A Appendix

We observe N_{id} samples from each signal d of every individual i; $\mathbf{y}_{id} = y_{id}(\mathbf{t}_{id}) = \{y_{id}(t_{idn}), \forall n = 1, 2, ..., N_{id}\}$. We denote the collection of observations of D longitudinal signals by $\mathbf{y}_i = \{\mathbf{y}_{i1}, ..., \mathbf{y}_{iD}\}, \forall i$. We also observe the treatments given to each individual i, $\mathbf{x}_{ijd}(t), \forall t, j, d$. We let $\mathbf{x}_i = \{\mathbf{x}_{i1}, ..., \mathbf{x}_{iJ}\}, \forall i$, be the collection of treatment inputs of all types.

For each latent function, we define a set of inducing input-output pairs $\mathbf{Z}, \mathbf{u}$, where $\mathbf{Z}$ are some pseudo-inputs, known as inducing points, and $\mathbf{u}$ are the values of the Gaussian process at $\mathbf{Z}$. We place the inducing points $\mathbf{Z}$ on a grid. We define a variational distribution for $\mathbf{u}$, $q(\mathbf{u}) = \mathcal{GP}(\mathbf{m}, \mathbf{S})$, where $\mathbf{m}$ and $\mathbf{S}$ are variational parameters. Using $q(\mathbf{u})$ we compute a variational GP distribution for each shared latent function: $q(\mathbf{g}) = \mathcal{GP}(\boldsymbol{\mu}_g, \boldsymbol{\Sigma}_g)$, where $\boldsymbol{\mu}_g = \mathbf{K_{NZ}}\mathbf{K_{ZZ}^{-1}}\mathbf{m}$ and $\boldsymbol{\Sigma}_g = \mathbf{K_{NN}} - \mathbf{K_{NZ}}\mathbf{K_{ZZ}^{-1}}(\mathbf{I} - \mathbf{S}\mathbf{K_{ZZ}^{-1}})\mathbf{K_{ZN}}$, with $\mathbf{K_{NZ}} = K(\mathbf{t}, \mathbf{Z})$. We similarly obtain $q(\mathbf{v}_d) = \mathcal{GP}(\boldsymbol{\mu}_{v_d}, \boldsymbol{\Sigma}_{v_d}), \forall d$, for signal-specific latent functions.

Here, to simplify the notation, we assume $\mathbf{t}_{id} = \mathbf{t}_i, \forall d$, and write $\mathbf{K_{NN}} = K_i(\mathbf{t}_i, \mathbf{t}_i')$. We emphasize that the observations from different signals need not be aligned for our learning and inference algorithm.

We obtain the variational distribution $q(\mathbf{f})$ by taking the linear combinations of the variational distributions for individual latent GPs: $q(\mathbf{f}_d) = \mathcal{GP}(\boldsymbol{\mu}_d, \boldsymbol{\Sigma}_d)$, where $\boldsymbol{\mu}_d = \omega_d \boldsymbol{\mu}_g + \kappa_d \boldsymbol{\mu}_{v_d}$ and $\boldsymbol{\Sigma}_d = \omega_d^2 \boldsymbol{\Sigma}_g + \kappa_d^2 \boldsymbol{\Sigma}_{v_d}$.

The log-likelihood of the observations and local model parameters for each individual is $\log p(\mathbf{y}_i, \Theta_i) = \log p(\mathbf{y}_i|\Theta_i) + \log p(\Theta_i)$, where we dropped the explicit conditioning on $\mathbf{x}_i$. Using sparse GP approximations and Jensen's inequality, we compute a variational lower bound for $\log p(\mathbf{y}_i|\Theta_i)$:

$$
\begin{aligned}
\log p(\mathbf{y}_i|\Theta_i) &= \log \int p(\mathbf{y}_i|\mathbf{f}_i)p(\mathbf{f}_i|\mathbf{u}_i)p(\mathbf{u}_i)\, \mathrm{d}\mathbf{f}_i\, \mathrm{d}\mathbf{u}_i \\
&\leq E_{q(\mathbf{f}_i)} \log p(\mathbf{y}_i|\mathbf{f}_i, \Theta_i) - \mathrm{KL}(q(\mathbf{u}_i)||p(\mathbf{u}_i)) \\
&= Q_i(\mathbf{y}_i; \Theta_i)\,, \qquad (6)
\end{aligned}
$$

where, $q(\mathbf{f}_i) = E_{q(\mathbf{u}_i)} p(\mathbf{f}_i|\mathbf{u}_i)$. Also, $\mathrm{KL}(q(\mathbf{u}_i)||p(\mathbf{u}_i))$ is the Kullback-Leibler divergence between $q(\mathbf{u}_i)$ and $p(\mathbf{u}_i)$ which we compute analytically. We note that conditioned on $\mathbf{f}_i$, the distribution of $\mathbf{y}_i$ factorizes over all signals. Thus, we have $E_{q(\mathbf{f}_i)} \log p(\mathbf{y}_i|\mathbf{f}_i, \Theta_i) = \sum_d [\log m_{id}(\mathbf{t}_{id}) + E_{q(\mathbf{f}_{id})} \log p(\mathbf{y}_{id}|\mathbf{f}_{id})]$, where m_{id} is the sum of the treatment response component and the fixed-effect terms. The expectation $E_{q(\mathbf{f}_{id})} \log p(\mathbf{y}_{id}|\mathbf{f}_{id})$ is also available in closed-form (Titsias, 2009; Hensman et al., 2013). Given (6), we compute the evidence lower bound (ELBO) for each individual: $\mathrm{ELBO}_i = Q_i(\mathbf{y}_i; \Theta_i) + \log p(\Theta_i)$.