2019

The Role of Glucose Level on the Performance of the Framingham Risk Score

Uohna June Thiessen
Walden University

Walden University

College of Health Sciences

This is to certify that the doctoral dissertation by

Uohna June Thiessen

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Shingairai Feresu, Committee Chairperson, Public Health Faculty
Dr. Simone Salandy, Committee Member, Public Health Faculty
Dr. Mountasser Kadrie, University Reviewer, Public Health Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2019

Abstract

The Role of Glucose Level on the Performance of the Framingham Risk Score

by

Uohna June Thiessen

BS, Oakwood University, 2003

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

August 2019

Abstract

Cardiovascular diseases (CVD) are responsible for more deaths than any other disease, continue to threaten the quality of life for many, and is a major burden to the health care system. The Framingham Heart Study (FHS) identified the major CVD risk factors that became essential to effective CVD screening strategies and the Framingham Risk Score (FRS), is used to assess CVD risk. Based on the concepts of the health behavior model and CVD as a cardiometabolic disorder, multivariate logistic regression analysis was used to evaluate the association between fasting blood glucose (FBG) levels and a CHD event, and to determine the value of FBG replacing a diagnosis of diabetes (DM2) in the FRS. The data set consisted of the 2,677 subjects of the FHS III cohort. In the univariate analysis, both DM2 and FBG were statistically significant (both p =.000), but the association was stronger for DM2, b = 2.138, OR = 8.483 (95% CI: 4.229, 17.105) than for FBG, b = .015, p = .000, OR=1.015 (95% CI: 1.009, 1.022). When adjusted for age, blood pressure, cholesterol, and smoking status, only DM2 remained statistically significant, OR = 2.295, p = .041, (95% CI: 1.035, 5.087) in the model. The FBG version of the FRS did not provide any improvement in performance, as it was marginally inferior to the DM2 version. Furthermore, the interactions between FBG and the metabolic risk factors were not statistically significant for this given data set. The results imply that a diagnosis of diabetes remains the factor of choice for inclusion in the FRS model for predicting the 10-year risk of CHD and replacing it with FBG provides little to no practical benefit. These findings support the use of CVD risk factor reduction and the use of effective screening tools in CVD prevention and promotion heart health.

The Role of Glucose Level on the Performance of the Framingham Risk Score

by

Uohna June Thiessen

BS, Oakwood University, 2003

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Public Health

Walden University

August 2019

Dedication

This dissertation document is dedicated to my dear mother, Marguerite Antoine-Roberts, who means so much to me. Her love, sacrifice, dedication, and hard work on my behalf and that of all her children have always been an inspiration to me. My hope is that she will always be remembered and honored for her steadfastness and perseverance.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Introduction

Diseases involving the heart and/or blood vessels are known collectively as cardiovascular diseases (CVDs) and include disorders such as angina pectoris (AP); myocardial infarction (MI); stroke; heart failure (HF); peripheral artery disease (PAD); and coronary heart disease (CHD, or coronary artery disease or CAD). Every year over 800,000 persons in the United States, 4 million in Europe, and another 11 million around the globe die from CVDs (American Heart Association, Center for Disease Control, and Prevention, & National Institutes of Health, 2015; Ferreira-GonzáLez, 2014). In particular, CHD, the most common CVD, is responsible for approximately 70% of all CVD deaths, and it is the leading cause of morbidity and mortality worldwide (Grimes, 2012; Jones & Greene, 2013).

Not only are cardiovascular diseases responsible for more deaths than any other disease type, they are a major burden to the society. CVDs cause more physical disabilities, are a major financial drain on the healthcare system, and they also contribute to many health disparities (World Health Organization, 2016). In the United States alone, CVDs claim the lives of 2,200 Americans every day, while another 92 million remain alive to deal with the disease and its many related complications (Benjamin et al., 2017). In 2012, those living with cardiovascular disease and its physical, mental and social impact, generated $316 billion, in direct health care cost and loss of productivity combined (Benjamin et al., 2017). Consequently, responding to the epidemic of heart disease has been the focus of many national and international healthcare organizations.

Given the severity of the threat they pose, there is a concerted effort at reducing the incidence and prevalence of CVDs. The American Heart Association uses records of risk factors from epidemiological and clinical data to monitor the prevalence and incidence of CVD and track the results of prevention and treatment efforts (Goff et al., 2014). The 2020 Impact Goal remains committed to reducing CVD deaths by 20% and to a general improvement in cardiovascular health by 20% by the year 2020 (Benjamin et al., 2017). The Healthy People 2020 initiative uses early risk identification and treatment to "improve cardiovascular health and quality of life through prevention, detection, and treatment of risk factors" (U.S. Department of Health and Human Services, 2018). Critical to any effective early identification and prevention strategy is effective risk factor detection and classification.

Background

Over half a century ago, the Framingham Heart Study (FHS) expanded the understanding of CVDs by identifying the major risk factors involved (Dawber et al., 1959; Dawber, Meadors, & Moore, 1951). The FHS research continues to today and now includes other chronic diseases but CVD remains its primary focus (D'Agostino, Pencina, Massaro, & Coady, 2013; Kannel & McGee, 1979; Mahmood, Levy, Vasan, & Wang, 2014). The major CVD risk factors uncovered by the FHS include gender, age, cholesterol level, smoking status, blood pressure, and history of heart disease (Dawber et al., 1959; Dawber & Lansing, 1966; Kannel, Castelli, Gordon, & Mcnamara, 1971). Subsequent epidemiological and clinical research have confirmed and further elucidated

the role of these risk factors in the long-term development of CVD (Centers for Disease Control and Prevention, 2013)

Most of the established CVD risk factors are modifiable, and it is for this reason mainly that CVD is classified as a preventable disease. The CDC reported that 80% of all CVD deaths in 2010 were the direct result of these modifiable factors, and as such those deaths were considered avoidable (Centers for Disease Control and Prevention, 2013). In fact, prevention strategies focused on risk factor reduction have resulted in significant decreases in the prevalence, the incidence, the morbidity, and the mortality of CVD (Goff et al., 2014; Grundy et al., 2002; Stone et al., 2013). Effective CVD prevention is dependent on early detection and prediction, for which accurate and reliable risk formulas are essential (Eichler, Puhan, Steurer, & Bachmann, 2007; Fuster & Kelly, 2010; Keaven Anderson, Wilson, Odell, & Kannel, 1991;Wilson et al., 1998).

One of several CVD risk formulas is the framing risk score (FRS) formulated by the Framingham heart study researchers. The initial version, published in 1998, was developed using many of the risk factors discovered several decades earlier (Wilson et al., 1998). The original FRS formula was used to determine the likelihood of an individual developing CVD over a given period, most commonly, 10-year period (Wilson et al., 1998). A revised version, released 4 years later, replaced *diabetic status* with *dyslipidemia* and added *hypertension medication* to represent a history of heart disease (Grundy et al., 2002). This was based on diabetes now being considered a CVD risk equivalent, and with the omission, diabetics were now regarded as having a *history of CVD* (D'Agostino et al., 2008). In addition, the third and most recent version of the FRS

is now gender specific, and male and female FRS includes measures of age, total and HDL cholesterol, systolic blood pressure, and cigarette smoking (Dahlöf, 2010; O'Donnell & Elosua, 2008).

Presently, almost anyone with relevant medical information can obtain a measure of their 10-year risk of CVD. The FRS is available in many formats including survey questionnaire printouts, as tables and charts for readout, or as computerized programs for healthcare professionals, and even as smart phone applications for lay persons (D'Agostino et al., 2008; "Framingham Risk Score Calculator for Coronary Heart Disease," 2018). The FRS is the most widely used CVD screening tool in the United States (Steyerberg et al., 2010), and has become the heart of the public health response to the CVD epidemic (Kones, 2011; Mahmood et al., 2014; Wilson et al., 1998). Since its introduction in the late 1990s, alternative CVD risk screening approaches have been developed, but the FRS remains the most popular (Günaydın et al., 2016; Tsao, 2015).

Although the FRS has been an invaluable tool in the task of screening for individual CVD risk, it has several noteworthy limitations. First, the experts have confirmed that the FRS's predictive ability is limited to future coronary heart disease (CHD) only and does not extend to other coronary events, such as stroke, ischemic attack or heart failure (D'Agostino et al., 2013; Root, Hu, & Duncan, 2014). Second, the FRS overestimates CVD risk in some populations and underestimates CVD risk in others, impeding its usefulness in targeting high risk individuals for prevention treatment (Günaydın et al., 2016; Kones, 2011; Van Staa et al., 2014). Third, the FRS's inability to

track changes in risk levels over time, also presented opportunities for formula optimization (Dahlöf, 2010).

One of the more popular areas of formula optimization for CVD risk models has been the reevaluation of the role of diabetes as a risk factor. There are several reasons for this reevaluation but the main two are (a) the changes in the working definition of the disease diabetes (new glucose threshold and new diagnostics), and (b) the significant increase in the incidence and prevalence of diabetes over the last 50 years (Mayfield, 1988; Shaw, Zimmet, McCarty, & de Courten, 2000; Wareham, 1998). Reevaluation of diabetes was also supported by the discrepancy relationship between observational studies, with respect to CVD and diabetes treatment, and findings in clinical trials, at the end of the last decade (Qazi & Malik, 2013a). The efforts to abate the increased CVD mortality and morbidity of diabetics by treating hyperglycemia has proven ineffective (Duckworth et al., 2009; Kelly, Bazzano, Fonseca, & al., 2009; Miller et al., 2008; Reaven et al., 2009). This led to the call for more research into the interaction between diabetes and cardiovascular diseases and the level of CVD risk possessed by those with diabetes.

The role of the diabetic status as factor in the FRS has undergone several changes over the years. Diabetic status was included as binary yes/no variable in the first published FRS formula several decades after the FHS was launched (Wilson et al., 1998). And, despite subsequent augmentation, namely replacing dichotomous or categorical variables (e.g. cholesterol and blood pressure) with their continuous counterparts, diabetes remained dichotomous, only to be eventually omitted from the formula (Grundy

et al., 2002). However, a diagnosis of diabetes based simply on an arbitrary threshold for blood glucose ignores the complexity of the association between glucose levels and CVD risk, a complexity that is confirmed in clinical research (Faeh, Rohrmann, & Braun, 2013; Fawwad, Moin, Siddiqui, Hydrie, & Basit, 2016; Park et al., 2013). But the eventual exclusion of diabetes from the FRS has resulted in the loss of significant risk information and may have rendered it less effective in predicting heart disease.

The relationship between blood glucose and the risk of CVD provides important CVD risk prediction. Previous research show that blood glucose levels both above and below the diabetes threshold are associated with the probability of CVD development (Dahlöf, 2010; Kadowaki et al., 2008; Park et al., 2013; Perreault et al., 2014; Valentino et al., 2015). It is now accepted that hyperglycemia, hypoglycemia, and glycemic variability are all implicated in the macrovascular dysfunction common in diseases of the heart and circulatory system (Benjamin et al., 2017; Saisho, 2014; The Emerging Risk Factors Collaboration, 2010; Zhang et al., 2014). Moreover, recent research claims that glucose level may be a better predictor of CVD development than is cholesterol level and supports its inclusion in CVD risk formulas (Braun, Bopp, & Faeh, 2013; Clark, Perkins, Carson, Boyd, & Jefferson, 2015; Lammertyn et al., 2011).

Unlike a decade ago when the clinical evidence was contradictory, several studies now show that the control of blood glucose levels does improve symptoms of CVD (Chi, Snaith, & Gunton, 2017; Coch & Green, 2016; Kelly et al., 2009; Xu & Rajaratnam, 2017). Carter et al (2016) recently found that among the primary risk indicators (hypertension, cholesterolemia, and hyperglycemia) hyperglycemia is the most effective

cardiometabolic marker and has the strongest association to high CVD risk in African Americans. Prediction models that exclude glucose levels, even at the prediabetes state, ignore important CVD risk prediction and stratification information (Huang, Cai, Mai, Li, & Hu, 2016). More accurate risk prediction information could ultimately lead to more effective early prevention strategies. And it is this theory that was the basis of the statistical analysis in my research project.

Problem Statement

Despite decades of clinical and epidemiological research, estimating the risk of CVD in the general population remains challenging. The FRS, the most commonly used prediction tool, has many known limitations with regard to precision and accuracy (Brindle et al., 2005; Eichler, Puhan, Steurer, & Bachmann, 2007). Even with its widespread use, the FRS is far from infallible, frequently generating overestimations or inconsistent results, with significant classification errors, which has been reported as high as 37% on one occasion (D'Agostino et al., 2008; Kones, 2011). A more effective and accurate risk assessment tool is critical to an improved public health response to the burden of CVDs (Goff et al., 2014; Kones, 2011).

Recognition of CVD as a cardio-metabolic disease can lead to important improvements in the risk prediction effort. Some of the most key CVD risk factors (BMI, high cholesterol, and high blood pressure), along with insulin resistance and hyperglycemia, are all symptoms of metabolic syndrome (MetS), which is itself a most effective predictor CVD risk (Clark et al., 2015; D'Agostino et al., 2008; Goff et al., 2014). Treating CVD as a cardiometabolic disorder, and using metabolism indicators

such as blood glucose levels as a risk factor, may provide valuable risk prediction information (Cockram et al., 2001; D. Lloyd-Jones et al., 2010). In this research I aimed to demonstrate the utility of blood glucose levels in improving the discriminatory power and classification ability of the FRS predicting the most injurious of CVDs, CHD.

Purpose

This study investigated the effect of blood glucose (BG) levels on the performance of the FRS formula. The approach involved generating an alternative version of the FRS formula, in which the binary variable diabetic status was replaced with continuous variable, BG levels values and evaluating the prediction performance of the FRS version. The coefficients or odds ratio (OR) of the two risk factors (BG and diabetic status), as well as the overall performance of respective multiple logistic regression (MLR) models was assessed for both. The role of the comorbid, metabolic risk factors (BMI, cholesterol, blood pressure), was also examined for any interaction with and then their effect on the predictive role of blood glucose level on the model's performance. The objective was to produce a more accurate, and therefore reliable algorithm for predicting CHD event probability among asymptomatic individuals (see D'Agostino et al., 2013; Van Staa et al., 2014). The formulas were based on the 10-year CHD risk algorithm the FRS generated from the original Framingham heart study data set.

Research Questions

I sought to answer the following three research questions:

RQ1: Is the relative risk for glucose level higher than that for diabetic status in the CHD 10-year risk prediction?

$H_0$1: There is no difference between the relative risk for glucose level and that for diabetic status in the CHD 10-year risk prediction.

H_a1: The relative risk for glucose level is higher than that for diabetic status in the CHD 10-year risk prediction.

