

Chapter 7

The Framingham Heart Study

You can observe a lot just by watchin'.

– Yogi Berra

Contents

- 7.1. Early Misconceptions
- 7.2. The Beginning of the Study
- 7.3. The Framingham Risk Score
- 7.4. Other Impacts of the Study
- 7.5. A Different World
- 7.6. Notes and Sources

In 1948, Americans were ready for a time of luxury and indulgence after surviving the Great Depression and World War II. Refrigerators were full of whole milk, steaks, and cream. One million households owned televisions, compared to just 5,000 households in 1945. The first McDonald's was opened in San Bernardino, California, and almost half of all Americans smoked.

But amidst all this luxury and indulgence, heart disease had emerged as an epidemic. One in three deaths in the United States was caused by heart disease, more than four times the rate in 1900. Physicians were puzzled as to what was causing heart attacks and strokes, but even worse, there was nothing to be done if heart disease was discovered. High blood pressure and heart disease were being accepted as inevitable consequences of aging.

This chapter is about the Framingham Heart Study, and how all of this changed with the help of analytics. The Framingham Heart Study started in 1948, with the goal of observing a large population of healthy adults over time to better understand the factors that led to heart disease. It was a study that revolved around data. Over 80 variables were collected about 5,209 people at a time when computers did not exist and everything had to be recorded on paper. This data has now been used in more than 2,400 studies, and the *Texas Heart Institute Journal* declared the Framingham Heart Study one of the top 10 cardiology advances of the twentieth century, alongside the electrocardiogram and open-heart surgery.

7.1 Early Misconceptions

In 1948, physicians were trained to believe that high blood pressure was a natural part of aging. Blood pressure measures the amount of pressure exerted by circulating blood on the walls of the arteries, and is measured in two numbers, like 120/80. The first number is systolic blood pressure, which measures the pressure when the heart beats, pushing blood through the arteries. The second number is diastolic blood pressure, which measures the pressure between beats, when the heart is at rest. In 1948, physicians believed that a systolic blood pressure of 100 plus one's age was healthy. Today, a systolic blood pressure of 140 is considered high, and medication is prescribed to lower blood pressure and thereby reduce the stress on artery walls.

In 1945, three years before the Framingham Heart Study started, President Franklin Delano Roosevelt, the 32nd President of the United States, died from heart disease at age 63. This stunned the American people, and surprised the President's closest friends and family. Little did they know that while President Roosevelt was leading the country through the Great Depression and World War II, his blood pressure was steadily increasing to alarming levels. In 1931, his blood pressure was already 140/100. Today, a healthy blood pressure is less than 120/80, so by today's standards this is already considered high blood pressure. One year before President Roosevelt's death, his blood pressure was 210/120, which is today called hypertensive crisis, and emergency

care is needed. At the time, the President's personal physician said that this was "a moderate degree of arteriosclerosis, although no more than normal for a man of his age." Two months before his death, his blood pressure was 260/150, and the day he died his blood pressure was 300/190. These numbers are very rarely even seen today, and if they are, it is considered a crisis and treatment is immediately given.

The President's personal doctor hid Roosevelt's deteriorating health from his family and the public. It's unclear if this deception was medical ignorance or a deliberate cover-up, but even if the President was properly diagnosed, there was no treatment at the time to give. The death of the President of the United States due to heart disease served as a wake-up call, bringing together physicians and politicians behind a massive research effort to understand the causes of heart disease.

7.2 The Beginning of the Study

When the Framingham Heart Study began in 1948, it was unlike anything that had ever been conducted before. The plan was to track a large cohort of initially healthy patients over time, and the city of Framingham, Massachusetts was selected. This particular town was selected for several reasons: it was of an appropriate size; the population was very stable, and they expected the participants to be around to observe for many years; the doctors and residents of the town were willing to participate; and the town was close to Boston, a city that had several prestigious medical schools and research institutions.

In total, 5,209 men and women between the ages of 30 and 59 volunteered to participate in the study. The inclusion of women is surprising for a study that began in 1948, and therefore many people believe that the study was ahead of its time in recruiting men and women. However, the reason for including women in the study was because medical science believed at the time that women were immune to heart disease, and they wanted to discover what was protecting women. Researchers later discovered that women were not immune to heart disease, they just acquired it at a lower rate than men (one in three women will develop heart disease in their lifetime, compared to one in two men). In addition to discovering if men were truly more susceptible than women, the leaders of the heart study hoped to discover if family history played a role in heart disease, and if smoking, drinking, obesity, inadequate sleep, high blood pressure, or high cholesterol had anything to do with the development of heart disease.

