{p,v,b}|m^G([\alpha_{p,v,b}]^{(j)}), \Sigma^G([\alpha_{p,v,b}]^{(j)}))$$
$$[D_{p,v,b}^{(k)}]^{(j)} = \mathcal{N}(V_{p,v,b}|[\boldsymbol{\mu}_{v,b}^{(k)}]^{(j)}, [\boldsymbol{\Sigma}_{v,b}^{(k)}]^{(j)}), k = 1, 2, 3$$
$$[w_{p,v,b}^{(k)}]^{(j)} = \frac{[C_{p,v,b}^{(k)}]^{(j)}}{\sum_{i=1}^{3}[C_{p,v,b}^{(i)}]^{(j)}}, k = 1, 2, 3$$

Let $Z_{v,b} = (x_{p,-v,b})_{p \in P_{v,b}^{tr}}$ and $Y_{v,b} = (x_{p,v,-b})_{p \in P_{v,b}^{tr}}$. In the M (Maximization) step, we re-estimate the parameters in iteration $(j+1)$ using the jth responsibilities:

$$[\pi_{v,b}^{(k)}]^{(j+1)} = \frac{1}{|P_{v,b}^{tr}|}\sum_{p \in P_{v,b}^{tr}}[w_{p,v,b}^{(k)}]^{(j)}, k = 1, 2, 3$$
$$[\boldsymbol{\mu}_{v,b}^{(k)}]^{(j+1)} = \frac{\sum_{p \in P_{v,b}^{tr}}[w_{p,v,b}^{(k)}]^{(j)}V_{p,v,b}}{\sum_{p \in P_{v,b}^{tr}}[w_{p,v,b}^{(k)}]^{(j)}}$$
$$[\boldsymbol{\Sigma}_{v,b}^{(k)}]^{(j+1)} = \frac{1}{\sum_{p \in P_{v,b}^{tr}}[w_{p,v,b}^{(k)}]^{(j)}}\sum_{p \in P_{v,b}^{tr}}[w_{p,v,b}^{(k)}]^{(j)}[U_{p,v,b}^{(k)}]^{(j+1)}$$
$$[U_{p,v,b}^{(k)}]^{(j+1)} = \{V_{p,v,b} - [\boldsymbol{\mu}_{v,b}^{(k)}]^{(j+1)}\}\{V_{p,v,b} - [\boldsymbol{\mu}_{v,b}^{(k)}]^{(j+1)}\}'$$
$$[\beta_{v,b}^{(1)}]^{(j+1)} = \{\{Z_{v,b}'[\mathbf{w}_{v,b}^{(1)}]^{(j)}Z_{v,b}\}^{-1}Z_{v,b}'[\mathbf{w}_{v,b}^{(1)}]^{(j)}x_{:,v,b}^{obs}\}'$$
$$[{\sigma_{v,b}^{(1)}}^2]^{(j+1)} = \frac{1}{\sum_{p \in P_{v,b}^{tr}}[w_{p,v,b}^{(1)}]^{(j)}}[S_{v,b}^{(1)}]^{(j+1)}$$
$$[S_{v,b}^{(1)}]^{(j+1)} = \sum_{p \in P_{v,b}^{tr}}[w_{p,v,b}^{(1)}]^{(j)}\{x_{p,v,b} - x_{p,-v,b}[\beta_{v,b}^{(1)}]^{(j+1)}\}^2$$
$$[\beta_{v,b}^{(2)}]^{(j+1)} = \{\{Y_{v,b}'[\mathbf{w}_{v,b}^{(2)}]^{(j)}Y_{v,b}\}^{-1}Y_{v,b}'[\mathbf{w}_{v,b}^{(2)}]^{(j)}x_{:,v,b}^{obs}\}'$$
$$[{\sigma_{v,b}^{(2)}}^2]^{(j+1)} = \frac{1}{\sum_{p \in P_{v,b}^{tr}}[w_{p,v,b}^{(2)}]^{(j)}}[S_{v,b}^{(2)}]^{(j+1)}$$
$$[S_{v,b}^{(2)}]^{(j+1)} = \sum_{p \in P_{v,b}^{tr}}[w_{p,v,b}^{(2)}]^{(j)}\{x_{p,v,b} - x_{p,v,-b}[\beta_{v,b}^{(2)}]^{(j+1)}\}^2$$
$$[\theta_{v,b}]^{(j+1)} = G([\mathbf{w}_{v,b}^{(3)}]^{(j)}, [\theta_{v,b}]^{(j)}, x_{:,v,b}^{obs}, Y_{v,b}, t_{:,v,:})$$

where $[\mathbf{w}_{v,b}^{(k)}]^{(j)}$ is the vector of $[w_{p,v,b}^{(k)}]^{(j)}$ for $p \in P_{v,b}^{tr}$ in iteration j. The kernel parameters $\theta_{v,b}$ of GP models are evaluated by function G, a gradient descent method that calculates the estimates of $[\theta_{v,b}]^{(j+1)}$ to maximize $\mathcal{L}_{v,b}(\gamma)$, using $[\theta_{v,b}]^{(j)}$ as the starting point. The first order derivatives of $\mathcal{L}_{v,b}(\gamma)$ with respect to $\theta_{v,b}$ that are used in G are given in Appendix C.