RQ2: Is the measure of accuracy for glucose level higher than that for diabetic status in the CHD 10-year risk prediction formula?

$H_0$2: There is no difference between the measure of accuracy for glucose level and diabetic status in the CHD 10-year risk prediction formula.

H_A2: The measure of accuracy for glucose level is higher than that for diabetic status in the CHD 10-year risk prediction formula.

RQ3: Is the measure of accuracy for glucose level independent of 'BMI', cholesterol level or blood pressure level in the CHD 10-year risk prediction formula?

$H_0$3: The measure of accuracy for glucose level is independent of 'BMI', cholesterol level or blood pressure level in the CHD 10-year risk prediction formula.

H_A3: The measure of accuracy for glucose level is dependent on BMI, cholesterol level or blood pressure level in the CHD 10-year risk prediction formula.

Theoretical Framework

Health Behavior Model

The concepts on which this study is based are the health behavior model (HBM) and the cardiometabolic model of heart disease. The former is borrowed from the social sciences and is very commonly used in health care research and in developing public health strategies (Esparaza-del Villar et al., 2017). There are three main supposition on which the HBM is based: perceived vulnerability, perceived severity, and perceived efficacy (Janz & Becker, 1984). According this model, an individual's health behavior is motivated by their belief and that belief is based on the information and level of health education to which they have been exposed or are in their awareness. HBM suggests that proper knowledge of the association between risk factors and the related diseases is critical to heart healthy behavior (Khorsandi, Fekrizadeh, & Roozbahani, 2017). The identification of the CVD risk factor, dysglycemia, and the level of risk they present is the goal of this study.

Cardiometabolic Model of CVD

In the cardiometabolic model, there is a close relationship between the elements involved in metabolic disorders and those involved in heart disease (Brunzell, Davidson, & Fuberg, 2009). It is already accepted that a diagnosis of diabetes incurs a two-four-fold increase risk of heart disease, depending on gender (Booth, Kapral, Fung, & Tu, 2006; Conroy et al., 2003; Fox, 2010; Sarwar et al., 2010). I focused on one of the overlapping risk factors, namely dysglycemia, and its inclusion in the CVD risk prediction formula as a predictive factor in predicting the outcome of a CHD event. Dysglycemia t damages the

blood vessels of the heart and this is usually the initiation of or exacerbation of other factors that lead to CHD (Jackson et al., 2016; Kozakova et al., 2017; Saisho, 2014). This analysis attempts to take a closer look at this relationship between glucose level and CHD by measuring its strength the relationship and testing its utility in the CHD prediction formula.

Nature of the Study

This research is a quantitative study using statistical analysis of an existing data set to evaluate the answers to the research questions. The data set is from the third cohort of the FHS clinic data, collected from over 4,000 of the descendants of the first cohort that began in 1964 (Framingham Heart Study Longitudinal Data Documentation, 2004). More details will be included in Chapter 3. The variables of interest are fasting blood glucose (FBG) and diabetes diagnosis (DM2) and they were compared for their predictive ability and their interaction other metabolic comorbidities (cholesterol, blood pressure, BMI), for their effect on the relation to the outcome of a CHD event (Rudestam & Newton, 2014). The methodology involves both univariate and MLR analyses of the FHS data set as secondary data. The Framingham risk score (FRS) is the formula that describes the relationship between the variables, developed by the FHS researchers based on the many risk factor they discovered, and is a multivariable logarithm of the probability of developing coronary heart disease. The same latest version of the 10-year FRS was used in this research to evaluate and compare the predictive value of FBG versus DM2.

Assumptions, Limitations and Delimitations

Assumptions

The FHS is a seminal epidemiological study in the fields of public health and of medicine and has provided invaluable insight into the factors involved in the development of cardiovascular diseases (O'Donnell & Elosua, 2008). The subsequently derived FRS has proven instrumental in predicting cardiac events for many at risk but asymptomatic members of the population (O'Donnell & Elosua, 2008). I assumed that the data FHS collected was accurate and included objective responses from the participants, competence on the part of the health care workers' in conducting surveys and in performing examinations, and accurate recording and documenting of the relevant information.

Despite the limitations and the challenges to the internal and external validity of the original Framingham cohort, the study successfully identified several major CVD risk factors. The researchers were careful to point the steps needed for internal and external validation that other researchers should put in place (O'Donnell & Elosua, 2008; Pencina et al., 2009). These factors do provide general applicability to similar demographics, even if calibration is required in cases of the nonuniform distribution of variable among certain groups or for groups with risk factor prevalence that are different from the FHS (Hermansson & Kahan, 2018; Tsao, 2015)(Moons et al., 2012). The foundation laid by the FHS continues to be the benchmark for CVD research, but the limitations of the tools and the persistence of the disease necessitates an ongoing effort at improving on CVD risk formulas' performance.

Limitations

Despite the great successes of the FHS, the usefulness of the FRS, and the many effective succeeding projects, there are shortcomings that must be addressed (Schlendorf, Nasir, & Blumenthal, 2009). First, the outcome of the original cohort focused solely on CHD and therefore the data does not extrapolate well in the prediction of other coronary-related events (Mahmood et al., 2014). Second, given that several groups were underrepresented in the original cohort, use of the FRS is less than ideal for certain groups like for young people or for ethnicities other that are not whites of European descendent (Hemann, Bimson, & Taylor, 2007). Thirdly, the original delineation of no prior CVD events and the challenges to internal validity (the response rate, recall bias, mortality/attrition, etc.) resulted in a sample that was healthier than the general population (Mahmood et al., 2014).

Notably, the initial FHS was conducted when diabetes was less prevalent, and the diabetes protocol used was different from the present. The previous criterion for a diagnosis of diabetes was a random blood glucose level of 200 mg/dL or more, but the current criterion is a fasting blood glucose of 125 mg/dL or more (Framingham Heart Study Longitudinal Data Documentation, n.d.). As a result of the difference in diabetes diagnosing, the FRS's discriminatory powers within current populations, where diabetes is now more prevalent, is compromised. Other adverse features of the original cohort include the absence of factors like metabolic risk factors such as C-reactive protein (CRP) levels and coronary artery calcium (CAC) (Kones, 2011). These issues, combined with restriction of risk estimation to only a 10-year period, as opposed to a 5-year or 30-

year period, all leave opportunity for optimization of the FRS formula (Schlendorf et al., 2009). This optimization can take many different directions, but this study focuses on the role of fasting blood glucose levels (FBGs) in CVD risk estimation.

Delimitations

The FHS researchers considered the long-term epidemiological study design most suitable for studying the lifestyle and environmental factors involved in the development of cardiovascular disease (O'Donnell & Elosua, 2008). The participants were restricted, due to the logistics of the early stages of the study, to adults living in the town of Framingham, Massachusetts (Dawber & Lansing, 1966; Mahmood et al., 2014) The middle-class white, mostly of European descent, residents of this town were believed (erroneously) to be geographically, socioeconomically, and environmentally representative of the rest of the country (Mahmood et al., 2014). The variables measured and recorded, including age, gender, serum cholesterol, diastolic and systolic blood pressure, glucose, and diabetes, were predetermined by the study leaders as potential risk factors relevant to the outcome (Dawber et al., 1951).

The original cohort sampling spanned 5 years from October 1948 to 1953 and potential candidates were between the ages of 20 and 70 years (Dawber & Lansing, 1966). From that group, two-thirds of the all the families in the town were selected at random, and all eligible members in those families were invited to participate. Subsequently, there were only two edibility criteria- being 30 to 59 years old and having had no history of cardiovascular disease (Dawber & Kannel, 1966). Of the 6,507 invited to participate, some 5,209 were interested in participating, resulting in a 68.7% response

rate (Tsao, 2015). The researchers believed that this was an adequate sample size for generating reliable results, but there were still other challenges to the internal and external validity of the results (Mahmood et al., 2014). The internal and external validity of the data collected will be discussed further in chapter three.

Significance of the Study

The uncertainty surrounding the pathophysiology of CVD, and its often-subclinical symptomology, serve to complicate most detection and prevention strategies (K. M. Anderson, Wilson, Odell, & Kannel, 1991; Kones, 2011). Comorbidity between the risk factors and the varying prevalence rates among the different populations both challenge any effort at risk assessment (D'Agostino et al., 2013; Haregu et al., 2016; Scheerbaum et al., 2017). The limitations of the current screening tools —inadequate scope, insufficient sensitivity, and sometimes, imprecise results— continue to underscore the need for risk estimation augmentation (Kones & Rumana, 2014; D. Lloyd-Jones et al., 2010). This research is a part of the need for augmentation research that advances the effectiveness of risk prediction formula.

The increased prevalence of metabolic diseases, which do strongly predispose to heart disease, makes both metabolic and diabetic factors essential CVD risk factors. Glucose is a most fundamental maker for cardio-metabolic given its significant relationship to both developing diabetes and developing CVDs (Haregu et al., 2016; Huang et al., 2016; Park et al., 2013; Valentino et al., 2015). As such, glucose should be an important predictor of cardiovascular health as well. This dissertation research will provide additional information on the effect that blood glucose levels predictive potential

and its potential for improving the performance of the FRS-CHD risk assessment formula.

Positive Social Change

This research focuses on glucose as a major risk factors associated with cardiovascular disease and aims to further elucidate its role in CVD risk prediction. The findings will impact the way cardiovascular disease is screened for in the field and the way it is treated in the both the outpatient and hospital setting (Hosseini, 2015). Secondly, it will help guide the physicians' decisions about where and how to direct often limited resources in CVD prevention and treatment efforts, especially for those at high risk (Hosseini, 2015). Finally, the information generated may influence health education and literacy efforts among the low and moderate-risk population, providing greater incentive for behavioral change adoption (see Rosenstock, 1974).

I was motivated to complete this study by the burden that exists for more effective screening tools and for a greater efficiency and wider applicability of the CVD screening protocols (see Kones, 2011; Lloyd-Jones et al., 2010; Roger et al., 2012). The results could be used to support treatment regimen and the education of the public about the effectiveness of the positive lifestyle and behavioral changes effective in protecting against this deadly disease (see Goong et al., 2016). This research may provide much needed insight about cardiometabolic pathophysiology that informs future epidemiological and clinical research (see Dahlöf, 2010; Kones, 2011). The objective is to move the field closer towards the goal of reduced CVD mortality and morbidity and a

greater quality of life and heart health for those at risk, as well as the afflicted by CVD and those who care for them.

Summary

CVD is currently the deadliest of all diseases and it is a major burden on the healthcare system. The FHS uncovered the many CVD risk factors and shifted the emphasis from treatment to prevention. The FRS is a multivariate logistic regression formula used to predict the probability of developing CHD and has been used effectively in reducing the incidence and prevalence of the CHD. Despite its effectiveness and widespread use, the FRS has some notable limitations that necessitate optimization. By nature, CVD is believed to be a cardiometabolic disease, as many metabolic markers are strongly associated with the disease outcome. I hypothesized that replacement of the dichotomous variable of diabetic status with that of the continuous variable glucose levels in FRS should provide an improve stratification and classification power to the formula.

Chapter 2: Literature Review

Introduction and Literature Review Strategy

The purpose of this literature review is to detail the burden of CVD, both past and present, and to establish the need for ongoing CVD research, specifically in area CVD risk assessment. Although, there has been a decline in the incidence and prevalence of CVD over the last 7 decades, it remains the leading cause of death in most nations (Emelia J. Benjamin et al., 2017). This decline is largely the result of the identification of the major risk factors followed by concerted and successful efforts in reducing the prevalence of these risk factors, mainly that of cigarette smoking . However, the increase in obesity and other metabolic disorders pose new threats to cardiovascular health . Other studies have looked at the role of various risk factors in risk assessment models, but none have focused the role of glucose level in the most common risk assessment formula, the FRS.

This second chapter of the dissertation will begin with the delineation of the pathophysical model and the conceptual framework on which this research is based. This will be followed by a discussion of the history and epidemiology of heart disease and how it went from being a rare disease to an epidemic. Next will be an account of the research into heart disease and how the FHS got its start and how it became a seminal study, not just in heart disease but in public health in general is then presented. The first half of chapter will close out with a description of how the CVD risk factors were discovered and how their discovery influenced the CVD screening and prevention therapy in use today.

The second half of this chapter will focus on the multivariate risk model that came out of the FHS FRS. Each of the risk factors included in the FRS and their role in CVD screening is described in detail. The next section will focus on the disorders of carbohydrate metabolism and how they are related to the development of heart disease. Glucose level, as the main metabolic marker, will be a highlighted with information about how it is measured, how it affects the metabolic system, and its role in screening for CHD. The discussion will also report on the current state of glucose lowering therapy in CHD prevention therapy. The chapter will end with a discussion on the main risk assessment tools, their formula performance evaluation and their optimization over the years.

The literature search for this chapter was conducted using several electronic databases that cover research in the areas of medicine and health. These included PubMed, Medline, CINAHL, ProQuest, Springer, Google Scholar, and Elsevier. The keywords used for searching these databases include *cardiovascular disease, CVD, coronary heart disease, CHD, CVD risk factor, diabetes, blood glucose, blood sugar, pre-diabetes, metabolic syndrome, CVD risk assessment,* and *CVD risk tools.* The Boolean operators (AND, OR, NOT) were used to ensure the relevance of the returned journal articles. For articles not available online, I made requests to the Walden University Library staff and was either provided with the article or directed to an alternative available source. For current epidemiological data on heart disease and related risk factors, I used the websites of scientific and governmental intuitions such as the CDC, WHO, AHA, NIH and NCLBI.

Given the nature and importance of the topic of heart disease, there was an abundance of available sources. However, a critical review allowed for restriction to articles that had relevance to the objective of the research. Articles chosen focused on the most common and the most dangerous of the CVDs, CHD and its risk factors, the tools for CHD assessment, and the role of blood glucose as a risk factors and in risk assessment tools. Efforts were made to obtain the primary research article or the original report of the findings, and to restrict use of only research conducted by or associated with reputable institutions. On obtaining access to the articles, the abstract was perused to determine the relevance of the article and if it appeared relevant then the full text was reviewed. Additional articles were found in the references of the articles returned from the original search. All articles deemed important were then saved to a Mendeley account where they were tagged and filed for later reference and citation in the document.

Theoretical Framework

Health Behavior Model

The theoretical framework used for this research is the health behavior model (HBM), which was developed using theories from social and behavioral sciences (Esparaza-del Villar et al., 2017; Janz & Becker, 1984). HBM was first used in the 1950s by the U. S. Public Health Service in their efforts to curb the spread of tuberculosis (TB) (Esparaza-del Villar et al., 2017; Hosseini, 2015; Rosenstock, 1974). It remains one of the most popular conceptual frameworks for understanding changes in or explaining and predicting acceptance of new health behaviors (Janz & Becker, 1984). It has been successfully applied to several public health issues such as TB, breast self-exams,

osteoporosis prevention, Hepatitis A & B vaccination, and Pap smear testing (Esparaza-del Villar et al., 2017).