The Data

The 5,209 volunteers were given a questionnaire and an exam every 2 years, starting in 1948. In the beginning, the physicians collected data on 80 different variables, all recorded on paper. The variables collected included: family and

personal history of disease; present conditions, complaints, and drugs taken; height, weight, temperature, respiratory rate, and mental state; an x-ray of the heart; electrical activity traced on an electrocardiogram; samples of urine and blood; examination of the pupils, retinas, thyroid, and liver; and body type measurements (like body-mass index, or BMI). The exams and questions expanded over time to include additional risk factors, such as physical activity and genomic information.

The Framingham Heart Study is now in its seventh decade of collecting and analyzing this data. In 1971, the Framingham Offspring Study was started, which follows 5,124 sons and daughters of the original Framingham Heart Study volunteers. In 2002, the Third Generation Study was started, which follows 3,500 grandchildren of the original Framingham Heart Study volunteers.

First Findings

The first paper of key findings of the Framingham Heart Study, “Coronary Heart Disease in the Framingham Study,” was published in 1957, almost ten years after the study began [41]. The paper published results that were surprising at the time, including the unexpected rate of heart disease in women, and evidence that both high blood pressure and high cholesterol are risk factors for heart disease. In 1961, the researchers leading the Framingham Heart Study published a groundbreaking paper showing that by detecting and treating certain *risk factors*, like high blood pressure and high cholesterol, heart disease could be prevented [74].

7.3 The Framingham Risk Score

Dealing with the large amount of data collected in the Framingham Heart Study was a huge hurdle for statisticians working on the project. They responded by developing *multivariate analysis*, a method that is at the core of virtually every epidemiology study done today. Multivariate, or multiple variable, models are used to capture the effects of multiple risk factors on a single outcome. In this particular application, the researchers were trying to understand how risk factors such as age, blood pressure, cholesterol, smoking, and diabetes affect a person’s risk of developing heart disease.

In 1998, a landmark paper describing the Framingham Risk Score was published [152]. The Framingham Risk Score is computed using a simple prediction algorithm involving a few key risk factor categories that allows physicians to predict coronary heart disease risk in patients. It has been adjusted several times since the original publication, and in 2004 a paper was published that developed a “points system” to make the models more useful to practitioners [135]. Today, individuals can go online and predict their 10-year risk of having a heart attack (<http://cvdrisk.nhlbi.nih.gov/>) using a simple online form.

In this section, we describe a multiple logistic regression model to predict an individual's ten year risk of developing coronary heart disease (CHD) that we developed using the Teaching Dataset provided by the National Heart, Lung, and Blood Institute. This dataset is an anonymized version of a large subset of the original Framingham Heart Study data. This data can be downloaded from the online companion for this book and can be used to build the models described in this chapter (www.dynamic-ideas.com/analytics-edge).

After building and evaluating the model, we will describe how the model translates into a points system. The Framingham Risk Score is calculated using a related method, called *Cox proportional hazards*, but we focus here on a logistic regression model because it is simpler and is commonly used in the medical literature. To learn more about logistic regression, see Chapter 21.

The Data, Variables, and Model

The data used to develop this model is an anonymized sample of 4,240 participants from the original Framingham Heart Study. The dataset includes exam results and questionnaire data from the examination in 1956, and whether or not each participant developed heart disease in the ten year period from 1956 to 1966. The participants were between the ages of 32 and 70 and free of heart disease at the start of the time period.

When building the Framingham Risk Score models, researchers only used a limited number of the available risk factors, due to correlations between the risk factors and because some variables (like obesity, family history, or physical activity) can be difficult to quantify. They also wanted to make sure that all of the risk factors used were readily available in clinical practice. Likewise, we use a few simple risk factors in the model built in this chapter: age, male (1 if the participant is male, 0 otherwise), systolic blood pressure, diabetes diagnosis (1 if the participant has diabetes, 0 otherwise), and current smoking status (1 if the participant smokes, 0 otherwise). We will denote these variables by Age, Male, SystolicBP, Diabetes, and Smoker.

