1 Framingham Heart Study: health risk analytics

<u>Tool</u>: Logistic regression

<u>The Analytics Edge</u>: The risk of having coronary heart disease (CHD) ten years from now is predicted using data available on patients such as their current blood pressures, cholesterol level, smoking habits, age and other such relevant variables. Using a logistic regression model, it is possible to develop good predictive rules. Such models have spawned the use of clinical decision rules and correspondingly new markets for drugs and intervention programs.

2 Overview

The Framingham Heart Study (FHS) (`www.framinghamheartstudy.org`): In 1948, the FHS under the direction of the National Heart Institute embarked on an ambitious and landmark project in health research. At that time little was known about the causes of heart diseases and strokes. Furthermore, the death rates from heart diseases were increasing.

The objective of the FHS study was to identify the common factors that contributed to heart diseases by following its development over a long period of time in a large group of patients who had not yet developed overt symptoms of heart diseases.

2.1 Origins of FHS

The origins of FHS is closely linked to the cardiovascular health of the US president Franklin D Roosevelt and his premature death from hypertensive heart disease and stroke in 1945.

1. 1932: Roosevelt's blood pressure was 140/100.

2. 1935-1941: There was gradual rise in his blood pressure from 136/78 to 188/105. Despite the rising blood pressure, his personal physician insisted that the president's health was fine and his blood pressure was normal for a man of his age.

3. 1945: Leading up to his death at the age of 63, his blood pressure rose from 186/108 to 240/130 to 300/190 when he passed away.

Harry Truman who was vice president under Roosevelt, became president and signed into law the National Heart Act. This law allocated a US $500000 seed grant for a 20 year epidemiological study of heart disease. Today, our belief is that normal blood pressure (systolic/diastolic) is in the range of (90 to 120)/(60 to 80).

Year	Tuberculosis	Other Communicable Diseases	Neoplasms	Cardiovascular Diseases	External Causes of deaths	Disease of Early Infancy	Other Causes of Deaths	Per Cent Total
	TOTAL DEATHS BY BROAD CAUSES							
1950	12.0	32.5	2.8	6.3	4.0	7.2	35.3	100.0
1955	9.1	21.7	6.5	8.5	4.9	11.1	38.3	100.0
1960	6.3	18.7	10.4	10.6	5.0	11.2	37.8	100.0
1965	6.2	12.6	13.9	13.7	6.5	9.9	37.2	100.0
1970	4.2	12.7	15.1	27.0	7.9	5.9	27.1	100.0
1975	3.7	11.7	18.5	29.4	4.1	3.5	29.0	100.0
1980	1.8	11.4	21.0	34.4	7.2	3.3	20.9	100.0
1985	1.3	11.6	22.0	34.8	8.1	2.5	19.7	100.0
1990	0.8	10.3	23.9	37.1	7.3	2.2	18.4	100.0
1995	0.8	14.3	25.2	35.7	7.1	1.3	15.6	100.0
2000	0.6	13.9	27.0	36.6	7.2	0.8	13.8	100.0
2005	0.4	17.9	26.5	33.3	6.3	0.7	15.1	100.0
2006	0.4	15.3	28.8	33.2	6.3	0.7	15.4	100.0
2007	0.5	16.1	27.7	34.0	6.0	0.5	15.2	100.0
2008	0.5	15.3	29.3	33.6	5.8	0.6	14.9	100.0
2009	0.4	16.8	29.3	32.8	5.7	0.6	14.3	100.0
2010	0.4	17.2	28.8	33.0	5.5	0.5	14.5	100.0
2011	0.4	17.7	30.4	31.7	4.6	0.5	14.6	100.0
2012	0.4	18.0	30.6	31.1	5.6	0.5	13.9	100.0
2013	0.3	19.7	30.9	30.4	4.9	0.5	13.3	100.0
2014	0.3	20.0	29.9	30.9	4.7	0.5	13.8	100.0
2015	0.2	20.4	30.1	30.7	4.5	0.5	13.6	100.0
2016	0.2	20.4	29.9	30.5	4.4	0.5	14.0	100.0
2017	0.2	21.3	29.8	31.3	4.0	0.4	13.0	100.0
2018	0.1	21.8	29.1	30.2	4.3	0.5	14.1	100.0

