# framingham datsets

## Bhattacharyya, Timothy (NIH/NIAMS) [C]

Wed 9/27/2017 9:41 PM

To:Dasgupta, Abhijit (NIH/NIAMS) [C] <abhijit.dasgupta@nih.gov>;

 3 attachments (2 MB)

original_full_columns_renamed.csv; offspring_full_2000_redone_columns_bmi.csv; python assembling framingham databases.ipynb;

Dear Abijit

Finally! Attached please find 2 excel spreadsheets containing data on the Framingham original and offspring cohorts, and their risk factors for hip fracture that I assembled.  Please make a plan to analyze them.

Just a quick review since its been so long.  We know that the incidence of hip fracture in the cohorts declined from 1980 to 1990 to 2000.  The question is, what changes in risk factors correlate with the drop.   I believe a drop in smoking is the main reason.  Here is a list of the risk factors:  Age, Sex, BMI, current smoking, steroid use, alcohol>3drinks per day, taking estrogen, taking bisphosphonates (only measured in 2000), beta blockers, diabetes, menopause <age 45

Here are the next steps in the analysis.

1.   Alcohol is considered a risk factor for hip fx if more than 3 drinks per week are consumed.  IN the datasets, you'll find colums for Beer_1980, wine_1980, and Cocktail 1980.  These represent the number of drinks per week of each.  Please add together, and if >3, then RF_ETOH would be 1.  Else RF_ETOH=0.  As time goes on, wine is broken in to white and red.  [I couldn't do this because when I added the columns, I got concatenation instead of sum]
2.   Menopause is a risk factor if the subject developed menopause <age 45.  In Framingham they measured the age (if any) that menopause occurred.  So please create three variables, one for each decade.  They will be 0 if patient is male.  0 if patient has not yet achieved menopause at that visit year.  And 1 if by that visit year, she has achieved menpausse and it occurred before age 45.
3.   Smoking in 1980 was measured as number of cigarettes per day.  Please convert this to a dichotomous variable.  (Smoking_1980= 1 if >1 cig smoked per day.)
4.   Then, I believe we should see how the prevalence of risk factors trended over the decade, prior to creating models.

I'm planning to be there Friday if we need to talk or if anything is unclear.  I thought about making a data dictionary but the columns headers are pretty clear.

tim