The HBM is based on three main components: (a) perceptions of susceptibility or vulnerability to the condition or disease; (b) perceptions of the severity of the disease and its consequences; and (c) perceptions of the effectiveness of action to avert or ameliorate the disease condition (Janz & Becker, 1984). The other tenets of the model include the perceived barriers to selected actions or action, cues-to or triggers-to these particular actions, and self-efficacy or confidence to carry out the chosen actions (Hosseini, 2015). The energy or force driving the new behavior comes from the perceptions of severity and susceptibility, while the choice of behavioral path is influenced by subjective barriers and any feelings or lack thereof, of self-efficacy (Janz & Becker, 1984). Essentially, a person health behavior is driven by their own perceived susceptibility and whether acting will relieve their susceptibility.

The general tenet of the HBM is that people are only motivated to make changes in their behavior, as it relates to health, in response to some perceived personal threat or impending illness (Esparaza-del Villar et al., 2017; Janz & Becker, 1984). Knowledge of the threat, and the belief that it will be averted by certain actions, is the impetus for the behavioral changes. Conversely, ignorance of the association between the disease and its risk factors or the impact of related behavioral changes causes skepticism about the actions being promoted (Khorsandi et al., 2017). According to HBM, subjects who are provided with appropriate disease risk information will act to mitigate that risk.

In the HBM model, health behavior is closely related to health belief, making knowledge the most important factor in motivation towards change (Rosenstock, 1974; Janz & Becker, 1984). The HBM guides much of the research and the communication of relevant information about the susceptibility to CVD, the severity of its consequences, and the effectiveness of risk factor averting behavior. Behavioral changes and lifestyle modifications have been shown to reduce the risk of developing CVD as well as minimize the severity of the symptoms (Esparaza-del Villar et al., 2017; Goong et al., 2016; Khorsandi et al., 2017). The research objective for my study is to strengthen the link between glucose level and heart disease will increase attention to this risk factor and promote the adoption of related health behaviors.

Cardiovascular Framework

The conceptual model that undergirds this study is the relationships between the established risk factor and the prediction of CVD risk (see Mozaffarian, Benjamin, Go, et al., 2015). The covariates and commodifiers that have an impact on the outcome are included in the model (see Figure 1) as well as the complex interactions among these health determinants and how they are mediated through the specific metabolic risk factors (see Brunzell, Davidson, & Fuberg, 2009; Nguyen et al., 2012; Wilson & Meigs, 2008).. Understanding what biochemical factors predates CVD can provide insight into the etiology of the disease development. Insight into the nature of these factors, their symptoms, and their role can be used to improve effort at predicting the likelihood of the disease.

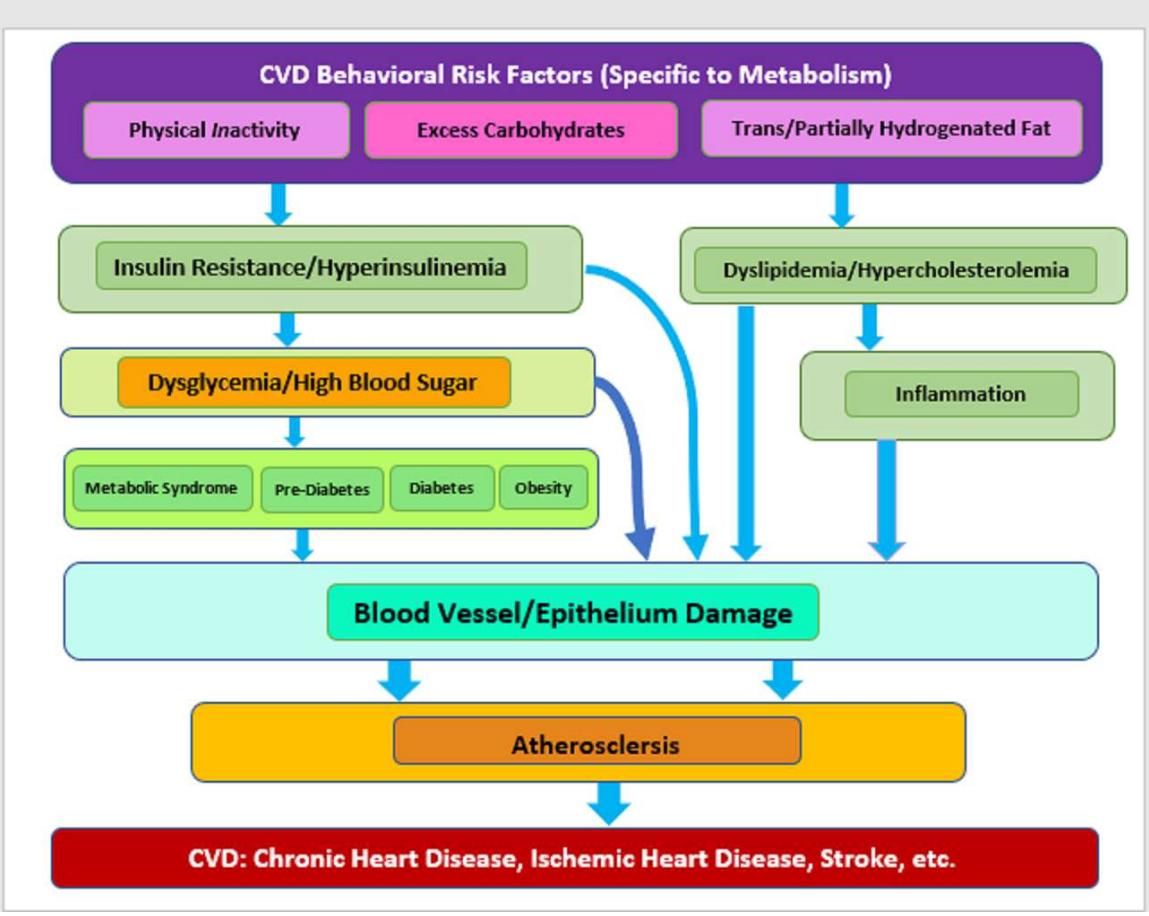


Figure 1. Cardiometabolic model of CVD development.

Heart disease is a disease of the vascular system, detected as the arteries develop atherosclerosis and become clogged with fatty substances or plaques made up of low-density lipoproteins (LDL) and other material . However, it is now accepted knowledge that the atherosclerosis is preceded by damage to the walls of the blood vessels, and the cause of this damage may be the root cause of the disease (Cockram et al., 2001; Jackson et al., 2016; Kozakova et al., 2017; Saisho, 2014). Increases in blood sugar concentration have been shown to be associated with significant damage to the blood vessel walls (Jackson et al., 2016; Tostes & et al., 2009; Zhang et al., 2014) and may therefore be a

major contributor to atherosclerosis and the subsequent heart disease. This process explains the connection between diabetes, and more specifically, hyperglycemia and cardiovascular disease and is basis for cardiometabolic disease theory of CVD.

Based on the medical model of cardiometabolic disease development combined with the principles of the HBM, I investigated the association between the glucose level and the development of CVD with the objective of informing and promoting heart healthy behavior. I focused on the role of blood glucose as a CVD risk factor and how it affects the performance of a risk assessment algorithm, the FRS. These findings provided evidence in support of an emphasis on dysglycemia as a marker for metabolic disorder and a predisposition to CVD. Ultimately, the purpose of the study is to advocate for the inclusion of glucose level in risk detection formula tools and to encourage glucose reduction strategies through behavior that foster cardiovascular health.

History and Epidemiology of CVD

Discovery and Rapid Rise of CVD

A little less than 100 years ago a new era of public health emerged (citation), and it presented novel challenges to the field. The patterns of old diseases were replaced by unfamiliar ones and a different set of problems confronted public health practitioners (U.S. Department of Health and Human Services, 2005). No longer were infectious and communicable diseases the major global threat they had been for centuries; the greater danger now lay in noncommunicable diseases (Rosen , 1958). Optimism over the success of vaccination, improved sanitation, and antibiotics eradication of diseases like small pox and tuberculosis was replaced by the confusion over a new epidemic of chronic non-contagious health issues (Nieto, 1999a).

It all began with shifts in the causes of morbidity and mortality brought on by the emergence of new noncommunicable disease. The cause of the increase of these diseases was unhealthy behaviors and practices . These newly-discovered diseases would eventually become known as *lifestyle diseases* due to their most common trait being a susceptibility to lifestyle changes and behavior modifications (Moore, Chaudhary, & Akinyemiju, 2017). One of those diseases, heart disease, quickly became seen as the deadliest (Nieto, 1999; Ockene, Daley, & Tran, 2014). But this shift in the cause of morbidity and mortality occurred gradually.

Physicians became aware of heart disease somewhere between the middle of the 19th century and the end of World War I (citation). Between 1900 and 1930, CVD moved from fourth to first on the list of causes of death in the United States (Dalen, Alpert,

Goldberg, & Weinstein, 2014). Today, CVD accounts for one out of every three deaths in the United States and in Europe Benjamin et al., 2017; Go et al., 2014; (Kannel & Boston, 1990; Mozaffarian et al., 2015).

The first anecdotal case of heart disease was mentioned in a 1859 report at the Swedish Medical Society meeting, when the pathological features of myocardial infarction (MI) were described (Jones & Greene, 2013). By 1923, angina pectoris, still considered very rare with roughly one case per month, was being documented as resulting from the later stages of infectious diseases like rheumatoid fever and syphilis (Jones & Greene, 2013). But by the 1920s, cases specifically identified as CVD were now ubiquitous and the mortality rate for these diseases rose so sharply that the death rate was doubling approximately every 5 years (Grimes, 2012).

During the 1930s and the 1940s, more cases of CVD, specifically chronic heart disease (CHD), were being readily recognized and classified Grimes, 2012; O'Donnell & Elosua, 2008; (Stehbens, 1995). At around the middle of the 20th century, CVD deaths numbered twice those of TB and it was fast becoming a major public health concern (Grimes, 2012). By the last quarter of the century it was so prevalent and deadly that 650,000 CHD deaths were reported in 1978 alone (Levy, 1981,). It was this dramatic change in the main causes of death, led by CVD, that caught the attention of public health professionals and epidemiological research efforts on CVD were subsequently launched (O'Donnell & Elosua, 2008).

Top on the list of research objectives was identifying the causes of heart disease. Some believed that the cause and the widespread nature of the CVD was due to things

like the increase in smoking, more sedentary lifestyles, and increased consumption of processed food, especially sugar (Rosen, 1958). Most notably, smoking alone saw an eight-fold increase simultaneous to the rapid increase in CVD deaths . The incidence of smoking in the American population jumped from 5% in 1930 to 42% in 1965 and was believed by many to be the main reason for the of CVD rise (Dalen et al., 2014). Others believed the eradication of other deadly diseases and changes in the diagnosis and reporting practice for CVD resulted in an apparent increase (Nieto, 1999). Later research would confirm the role of smoking (cigarette and tobacco) as a major risk factor in the development of cardiovascular and in other lifestyle diseases . Smoking cessation would mark the beginning of heart disease prevention and the curb of the rising CVD rates .

The Declining Rates of CVD

The rapid increase in the prevalence of cardiovascular diseases during the 1940s and 1950s began slowing down in the 1960s (Dalen et al., 2014; Dawber & Lansing, 1966). By the late 1964 the disease incidence peaked and then slowly declined in the latter half of the decade (Dalen et al., 2014; Dawber & Lansing, 1966). The CVD death rates in the US continued to decline, resulting in a 50% decrease over the next 30 years (Jones & Greene, 2013). As the CVD rates declined, there was much speculation about the reason for the decline. The combined effect of personal and communal behavioral changes, the increase in health literacy, the improved medical therapies, and supportive public policy were all credited (Dalen et al., 2014). Some believed that it was the successful combination of primary (risk factor reduction) and secondary (medical intervention) prevention efforts (Fuster & Kelly, 2010; Nieto, 1999b; Pedersen, 2002).

Others attributed it to the reduced prevalence of atherosclerosis that coincided with a drop in cigarette smoking (from 42% to 18% between 1965 and 2012) following the 1964 Surgeon General's report on the dangers of smoking (Dalen et al., 2014).

The CVD rate decline in the US and Europe was hailed as a victory for public health. In 2011 some experts predicted, based on age-standardized extrapolations, that the disease would come to an end in another 5 years (Grimes, 2012)(Grimes, 2012). But that did not happen, as the declining CVD rates slowed considerably and eventually reversed, with CVD rates rising in most developing countries (Bragg et al., 2014; Mensah et al., 2015; Reddy & Satija, 2010). Surprisingly, the continued downward trend in cigarette smoking was not matched by any further decreases in incidence of CVD (Jones & Greene, 2013). The experts, baffled by the trend reversal, redoubled their efforts on continued surveillance and risk factor modeling for the cardiovascular disease complex (Grimes, 2012).

Origins of the Framingham Heart Study

In the early days of the disease, circa 1930s, the rapid rise in CVD mortality and morbidity made the disease a top priority in the public health discourse. The consensus among both health and political leaders was that the situation critically demanded action (Tsao, 2015). That period in history was marked by very little understanding of the cause of the disease and even less knowledge of how to respond to the crisis. During this period of confusion, almost every other death was in some way related to CVD (Mahmood et al., 2014). It was evident to those in the know that a CVD epidemic was in full force and it was seemingly getting worse.

During that time, it seemed no one was protected from the threat of heart disease. Not only were average individuals being affected, but even the most powerful members of the community were at risk as the disease was also impacting the top echelons of society. In 1944, a diagnosis of CVD was confirmed in the nation's leader, Franklin D. Roosevelt, who along with his disability, had been living with the major indicators for years (Rosen G, 1958). President Roosevelt endured many symptoms of CVD during the later years of his life, eventually succumbing to a premature death in 1945 when his blood pressure rose dangerously to 300/190 and he suffered a fatal stroke at the age of 63 (Mahmood et al., 2014).

While still in power however, it was President Roosevelt's administration who established the National Institutes of Health (NIH) as a federal agency in 1940; an act contemporaneous with the growing threat to the nation's heart health (Mahmood et al., 2014). In 1948, Roosevelt's successor Harry S. Truman established the National Heart Act (NHA), which funded the establishment of the National Heart, Lung & Blood Institute (NHLBI) that same year (Mahmood et al., 2014). The hope was that the NHLBI, in response to the dramatic increase in CVD mortality, would somehow be able to unravel the complex etiology of CVD and devise a solution to the epidemic (Lansing, 1961; Jones & Greene, 2013; Mahmood et al., 2014). The benefits of FDR's foresight in establishing these agencies are still be reaped today.