Figure 1: Total death by broad causes: Source: MOH Singapore

DAYSPRING

HDL (ie 'good cholesterol').
- Replace food rich in saturated fats (pork, beef, mutton, cheese, coconut milk) and cholesterol (egg yolk, liver, kidney, brain) with skinless poultry, fish and low fat milk. Avoid oily, fatty and fried food as well as sugary food and starches.
- Increase intake of fruits and vegetables.
- Quit cigarette smoking, if you are a smoker.
- Repeat your cholesterol test in 3-6 months.

Glucose

Your blood glucose level is normal. There is no evidence of diabetes noted.

CORONARY HEART DISEASE (CHD) RISK ASSESSMENT

Based on the results of the health questionnaire and screening parameters, your risk of developing myocardial infraction or coronary death in the next 10 years is *Low (3%)*.

Lipid goal for Low risk group	
Total Cholesterol	<200
LDL	<160
HDL	≥40

ADVICE:

- Do take special effort to minimize your modifiable risk factors with healthy dietary, exercise and lifestyle habits. Find out more at *http://www.dayspring.sg/results#cdra*.

Note: The 10-Year CHD Risk Score for Chinese, Malay and Indian males and females in Singapore is derived from the Framingham-based NCEP ATP III 10-Year Risk Score Tables which have been modified taking into account the Singapore cardiovascular epidemiological data. This modification was carried out as part of a collaboration between investigators at the Singapore Ministry of Health, Singapore General Hospital, National University of Singapore and Prof. Ralph B D'Agostino from the Framingham Heart Study, USA. Since there are insufficient data for other ethnic minorities, it is recommended that the 10-Year CHD Risk Score for the lowest risk group (i.e. Chinese) be used for these individuals.

CONCLUSION

Once again, thank you for screening with Dayspring. For more information on the screening results, please refer to *http://www.dayspring.sg/results*.

No definite diagnoses may be made from the test results alone. If there are abnormal findings, please consult your doctor for follow-up. In addition, normal test results may not necessarily mean the absence of a medical condition.

In accordance with a Ministry of Health (MOH) advisory, any person undergoing general health screening is to be attended by a medical practitioner when the screening results are available. A referral letter is attached for your convenience.

Figure 2: Sample health report

2.2 Timeline

1. 1948: 5209 men and women between the ages of 30 and 62 were enrolled in the FHS from the town of Framingham, Massachusetts, United States of America. The first round of extensive physical examination and lifestyle interviews were conducted with participants returning every two years for examinations and tests.

2. 1971: The second generation of 5124 original participant's children and spouses were enrolled into the program.

3. 1994: An Omni cohort consisting of 507 men and women of African-American, Hispanic and Asian origins who were residents of Framingham were enrolled to reflect the increasing diversity of residents in the area.

4. 2002: Third generation of participants were enrolled.

5. 2003: A second group of Omni cohort was enrolled.

Most of the knowledge concerning heart diseases such as the effect of diet, exercise and blood pressure that we know today is based on this longitudinal study.

3 Data and analysis

The dataset and the model that we will study is inspired from an influential paper in 1998 titled "Prediction of coronary heart disease using risk factor categories" by Wilson, Agostino, Levy, Belanger, Silbershatz, Kannel in the journal Circulation published by the American Heart Association. This paper has been cited over 9800 times as of September 2019.

The main contribution of the paper lies in discussing the development of a clinical decision rule (predictive algorithm) that predicts the 10 year CHD (coronary heart disease) risk using risk factors of blood pressure, total cholesterol and LDL cholesterol in a white middle aged population sample from the Framingham Heart Study. This allows physicians to predict the CHD risk in patients who do not currently have CHD.

