**Progress on Basecase 3 analysis of the duration of symptomatic GERD, obesity, age, sex, and race on the risk for Barrett’s esophagus (BE) and esophageal adenocarcinoma (EAC).**

US population data on obesity are needed to move forward with our Basecase 3 analysis of the role of obesity (as measured by body mass index, or BMI), history of gastroesophageal reflux, age, race, and sex on the risk for Barrett's esophagus (BE) and esophageal adenocarcinoma (EAC). These data will be used in conjunction with case control data from the International Barrett’s and Esophageal Adenocarcinoma Consortium (BEACON).

**Data**

Data sources include longitudinal and cross sectional demographic and medical examination data from the National Health and Nutrition Examination Survey (NHANES). NHANES was designed to periodically assess the health and nutritional status of US adults and children.

Cross sectional data were downloaded for NHANES surveys conducted between 1971 and 2014, and from the first National Health Examination Survey (NHES I) conducted between 1959-1962. These data include Continuous NHANES for years [2013-14 (n=10,175), 2011-12 (n=9,756), 2009](callto:2013-14,%202011-12,%202009)-10 (n=10,537), 2007-08 (n=10,149), 2005-06 (n=10,348), 2003-04 (n=10,122), 2001-02 (n=11,039) and 1999-2000 (n=9,965), NHANES\_III surveyed in 1982-85 (n=31,311), NHANES\_II in 1976-80 (n=20,322), NHANES\_I in 1971-74 (n=23,808),  and NHES\_I in 1959-62 (n=6,672).

Longitudinal data were downloaded from NHANES I Epidemiological Followup Studies (NHEFS) beginning with NHANES 1 with followup in years [1982-84, 1986, 1987](callto:1982-84,%201986,%201987), and 1992. Longitudinal data include all persons 25-74 years of age who completed a medical examination at NHANES I in 1971-75 (n = 14,407), with followup in years [1982-84 (n=12,220), 1986 (n=3,980 non-deceased individuals age 55-74 at baseline), 1987](callto:1982-84,%201986,%201987) (n=11,750), and 1992 (n=11,195).

**Methods**  
Data elements related to BMI, age, race, sex, educational attainment and US population weights were harmonized across all surveys. Data were combined across files for children and adults, but children under 12 years of age were excluded from further analysis.

Individual BMI histories were constructed using the longitudinal data from the NHANES I and four NHEFS followup studies. Each individual history was smoothed using interpolating splines, then BMI quintiles were calculate by age, sex, and race (black, white). Each individual history was then traced to determine its BMI quintile on an annual basis. Using these data, Markov transition rates were calculated for transitioning between quintiles (see figures below).

Thin-plate splines are being used to combine all of the cross-sectional and longitudinal data by race and gender to model each BMI quintile as functions of age, period, and cohort. Educational attainment or other exposures (smoking history, alcohol use, etc.) are not currently included in the model, but they may be included later as part of an individual history generator.





 