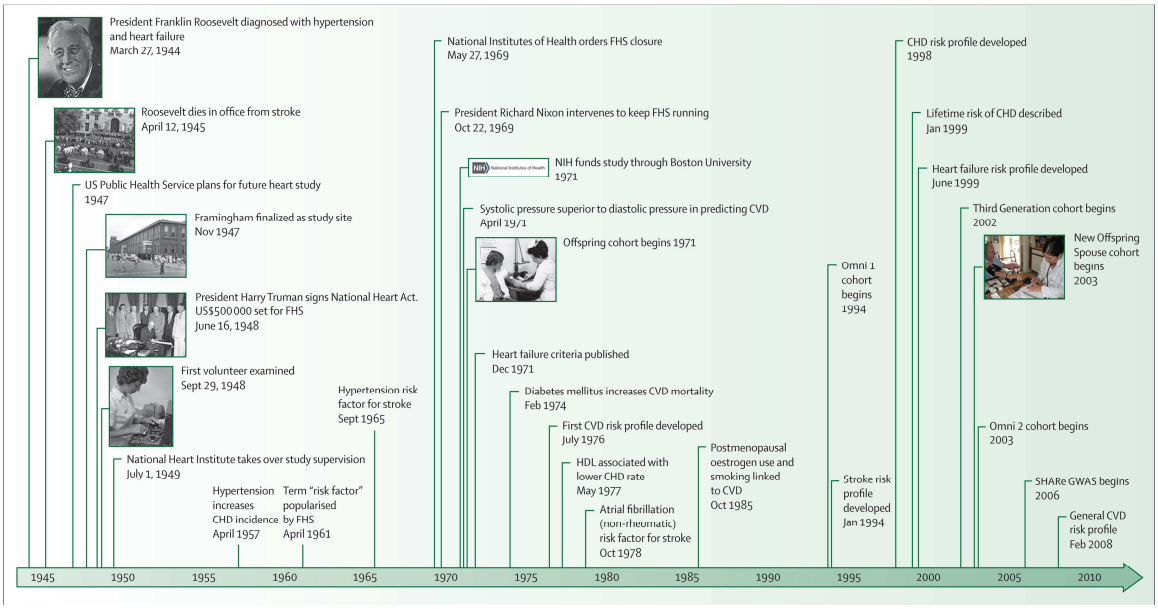


Figure 2. The history of FHS. Graphical representation of the history of FHS. Reprinted from "The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective." by S. S. Mahmood, D. Levy, R. S. Vasan, & T. J. Wang, 2014, *The Lancet*, 383(9921), p 999-1008.

Not long after its formation in 1948, the federally funded NHLBI joined with academic researchers, first from Harvard University and then later from Boston University, to launch the Framingham Heart Study (Tsao & Vasan, 2015a). Improving insight into the etiology, finding ways to effectively treat, and curb the spread of the disease were the objectives (Mahmood et al., 2014; Tsao, 2015). The long natural history of the disease necessitated a prospective, observation-type, cohort study. The data collection included, but was not limited to, information on: consent, family history, medical history, current symptoms, medication, personal and dietary habits, weight, biometrics, cardiac exam, x-ray, ECG, blood (Hb, cholesterol, uric acid, glucose,

syphilis), and urinalysis (Dawber et al., 1951). For each case, collected data were eventually compared to the original clinical findings of the attributes associated with the development of CVD, with attention paid to the most common CVD subtype- CHD (Tsao & Vasan, 2015a).

FHS Through the Years

The FHS began in 1948 with 5,209 participants. The off springs' of the 1st of the 2nd generation (children and grandchildren of the original cohort participants) made up the 2nd and 3rd FHS cohort (Mahmood et al., 2014; Tsao & Vasan, 2015). These groups provided additional understandings, including which individuals were susceptible, the effects of shared environments, and most importantly, familial and multigenerational trait aggregations (Mahmood et al., 2014; Long & Fox, 2016). The smaller OMNI1 and OMNI2 groups, started in 1994 and 2003 respectively, focused on increasing the racial and ethnic diversity of the participant pool beyond the mostly white (of European descent) participants that made up the first three (Long & Fox, 2016; Tsao & Vasan, 2015a)

The first published report from the biennial FHS examinations described the extent of the CVD disease, as well as the associated personal and environmental traits (Dawber, George, & Mann, 1957). The second FHS report, based on data from 6 years of observation (3 biennial examinations), looked at the role of nationality, education, smoking and drinking habits (Dawber et al., 1959). By the 3rd cohort, more tests like the 12-lead ECG, cardiovascular imaging procedures, and biomarkers were added to the assessment (Woodruff, 2012). The main influential factors discovered first were the

metabolic factors: elevated blood pressure, obesity, and cholesterol level (Tsao & Vasan, 2015). The second follow-up on the original group, implicated the additional factors of age and gender (Tsao & Vasan, 2015; Long & Fox, 2016). In the late 1950s, more data analysis determined that cigarette smoking had a major effect on the CVD mortality rate (Nieto, 1999b), and that it synergistically amplified other factors, as smokers tended to also have higher cholesterol levels (Tsao, 2015).

The FHS revolutionized the field of epidemiology and birthed the term "risk factors" (Tsao, 2015). The earliest findings of the FHS study indicated that CVD was multifactorial (Dawber, George, & Mann, 1957; Dawber et al., 1959). It also caused the shift in emphasis from secondary prevention and treatment to that of screening and of primary prevention efforts (Ockene et al., 2014). The FHS results served as the foundation for primary prevention projects and became an integral part of the CVD practice guidelines and treatment strategies (D'Agostino et al., 2013). The FHS originated the multivariate risk factor approach and its tool eventually became known as the Framingham Risk Score (FRS) risk formula (Mahmood et al., 2014; Tsao, 2015).

The FHS has served as a template for many longitudinal cohort studies and its protocol has become a widely accepted standard in the field of epidemiology (Kim, 2016; Tsao & Vasan, 2015a). The FHS project has grown steadily over the last five decades from the original 5,200+ to over 13,000 participants today in the core cohorts and several auxiliary projects. The expansion has included 'risk marker' measurements such as homocysteine, fibrinogen, lipoprotein lipase, and c-reactive protein and investigation of social and psychological factors, most notably stress (Tsao & Vasan, 2015a). But despite

all the novel projects thus far, the most effective CVD health promotion has been those that focus on identification and reduction of the traditional major risk factors.

The Creation of the Framingham Risk Score

Primary Prevention and CVD Risk Reduction

Over the years, evidence has confirmed that primary prevention focused on risk factor reduction lowered the frequency of coronary arteriosclerosis and decreased the CHD death rate (Dalen et al., 2014; Ray et al., 2014). Primary prevention, unlike secondary prevention, is administered prior to the onset of the disease with the goal of preventing of the disease from occurring. Secondary prevention, on the other hand, can involve drug administration, surgery to reverse an already existing condition, or any treatment administered after a confirmed diagnosis to minimize complications, reoccurrence, or death (Stone et al., 2013). The most common and most effective primary prevention effort focused on smoking cessation, but primary prevention can also include healthy eating and exercise programs (Grundy et al., 2002; Roger et al., 2012). The importance of primary prevention and early intervention is underscored by the report that over 1/3 of the victims of myocardial infarctions die within 24 hours or they develop debilitating chronic heart failure, angina or have a cardiac arrest (Grundy et al., 2002). Primary prevention involves screening and early detection of the antecedent risk factors and applying specific therapies that have been shown to significantly reduce the risk of developing heart disease and CVD (Jones & Greene, 2013). Primary prevention starts with searching for and detecting the presence of the certain traits in individuals at risk,

and for this CVD risk models are most effective tools (Tomasik, Krzysztoń, Dubas-Jakóbczyk, Kijowska, & Windak, 2015).

Disease Modeling and the Framingham Risk Score

The science of CVD risk modeling began in the 1950s with the introduction of the term 'risk factors' in the first FHS report and has progressed to the several dozen models in use today. The original risk estimation was based on a basic combination of risk factors into discriminant analysis tables, without any account for the severity or relative role of each risk factor. This rudimentary tool focused mainly on the four major traits: hypertension, high cholesterol and diabetes (D'Agostino et al., 2013; Mahmood et al., 2014). Risk assessment models subsequently progressed to the use of cross-classification combinations, but they were difficult to understand and even more challenging for the physicians to use (Mahmood et al., 2014). By 1976, the first Framingham Risk formula being used were based on seven factors (age, total cholesterol, weight, ECG abnormality, hemoglobin, cigarettes smoked, and systolic blood pressure) with crude relative risk values for blood pressure and cholesterol (Truett, 1976 in Mahmood, 2014). By 1998, in the second iteration of the FRS, a more sophisticated multivariable logistic regression formula had been devised to calculate an individual's risk score (Mahmood, 2014). The critical importance of these tools in dealing with CVD morbidity and mortality and the millions of dollars in healthcare cost potentially saved, resulted rapid distribution of the different versions of this formula. Many versions are still in use and frequently updated in the continually evolving the science of CVD risk estimation (Ray et al., 2014).

CVD is a multifaceted disease and as such requires a complex risk assessment formula that incorporates several risk factors. Using the FRS, an individual's calculated risk score (as a percentage) can be in either four categories: low, moderate or intermediate, moderately high, and high, depending on the established threshold values. The ATP III (Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) guidelines classify people into categories that depend on both their 10-risk score and the presence of certain risk factors. The four are: high risk (FRS over 20% and CHD or a CHD equivalent or 2 or more risk factors); moderately high risk (FRS between 10% and 20% and 2 or more risk factors); moderate risk (FRS below 10% and 2 or more risk factors), and low risk (0 or 1 risk factors) (Pedersen, 2002). Over the years, the FRS has proven to be an important tool in CVD research, especially the ones with disease incidence and risk factor similar to that of Framingham, MA, 60 years ago (Eichler et al., 2007; Fawwad et al., 2016; Feresu, Zhang, Puumala, Ullrich, & Anderson, 2008; Rodondi et al., 2012; Yeung, Yuan, Hui, & Feresu, 2016; Yosaputra, Kholinne, & Susanto Taufik, 2010).

Unfortunately, in instances where the disease pattern changes (e.g., increase in diabetes prevalence) or the risk frequencies are different (e.g., low-income, minority, etc.) the FRS may fail to identify those who need preventative therapy (P Brindle et al., 2005). Refining the FRS has taken many forms mainly the revaluation of current variables or the addition of novel risk factors and biological markers, many of which require expensive equipment. The changing (arbitrarily chosen) thresholds, , risk factor removal or inclusion, choosing new time periods for testing, and the cost versus benefit

ratio of any of those changes are important considerations (D'Agostino et al., 2013). Finally, simply using the disease incidence and prevalence data for a given population to recalibrate and improve the formula's performance is the most common ways of remodeling CVD risk formula to improve its accuracy and precision (D'Agostino, Grundy, Sullivan, & Wilson, 2001a).

The Most Common Outcome for the FRS: CHD

Cardiovascular disease can be divided into three main classes: congenital heart disease (i.e., genetic factors or birth defects); disease caused by microorganism (e.g., rheumatoid heart disease); and heart disease resulting from atherosclerotic or hypertensive disorders. Of the many forms of atherosclerotic heart disease, coronary heart disease (CHD) is the most common and is the leading cause of sudden cardiac arrest (Dalen et al., 2014). Coronary heart disease (CHD) is the term used to describe the narrowing of the arteries that supply blood to the heart and is leading cause of CVD morbidity and mortality, as 6.2 million are hospitalized and 400,000 die from CHD every year (Kones, 2011). The prevalence of CHD, also referred to as coronary artery disease (CAD), equals the sum of the prevalence of all other CVDs and accounts for 64% of all CVD related deaths (Dalen et al., 2014). It is because of numbers like these that point to the greater threat of CHD, that most CVD risk formulas focus of CHD prediction.

CHD is caused by the narrowing of the arteries when plaque accumulates on the vessel walls. Plaque, made up mostly of cholesterol, builds up on the walls of the arteries and reduces the supply of blood to the muscles of the heart (Kannel, Castelli, et al., 1971). As blood flow decreases, the muscle tissues are deprived of sufficient oxygen and

begin to die. The death of cardiac tissue is experienced as chest pain and shortness of breath, the symptoms of angina pectoris (Bennet, Di Angelantonio, et al., 2008). Ultimately, if the blood flow is completely occluded by fatty deposits or clotted blood, nutrients and oxygen supply ceases, the muscle tissue dies, and a heart attack occurs. Unfortunately, some 735,000 Americans experience a heart attack every year (Mozaffarian, Benjamin, Go, et al., 2015).

The FHS group of researchers provided the very important insight that CHD disease is always preceded by a condition of atherosclerosis (Dawber & Lansing, 1966;. The belief then became that CHD was caused by a buildup of plaque in the arteries of the circulatory system, a process usually starting in the second and third decade of life and progressively worsening with time (Ockene et al., 2014). Later research revealed, when this atherosclerosis is exacerbated, the vessels walls become inflamed and the plaque ruptures; the debris forms a clot that completely occludes blood flow, resulting in a heart attack or a stroke (Cockram et al., 2001; Hosseini, 2015; Ockene et al., 2014). This had been the main etiological concept of heart disease for many decades

However, due to the invention of high-tech imaging tools, a better understanding into the development of CHD was been gained. It is now known that the disease does not begin with atherosclerosis, but is accurately is preceded by damage to the endothelial walls of the coronary arteries (Grundy et al., 2002; Pedersen, 2002; Zhang et al., 2014). The plaque deposits at the injury site are attempts by the body to repair the damage to the endothelium and is it made up of cellular waste, calcium, fibrinous clotting material, and the lipid cholesterol at the core (Pedersen, 2002). When this plaque builds up on the walls

of the body's arteries, it is referred to as atherosclerosis and plaque accumulation specifically in the arteries of the heart, is diagnosed as coronary artery disease (CAD) or CHD (Hutcheson & Rocic, 2012). So, heart disease is more a disease of blood vessel damage than it is of atherosclerosis. But what causes the damage to the walls and what risk factors account for this threat?

Components of FRS: The CHD Risk Factors

The risk factors that are involved in the development of CHD fall into four main categories: genetic, physiological, behavioral, and environmental (Kones, 2011). The non-modifiable factors are genetic or physiological and include age, gender, and family history of CHD. The other factors are behavioral or environmental and therefore modifiable, and they include: smoking status, cholesterol levels, blood pressure, diabetes, and obesity (Hobbs & Hobbs, 2004). The major predictive risk factors such as high cholesterol, diabetes, and high blood pressure are linked to diet and level of physical activity and smoking directly to the individual's choices, making them all behavioral.