In our analysis, we will use a dataset from the website http://biolincc.nhlbi.nih.gov/home/ which is an anonymized dataset. The data on 4434 participants with data collected during three examination periods, approximately 6 years apart from 1956 to 1968 with participants followed for 24 years.

4 Intervention strategies and impact

The results from the FHS has been validated with external validation methods by generalizing to different populations such as African-Americans, Asians.

The advantage of such an approach is that it can be used to develop intervention strategies for example drugs to lower blood pressure, drugs for lower cholesterol. The effects of these on reduced chances of coronary heart diseases can be tested by doing clinical trials. The market for diuretics (to reduce blood pressure) and statins (to reduce cholesterol levels) are now in billions of dollars.

The FHS has also led to an increase in clinical decision rules in many areas of healthcare that predict clinical outcomes using patient data and results. These models are unbiased, unemotional and can assist new physicians with little experience to make decisions.

5 Points system

While this approach predicts the risk of patients getting a heart disease, it is not particularly easy for patients and physicians to uses. A points system is often implemented to make these results more interpretable and usable in practice. We can develop such a points system from the logistic regression results as follows.

Assign the lowest risk to a person who is in the low age category, is female, has a systolic blood pressure of lower than 120 (from clinically meaningful states), smokes zero cigarettes per day and has a smaller value for the glucose level. Based on this by appropriately normalizing and scaling the results, we can generate a points system that is particularly useful in telling patients on how improving on a certain aspect can help decrease the risk of CHD.

Example: Here we divide by 0.059×5 and then round to nearest integers to get points (note we have negative points (bonus) here).

Variables	Category	Reference	Base difference	Logit units	Points
Age	30-39	34.5	0	0	0
	40-49	44.5	10	0.59	2
	50-59	54.5	20	1.18	4
	60-69	64.5	30	1.77	6
	70-79	74.5	40	2.36	8
Sex	Male	1	-1	0.54	2
	Female	2	0	0	0
Systolic blood pressure	< 120	106.5	-13	-0.208	-1
	120-139	129.5	0	0	0
	140-159	149.5	20	0.32	1
	≥ 160	174.5	45	0.72	2

Table 1: Points system

6 Summary

Data: The data is a larger dataset that provides information on patients including lifestyle data and clinical measurements. Such datasets were pioneered by the Framingham heart study and is unique in that it is longitudinal and monitors a patient over a long period of time.

Model: A logistic regression model helps characterize the probability of getting a heart disease within 10 years. The results identify the variables that are significantly correlated with the chances of getting a heart disease.

Decision: The model provides an estimate of the probability of getting a heart disease for any new patient. Furthermore it provides intervention strategies that can be used to provide recommendations to patients on better reducing the risk of a heart disease

Value: Such models have paved an entire industry of clinical decisions rules to aid doctors and identify new variables that correlate to an ailment. This affect the lives of almost all of us.

Age	Points
20 – 34	- 9
35 – 39	- 4
40 – 44	0
45 – 49	3
50 – 54	6
55 – 59	8
60 – 64	10
65 – 69	11
70 – 74	12
75 – 79	13

Smoker	Points				
	Age 20 – 39	Age 40 – 49	Age 50 – 59	Age 60 – 69	Age 70 – 79
No	0	0	0	0	0
Yes	8	5	3	1	0

Total cholesterol mmol/L (mg/dL)	Points				
	Age 20 – 39	Age 40 – 49	Age 50 – 59	Age 60 – 69	Age 70 – 79
< 4.1 (160)	0	0	0	0	0
4.1 – 5.1 (160 – 199)	4	3	2	1	0
5.2 – 6.1 (200 – 239)	7	5	3	1	0
6.2 – 7.2 (240 – 279)	9	6	4	2	1
≥ 7.3 (280)	11	8	5	3	1

HDL cholesterol mmol/L (mg/dL)	Points
≥ 1.6 (60)	-1
1.3 – 1.5 (50 – 59)	0
1.0 – 1.2 (40 – 49)	1
< 1.0 (40)	2

Systolic BP (mmHg)	Points	
	If untreated	If treated
< 120	0	0
120 – 129	0	1
130 – 139	1	2
140 – 159	1	2
≥ 160	2	3

Allocate points based on person's age, total and HDL cholesterol levels, smoking status and systolic blood pressure as indicated in the tables to the left.