Our ultimate goal is to predict an individual's 10-year risk of developing coronary heart disease (CHD), which is indicated by a heart attack or coronary death. A total of 644 participants developed CHD in the 10-year period (15% of all participants in our dataset).

To build the model, we first randomly split the data into two pieces: 70% of the data to train the model (the training set), and 30% of the data to test the model out-of-sample (the testing set). This is important to evaluate how the model performs on new patients. All patients in the test set will not be used to construct the model, so when we make predictions on these patients, we are able to estimate how good the model will do at predicting the 10-year risk of new patients.

Using the training set, the following logistic regression model was developed to predict the probability of a participant developing CHD:

$$\text{Logit(CHD)} = -7.7013 + 0.0524 \times (\text{Age}) + 0.6555 \times (\text{Male})$$

$$\begin{aligned} & + 0.0205 \times (\text{SystolicBP}) + 0.6723 \times (\text{Diabetes}) \\ & + 0.2991 \times (\text{Smoker}) \end{aligned}$$

All of the five independent variables are significant in the model. This equation describes the “logit” of the logistic regression model, and can be translated to a probability of developing CHD with the following equation:

$$\text{Probability of CHD} = \frac{1}{1 + \exp(-\text{Logit(CHD)})}$$

For example, suppose that we have a patient who is 65 years old, male, has a systolic blood pressure of 145, has diabetes, and is not a smoker. We can use these equations as follows:

$$\begin{aligned} \text{Logit(CHD)} &= -7.7013 + (0.0524 \times 65) + (0.6555 \times 1) \\ &+ (0.0205 \times 145) + (0.6723 \times 1) + (0.2991 \times 0) = 0.005, \end{aligned}$$

$$\text{Probability of CHD} = \frac{1}{1 + \exp(-0.005)} = 0.501.$$

So, according to our model, this patient has a 50.1% chance of developing CHD in the next 10 years. As another example, suppose we have a patient who is 40 years old, female, has a systolic blood pressure of 155, does not have diabetes, and is a smoker. Then we can calculate for her that $\text{Logit(CHD)} = -2.1287$, and her probability of developing CHD in the next 10 years is 0.106. The treatment guidelines indicate that a 10-year risk above 20% should be handled with aggressive treatment, while a 10-year risk between 10% and 20% indicates that more moderate treatment is needed.

How good is this model? The Receiver Operator Characteristic (ROC) curve for the model is shown in Figure 7.1. An ROC curve gives the false positive rate (the percentage of patients that *did not* develop CHD that the model predicted *incorrectly*) and the true positive rate (the percentage of patients that *did* develop CHD that the model predicted *correctly*) for varying threshold values on the predicted probabilities. If we select a higher threshold, we can achieve a lower false positive rate, but we also have a lower true positive rate. For this data, we can select a threshold value that gives us a true positive rate of about 60%, while only having a false positive rate of about 25%. This means that we will make correct predictions for 60% of the patients that do end up developing CHD, while incorrectly flagging 25% of the patients who do not actually end up developing CHD. If we want a higher true positive rate, we can select a lower threshold value. The ROC curve tells us that we can select a threshold value that will yield a true positive rate of about 85%, with a false positive rate of about 60%. In this case, we would make correct predictions for

85% of the patients that end up developing CHD, but we also incorrectly flag 60% of the patients who do not end up developing CHD.

The area under the curve (AUC) of the model is 0.72. The AUC captures the model's ability to distinguish between the two outcomes. So while the model does make mistakes, the AUC tells us that the model can correctly distinguish between a randomly selected patient who will develop heart disease in 10 years and one who will not develop heart disease in ten years 72% of the time. For clinical prediction models, which are often very challenging to develop, this is a promising result. For more about ROC curves and AUC, see Chapter 21.

In this particular situation, there is an important tradeoff between false positives and false negatives. Remember that a false positive prediction means that for a person who actually *did not* develop CHD in the next ten years, the model predicted that they *would*. This prediction might unnecessarily alarm the patient, and unnecessary treatment might be applied. However, the patient remains healthy. A false negative prediction means that, for a person who actually *did* develop CHD in the next ten years, the model predicted that they

Figure 7.1: The Receiver Operator Characteristic (ROC) Curve for the model on the training set.

