**Framingham Heart Study
dataset description**

The Framingham Heart Study is a long term prospective study of the etiology of cardiovascular disease among a population of free living subjects in the community of Framingham, Massachusetts (US). The Framingham Heart Study was a landmark study in epidemiology in that it was the first prospective study of cardiovascular disease and identified the concept of risk factors and their joint effects. The study began in 1948 and 5,209 subjects were initially enrolled in the study. Participants have been examined biennially since the inception of the study and all subjects are continuously followed through regular surveillance for cardiovascular outcomes.
Clinic examination data has included cardiovascular disease risk factors and markers of
disease such as blood pressure, blood chemistry, lung function, smoking history, health
behaviors, ECG tracings, Echocardiography, and medication use. Through regular surveillance of area hospitals, participant contact, and death certificates, the Framingham Heart Study reviews and adjudicates events for the occurrence of Angina Pectoris, Myocardial Infarction, Heart Failure, and Cerebrovascular disease.

The enclosed dataset is a subset of the data collected as part of the Framingham study and
includes laboratory, clinic, questionnaire, and adjudicated event data on 4,434 participants.
Participant clinic data was collected during three examination periods, approximately 6 years
apart, from roughly 1956 to 1968. Each participant was followed for a total of 24 years for the
outcome of the following events: Angina Pectoris, Myocardial Infarction, Atherothrombotic
Infarction or Cerebral Hemorrhage (Stroke) or death.

(NOTE: Although the enclosed dataset contains Framingham data ‘as collected’ by Framingham investigators, specific methods were employed to ensure an anonymous dataset that protects patient confidentiality; therefore, this dataset is inappropriate for publication purposes).

The data is provided in Longitudinal („long”) form. Missing values in the dataset are indicated by an empty cell (), for decimal separator dot was used (.).

The dataset contains the following „risk” variables.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable  | Description  | Units  | Range orcount |
| RANDID  | Unique identification number for eachparticipant |  | 2448-9999312 |
| SEX  | Participant sex  | 1=Men2=Women | n=5022n=6605 |
| PERIOD  | Examination Cycle  | 1=Period 12=Period 23=Period 3 | n=4434n=3930n=3263 |
| TIME  | Number of days since baseline exam  |  | 0-4854 |
| AGE  | Age at exam (years)  |  | 32-81 |
| SYSBP  | Systolic Blood Pressure (mean of lasttwo of three measurements) (mmHg) |  | 83.5-295 |
| DIABP  | Diastolic Blood Pressure (mean oflast two of three measurements)(mmHg) |  | 30-150 |
| BPMEDS  | Use of Anti-hypertensive medicationat exam | 0=Not currently used1=Current Use | n=10090n=944 |
| CURSMOKE  | Current cigarette smoking at exam  | 0=Not current smoker1=Current smoker | n=6598n=5029 |
| CIGPDAY  | Number of cigarettes smoked eachday | 0=Not current smoker1-90 cigarettes per day |  |
| EDUC | Attained Education | 1=0-11 years2=High School Diploma, GED3=Some College, Vocational School4=College (BS, BA) degree or more | n=4690n=3410n=1885n=1347 |
| TOTCHOL  | Serum Total Cholesterol (mg/dL)  |  | 107-696 |
| HDLC  | High Density Lipoprotein Cholesterol(mg/dL) | available for period 3only | 10-189 |
| LDLC  | Low Density Lipoprotein Cholesterol(mg/dL) | available for period 3only | 20-565 |
| BMI  | Body Mass Index, weight inkilograms/height meters squared |  | 14.43-56.8 |
| GLUCOSE  | Casual serum glucose (mg/dL)  |  | 39-478 |
| DIABETES  | Diabetic according to criteria of firstexam treated or first exam withcasual glucose of 200 mg/dL or more | 0=Not a diabetic1=Diabetic | n=11097n=530 |

|  |  |  |  |
| --- | --- | --- | --- |
| Variable  | Description  | Units  | Range orcount |
| HEARTRTE  | Heart rate (Ventricular rate) inbeats/min |  | 37-220 |
| PREVAP  | Prevalent Angina Pectoris at exam  | 0=Free of disease1=Prevalent disease | n=11000n=627 |
| PREVCHD  | Prevalent Coronary Heart Diseasedefined as pre-existing AnginaPectoris, Myocardial Infarction(hospitalized, silent or unrecognized),or Coronary Insufficiency (unstableangina) | 0=Free of disease1=Prevalent disease | n=10785n=842 |
| PREVMI  | Prevalent Myocardial Infarction  | 0=Free of disease1=Prevalent disease | n=11253n=374 |
| PREVSTRK  | Prevalent Stroke  | 0=Free of disease1=Prevalent disease | n=11475n=152 |
| PREVHYP  | Prevalent Hypertensive. Subject wasdefined as hypertensive if treated orif second exam at which meansystolic was >=140 mmHg or meanDiastolic >=90 mmHg | 0=Free of disease1=Prevalent disease | n=6283n=5344 |

For Each participant the following event data is provided. For each type of event, ‘0' indicates the event did not occur during followup, and ‘1' indicates an event did occur during followup. Only the first event occurring during the interval of baseline (PERIOD=1) to end of followup is provided:

|  |  |
| --- | --- |
| Variable name  | Description |
| ANGINA  | Angina Pectoris |
| HOSPMI  | Hospitalized Myocardial Infarction |
| MI\_FCHD  | Hospitalized Myocardial Infarction or Fatal Coronary Heart Disease |
| ANYCHD  | Angina Pectoris, Myocardial infarction (Hospitalized and silent orunrecognized), Coronary Insufficiency (Unstable Angina), or FatalCoronary Heart Disease |
| STROKE  | Atherothrombotic infarction, Cerebral Embolism, IntracerebralHemorrhage, or Subarachnoid Hemorrhage or Fatal CerebrovascularDisease |
| CVD  | Myocardial infarction (Hospitalized and silent or unrecognized), FatalCoronary Heart Disease, Atherothrombotic infarction, CerebralEmbolism, Intracerebral Hemorrhage, or Subarachnoid Hemorrhage orFatal Cerebrovascular Disease |
| HYPERTEN  | Hypertensive. Defined as the first exam treated for high blood pressureor second exam in which either Systolic is $ 140 mmHg or Diastolic $90mmHg |
| DEATH  | Death from any cause |
| TIMEAP  | Number of days from Baseline exam to first Angina during the followupor Number of days from Baseline to censor date. Censor date may beend of followup, death or last known contact date if subject is lost tofollowup |
| TIMEMI  | Defined as above for the first HOSPMI event during followup |
| TIMEMIFC  | Defined as above for the first MI\_FCHD event during followup |
| TIMECHD  | Defined as above for the first ANYCHD event during followup |
| TIMESTRK  | Defined as above for the first STROKE event during followup |
| TIMECVD  | Defined as above for the first CVD event during followup |
| TIMEHYP  | Defined as above for the first HYPERTEN event during followup |
| TIMEDTH  | Number of days from Baseline exam to death if occurring duringfollowup or Number of days from Baseline to censor date. Censor datemay be end of followup, or last known contact date if subject is lost tofollowup |

Note that defining Hypertensive requires exam participation and bias can therefore occur. Subjects attending exams regularly have a greater opportunity to be defined as hypertensive. Subjects not attending exams would be assumed to be free of hypertension. Since Hypertension is highly prevalent, this misclassification could potentially be large.