**Data Dictionary for Group 4 Framingham Heart Study**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variable Name | Data Type | Measurement Units | Description | Data Source |
|
| RANDID | Integer |  | Unique identification number for each participant | M & D |
| Age | Integer |  | Age in years at exam | M & D |
| BMI | Continuous variable | (weight in kilograms) / (height in meters squared) | Body Mass Index | M & D |
| Cursmoke | Binary variable |  | 0 = Not a current smoker 1 = Current cigarette smoker | M & D |
| Sex | Binary variable |  | 0 = Male 1 = Female | M & D |
| SysBP | Continuous variable | mmHg | Systolic Blood Pressure in mmHg | M & D |
| TotalChol | Integer | mg/dL | Serum Total Cholesterol | M & D |
| MISTRK | Binary variable |  | 0 = Free of disease 1 = Prevalent disease | M |
| Diabetes | Binary variable |  | 0 = Not a diabetic 1 = Diabetic | D |

\*M = MISTRK\_nona\_casecontrol.csv, the dataset where the missing values of BMI and TotalChol are excluded and has Myocardial Infarction (MI) or Stroke case and control (total 876 observations thus 292 subjects).

\*D = Diabetes\_nona\_casecontrol.csv, the dataset where the missing values of BMI and TotalChol are excluded and has diabetes case and control (1038 observations, 346 subjects)