Appendix B
GP Model

We assume the GP model discussed here in a mixture model for a certain variable and time, and thus we exclude the subscripts v and b. We use $x_{p,t}$ to denote a measurement of the time series x_p at time t for patient p of a certain variable. We use $x_{p,-t}$ to denote a time series without the measurement at time t. The GP model is given by

$$x_{p,t} = \mu_{p,t} + f(t),$$
$$f(t) \sim \mathcal{GP}(0, \mathcal{K}(t, t'))$$

where $\mu_{p,t}$ is the overall mean of the model and $f(t)$ is a Gaussian process with mean of 0 and covariance of $\mathcal{K}(t, t')$. Following the maximum likelihood approach, the best linear unbiased predictor (BLUP) [1] at t and the mean squared error are

$$m^{(3)}(\theta, x_{p,-t}, \bar{t}) = (\frac{1 - r^T R^{-1} 1_n}{1_n^T R^{-1} 1_n} 1_n^T + r^T) R^{-1} x_{p,-t}$$
$$\Sigma^{(3)}(\theta, x_{p,-t}, \bar{t}) = \sigma_f^2 [1 - r^T R^{-1} r + \frac{(1 - 1_n^T R^{-1} r)^2}{1_n R^{-1} 1_n}]$$

where $r_t(t') = corr(f(t), f(t'))$, r is the vector of $r_t(t')$ for all possible t, $\bar{t}$ is a vector of time except for time t, R is the $(B-1) \times (B-1)$ correlation matrix and the correlation function is given by

$$R_{t,t'} = \exp(-\theta |t - t'|^2).$$

The estimator σ^2 is given by

$$\sigma_f^2 = \frac{C^T R^{-1} C}{n}, C = x_{p,-t} - 1_n (1_n^T R^{-1} 1_n)^{-1} (1_n^T R^{-1} x_{p,-t})$$

where 1_n is a vector with length $(B-1)$ of all ones.

Appendix C
Partial Derivatives in GP

To simplify the notations, we assume that the likelihood function L under consideration is for a mixture model for a certain variable and time. The partial derivative with respect to Gaussian process parameters θ is

$$\frac{\partial L}{\partial \theta} = \sum_{p=1}^{|p^{tr}|} w_p \frac{\partial}{\partial \theta} \ln \mathcal{N}(x_{p,t}; m^{(3)}(\theta, x_{p,-t}, \bar{t}), \Sigma^{(3)}(\theta, x_{p,-t}, \bar{t})).$$

[1] Sacks, Jerome, et al. "Design and analysis of computer experiments." Statistical science (1989): 409-423.

Letting $g_p(\theta) = m^{(3)}(\theta, x_{p,-t}, \bar{t})$ and $h_p(\theta) = \Sigma^{(3)}(\theta, x_{p,-t}, \bar{t})$, we have

$$\begin{aligned}
\frac{\partial L}{\partial \theta} &= \sum_{p=1}^{|p^{tr}|} w_p \frac{\partial}{\partial \theta} \ln \mathcal{N}(x_{p,t}; g_p(\theta), h_p(\theta)) \\
&= \sum_{p=1}^{|p^{tr}|} w_p \frac{\partial}{\partial \theta} \{\ln \frac{1}{\sqrt{2\pi h_p(\theta)}} - \frac{[x_{p,t} - g_p(\theta)]^2}{2h_p(\theta)}\} \\
&= \sum_{p=1}^{|p^{tr}|} w_p \{-\frac{1}{2h_p(\theta)} \frac{\partial h_p(\theta)}{\partial \theta} - \frac{\partial}{\partial \theta} \frac{[x_{p,t} - g_p(\theta)]^2}{2h_p(\theta)}\} \\
&= \sum_{p=1}^{|p^{tr}|} w_p \{-\frac{1}{2h_p(\theta)} \frac{\partial h_p(\theta)}{\partial \theta} \\
&- \frac{1}{2h_p^2(\theta)} \{2[x_{p,t} - g_p(\theta)][-\frac{\partial g_p(\theta)}{\partial \theta}] h_p(\theta) \\
&- \frac{\partial h_p(\theta)}{\partial \theta} [x_{p,t} - g_p(\theta)]^2\}\} \\
&= \sum_{p=1}^{|p^{tr}|} w_p \{-\frac{1}{2h_p(\theta)} \frac{\partial h_p(\theta)}{\partial \theta} \\
&+ \frac{[x_{p,t} - g_p(\theta)] \frac{\partial g_p(\theta)}{\partial \theta}}{h_p(\theta)} + \frac{\frac{\partial h_p(\theta)}{\partial \theta} [x_{p,t} - g_p(\theta)]^2}{2h_p^2(\theta)}\}.
\end{aligned}$$

