**Introduction**

In biomedical studies, investigators are often interested in predicting future or missing observations of subjects based on their historical data, referred to as dynamic prediction. Traditionally, mixed models or joint modeling have been used for these problems. However, these methods are often limited in flexibility when dealing with densely measured functional data. In addition, out-of-sample prediction is challenging under model frameworks above, especially with non-Gaussian outcomes (e.g., binary and count data), because the estimation of individual random effect can be computationally intensive. To address these problems, we propose a novel method for dynamic prediction based on functional data analysis methods, which allows for more flexible correlation structure, can be generalized to non-Gaussian outcomes and is computationally feasible.

**Method**

The proposed new method is an extension of Generalized Functional Principal Component Analysis (FPCA) and Functional Mixed Models. Assume that the generalized functional outcome can be characterized by a latent Gaussian function. Specifically, the outcome at a specific time follows an exponential family distribution parameterized by the value of latent function at that time. Such latent functions can be approximated with the linear combination of a set of basis functions, which naturally lead us to dimension reduction methods such as FPCA. Here, each latent function can be characterized with a set of eigenfunctions and their corresponding subject-specific coefficients, often referred to as PC scores. Since eigenfunctions does not change across population, any point in the functional domain can then be estimated with these subject-specific scores.

Practically, the prediction procedure above consists of the following steps: 1) Binning the data across functional domain into small, equal-length, non-overlapping intervals; 2) Fitting local generalized mixed models (GLMM) at every bin to obtain an estimated latent function for each subject on the binned grid; 3) Use FPCA to smooth the estimated latent functions, also obtain eigenfunctions and estimates of variance components for prediction. With this information from FPCA model, we can obtain estimates of subject-specific PC scores based on partial observations, thus recover the unobserved, latent Gaussian function on the binned functional domain.

The recovery of latent Gaussian functions includes two different cases: in-sample estimation and out-of-sample prediction. The former refers to situations when the full functional tracks are observed and used for model fitting. The individual functions are then “in-sample” subjects. The latter, on the other hand, refers to new function tracks that have not been included in steps 1-3, namely “out-of-sample” subjects. If the out-of-sample function tracks are incomplete, we will be able to recover their full latent function tracks with information obtained from step 3. This procedure is called out-of-sample prediction and is more in line with the “dynamic prediction” procedure in practice.

**Results**

We have designed and implemented a simulation study with 500 subjects, with binary outcomes generated from a latent Gaussian function with four eigenfunctions. The preliminary results reveal that the method above led to unbiased estimates of PC scores. For new, partially observed subjects, the out-of-sample prediction recovered the full latent track as well as in-sample estimation where the full function track is observed. Both in-sample and out-of-sample prediction also improves significantly with longer observed tracks.

**Discussion**

We are not clear about the effect of arbitrary decisions in the binning and local GLMM steps on the final predictive performance, such as the effect of bin width, number of observations in each bin, and the model form of GLMMs. While some of them can be treated as hyperparameters and cross-validated, others have more complicated indications. In addition,

the procedure above allowed us to make predictions on the binned grid, but not the original grid. One potential way to extend prediction back to the original grid is through interpolation. Since the bins are set up to me small so that binned observations points are also dense, interpolation should perform well filling in function values between two discrete points.

The out-of-sample prediction interval is another challenging issue that needs further exploration. Just like random effects of generalized mixed models, there is no closed-form solution for the estimates of PC scores of generalized functions, neither for their covariance. A general strategy to address this issue is to use resampling procedures, such as bootstrap, to estimate its distribution. We may also be able to get conditional variance estimates using observed Fisher information. However with the additional constraints, such estimates can be biased and not interpretable.