**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 or count |
| RANDID | Unique identification number for each participant |  | 2448- 9999312 |
| SEX | Participant sex | 1=Men 2=Women | n=5022 n=6605 |
| PERIOD | Examination Cycle | 1=Period 1 2=Period 2 3=Period 3 | n=4434 n=3930 n=3263 |
| TIME | Number of days since baseline exam |  | 0-4854 |
| AGE | Age at exam (years) |  | 32-81 |
| SYSBP | Systolic Blood Pressure (mean of last two of three measurements) (mmHg) |  | 83.5-295 |
| DIABP | Diastolic Blood Pressure (mean of last two of three measurements) (mmHg) |  | 30-150 |
| BPMEDS | Use of Anti-hypertensive medication at exam | 0=Not currently used 1=Current Use | n=10090 n=944 |
| CURSMOKE | Current cigarette smoking at exam | 0=Not current smoker 1=Current smoker | n=6598 n=5029 |
| CIGPDAY | Number of cigarettes smoked each day | 0=Not current smoker 1-90 cigarettes per day |  |
| EDUC | Attained Education | 1=0-11 years 2=High School Diploma, GED 3=Some College, Vocational School 4=College (BS, BA) degree or more | n=4690 n=3410  n=1885  n=1347 |
| TOTCHOL | Serum Total Cholesterol (mg/dL) |  | 107-696 |
| HDLC | High Density Lipoprotein Cholesterol (mg/dL) | available for period 3 only | 10-189 |
| LDLC | Low Density Lipoprotein Cholesterol (mg/dL) | available for period 3 only | 20-565 |
| BMI | Body Mass Index, weight in kilograms/height meters squared |  | 14.43-56.8 |
| GLUCOSE | Casual serum glucose (mg/dL) |  | 39-478 |
| DIABETES | Diabetic according to criteria of first exam treated or first exam with casual glucose of 200 mg/dL or more | 0=Not a diabetic 1=Diabetic | n=11097 n=530 |

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Description | Units | Range or count |
| HEARTRTE | Heart rate (Ventricular rate) in beats/min |  | 37-220 |
| PREVAP | Prevalent Angina Pectoris at exam | 0=Free of disease 1=Prevalent disease | n=11000 n=627 |
| PREVCHD | Prevalent Coronary Heart Disease defined as pre-existing Angina Pectoris, Myocardial Infarction (hospitalized, silent or unrecognized), or Coronary Insufficiency (unstable angina) | 0=Free of disease 1=Prevalent disease | n=10785 n=842 |
| PREVMI | Prevalent Myocardial Infarction | 0=Free of disease 1=Prevalent disease | n=11253 n=374 |
| PREVSTRK | Prevalent Stroke | 0=Free of disease 1=Prevalent disease | n=11475 n=152 |
| PREVHYP | Prevalent Hypertensive. Subject was defined as hypertensive if treated or if second exam at which mean systolic was >=140 mmHg or mean Diastolic >=90 mmHg | 0=Free of disease 1=Prevalent disease | n=6283 n=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 or unrecognized), Coronary Insufficiency (Unstable Angina), or Fatal Coronary Heart Disease |
| STROKE | Atherothrombotic infarction, Cerebral Embolism, Intracerebral Hemorrhage, or Subarachnoid Hemorrhage or Fatal Cerebrovascular Disease |
| CVD | Myocardial infarction (Hospitalized and silent or unrecognized), Fatal Coronary Heart Disease, Atherothrombotic infarction, Cerebral Embolism, Intracerebral Hemorrhage, or Subarachnoid Hemorrhage or Fatal Cerebrovascular Disease |
| HYPERTEN | Hypertensive. Defined as the first exam treated for high blood pressure or 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 followup or Number of days from Baseline to censor date. Censor date may be end of followup, death or last known contact date if subject is lost to followup |
| 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 during followup or Number of days from Baseline to censor date. Censor date may be end of followup, or last known contact date if subject is lost to followup |

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.