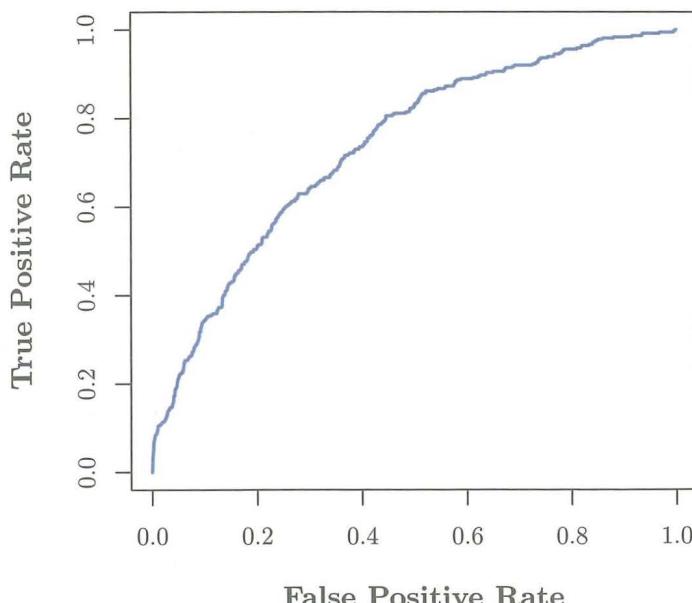


Table 7.1: Classification matrix for the logistic regression model on the training set. A threshold of 0.25 is used. The columns are labeled with the predicted outcome, and the rows are labeled with the actual outcome.

	Predicted No CHD	Predicted CHD
Actual No CHD	2,204	313
Actual CHD	289	162

would not. This prediction could result in no preventative action being taken with the patient, and so the model did not help this patient improve their health. For this reason, false negatives are often seen as more costly long-term than false positives in this application. While the immediate cost of a false positive might be higher due to the unnecessary treatment, the long-term cost of a false negative prediction can be drastic due to the expense of treating CHD (not to mention the emotional cost to the patient). Therefore, we are more inclined to select a lower threshold.

Table 7.1 shows the classification matrix on the training set with a threshold of 0.25. We predict that 475 participants will develop heart disease in the next 10 years - these patients would receive preventative treatment according to our model. If clinicians had used this model in 1948, they would have unnecessarily treated 313 people who did not develop CHD (or 12.4% of the “no CHD” patients), but they would have correctly treated 162 people who did develop CHD (or 35.9% of the “CHD” patients). If we want to decrease our false negative rate even further, we could lower the threshold more.

On the test set, again using a threshold of 0.25, we correctly predict CHD for 68 out of the 193 patients who develop CHD, and we incorrectly predict CHD for 130 out of the 1,079 patients who did not develop CHD. Our AUC on the test set is also 0.72. Overall, this model can give physicians an edge in deciding which patients to prioritize for treatment. Using even more advanced methods and more detailed data, the model can be improved further.

A Points System

The equations developed here make predictions possible. However, a points system was developed to make the calculation even easier for physicians and patients. To implement the points system, each of the risk factors is first organized into meaningful categories and assigned a reference value, which is just the mid-point of values in that category. These numbers are given in Table 7.2.

The categories for age and blood pressure are meaningful because they are designed to mirror clinically meaningful states. In particular, the blood pressure categories are defined by using the Joint National Committee’s blood pressure

Table 7.2: Variable categories and reference values.

Variable	Category	Reference Value
Age	30-39	34.5
	40-49	44.5
	50-59	54.5
	60-69	64.5
	70-79	74.5
Sex	Female	0
	Male	1
Systolic Blood Pressure	< 120	107
	120-129	125
	130-139	135
	140-159	150
	≥ 160	175
Diabetes	No	0
	Yes	1
Current Smoker	No	0
	Yes	1

categories. Note that the age categories only include values that were in the original data. More validation would be necessary if we wanted to extrapolate the model beyond the values seen in the training set.

The next step to develop the points system is to make a “base” computation, by adding up all of the reference values that represent low risk, according to the risk prediction model. If we look at the logistic regression equation, we can see that the lowest risk is for a person in the age category 30-39, who is female, has a systolic blood pressure of less than 120, does not have diabetes, and is not a current smoker. We will actually use the blood pressure category 120-129 for the base model, and give “bonus points” to people who have an even lower blood pressure.