The more recently discovered, less established, factors include biological markers such as C-reactive protein (CRP), lipoprotein A (lpA), low density lipoprotein (LDL), fibrinogen, and triglyceride (TG) levels (D. Lloyd-Jones et al., 2010) (Bennet, Di Angelantonio, et al., 2008). Also included in this category are imaging factors like coronary artery calcification (CAC), carotid intima media thickness (CIMT), and coronary angiography- all measured by MRI or CT scans (D. Lloyd-Jones et al., 2010) (Ray et al., 2014). The sophisticated technology required for these markers makes testing for them expensive and inconvenient. Additionally, research indicates that these markers

offer very little improvement in the performance of the predictive models based on established risk factors (Tsao & Vasan, 2015a).

In the US, screening for primary prevention purposes uses the latest version of the multivariable FRS to estimate the risk of CHD. The FRS combines measures of the more classical risk factors (i.e., age, gender, total cholesterol, tobacco use, HDL-C, TC, systolic blood pressure, and treatment status or CHD history) into an equation that calculates the risk score. Of the modifiable risk factors, high cholesterol is considered the strongest predictive factor, (after smoking), followed by diabetes, and high blood pressure. These modifiable risk factors are the focus of primary prevention efforts that include smoking cessation, and the lowering of blood pressure and cholesterol levels and reducing measures of obesity. This research aims to show that blood glucose level stabilization may also be critical to CVD prevention.

Cigarette Smoking

Smoking cigarette or tobacco is a major CVD risk factor, which was overlooked in the first report, but later confirmed as having the strongest association to CVD in the second FHS report (Dawber et al., 1959). Clinical research showed that the effect of smoking habits are associated with reduced heart rate, decrease oxygen capacity of the lung, increase blood clotting and chronic damage to the endothelium of the blood vessels (Ockene et al., 2014). Nicotine, the accompanying toxins, and the free-radicals generated from smoking all cause direct damage to the tissues, including those of the cardiovascular system, resulting in reduced cardiac function and an increased propensity to the

formation blood clots (Ockene et al., 2014). These structural and functional damage to the blood vessels are what lead to atherosclerosis and eventually cardiovascular disease.

After much debate between the government, the policy makers, and the tobacco industry, the Surgeon General officially reported, in 1964, that smoking was dangerous for the smoker's health (Dalen et al., 2014). The report stated that smoking causes lung cancer and was linked to heart disease, but the impact of this information on subsequent CVD surveillance measures was not immediate. Slowly, over many decades, the prevalence of smoking dropped from 42% in 1965 down to 18% by 2012 (Dalen et al., 2014). This decline was eventually reflected in the decline in CVD prevalence, and by an even larger decline in the incidence of heart disease in the population (Tsao & Vasan, 2015a). Now cigarette smoking, though the strongest risk factor, is no longer the most prevalent as other risk factor have become more common.

Unlike hypertension and hypercholesterolemia, cigarette smoking was found to be independent of other risk factors (Lansing, 1961). Of the many reversible risk factors, smoking has the biggest impact as it confers a 6-fold increase in men, and a 3-fold increase in women in the risk for MI, compared to non-smokers (Amsterdam, 2011). As with other risk factors, the duration and the intensity of smoking impact the degree of the risk, as former smokers who are now non-smokers have worse risk profile than those who never smoked, and the more packs per day the more severe the risk (Fuster & Kelly, 2010). A habit of cigarette smoking is still associated with the risk of Myocardial Infarction and sudden death due to CVD and smokers are 2-4 times as likely to have a stroke (Roger et al., 2012) (Mendis, 2010).

Over the last four decades, CVD mortality has decreased by one third and the related symptoms have decreased by one half, due in large part to the reduced prevalence of smoking (Dalen et al., 2014). When a person stops smoking the damaging substances are eliminated and the many hazardous effects minimized and the risk is gradually reversed (Jones & Greene, 2013). Consequently, widespread smoking cessation leads to a considerable reduction in heart attacks and cardiac death rates. Reduced prevalence of smoking when accompanied by dietary changes was found to amplify the benefits of smoking cessation (Huxley, Woodward, Huxley, & Woodward, 2011).

Hypercholesterolemia

Although smoking and hypertension were among the first factors found to have a high correlation to CVD, cholesterol was the first predictive factor uncovered in analyzing the data (Dawber & Lansing, 1966). Cholesterol values describes the concentration of fat molecules or lipoprotein dissolved in the blood. In the early days, when the cross-classification tables were used for risk assessment, total serum cholesterol had three levels- low, moderate, and high, corresponding to <225 mg/dL, 225-274mg/dL, and > 275 mg/dL respectively (Truett, Cornfield, & Kannel, 1967). More recent risk assessments use slightly different values: normal cholesterol levels are <200 mg/dL, borderline is 200-239 mg/dL, and high is >240 mg/dL (Wilson et al., 1998). According to the latest ATP III report 39.7% of Americans have a total cholesterol >200 mg/dL with 11.9% >240 mg/dL (Emelia J. Benjamin et al., 2017). In modern risk calculations, the cholesterol level value, instead of the categories, are entered into the formula for risk score computation. Hypercholesterolemia is associated with an elevated CVD death risk

as the RR ranges from 2.83 t0 4.46 depending on the cholesterol level and is even higher if diabetes is also present (Bertoluci & Rocha, 2017; Dahlöf, 2010)

During the 2nd generation of the FHS in 1971, the FHS lab was technologically advanced enough to separate the blood lipids in to high density lipoproteins (HDL), low density lipoproteins (LDL) and lipoprotein A (LpA) (Tsao & Vasan, 2015a). This allowed for the determination that LDL blood lipids are the pre-disposing risk factor contributing to plaque buildup in the coronary arteries. The discovery that LDL (and not HDL) was the blood lipid most predictive of the development of CVD, resulted in the substitution of the ratio of "total cholesterol to HDL" (Tot Chol/HDL) measure instead of the "total cholesterol" in risk calculations (Tsao & Vasan, 2015a). When assessing LDL separately, the optimal level is <100 mg/dL, moderate is 100-129 mg/dL, borderline high is 130-159 mg/dL, high is 160-189 mg/dL, and very high is >190 mg/dL (Hosseini, 2015). High or very high levels of unhealthy LDL are found in 33% of Americans (Pedersen, 2002).

More recent research led to the understanding that not only was the HDL component of total cholesterol not implicated in the development of atherosclerosis, but it was in fact beneficial to CVD health. HDL has been shown to protect of the endothelial surface of the cell wall, to reduce LDL oxidation, and to help maintain the cells sensitivity to insulin of the cells (Kones, 2011; Ray et al., 2014) HDL lowers CVD risk and is recommended to be ideally > 60 mg/dL, while <40 mg/dL is considered low and unhealthy (Hosseini, 2015). Epidemiological research has indicated that some 18.7% of Americans have low levels of the healthy cholesterol HDL (E J Benjamin et al., 2017).

The difference in the roles of HDL and LDL may explain why total cholesterol as a risk factor only accounted for 50% or the CVD prevalence in previous data analysis. Some individuals had the 'high' cholesterol, but had no CVD and others had CVD yet were free of traditional hypocholesterolemia (Nieto, 1999b). High HDL concentration is now known to be associated with a lower incidence of CVD and is regarded as a *negative* risk factor (Pedersen, 2002). To account for this discovery, researchers decided to use the Total Cholesterol to high density lipoprotein (Tot Chol/HDL) ratio in the Framingham Risk Score. Eventually, the HDL/TC ratio was replaced by two distinct variables, total cholesterol (TC) and the high-density lipid cholesterol (HDL) (Wilson et al., 1998).

The drop in the incidence of smoking caused hypercholesterolemia to the become the most important CVD risk factor, but the prominence of cholesterol as a risk factor has been challenged by recent research. According to Grimes (2012) the administration of statin drugs did not cause the drop in the CVD death rate as the decline started before statins were introduced 1994, and clinical trials showed benefits in only 2% of the sample, with aspirin and streptokinase therapy benefiting only 5%. Similarly, in clinical trials, aspirin, streptokinase, ezetimibe, and other cholesterol/plaque reducing pharmaceuticals lowered cholesterol levels, but had very little (< 5%) or no clinical benefit (Grimes, 2012). Additionally, the association between lowering of high cholesterol and the reduction in CHD deaths was only experienced for young men, but not for the older men or women of any age group (Grimes, 2012).

Hypertension

There is a powerful association between hypertension and heart disease, and it was one of the earlier traits identified by the FHS. High blood pressure (HBP) is diagnosed as systolic blood pressure ≥140 mm Hg and/or diastolic blood pressure of ≥90 mm Hg (Goff et al., 2014). HBP is regarded as a leading independent risk factor for CVD and is believed to result from the efforts of the heart to move blood through blood vessels that have been narrowed by plaque buildup (Xu, 1991). As the condition of atherosclerosis progresses and the narrowing of the arteries worsens, the vessel diameters decrease, the heart is required to work harder and blood pressure goes up in response (Hollander, 1976). This continual over exertion of the cardiac muscles sets up the conditions for heart failure (stressed enlarged heart muscles), which can progress to heart attack or stroke if the blood supply is eventually obstructed. (Renna, 2013).

Hypertension or HBP has a prevalence of 34% in adults, affecting over 90 million in the US, and 972 million globally (E J Benjamin et al., 2017). Rates in subpopulations can vary- African Americans have HBP rates of almost 46%, while children have 11% HBP rates. Of those with a first heart attack, 69% were hypertensive and 77% of stroke patients and 74% of those with congestive heart failure had blood pressures over 140/90 mm Hg (Mozaffarian, Benjamin, Go, et al., 2015). While hypercholesterolemia may be regarded as the leading risk factor for CHD, hypertension is significantly correlated with the other subtypes of CVD and is the leading risk factor for MI, HF, AF, PAD, stroke and kidney failure (Ockene et al., 2014).

The finding that an elevated *systolic* blood pressure (Sys BP) was more predictive than elevated *diastolic* blood pressure (Dia BP) of a CHD event was a most important FHS discovery (Dawber & Lansing, 1966). As a result, systolic and diastolic BP was replaced by only 'Sys BP' as a risk factor in the FRS formula (Kannel, 1971). The FHS researchers determined that the association between Sys BP and the risk of developing CVD was continuous and graded no matter the level of BP or the age of the subject (Kannel, Vasan, & Levy, 2003). Elevated Sys BP (> 140 mg Hg) is directly related to both obesity and hypercholesteremia and the severity of the risk is dependent on the presence of other CVD risk factors (Tsao & Vasan, 2015a). One study found that for blood pressure measures that are above normal, every reduction in the BP by 10 mmHg reduced the risk of CHD or stroke mortality by 15% (Bertoluci & Rocha, 2017).

Type 2 diabetes (T2DM).

Type 2 diabetes mellitus (T2DM) is a condition of impaired carbohydrate metabolism. T2DM caused by an inability of cells to respond to insulin, results in elevated levels of glucose concentrations in the blood. Insulin is a pancreatic hormone that facilitates the entry of glucose into cells to be used for energy production. When the supply of glucose is excessive, the high glucose concentration triggers an overproduction of insulin from the pancreas, in an attempt to shuttle the surplus glucose from the blood into the cells (Faerch, Bergman, & Perreault, 2012; Roberts, Hevener, & Barnard, 2014). But this excess of glucose is harmful to the cells and they respond by removing their insulin receptors from their cell membrane to minimize glucose influx (Roberts et al., 2014; Zhang et al., 2014) The removal of the insulin receptors renders the cells less able

to respond to insulin and they are become effectively ‘insulin resistant’ (Roberts et al., 2014). At the same time the pancreas, which has been pumping out extra insulin to deal with the rising glucose concentrations, becomes overworked and eventually shuts down (Sah, Singh, Choudhary, & Kumar, 2016). With less insulin being produced and cells in a state ‘insulin resistance’, dissolved glucose accumulates in the blood and abnormally high blood glucose levels or hyperglycemia persists.

Epidemiological and clinical studies now indicate that one of the consequences of prolonged hyperglycemia is CVD. High concentrations of glucose are damaging to the cells and tissues of the body (Faerch et al., 2012; Laakso, 2015; Tostes & et al., 2009) Excessive glucose causes glucose molecules to attach to the protein structures of cells, the severity and duration of which can be measured by the hemoglobin A1 C (HbA1C) test (Huang et al., 2016). High concentrations of glucose also cause the cells of the epithelium of the blood vessels to lose their integrity and become damaged (Laakso, 2015). It is this damage to coronary blood vessels that initiates cardiovascular disease.

Advances in imaging technology have revealed that CVD is preceded by damage to the arteries that carry blood to the organs and tissues of the body. One of the ways the body tries to deal with this damage is by generating a blood clotting mechanism with patches of fatty deposits at the site of injury (D. Lloyd-Jones et al., 2010). Unfortunately, this ‘patch work’ becomes excessive with ongoing damage (persistent insulin resistance and resulting hyperglycemia) and it leads to a build-up of plaque on the walls of the arteries (Clark et al., 2015; Kishore, Kim, & Crandall, 2012) The plaque buildup may reduce blood flow, comprise the supply of oxygen and nutrients to the tissue, occlude

blood flow at the site, or break off and travel to other smaller spaces and occlude blood flow there (Hutcheson & Rocic, 2012; Long & Fox, 2016). If this happens in the heart, angina and heart attacks results; if the clot travels to the brain and obstructs blood there, the person suffers an ischemic stroke. Hyperglycemia, and the accompanying hyperinsulinemia, is believed underly the exacerbation of the many pathologies and morbidities of CVD (Hobbs & Hobbs, 2004).

CHD Risk Assessment Tool

CHD Risk Tool Performance

Accurate prediction of threat and effective disease prevention and treatment strategies depends on well-designed risk estimation models. The widespread use of tools like the FRS and others is a testament to this. A risk formula combines the major risk factors into an equation that returns an estimation of the risk as the probability of a disease event happening. The CVD risk formulas typically calculate the probability of developing CVD events within a given period, usually "10-years" but sometimes "20-years", "30-year" or even "life-time". Ideally, the use of baseline relative risk and the levels of exposure of the risk factors used in these tools should be specific to the population under investigation, otherwise additional calibration adjustment is required (Eichler et al., 2007).

A more trustworthy stratification can be achieved by changing the disease range and using alternative thresholds for the categorization of different levels of predicted risk. Incorporating additional variables or replacing old ones with newer, more strongly correlated ones, can improve accuracy and stratification power. To ensure the continued

applicability and utility of prediction tools, the ability to adequately discriminate (to differentiate between those with CVD outcome from those without) and to sufficiently calibrate (agreement between the prediction and the observation) are essential. In most cases, regression statistics are used to generate the model and the corresponding AUC or c-statistic are used evaluated the model's performance (Elosua, 2014).

