A Appendix

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

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

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

We obtain the variational distribution \( q(f) \) by taking the linear combinations of the variational distributions for individual latent GPs: \( q(f_d) = \mathcal{GP}(\mu_d, \Sigma_d) \), where \( \mu_d = \omega_d \mu_g + \kappa_d \mu_v \) and \( \Sigma_d = \omega_d^2 \Sigma_g + \kappa_d^2 \Sigma_v \).

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

\[
\log p(y_i|\Theta_i) = \log \int p(y_i|f_i)p(f_i|u_i)p(u_i) df_i du_i \\
\leq E_{q(f_i)} \log p(y_i|f_i, \Theta_i) - KL(q(u_i)||p(u_i)) \\
= Q_i(y_i; \Theta_i), \tag{6}
\]

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