Framingham Heart Study Description

Background

The National Heart, Lung, and Blood Institute (NHLBI)[1] created a teaching dataset that includes real but anonymized data collected as part of the Framingham Heart Study. The Framingham Heart Study is one of the most influential and longest running epidemiological studies of risk factors for cardiovascular disease ever run. The study started in 1948 and continues today to collect extensive data from original participants, their children, and their children's children. Much of what we know about cardiovascular disease was discovered by investigators involved in the Framingham Heart Study. In fact, studies to date using data collected in the Framingham Heart study have resulted in over 3000 publications in high impact, peer-reviewed medical journals.

The Framingham Heart Study has been widely discussed in the media. WGBH in Boston produced a video documentary for PBS entitled "The Hidden Epidemic: Heart Disease in America" that details the history of heart disease in this country and highlights the Framingham Heart Study.[2] In 2007, CBS News did a story on the study, its participants, and its impact.[3] Additionally, research results from the Framingham Heart Study are communicated widely, most recently highlighting the discovery of a gene that may promote obesity[4] and new data showing declining rates of dementia.[5] Interested readers can visit the Framingham Heart Study website for a detailed history of this incredible study and its many contributions to preventive medicine.[6]

Datasets for Analysis

NHLBI created a longitudinal teaching dataset includes clinical, laboratory, and outcome data on *n* = 4434 participants. Each participant has between one and three observations—which represent examinations held approximately 6 years apart. There are a total of 11,627 observations in the full dataset. A detailed description of the Framingham Heart Study dataset and other public use datasets available from NHLBI are available on the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (BioLINCC) website.[7]

Two datasets are available for analysis—one is the complete dataset with *n* = 11,627 observations (or person-exams), and the second includes only data collected at the first examination for each participant (*n* = 4434). FHS_all.csv contains *n* = 11,627 observations and FHS_exam1.csv contains *n* = 4434 observations.

---

[1] http://www.nhlbi.nih.gov/
[2] http://www.pbs.org/wgbh/takeonestep/heart/
[3] http://www.cbsnews.com/videos/landmark-heart-study/
[4] http://www.cbsnews.com/news/how-a-fat-gene-may-influence-your-weight/
[5] http://www.cbsnews.com/news/dementia-alzheimers-risk-signs-of-hope-study/
[6] https://www.framinghamheartstudy.org/about-fhs/history.php/
[7] https://biolincc.nhlbi.nih.gov/static/studies/teaching/framdoc.pdf?link_time=2016-07-06_14:21:55.514359

Variables

The following variables are available in each dataset for analysis (extracted from the complete documentation file, available on the NHLBI BioLINCC website[8]).

Variable Name	Description	Coding Details/Range
RANDID	Unique identification number for each participant	2248-9999312
SEX	Participant sex	1 = Male, 2 = Female
PERIOD	Exam cycle	1, 2, 3
TIME	Number of days since first (baseline) exam	0–4854
AGE	Age at exam, years	32–81
SYSBP	Systolic blood pressure, mmHg	83–295
DIABP	Diastolic blood pressure, mmHg	30–150
BPMEDS	Use of anti-hypertensive medication	0 = No, 1 = Yes
CURSMOKE	Currently smoking cigarettes	0 = No, 1 = Yes
CIGPDAY	Number of cigarettes smoked per day	0 (non-smoker)–90
TOTCHOL	Total serum cholesterol, mg/dL	107–696
HDLC*	High density lipoprotein cholesterol, mg/dL	10–189
LDLC*	Low density lipoprotein cholesterol, mg/dL	20–565
BMI	Body mass index = weight (kg)/height $(m)^2$	14–57
GLUCOSE	Serum glucose, mg/dL	39–478
DIABETES	Diabetes (glucose > 200 mg/dL or on treatment)	0 = No, 1 = Yes
HEARTRTE	Heart rate, beats/minute	37–220
PREVAP	Prevalent angina pectoris	0 = No, 1 = Yes
PREVCHD	Prevalent coronary heart disease (CHD)	0 = No, 1 = Yes
PREVMI	Prevalent myocardial infarction (MI)	0 = No, 1 = Yes
PREVSTRK	Prevalent stroke	0 = No, 1 = Yes
PREVHYP	Prevalent hypertension	0 = No, 1 = Yes
The following are outcome events coded 1 if the event occurred during the follow-up (only the first event is recorded).		
ANGINA	Angina pectoris	0 = No, 1 = Yes
HOSPMI	Hospitalized for MI	0 = No, 1 = Yes
MI_FCHD	Hospitalized for MI or fatal CHD	0 = No, 1 = Yes
ANYCHD	Any coronary heart disease event	0 = No, 1 = Yes
STROKE	Stroke	0 = No, 1 = Yes
CVD	Cardiovascular disease	0 = No, 1 = Yes
HYPERTEN	Hypertension	0 = No, 1 = Yes
DEATH	Death from any cause	0 = No, 1 = Yes
The following are numbers of days from the first (baseline) exam to the first event during the follow-up. If no event occurred, time is end of follow-up, death, or last known contact date.		

[8] https://biolincc.nhlbi.nih.gov/static/studies/teaching/framdoc.pdf?link_time=2016-07-06_14:21:55.514359

TIMEAP	Time from baseline to first angina	
TIMEMI	Time from baseline to first myocardial infarction	
TIMEMIFC	Time from baseline to first MI or fatal CHD	
TIMECHD	Time from baseline to first CHD	
TIMESTRK	Time from baseline to first stroke	
TIMECVD	Time from baseline to first cardiovascular disease	
TIMEHYP	Time from baseline to first hypertension	
TIMEDTH	Time from baseline to death	

*Available only at period = 3 exam, missing otherwise