The FRS Tool and its Performance Limitations

The Framingham Risk Score tool has endured the test of time and is credited with popularizing the concept of risk prediction (Bitton & Gaziano, 2010). Because of these tools like the FRS physicians no longer must rely on their experience and judgment in assessing the CVD risk of their patients. Instead they can use the risk prediction function to more adequately screen and sort patients, more responsibly allocate resources and more effectively manage the treatment modalities (Günaydın et al., 2016). For a great many people, the FRS is still the most effective way for predicting the likelihood of developing CHD as well as screen for the presence and severity of the disease (D'Agostino et al., 2013).

However, as beneficial as the FRS is, it has its some limitations and these limitations serve as the main driving force behind of risk score optimization efforts. Earlier research on the risk factors led to changes such as the removal of HDL from cholesterol measures and the use of only diastolic blood pressure values. Other changes such as generating gender specific versions of the formula or the labelling of diabetes as a CHD equivalent were all done to improve the FRS prediction performance. But as the use

of the tool became more widespread there were reports of more fundamental issues with the tool's performance.

The main limitations of the FRS relate to the less than ideal discrimination and classification power. A comprehensive study conducted by the FHS team evaluated the performance of the FRS amongst six ethnically different groups across the US. Evaluating the classification and calibration ability of the FRS to predict the relative risk associated with CHD revealed that the tool had an AUC score between 60-70% (D'Agostino et al., 2001a). By their own standard any risk assessment tool with an AUC score below 70% is considered suboptimal and should be amended (D'Agostino et al., 2013). And although, the FRS has been updated since that 2001 report, the issues with under or overestimation and inadequate classification persist (Çevik, Özcan, & Satman, 2015; Günaydın et al., 2016; Van Staa et al., 2014).

The white middle-class cohort from which the original formula was designed limited its generalizability to more ethnically diverse populations. Studies conducted by the NHLBI have determined that the FRS, even after recalibration, tended to underestimate the risk in populations that have very low incidences of CVD, such as Japanese, East-Indians and Native-Americans (D'Agostino et al., 2008; N. Garg et al., 2017; Kones, 2011; Reddy & Satija, 2010). Additional studies reported on the misclassification and overestimation in high-risk groups or those in the lower socioeconomic class in the US and among European populations in general (Eichler et al., 2007; Günaydın et al., 2016; Pike et al., 2016). These findings were concerning and have reduced the confidence of US and European physicians in using risk equations.

Other Tools Developed in Response to the FRS Limitations

The next section describes the three most common (after the FRS) CVD risk estimation models being used in the field of CVD screening and prediction. These tools were developed from the FRS and came about in response to its limitations. Despite the recalibration and other adjustments made to the FRS, the inaccuracy was problematic, and researchers wanted risk formulas that was more representative of how the disease manifested for their patients. The desire for tools that specifically represented the CHD risk in their respective countries and communities resulted in the development of the QRISK, the SCORE, and the ASSIGN.

QRISK: the QResearch database CVD risk score. QRISK was in response to the challenge of using the FRS in Europe, namely European ethnic diversity and differences in CVD prevalence. The UK-NICE (National Institute for Health and Clinical Excellence) developed a revised version of the FRS, called FRS-NICE. The FRS-NICE added the risk factors BMI, family history of CVD, treatment for hypertension, and social deprivation to the FRS and raised the threshold for the high-risk category (Collins & Altman, 2012). Applying the FRS-NICE reduced by half the number of patients being recommended for primary prevention and thereby reduced the expense and many-side-effects associated with statin, the cholesterol-lowering drug, usage (Collins & Altman, 2012).

The SCORE (Systematic Coronary Risk Evaluation). SCORE, the second most common European risk assessment tool, began in 1994 as a collaboration between the European Society of Hypertension and the European Society of Cardiology using data

from 12 European cohorts and their 205,178 patients (Tomasik et al., 2015). The six risk factors used were: total cholesterol, TC:HDL ratio, age, gender, smoking status, and systolic BP with a 10-year CVD outcome prediction. The larger cohort, the easy-to-read charts, the inclusion of endpoints other than mortality, and the use of the respective CVD rates for each region did improve the risk estimation and prediction. This simplified prevention method assignment and patients were more appropriately categorized for application of education, lifestyle changes, or medication (Conroy et al., 2003).

ASSIGN (Assessing cardiovascular risk, using SIGN). ASSIGN is the CVD screening tool developed by the Scottish Heart Extended Cohort (SHHEC) from survey data, collected from 1944 to 1995 from clinic patients between 25 and 74 without a history of CVD (Woodward, Brindle, & Tunstall-Pedoe, 2007). Like the QRISK, the outcome was based on mortality and morbidity measures as well as both CHD and CAD as this was the most common CVD and the leading cause of death among the Scotts. The ASSIGN successfully improved upon the sensitivity and specificity of the FRS and alleviated the disadvantages suffered by high-risk minorities who were previously insufficiently allocated for preventive care (Tunstall-Pedoe, 2011).

Comparing the Common CVD Risk Formulas. Choosing between models is an important task for many personnel and organizations involved in CVD screening. There are several studies comparing the performance of the more common models of risk estimation. One such study compared the performance of the FRS (US based), the ASSIGN (Scottish), and the QRISK2 (England and Wales) on the same data set. There was congruency among all three models, except for the high-risk classification, although

challenged by the different thresholds (Van Staa et al., 2014). The social-deprivation term in ASSIGN did not work for non-UK populations but the region-specific data and CVD history information made the QRISK most accurate overall (Van Staa et al., 2014).

Recently Gunyadin (2016) compared the performance of the SCORE relative to FRS to determine which was better at predicting CVD. They used the SYNTAX test, an angiograph measurement of the number of lesions inside the coronary arteries, for a non-formula estimation of CVD risk. According to the results, the SCORE provided better discrimination among men and diabetic patients, but not among the obese or those with a family history of CVD. Although, both models predicted a similar level of CHD severity risk (based on the AUC evaluation), the only factors found to be statistically significant in the SCORE model were 'blood pressure' and 'total cholesterol' (Günaydın et al., 2016).

Performance Evaluation of CHD Risk Tools

It is important that prediction models that have been developed, redesigned, or extended continue to be adequately reevaluated as these tools are used for making important decisions, both diagnostically and prognostically. In the field of medicine, where decisions are usually binary in nature and require an estimation of absolute risk, logistic regression models have proven effective (Sperandei, 2014). Determining whether these formulas' predictions can be extended to the other sample population is vitally important. The calibration ability or the difference between the predicted outcome and the actual outcome reflects most how well the model describes the data. Making changes to the formula to improve on its calibration power is central to formula optimization.

For CVD risk prediction, the models are formulas that combine the major risk factors generates prediction probability as an attributable risk value. There are many methods for evaluating the performance of these prediction formula and by convention, quantifying the performance with known-outcome data sets works best. For logistic regression formulas, where the outcome is dichotomous, the model's performance is evaluated based on four main metrics: (i) the coefficient of the independent variable (which reflects the strength of the association), (ii) the confidence interval of the independent variable (which gives an indication of the importance of the variable to the prediction relationship), (iii) the accuracy of probability of the dependent variable, and (iv) the goodness of fit of the formula in both its discriminating and its calibrating abilities (P Brindle et al., 2005; D'Agostino et al., 2001a; Giancristofaro & Salmaso, 2003).

For models whose performance evaluation determined that they need to be optimized, there are several ways of doing so. Models can be optimized by adding new markers that extend to the formula but adding this new variable must provide a higher degree of accuracy (Root et al., 2014). Models can also be optimized by altering the variables already in use so that there is an increase in the additional risk information provided (Rodondi et al., 2012). Ideally this altering of the variable should have a generate an increase in the predictive power. In either case, the new formula is evaluated against the performance of the old one based on the same principles used in the original model development.

Improving the performance of multivariable risk assessments is vital to patients' classification and treatment guidelines and users should be confident of the tools applicability and generalizability. This study will compare the predictive performance of the FRS model that uses 'diabetic status' as a marker against the one that uses 'glucose levels'. First the relative risk and the confidence intervals of each marker will be computed and compared, then the attributable risk as a probability for each the two formulas. The predicted probability will be compared with the observed probability of the data compared as discriminating ability in the AUC to determine which is the better fit of the data. Finally, the co-variates will be added to the analysis and the effect on the AUC of the 'glucose level' formula will be assessed.

Glucose and the Crisis of Metabolic Disorders

Spectrum of Metabolic Disorders

Diabetes describes persistent elevated blood glucose levels produced by impaired carbohydrate metabolism. Diabetes is, however, one of many states that are part of a continuum of metabolic abnormalities and body weight issues; beginning with mild insulin resistance, moving to being overweight, then to full-blown insulin resistance and obesity, followed by pre-diabetes, and finally ending with diabetes (Long & Fox, 2016). Collectively these disorders, referred to as 'diabesity' by Dr. Hyman, are known as metabolic disorders or impaired glucose tolerance problems, most of which go undiagnosed and therefore ignored (Hyman, 2012). The rapid rise of metabolic disorders has been projected to affect some 1.7 billion people worldwide by the year 2020 (Matfin,

2010). More recent research indicates that in 2012 one in every two Americans (52%) was already either pre-diabetic or diabetic (Menke, Casagrande, Geiss, & Cowie, 2015).

Cardiovascular diseases are part of a larger group of *cardiometabolic* diseases that share clinical markers, that can co-occur, and are interlinked in their development (Assmann, Schulte, & Seedorf, 2008; Bertoluci & Rocha, 2017; Long & Fox, 2016; Srikanthan, Feyh, Visweshwar, Shapiro, & Sodhi, 2016; Wilson & Meigs, 2008). The high probability of co-morbidity necessitates an integrated approach to risk assessment that combines 'cardio' factors and 'metabolic' factors of diseases such as CVD. One cross-sectional study of 5,190 Kenyans looked at the most common cardio-metabolic markers and concluded that central-obesity (i.e., abdominal obesity), hypertension, hypercholesterolemia, hypertriglyceridemia, and hyperglycemia were all inter-linked (Haregu et al., 2016). Assmann et al., (2008) argued that including metabolic markers significantly improved the accuracy of CVD risk assessment tools. A decade ago, the FHS researchers concluded that metabolic risk factors, not diabetes only, should be part of the screening and prevention strategies for CVD (Wilson & Meigs, 2008).

Overweight, Obesity, and BMI

Metabolic disorders usually begin with some form of insulin resistance and if unchecked progresses through to several stages of metabolic disorders eventually end in diabetes. The hormone insulin is responsible for the movement of glucose into the cells as described above. It is also involved in the production of fat- both in the liver and the fat cells around the body (Sah et al., 2016). Insulin production by the pancreas and increased insulin resistance of the body's cells promotes the conversion of the excess

glucose in the blood into fat cells where it can be stored (Roberts et al., 2014; Srikanthan et al., 2016). The accumulation of fat tissue around the vital organs begins the overweight status that eventually leads(Roberts et al., 2014) to central or abdominal obesity. Overweight is usually the first recognizable sign of hyperglycemia, insulin resistance and metabolic dysfunction (Hyman, 2012; Roberts et al., 2014; Srikanthan et al., 2016).

Overweight status is determined by the height to weight ratio or body mass index (BMI) and is calculated by dividing the subject's weight by the square of their height (Kg/m^2) (Çevik et al., 2015). A BMI over 25 but less than 29.9 is considered overweight and affects approximately 35% of the American population (Go et al., 2014). Sometimes the BMI measures are misleading, and waist to hip ratio is used to be a more accurate estimation of an individual's obesity level (Reddy & Satija, 2010). However calculated though, more than half (53%) of Americans have visceral or abdominal obesity, and 3 billion around the globe are dealing with excess body fat (Kones, 2011).

If overweight condition is not adequately managed the next state, in the metabolic disorder continuum- obesity, happens. Obesity, measured by a BMI of 30 or above, is considered the leading cause of preventable death and afflicts 35% of the Americans (Nichols, 2012). Obesity is a serious medical condition that causes hypertension and atherosclerosis making it a major CVD risk factor, but it also predisposes individuals to respiratory problems, sleep disorders, and even cancer (Long & Fox, 2016). Increased BMI is correlated with hypertension, T2DM, congestive heart failure, and the onset of atrial defibrillation, and obesity was found to be double the risk of CVD in the original FHS research (Wilson & Meigs, 2008). Obesity also increases the risk of Angina Pectoris

and sudden death by cardiac arrest (Long & Fox, 2016). Obesity serves an independent risk marker for glucose intolerance and a corresponding increased risk of CVD (Long & Fox, 2016). The rate of obesity and overweight status has been increasing - between 1950 and 2000 the proportion of Americans who are overweight rose from 15 to 30%, while obesity rose from 3.9 to 14% (Long & Fox, 2016). Today, as many as 70% of American adults are either overweight or obese (Kones, 2011) and over 80% of overweight patients have at least one of four major CVD risk factors- hypertension, elevated triglycerides, low HDL, or impaired glucose tolerance (Nichols et al., 2012).

Pre-Diabetes

The order of progression through the metabolic continuum is not pre-set, obesity can be preceded by or followed by conditions of pre-diabetes. Pre-diabetes, or intermediate hyperglycemia refers to the condition where the blood glucose is higher than normal but not high enough to be diagnosed as diabetic (Huang et al., 2016). Pre-diabetes is diagnosed with a fasting blood glucose level from 100 to 125 mg/dL (5.6 to 6.4 mmol/L) and is accompanied by an established insulin resistance (Farooq Al-Azzawi, 2015; Tabák, Herder, Rathmann, Brunner, & Kivimäki, 2012). Prediabetes is strongly associated with other CVD risk factors such as obesity (especially central or visceral obesity), high levels of LDL cholesterol, high triglycerides, and high blood pressure (Tabák et al., 2012). Pre-diabetes was found to confer a 12.7 to 22.3 fold higher risk of developing diabetes, depending on the duration, which is strongly associated with an increased incidence of CVD (Tsao & Vasan, 2015; Long & Fox, 2016)(Fox, 2010). Diabetes has been estimated in 13% of the population, but an additional 38% are

determined to be pre-diabetic (Menke et al., 2015). Currently, there are over 84 million people living with pre-diabetes and over 90% of them are undiagnosed (Center for Disease Control & Prevention, 2017).

Metabolic Syndrome

Another condition that is clustered with obesity and pre-diabetes is that of metabolic syndrome (MetS). MetS is defined as a combination of three or more of the seven risk factors that include prediabetes, diabetes, central obesity, high blood pressure, high triglycerides, high levels of LDL, and low levels of HDL (Grundy et al., 2008). A co-existence of these inter-related parameters produces a synergistic effect that significantly increases the likelihood of CVD (Farooq Al-Azzawi, 2015). MetS was found to confer a 7-fold higher risk of T2DM and to double the risk for CVD mortality, all-cause mortality, myocardial infarction and stroke (Hutcheson & Rocic, 2012). Despite the dominance of factors such as HDL and BP, when MetS is included in the prediction formula, the probability of CHD increases from 37.4% to 54.7% (Bertoluci & Rocha, 2017). MetS, found in 35% of the adult US population or 50% for those over 60 years, has a rapidly rising prevalence, driven mainly by the increase in insulin resistance and obesity (Aguilar, 2015).