Table 7.3 adds two new columns to Table 7.2: the difference in reference value from the base value for each variable (Base Diff), and the Base Diff column multiplied by the corresponding logistic regression coefficient for that variable (Logit Units).

The next step is to scale and round the “Logit Units” column to produce an integer number of points. We will use a five year increase in age as the scaling factor, or $5 \times 0.0524 = 0.262$, so each number in the Logit Units column of Table 7.3 is divided by 0.262, and rounded to the nearest integer (recall that

0.0524 is the coefficient for the variable “Age” in our logistic regression model). This produces the points values given in Table 7.4.

Now given these point values, each value can be assigned a risk estimate. The point value represents the value of the Logit (relative to the baseline) divided by 0.262. So we can compute an estimate of the Logit for a given points value P by using the Logit equation for a “baseline” person, and adding $0.262 \times P$:

$$\begin{aligned}\text{Logit(CHD)} &\approx -7.7013 + (0.0524 \times 34.5) + (0.6555 \times 0) \\ &\quad + (0.0205 \times 125) + (0.6723 \times 0) + (0.2991 \times 0) \\ &\quad + (0.262 \times P) \\ &= -3.331 + (0.262 \times P)\end{aligned}$$

So for each possible point value (a person can have a minimum point value of -1 and a maximum of 18) we can compute the probability of developing CHD, or the risk estimate, by using the following equation:

Table 7.3: Variable categories and reference values, with the difference from the base category and the logistic regression model units.

Variable	Category	Reference	Base Diff	Logit Units
Age	30-39	34.5	0	0
	40-49	44.5	10	0.5240
	50-59	54.5	20	1.0480
	60-69	64.5	30	1.5720
	70-79	74.5	40	2.0960
Sex	Female	0	0	0
	Male	1	1	0.6555
Systolic BP	< 120	107	-18	-0.3690
	120-129	125	0	0
	130-139	135	10	0.2050
	140-159	150	25	0.5125
	≥ 160	175	35	0.7175
Diabetes	No	0	0	0
	Yes	1	1	0.6723
Current Smoker	No	0	0	0
	Yes	1	1	0.2991

$$\text{Probability of CHD} \approx \frac{1}{1 + \exp(3.331 - 0.262 \times P)}$$

These values are given in Table 7.5.

While slightly complicated to develop, this points system significantly simplifies the process for a patient or physician to compute a risk of CHD. They only need to use Table 7.4 to compute the number of points for an individual patient given the variable categories, and Table 7.5 to match that point value with a risk estimate. The risk estimates in the table are just estimates, but they correspond well with the estimates made by the exact equation.

Alternatively, individuals can go online and predict their 10-year risk of having a heart attack (<http://cvdrisk.nhlbi.nih.gov/>) using a simple online form. The individual inputs their personal data, and then the form gives a prediction of CHD. This website might be just as simple to use as the points system, but the points system adds an extra level of transparency to the model that can be insightful since patients can easily see what they can do to reduce their risk in terms of blood pressure and smoking. For example, suppose a 65 year old male has a systolic blood pressure of 145, does not have diabetes, and is a current smoker. He would have a points total of 12, and a risk estimate

Table 7.4: Variable categories and points values.

Variable	Category	Points
Age	30-39	0
	40-49	2
	50-59	4
	60-69	6
	70-79	8
Sex	Female	0
	Male	3
Systolic Blood Pressure	< 120	-1
	120-129	0
	130-139	1
	140-159	2
	≥ 160	3
Diabetes	No	0
	Yes	3
Current Smoker	No	0
	Yes	1

of 0.4534. But if he was able to reduce his systolic blood pressure from 145 to 128, he could reduce his points total to 10, and reduce his risk to 0.3294. If he also quit smoking, his risk would further decrease to 0.2743. The points system allows the patient to understand how much each action reduces his risk.

The points system is an example of how analytics models can be simplified to become more useful to people who do not necessarily understand the mathematics behind the model. Providing patients and physicians with logistic regression equations would not be feasible, but two tables of results are much easier to understand and use.

