**Statistical Analysis Plan and Data Dictionary**

## **General Methods**

Shapiro-Wilk test will be used to assess normality of variables. Non-parametric statistical methods will be used to analyze non-normally distributed variables.

P-values will be generated where appropriate and any p-value < 0.05 will be considered statistically significant. If multiple p-values are generated within an analyses, Benjamini-Hochberg false discovery rate (FDR) correction will be used to determine significance. P-values will be rounded to 3 decimal places and therefore presented as 0.xxx; P-values smaller than 0.001 will be reported as ‘<0.001’.

### 

### **Baseline and Handling Missing Data**

All summaries and analyses will be based on observed data and missing data imputation is not planned.

**Primary Aim 1.**

**Analysis**: Group-based trajectory model. Latent growth mixture modeling (LGMM) based on superior model fit indices.

* LCGA-linear, LCGA-quadratic, LCGA-cubic, LGMM-linear, LGMM-quadratic

**Primary Aim 2.**

**Analysis:** Univariate multinomial logistic regression to evaluate the associations between baseline factors and trajectories. Adjusting for age, sex, and BMI.

# **Data Dictionary**