Type II Diabetes

At the end of the metabolic disorder spectrum is T2DM, a state of persistent hyperglycemia. Unlike Type I Diabetes, in which hyperglycemia is caused by an autoimmune-based insulin insufficiency, T2DM is caused by a acquired insulin resistance (American Diabetes Association, 2009; Mayfield, 1988). In the latest classification

T2DM is diagnosed by a fasting blood glucose (FBG) level above 125 mg/dL or 6.9 mmol/L (Center for Disease Control & Prevention, 2017). In 2009, there were over 110 million Americans, and another 130 million worldwide, living with T2DM and it was predicted to rise to 180 million by the year 2030 (Fuster & Kelly, 2010). The incidence of T2DM increased so rapidly, that the projection was changed to 239 million people affected worldwide by 2030 (Shaw, Sicree, & Zimmet, 2010). The latest report from the International Diabetes Federation (IDF), states that there are 425 million diabetics on the planet (half of which are undiagnosed) and it is predicted to increase to 552 million by 2030 and 629 million by 2045 (International Diabetes Federation, 2017).

T2DM, though very common, is a serious illness and is a major cause of chronic disease, disability, and death. T2DM is associated with heart disease and stroke, and is the leading cause of kidney failure, lower limb amputation and new cases of blindness (Dauriz et al., 2015; Sarwar et al., 2010) Despite the decline in CVD related deaths, diabetics included, there are still a large number who survive and continue suffer the many complications of the combination of both diseases. T2DM can double, triple or quadruple (depending on age and gender) the risk of CVD mortality as well as increase the risk of chronic heart disease and ischemic stroke (Bertoluci & Rocha, 2017). T2DM and pre-diabetes are associated with many issues arising from damage of the microvascular vessels which including, retinopathy, neuropathy, and nephropathy, but CVD remains is the primary cause of death in diabetics (Tabák et al., 2012).(Bertoluci & Rocha, 2017).

Diabetic Status in the FRS 10-year Risk Assessment

Early scrutiny of the FHS data identified diabetes is one of the strongest CVD risk factors, but it would take several rounds of revisions before it would be included as a risk factor in the FRS assessment formula. In the early 1930s and 1940s, before the FHS, evidence was already accumulating to implicate T2DM as a risk factor for CVD, but it was the FHS analysis that provided empirical confirmation of this relationship (Kannel & McGee, 1979a). Several round of analysis confirmed diabetes an independent risk factor for CVD, one that conferred a higher risk of CVD mortality and morbidity, with a substantial impact on all-cause mortality (Avitabile, Banka, & Fonseca, 2012). Later researcher would find that three out of every four diabetics died from heart disease and therefore diabetic status should be an important major risk factor in any FRS risk assessment function (Qazi & Malik, 2013a). But the strength of other risk factor such as cholesterol and hypertension overshadowed diabetes and it would be included 50 years after the FHS study began (Bennet, Angelantonio, et al., 2008; Kannel, Gordon, & Schwartz, 1971; National Heart & Boston University, 2017; Tsao, 2015).

Assessing the risk of developing CVD in asymptomatic groups began approximately two decades after the data collection for the first FHS cohort started. Careful investigation of the data revealed several associated risk factors but only the ones with a relatively strong correlation to CVD were used for CVD risk assessment (Dawber et al., 1957; Lansing, 1961). Initially, the risk factors were used in cross-classification tables to generate a risk level, but as the number of risk factors grew, the tables were replaced with risk equations that generated a score (Truett et al., 1967)(Keaven M

Anderson et al., 1991). These equation functions were first published in 1967 and produced an estimate the risk possessed by each individual subject (Friedman, Hyg, Kannel, Dawber, & McNamara, 1967). The risk factors used then were age, gender, total cholesterol, left ventricular hypertrophy (LVH), smoking status, and systolic blood pressure (SBP). Clinicians used the functions to evaluate each patient and assign those with moderate and high risk to relative preventative treatment.

During the 1970s and the 80s several updates to the FRS model were published and the risk equation functions were eventually replaced by a risk factor combination to generate a risk profile. This novel risk profile differed in several ways: it distinguished between general CVD and its most common type, CHD; there were now 5-year risk and 10-year risk versions; and it included only variables that had a strong association with a CHD event (Gordon & Kannel, 1982; Kannel & McGee, 1979; Kannel, Castelli, Gordon, & Mcnamara, 1971). The risk factors were now relegated to those that were independently correlated to CHD and whose values were easy to obtain during an office visit. By the end of the 1980s, variables like 'BMI' values was added but later replaced with 'glucose tolerance test' (GTT) values; 'family history' had been added but removed because of the difficulty to accurately determine in most cases; and 'LVH' requiring expensive, inconvenient test was also removed (D'Agostino et al., 2001a; Qazi & Malik, 2013a). The new stream-lined formula now included the six variables: age; gender; total cholesterol; smoking status; and GTT.

After several rounds of revisions, the end of the 20th century birthed the latest coronary disease prediction logistic regression algorithm. The modern version of the

Framing Risk Score would mark the first appearance of 'diabetes' (as a dichotomous 'yes/no') as a variable, having replaced 'GTT' levels (Wilson, 1998). The new FRS assessed the 10-year risk and would go on to have a major impact on the national guidelines for treatment for hypercholesterolemia and hypertension. Ironically, this 10-yr FRS version was noted for substituting continuous measures of blood pressure and cholesterol levels with their risk categories based on the fifth Joint National Committee on Detection, Evaluation, and Treatment of High Blood Pressure (JNC-V) (Cleeman, 2001).

Sometime later, this official 10-year version of the FRS, using the hard CHD (coronary deaths, MI and stroke) as the endpoint, was adopted by the third Adult Treatment Panel (ATP III) of the National Cholesterol Education Program (NCEP) (Grundy et al., 2002). The classification using the 10-year risk estimates of the FRS proved to be more convenient and more accurate, and soon became the basis for the cholesterol, blood pressure and dietary rules used by physicians throughout the nation (D'Agostino, 2013; Mahmood, 2013). To this basic formula, the American College of Cardiology/American Heart Association (ACC/AHA) added cRP and a history of CVD to generate the Reynolds risk model, and 'race' for the Atherosclerotic Cardiovascular Disease (ASCVD) calculator for working with those who have a history of atherosclerosis or high LDL (Stone et al., 2013). Despite the existence of other formulas, it is the 10-year FRS that is regarded as *the* official formula in the US and the standard used by clinicians and researchers for all their CHD risk assessment needs (Ford, Giles, & Mokdad, 2004).

A diagnosis of diabetes became a most important variable in the official 10-year FRS. This formula, developed from the first cohort data set, was intended for a relatively shorter (than life-time risk) horizon and to have the greatest applicability to the age (30-74 years) group for whom assessment and prevention therapy would be most beneficial. Diabetes testing was already a routine medical assessment, and knowing that positive diagnosis imparted a high CHD risk made it critical to the accuracy of the CHD risk estimation (Cockram et al., 2001). The 'diabetic status' version of the 10-year-FRS became a valuable tool to clinicians for classifying new patients and for guiding subsequent treatment plans (Coch & Green, 2016; Damkondwar, Rajiv, Suganeswari, Kulothungan, & Sharma, 2011; Qazi & Malik, 2013b; Wilson & Meigs, 2008). As one of the major modifiable risk factors, 'diabetic status' helped in correctly assigning patients for either moderate intervention such as education about diet and activity levels, lifestyle change; or for stronger intervention like glucose lowering or statin drug prescription (Alatawi, Kavookjian, Ekong, & Alrayees, 2016; Brindle et al., 2005).

It should be noted that the criteria used for a diagnosis of diabetes during the initial FHS data collection and analysis was different from that used today. For the first cohort, diabetes was diagnosed in one of three ways: (i) by history of hypoglycemic (insulin) use; (ii) a casual blood glucose (CBG) level above 150 mg/dL recorded on two visits; or (iii) a fasting blood glucose (FBG) level **over 140** mg/dL (Wilson et al., 1998). By the second cohort, GTT were now part of the examination, and abnormal GTT levels was now the fourth way of diagnosing diabetes. But eventually only the fasting blood glucose test (FBG) was used and a FBG level **over 125** mg/dL indicated the presence of

diabetes (Shaw et al., 2000; Wareham, 1998). Even though FBG levels were now more routinely obtained, the use of the dichotomous 'diabetic status' in the standard 10-year FRS, based on an arbitrary threshold, ensured the loss of valuable risk information (Bertoluci & Rocha, 2017). Additionally, the risk facing those dealing with 'prediabetes' would go unrecognized and the risk assessment compromised. Also**,** further classification among those diagnosed with diabetes is important for choosing the appropriate secondary treatment and for monitoring the progression (or regression) of the disease in response to any of the treatment administered (Kishore et al., 2012).

Like the other major risk factors, cholesterol and blood pressure, diabetic status had a different relative risk for CVD across the genders. FHS data analysis reported that T2DM was a significant trait of CVD susceptibility for men and the strongest risk factor for women (Qazi & Malik, 2013b). Additionally, the FHS determined that T2DM was associated with a 2 or 4 (male or female respectively) times greater risk of developing CVD and it was correlated with an increased risk of dying from CVD-related complications (Fox, 2010). We now know that diabetics die mainly from CVDs such as MI, CHF, PAD, and stroke, with the increased risk being greater in women than in men (Cockram et al., 2001; Huang et al., 2016; Long & Fox, 2016; Tsao & Vasan, 2015a). These gender differences were the impetus for separating the 10-year FRS profile into distinct gender formulas that improved the sensitivity of the assessment.

By the turn of the new century, T2DM was not only recognized as a major risk factor, but it was believed to be conferring the worst prognosis to CVD patients, with both CVD and T2DM having very high impact on the risk of death (Booth et al., 2006;

Cockram et al., 2001; Raggi, Shaw, Berman, & Callister, 2004). This information led the FHS researchers to declare T2DM a CVD risk equivalent and to remove diabetes as a risk factor from it's the official 10-year FRS (Bonow, 2002; NHLBI & NIH, 2002). Instead, all diabetic patients were now labelled as having already suffered a CHD event and automatically placed in the high-risk category. Subsequently, there were the new formulas for 'global-CHD' and 'general-CVD' risk assessment and they include 'diabetic status' as a variable (D'Agostino et al., 2008). However, as with previous versions of the use of 'diabetic status' as a binary limited the accuracy of the FRS.

Alternative models of the FRS, calculating 20-year, 30-years for long-term and 2-year and 5-year risk for short-term, were subsequently designed, but the 10-year model remained the most commonly used. The 2-year FRS formula is used for patients who had already experienced a coronary event (D'Agostino et al., 2000). The 5-year FRS formula, being more finely tuned with less generalizability, is usually reserved for the elderly (over 75 years) or infirmed as they require a more sensitive test (Laurier, Chau, Cazelles, & Segond, 1994; Rodondi et al., 2012; Westendorp et al., 2009). The risk models with longer prediction times (20-year, 30-year, and life-time risk) are used for younger patients and for those from demographics with a lower incidence of CHD, as the 10-yr tool can underestimate the risk for these groups (Levy, Walmsley, & Levenstein, 1996)(Mccormack, Levine, & Rangno, 1997). The variations of 30-year risk and life-time risk usually add the variables 'BMI' or 'use of hypertensive medication' to the basic 10-year FRS (Pencina et al., 2009). But the 10-year risk model is most adequate for general

screening for CHD and is therefore most common (PM Brindle et al., 2005; Cleeman, 2001; Ford et al., 2004; Jahangiry, Farhangi, & Rezaei, 2017; Westendorp et al., 2009).

Diabetes Over the Years

At that time there was little understanding of how diabetes was caused, and even less attention was placed on how it developed. In addition to the difference in T2DM diagnosis protocol and the new FRS inclusion criteria, there have been more volatile changes in diabetes prevalence over the years.

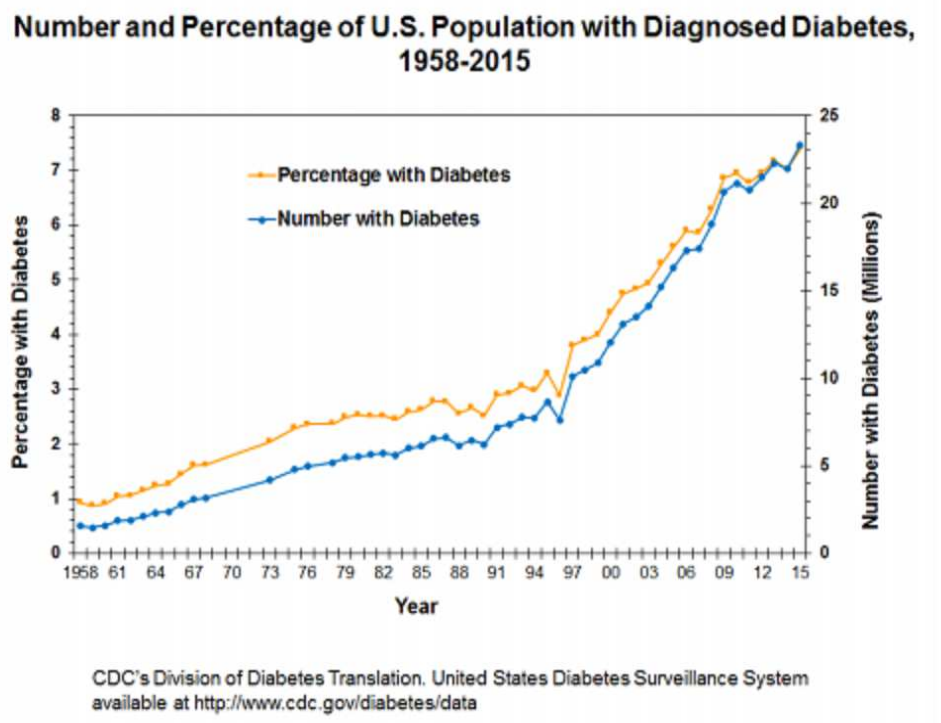


Figure 3. Diabetes over the years. Reprinted from Long-term Trends in Diabetes: CDC's Division of Diabetes Translation, 2017.*United States Diabetes Surveillance System* https://www.cdc.gov/diabetes/statistics/slides/long_term_trends.pdf

The disease was significantly less common 60 years ago than it is now, assuming that the prevalence of diabetes at the early stages of the FHS in 1948 was very lower than the 0.93% in 1958 (Centers for Disease Control and Prevention, 2017). By the publication of the first FRS, diabetes was only at 3.39% (1998) and at 6.29% by the second (2008), less than half the 13.5% prevalence it is today (Centers for Disease Control and Prevention, 2017). The higher prevalence, the different diagnostic protocols, and the rapidly increasing incidence rate are reasons for the complicated role of T2DM in the FRS.