Validating the Model on Other Populations

The Framingham Risk Score was shown to be predictive for the Framingham Heart Study population. However, there were concerns about whether or not the results could be extended to other populations. The Framingham Heart Study population was mostly white and middle-class. In 1999, researchers set out to better test the validity of the Framingham Risk Score in ethnically and racially diverse populations. Studies were performed on Native Americans

Table 7.5: Point values and risk estimates.

Points	Risk Estimate
-1	0.0268
0	0.0345
1	0.0444
2	0.0569
3	0.0728
4	0.0925
5	0.1170
6	0.1469
7	0.1829
8	0.2253
9	0.2743
10	0.3294
11	0.3896
12	0.4534
13	0.5187
14	0.5835
15	0.6454
16	0.7029
17	0.7545
18	0.7998

(The Strong Heart Study), Japanese American men (The Honolulu Heart Program), whites and blacks (The Atherosclerosis Risk in Communities Study), and Hispanic men (The Puerto Rico Heart Health Program).

In each study, the Framingham Risk Score was used to predict CHD risk, and the predictions were compared to actual outcomes. The model performed well in whites and blacks. Black participants had a higher diabetes rate, but the predictions handled it very well. Only minor recalibration was needed for other groups. For example, Japanese men had a much lower CHD rate, so the predictions needed to be scaled down to be accurate for this population. These studies provided strong evidence that the Framingham Risk Score could be used in other settings.

7.4 Other Impacts of the Study

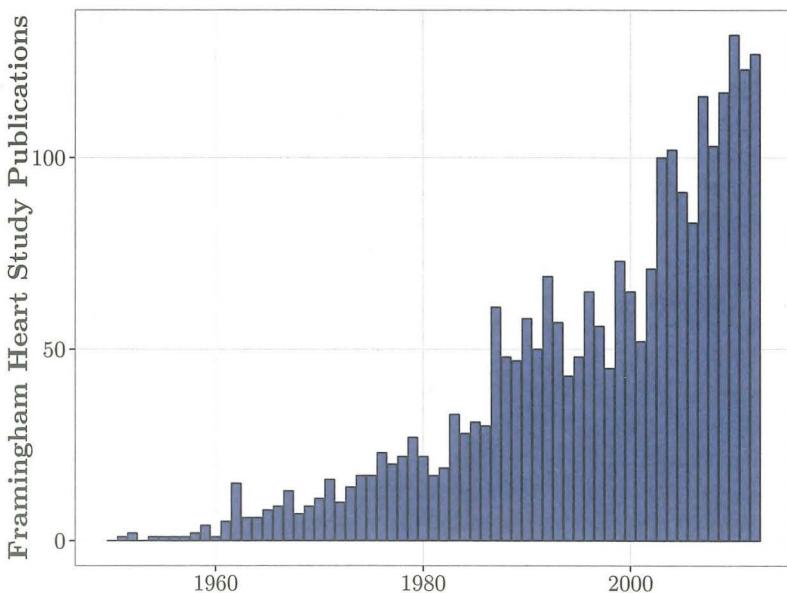
In addition to the Framingham Risk Score, the Framingham Heart Study had a very significant impact in other ways. More than 2,400 studies have used the Framingham Heart Study data, and the number of publications is increasing substantially (see Figure 7.2). In particular, the study was key to the development of drugs to lower blood pressure and hypertension, and it also paved the way for clinical decision rules in other areas of medicine.

Development of Medication

In 1948, drugs to lower blood pressure were too toxic for practical use. The side effects of the drugs were severe – one caused projectile vomiting, another blocked the whole autonomic nervous system, and a third turned the patient blue. But in the 1950s, the drug chlorothiazide was developed. Chlorothiazide belongs to a class of drugs called diuretics, drugs that are still used today to treat high blood pressure. The Framingham Heart Study gave Ed Freis, a leading researcher at the time on the treatment of hypertension, the evidence he needed to argue for testing the effects of blood pressure drugs in clinical trials. Since most physicians believed that hypertension was benign, he needed the justification provided by the Framingham Heart Study to pursue a clinical trial testing whether or not lowering blood pressure would reduce a patient's risk for cardiovascular disease. He started the Veterans Administration Trial, which ultimately established that lowering blood pressure can be used to decrease risk of heart disease. Now, there is a billion dollar market for diuretics worldwide, and there are dozens of drugs available to treat blood pressure.