Then $\frac{\partial g_p(\theta)}{\partial \theta}$ and $\frac{\partial h_p(\theta)}{\partial \theta}$ are given by

$$\begin{aligned}
\frac{\partial g_p(\theta)}{\partial \theta} &= (\frac{\partial H_1}{\partial \theta} R^{-1} + H_1 \frac{\partial R^{-1}}{\partial \theta}) x_{p,-t} \\
\frac{\partial h_p(\theta)}{\partial \theta} &= \sigma_f^2 \frac{\partial H_3}{\partial \theta} + \frac{\partial \sigma_f^2}{\partial \theta} H_3
\end{aligned}$$

where H_1, $\frac{\partial H_1}{\partial \theta}$, H_3 and $\frac{\partial H_3}{\partial \theta}$ are given as follows:

$$\begin{aligned}
H_1 &= \frac{[1 - (rR^{-1}1_n)]}{1_n^T R^{-1} 1_n} 1_n^T + r \\
\frac{\partial H_1}{\partial \theta} &= \frac{-(\frac{\partial r}{\partial \theta} R^{-1} + r \frac{\partial R^{-1}}{\partial \theta}) 1_n (1_n^T R^{-1} 1_n)}{1_n^T R^{-1} 1_n^2} \\
&- \frac{(1_n^T \frac{\partial R^{-1}}{\partial \theta} 1_n)[1 - (rR^{-1}1_n)]}{1_n^T R^{-1} 1_n^2} 1_n^T + \frac{\partial r}{\partial \theta} \\
fc &= (1 - 1_n^T R^{-1} r^T)^2 \\
gc &= 1_n^T R^{-1} 1_n \\
\frac{\partial fc}{\partial \theta} &= 2(1 - 1_n^T R^{-1} r^T)[-1_n^T (\frac{\partial R^{-1}}{\partial \theta} r^T + R^{-1} \frac{\partial r^T}{\partial \theta})] \\
\frac{\partial gc}{\partial \theta} &= 1_n^T \frac{\partial R^{-1}}{\partial \theta} 1_n \\
H_2 &= \frac{(1 - 1_n^T R^{-1} r^T)^2}{1_n^T R^{-1} 1_n} \\
\frac{\partial H_2}{\partial \theta} &= \frac{\frac{\partial fc}{\partial \theta} gc - \frac{\partial gc}{\partial \theta} fc}{gc^2}
\end{aligned}$$

$$\begin{aligned}
H_3 &= 1 - (rR^{-1}r^T) + H_2 \\
\frac{\partial H_3}{\partial \theta} &= -(\frac{\partial r}{\partial \theta} R^{-1} r^T + r \frac{\partial R^{-1}}{\partial \theta} r^T + rR^{-1} \frac{\partial r^T}{\partial \theta}) + \frac{\partial H_2}{\partial \theta} \\
H_4 &= x_{p,-t} - 1_n \frac{(1_n^T R^{-1} x_{p,-t})}{1_n^T R^{-1} 1_n} \\
\frac{\partial H_4}{\partial \theta} &= -1_n \frac{1}{(1_n^T R^{-1} 1_n)^2} [(1_n^T \frac{\partial R^{-1}}{\partial \theta} x_{p,-t})(1_n^T R^{-1} 1_n) \\
&- (1_n^T \frac{\partial R^{-1}}{\partial \theta} 1_n)(1_n^T R^{-1} x_{p,-t})] \\
\frac{\partial \sigma_f^2}{\partial \theta} &= \frac{1}{n} [(\frac{\partial H_4}{\partial \theta})^T R^{-1} H_4 + H_4^T \frac{\partial R^{-1}}{\partial \theta} H_4 + H_4^T R^{-1} \frac{\partial H_4}{\partial \theta}].
\end{aligned}$$