Our statistical approach is based on a mixed-membership framework. We use the data to determine a small number, K, of *canonical trajectories* that correspond to *idealized progressions* of patient outcomes as a function of age. We model individual outcomes by assuming that individuals borrow characteristics from the canonical profiles in different individual degrees. Our outcome of interest, Y, is the cognitive status of each individual at each time point in the study, where $Y = 1$ for *Normal*, $Y = 2$ for *Mildly Impaired*, and $Y = 3$ for *Severely Impaired*.

The basic MMTM implemented in [4], which we will describe in the remainder of this subsection, begins by modeling the canonical trajectories toward various levels of impairment in each profile K, $k = 1, \ldots, K$, as

$$\begin{aligned}\text{logit}\Pr(Y > 1|Age, \boldsymbol{\theta_k}) &= \beta_{0k} + \beta_{1k}Age \\ \text{logit}\Pr(Y > 2|Age, \boldsymbol{\theta_k}) &= \beta_{0k} + \beta_{1k}Age - c_k,\end{aligned}$$

where each trajectory $k = 1, \ldots, K$ has its own set of trajectory parameters $\boldsymbol{\theta_k} = (\beta_{0k}, \beta_{1k}, c_k)$. The probabilities of each outcome, $d_k(y|Age) = \Pr(Y = y|Age, \boldsymbol{\theta_k})$, can be computed from these two equations by subtraction. In our work to date, $K = 3$ profiles best fit the data. In the proposed project we will empirically test models with various numbers of canonical profiles, to determine the number of profiles that best reflects heterogeneity in the data.

The MMTM allows each individual, indexed by i, to have partial membership in all of the canonical profiles simultaneously. We specify the degree of membership through a vector $\mathbf{g_i} = (g_{i1}, ..., g_{iK})$, where component g_{ik} specifies the degree of membership of individual i in profile k.

For example, with $K = 3$, an individual who is perfectly aligned with the first canonical trajectory would have membership weights of (1, 0 , 0), whereas an individual who shared components of all three trajectories might, for example, have membership weights (.5, .2, .3). We construct individual trajectories toward impairment from the canonical ones by weighting them according to membership,

$$\Pr(Y_i = y|Age, \boldsymbol{\theta}, \mathbf{g_i}) = \sum_{k=1}^{K} g_{ik} d_k(y|Age).$$

The basic MMTM extends this impairment model, through joint modeling, to address the critical issue of survivorship. It is known that the presence of HAND, at least prior to the use of cART, is strongly associated with risk of death. Joint modeling allows us to incorporate the information provided by survival times to better reconstruct impairment trajectories—and vice-versa—and to address the complications derived from attrition due to death [3]. Following Chapter 6 of [3], and [2], we introduce a survival time s_i for each individual i, and expand the trajectory model for that individual to be

$$p(y_i, s_i|Age, \mathbf{g_i}) = \left[\sum_{k=1}^{K} g_{ik} d_k(y_i|Age)\right] \left[\sum_{k=1}^{K} g_{ik} h_k(s_i)\right], \tag{1}$$

where $h_k(s_i)$ is the density of the survival distribution for the k-th canonical profile, and the outcome is censored if $Age > s_i$. This model builds critically on two assumptions. First, we assume that each canonical profile comprises both an idealized trajectory to severe impairment, and an idealized mortality distribution; i.e. canonical profiles describe both impairment and survival. Second, given an individual's membership vector, $\mathbf{g_i}$, survival times and cognitive classification are independent. This is last assumption, sometimes referred to as local independence, is common and well-studied in factor analysis models, latent class models, and latent variable modeling generally [1]. Impairment and survival are dependent in the observed data, but the assumption of local independence forces $\mathbf{g_i}$ to be rich enough to fully explain this dependence when we condtion on it. We deal with the problem of right censoring—people who do not survive long enough to determine their final cognitive disposition—through a data-augmentation scheme at the time of estimation via Markov Chain Monte Carlo Simulation.

Finally, the basic MMTM implemented [4] incorporates recruitment cohort and other time-invariant covariates as predictors for membership scores. Specifically, we treat the mixed membership scores, $\mathbf{g}_i$, as random effects sampled from a distribution specified conditional on the individual covariates, $\mathbf{X}_i$,

$$\mathbf{g_i}|\mathbf{X}_i \sim F_{\boldsymbol{\alpha}}(\mathbf{X}_i). \tag{2}$$

We have implemented this specification through a multinomial logit model, and tested time invariant covariates including recruitment cohort, race, the presence of confounding conditions, infection with hepatitis C virus, a history of depression, HIV Disease, and AIDS.

References

[1] P. W. Holland and P. R. Rosenbaum. Conditional association and unidimensionality in monotone latent variable models. *The Annals of Statistics*, pages 1523–1543, 1986.

[2] F. Lecci. An analysis of development of dementia through the extended trajectory grade of membership model. In E. M. Airoldi, D. Blei, E.A. Erosheva, and S. E. Fienberg, editors, *Handbook of Mixed Membership Models*. Chapman & Hall/CRC, 2014.

[3] D. Manrique-Vallier. *Longitudinal Mixed Membership Models With Applications*. PhD thesis, Carnegie Mellon University Department of Statistics, Pittsburgh, PA, 2010.

[4] S.A. Molsberry, F. Lecci, L. Kingsley, B. Junker, S. Reynolds, K. Goodkin, A. Levine, E. Martin, Miller E., C. Munro, Ragin A., Sacktor N., and J. Becker. Mixed membership trajectory models of cognitive impairment in the multicenter AIDS cohort study. *AIDS*, 6:713–721, 2015.