Check the total points against Table 1 to estimate a person's ten-year CHD risk.

For example, if you are a 45-year-old Chinese male who smokes every day with a total cholesterol of 7.5 mmol/L, a HDL cholesterol of 1.1 mmol/L and a systolic BP of 135 mmHg, then your total score is >20. You are estimated to have a 'high' risk of heart attack or coronary death within the next ten years.

This would mean that more than 20 out of 100 persons in your risk category would experience a heart attack or coronary death within the next ten years.

Table 1. Estimation of ten-year CHD risk for men in Singapore

Total points	Ten-Year Risk (%)		
	Chinese	Malay	Indian
-1	< 1	< 1	1
0	< 1	< 1	1
1	< 1	1	1
2	1	1	1
3	1	1	2
4	1	1	2
5	1	1	3
6	1	2	3
7	2	2	4
8	2	3	5
9	3	4	7
10	4	5	9
11	5	6	11
12	6	8	14
13	8	11	18
14	11	13	> 20
15	13	17	> 20
16	17	> 20	> 20
≥17	> 20	> 20	> 20

Age	Points
20 – 34	- 7
35 – 39	- 3
40 – 44	0
45 – 49	3
50 – 54	6
55 – 59	8
60 – 64	10
65 – 69	12
70 – 74	14
75 – 79	16

Smoker	Points				
	Age 20 – 39	Age 40 – 49	Age 50 – 59	Age 60 – 69	Age 70 – 79
No	0	0	0	0	0
Yes	9	7	4	2	1

Total cholesterol mmol/L (mg/dL)	Points				
	Age 20 – 39	Age 40 – 49	Age 50 – 59	Age 60 – 69	Age 70 – 79
< 4.1 (160)	0	0	0	0	0
4.1 – 5.1 (160 – 199)	4	3	2	1	1
5.2 – 6.1 (200 – 239)	8	6	4	2	1
6.2 – 7.2 (240 – 279)	11	8	5	3	2
≥ 7.3 (280)	13	10	6	4	2

HDL cholesterol mmol/L (mg/dL)	Points
≥ 1.6 (60)	-1
1.3 – 1.5 (50 – 59)	0
1.0 – 1.2 (40 – 49)	1
< 1.0 (40)	2

Systolic BP (mmHg)	Points	
	If untreated	If treated
< 120	0	0
120 – 129	1	3
130 – 139	2	4
140 – 159	3	5
≥ 160	4	6

Allocate points based on person's age, total and HDL cholesterol levels, smoking status and systolic blood pressure as indicated in the tables to the left.

Check the total points against Table 2 to estimate a person's ten-year CHD risk.

For example, if you are a 40-year-old Chinese non-smoker female with a total cholesterol of < 4.1mmol/L, a HDL cholesterol of 1.3 mmol/L and a systolic BP of <120 mmHg, then your total score is 0. You are estimated to have a 'low' risk of heart attack or coronary death within the next ten years.

This would mean that less than one out of 100 persons in your risk category would experience a heart attack or coronary death within the next ten years.

Table 2. Estimation of ten-year CHD risk for women in Singapore

Total points	Ten-Year Risk (%)		
	Chinese	Malay	Indian
5	< 1	< 1	1
6	< 1	< 1	1
7	< 1	1	1
8	< 1	1	1
9	1	1	2
10	1	1	2
11	1	2	3
12	1	2	3
13	1	3	4
14	2	4	6
15	3	5	7
16	3	6	10
17	4	8	12
18	5	10	16
19	7	13	20
20	9	16	> 20
21	12	20	> 20
22	15	> 20	> 20
23	19	> 20	> 20
> 24	> 20	> 20	> 20