In addition to drugs for lowering blood pressure, the Framingham Heart Study was instrumental in the development of drugs to lower cholesterol. The study published evidence that high cholesterol was a risk factor for heart disease in the early 1960s, but it was not until the 1970s that the first cholesterol lowering drug was developed. This drug was the first statin, a class of drugs for which there is now a \$20 billion market. Patients now have a choice between more than a dozen cholesterol-lowering drugs.

Figure 7.2: Number of studies using the Framingham Heart Study data published each year, from 1960 - 2012.



Clinical Decision Rules

The Framingham Heart Study also paved the way for clinical decision rules in many different areas of medicine. Clinical decision rules are models that predict clinical outcomes using data such as patient characteristics and test results. More than 75,000 papers have been published across medicine regarding clinical decision rules, and the rate is still increasing. Figure 7.3 shows the number of clinical decision rules that have been published by year, since 1960. Today, over 7,000 clinical decision rules are published every year.

Some popular clinical decision rules predict the likelihood of strep, predict whether or not a patient with joint pain will develop Rheumatoid Arthritis in one year, and whether or not a knee injury requires the use of an x-ray to assess a fracture. These models are able to incorporate many different risk factors, and are unbiased and unemotional. One of the key arguments for the use of clinical prediction rules is that they can assist new physicians with little experience to draw from when making decisions.

While thousands of clinical prediction rules exist, they are not used as much in practice as they could be. One reason is the fact that most of these prediction rules are not presented in a user-friendly way. With approaches like the points system presented here, researchers hope to increase the use of clinical

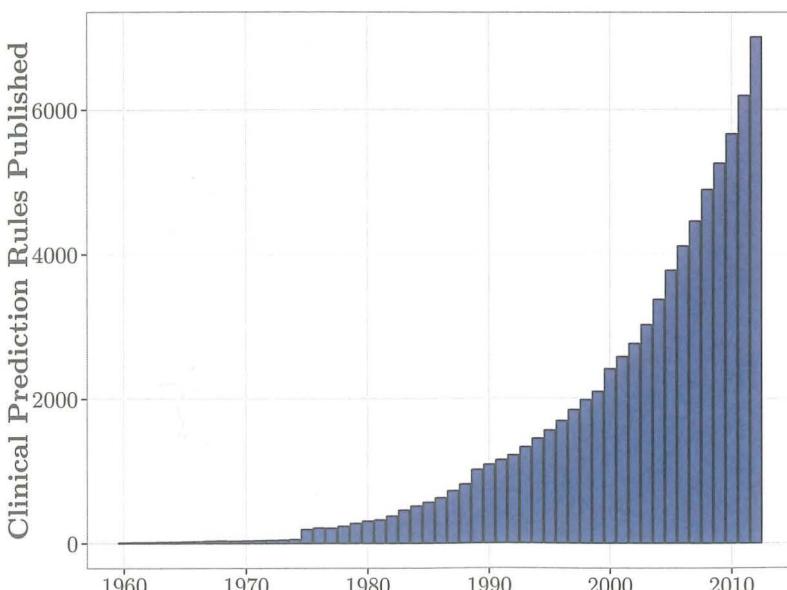
prediction rules.

7.5 A Different World

Since its peak in 1963, the death rate from coronary heart disease has fallen 60% and the death rate from stroke has fallen 66%. While heart disease is still the leading cause of death in the United States, we have made tremendous progress in understanding and treating this disease. The Framingham Heart Study and the studies that have followed from it have led to a better understanding of the risk factors involved in heart disease, and the hope is that more discoveries will occur in the coming years.

Sandra Merloni, one of the original volunteers for the heart study said that “it’s amazing what the whole country has learned from this study. It’s not just benefiting Framingham or Massachusetts. It’s benefiting the whole world, really.” With the continuation of the Framingham Heart Study and the analytical decision rules that it has inspired, the data collected has greatly influenced the way heart disease is treated today.

Figure 7.3: Number of clinical decision rules published each year, from 1960 - 2012.



7.6 Notes and Sources

Information about the history of the Framingham Heart Study in this chapter comes from the book *A Change of Heart: Unraveling the Mysteries of Cardiovascular Disease* by Levy and Brink [81].

- 7.3.** The Framingham Risk Score information comes from the papers Sullivan et al. [135] and Wilson et al. [152].