During the last quarter of the 20th century the incidence of diabetes doubled among individuals between the ages of 40 to 55 and this was mostly those who were obese (Long & Fox, 2016). However, even though the incidence rate of CVD among patients with diabetes has fallen by as much as 50% over that same period, the incidence of CVD among diabetics is still twice what it is among non-diabetics (Long & Fox, 2016). These numbers represent an increase in the attributable risk (AR), the difference in incidence of CVD for those dealing to diabetes and those who are not, while the relative risk (RR) which is a measure of the strength of the association remained the same. This increase in AR is an indication of complication and makes early detection and stratification even more critical (Fox, 2010). Early detection of T2DM, blood glucose control, and early aggressive action are all critical to managing subclinical CVD, and it is for this reason for comprehensively assessing the full range (i.e., hypoglycemia, pre-diabetes/metabolic syndrome and diabetes) of dysglycemia (Coutinho, Gerstein, Wang, & Yusuf, 1999; Fawwad et al., 2016; Perreault et al., 2014).

Need for Stratifying a Diagnosis of Diabetes in Risk Assessment

Stratification is necessary for clinicians to identify patients who are at highest risk and will therefore be more receptive to, and benefit most from, aggressive treatment. Stratification enables the clinicians to more accurately sort patients into additional risk categories such as: no-risk, low-risk, intermediate-risk, moderately high-risk and high-risk. It is also essential for the accuracy of the CVD risk prediction effort, especially since atherosclerosis and other CVD risks can remain asymptomatic for years (D'Agostino et al., 2013). This facilitates optimal use of time and resources such as intensive aggressive treatment including further testing, drug therapy, or surgery for those at high risk with moderate monitoring or treatments with a lifestyle emphasis for intermediate risk patients and no treatment for those in the low risk group.

Researchers in the CVD risk prediction field have determined that the risk among diabetics is far from homogeneous. In fact, many organizations involved in the diagnosis and treatment of heart disease and/or T2DM such as the ACC, the AHA, the ADA, and the ESC have rejected the notation that T2DM and CVD are risk equivalent (Bertoluci & Rocha, 2017)(Bertoluci & Rocha, 2017). In one study found that there were some diabetics who had CAC (coronary artery calcium) scores of zero, essentially the same risk level as that of a non-diabetic, especially among younger individuals (Raggi et al., 2004). Additionally, the AHA has officially recommended further stratification for diabetics based on age, particularly for those between 40 and 75 years, and on level of LDL in the blood (Ray et al., 2014). The ESC also recommends stratification for diabetics based on duration of diabetic condition, especially for those who have had the

disease more than 10 years, and on the presence or absence of renal dysfunction and the presence of other risk factors (Piepoli et al., 2016). Similarly, the ADA also recommends stratification for diabetics based on age, previous CVD status, and the presence of other risk factors; facilitating varying therapy recommendations such as intense statin therapy, moderate statin therapy, or simply lifestyle modification for some of those at risk (Coch & Green, 2016).

Age, as the strongest non-modifiable risk factor, is important to CVD risk stratification because of the synergistic effect it has on other risk factors. The older the individual, the greater the accumulation of risk factors and the increased duration and thus detrimental effect of the disease on the system (K. M. Anderson et al., 1991). A group of Canadian researchers has recommended important 'transition' ages that clinicians to pay attention to in CVD risk assessment. These age ranges are suspected to be when the risk of CVD can go from low/no to moderate or from low/moderate to high, and range between 33 to 48 for women and 45 to 52 for men (Booth et al., 2006). It is during these ages that newly developed risk factors can appear and serve as a warning or as new or additional focus for risk reduction therapy.

As indicated by the difference in the transition age, gender is another important non-modifiable risk factor. The earliest data indicated that men are at a greater risk of developing CVD than women (4X for men vs. 2X for women risk of CVD), however men have better CVD survival rates (D. Lloyd-Jones et al., 2010). Intriguingly, with a diagnosis of T2DM, those differences are reduced, with both genders having similar increased risk of developing CVD and dying from CVD related consequences (Booth et

al., 2006)(Amsterdam, 2011). This is believed to be the result of the higher risk profile of hyperlipidemia and hypertension in women, combined with the reduced probability of diagnosis and treatment of heart disease (D. Lloyd-Jones et al., 2010). Being female and having heart disease and diabetes has more severe outcomes for women heart disease is recognized less frequently.

Another variable that provides an opportunity for stratification of the risk of developing heart disease is 'family history of CHD'. The effect of having a family history of CHD is proportional to the number of biological relatives affected, and this is reflected in the hazard ratio for one blood relative being 1.5 but HR is 1.79 for two relatives with CHD (Ray et al., 2014). This increased risk is also dependent on the age of onset for each family member and the nature of the relationship (parents, sibling, or cousins, etc. (D. Lloyd-Jones et al., 2010; Tsao & Vasan, 2015a). However, although 'family history' is a better predictor than other biological risk factors, it did not improve the accuracy of the FRS and its use is only suggested in the case of older patients- males over 55 and females over 65 (Bertoluci & Rocha, 2017).

Using diabetes as a binary variable in the FRS results in a loss of valuable information. The risk level of among diabetics are not the same as there is a significant amount of heterogeneity and using the diagnosis as a risk equivalent reduces the accuracy power of the formula. Even among those who are non-diabetic there is also range of risk levels, as hypoglycemia can confer a higher CVD risk than does a normal glucose level. The use of glucose levels as a risk factor provides additional classification and discriminatory power thereby increasing the performance power of the FRS tool.

Glucose Level as CVD Risk Factor

Measuring Glucose Levels

Glucose level is an important CVD risk factor but there are many different methods for determining the quantity of glucose in the blood. The most common method is the fasting blood glucose (FBG) test, where a blood sample is collected, first thing in the morning after abstaining from food for at least 8 hours. The FBG concentration is reported either in milligrams (mg) per 100 mL (mg/100mL or mg/dL since 100ml = 1 dL) or in millimoles per liter (mmol/L). A FBG above **125** mg/dL (> 5.5 mmol/L) is diagnosed as diabetic, between **100-124** mg/dL (5-5.4 mmol/L) is considered pre-diabetic, **70-100** mg/dL (3.9-4.9 mmol/L) is normal, and below **70** mg/dL (<3.8 mmol/L) is hypoglycemic (American Diabetes Association, 2015).

Blood glucose levels can also be measured under non-fasting conditions by either of three ways: the casual blood glucose (CBG) or random blood glucose test; the post-prandial blood glucose (PPBG) test; or the oral glucose tolerance test (OGTT). The CBG is a test of the concentration of a blood sample taken any time of the day, before or after a meal, and values above 200 mg/dL (11 mmol/L) are diagnosed as diabetic (Kadowaki et al., 2008). However, the CBG test is not as reliable as the FBG since glucose levels are subject to normal fluctuations depending on the quantity of carbohydrates consumed in the last meal and the time since the meal was eaten. The CBG test is most useful for circumstances where the FBG is unavailable or obtaining it would be inconvenient.

The post-prandial blood glucose (PPBG) test or the oral glucose tolerance test (OGTT) test are both complicated procedures that involve blood samples taken at over

two-hour periods. For both tests, a sample of pre-prandial blood is taken, then the patient is made to eat a test meal in case of the PPBG or ingest a solution of 75 g of glucose dissolved in 8 oz water in the case of the OGTT. For the PPBG, a second blood sample is taken after two hours and the difference in the pre-prandial and post-prandial concentrations of glucose is determined. For the OGTT, additional blood samples are drawn every 30 minutes for the next two hours and plot of concentration over time is obtained (American Diabetes Association, 2009). Ultimately, the blood sugar level at the end of the two-hours should be <140mg/dl (<7.8 mmol/L) for a normal blood glucose response (Saisho, 2014). Values between 140–200 mg/dL (7.8–11.1 mmol/L) indicate impaired glucose tolerance (IGT) or prediabetes, while anything above 200 mg/dL (or >11.1 mmol/L) is diagnosed as diabetic (American Diabetes Association, 2015). These tests were commonly used to check for gestational diabetes in pregnant women, but their utility has recently been questioned.

Lastly, the average glucose concentration during fasting and non-fasting periods can be measured using the test of glycated hemoglobin, (the HbA1C) test. This newer biochemical test measures the percentage of particular protein molecules (hemoglobin in the red blood cells) that have been glycated (i.e., have glucose molecules attached to them) (Nolan, Damm, & Prentki, 2011). The higher the prevailing glucose concentration, the more glucose molecules bind to proteins, which negatively affects their structure and function (Tabák et al., 2012). The results from HbA1C tests reflect the average level of glucose in the blood over the previous 2-3 months and levels below 42 mmol/mol or

6.0% are considered normal; 42–47 mmol/mol (6.0–6.4%) is pre-diabetic; and anything above 48 mmol/mol (6.5%) is diabetic (American Diabetes Association, 2015).

Glucose Levels in Risk Assessment Model

The Full Range of Glucose Level is Associated with CVD Risk

The association between high glucose levels (i.e., hyperglycemia) and its exacerbation of the many complications of heart disease is well established. However, glucose levels just below and even far below the diabetic threshold also have an association with heart disease. Bertoluci and Rocha (2017), reported that hypoglycemia (FBG of 70-100 mg/dL) or low glucose levels (FBG of 101-125) have a hazard ratio of 2.64 for CVD and for 6.34 for CVD mortality, respectively. This is believed to be an effect of the high levels of insulin in the blood stream simultaneous to insulin resistance, that triggers the hypoglycemia and exacerbates the CVD symptoms (Faerch et al., 2012; Laakso, 2015; Park et al., 2013). This underscores the complex relationship between the predisposition to CVD and the full range of glucose levels from hypo- to normal to hyper-glycemia.

There are several studies that demonstrate the relationship between CVD and the full range of glucose levels. One such study is a cross-sectional analysis the data from a CVD prevention program that evaluated the CVD risk for 3,739 non-diabetics (Valentino et al., 2015). FBG values below 125 mg/dL (non-diabetic) were found to be a marker for CVD risk factor clustering and more specifically glucose levels below 100 mg/dL were found to be associated with low HDL, high triglycerides, hypertension, abdominal obesity, and the CVD inflammatory biological markers (Valentino et al., 2015). One

interesting aspect of this study was that in cases with similar, healthy cholesterol levels, it was the difference in glucose levels that identified the potential CVD risk. Findings like these that emphasize the need to include glucose levels in risk assessment as well as in risk factor reduction therapy for CVD health promotion.

Another study with middle-aged African Americans in Florida, looked at the role of modifiable and non-modifiable factors in CVD prediction (Carter, Ralston, Young-Clark, & Ilich, 2016). Information on diabetic indicators (blood glucose and insulin), apolipoproteins, adipokines, and lipid profile were collected. Using both the AHA and FRS assessment tools they analyzed the data and compared the results for the two genders. The results confirmed the higher CVD risk for men, and the negative association between the biological markers (except for HDL) and CVD outcomes. Additionally, the authors noted that glucose level was the strongest indicator, with the widest applicability across the groups, for CVD risk prediction (Carter et al., 2016).

An Icelandic meta-analysis of several Western prospective studies also indicated that glucose levels might be the strongest risk factor associated with vascular outcomes, CHD, and stroke (Sarwar et al., 2010). The analysis showed that the inclusion of glucose level, measured as HbA1C values, improved risk assessment as well as intervention and prevention strategies better than the more simplistic diabetic threshold concept. They also noted that the HbA1C test provides information on long-term glucose levels and does not require any fasting or post-prandial conditions, but that it is more expensive than standard glucose testing and is therefore not routinely measured.

A long-term follow-up study in Japan found that even random blood glucose or causal blood glucose (CBG) can be used to predict CVD mortality. Interestingly, borderline-high (140-200 mg/dL) and high CBG (above 200 mg/dL), the HR values were 2.43 to 2.62 respectively (Kadowaki et al., 2008). For normal range of CBG (126 – 140 mg/dL) the hazard ratio, although predictability lower, was still statistically significant. The data also showed that the relationship was not affected by time since the last meal, proving that CBG levels can also be used to predict CVD mortality and is a suitable substitute in settings where FBG is not available.

A very large cross-sectional study, involving 500,000 Chinese participants, looked at the association of diabetes and blood glucose levels on the risk of CVD development (Bragg et al., 2014). Participants age, height, weight, hip and waist circumference, and levels of blood pressure and non-fasting blood glucose were recorded. The results agreed with previous reports that diabetes doubled the risk of ischemia and stroke, but it also showed that among non-diabetics, there was a positive association between blood glucose levels and heart disease. For levels below the diabetic threshold, every 1 mmol/L increase in blood glucose correlated with a prevalence increase of 4% for ischemic heart disease and a 5% increase of ischemic heart attack (Bragg et al., 2014).

Glucose level vs Cholesterol level in Risk Assessment

In a South African study, Lammertyn et al. (2011) looked at the many aspects of cardiovascular dysfunction among Black Africans. One of the objectives was to compare the roles of glucose and cholesterol in predicting cardiovascular function. The measurements included: age, BMI, waist-to-hip ratio, electrocardiograms; systolic and

diastolic blood pressure, total HDL and LDL cholesterol, smoking, and the use of anti-hypertensive medication. Each of the 200 adults were subjected to an imaging test that recorded the carotid intima media thickness (CIMT). An increased CIMT reflects "impaired compensatory remodeling of the arterial wall and atheroma progression" following damage and is prevalent in diabetics, predisposing them to stroke and myocardial infarctions (Lammertyn et al., 2011). The analysis showed a negative correlation between CIMT and HDL:TC, a positive correlation between glucose and hypertension, but FBG proved to be the better candidate for CVD prediction.

A study conducted by Faeh et al. (2013) verified that replacing cholesterol levels with the long-term measure of hyperglycemia, produced a more accurate risk prediction model. Using the NHANNES III data set and the ESC SCORE formula, they compared four different models (HDL:TC ratio, cholesterol, and glucose) and found that the glucose values provided better prediction of CVD mortality compared with the other two variables, even when below the diabetes threshold values (Faeh et al., 2013). The researchers added that when the FRS and other models use only diabetes (yes/no) that it insufficiently maps the potential impact of blood glucose on CVD.

A swiss study that followed 6,095 adults for 32 years, compared the ability of glucose levels to that of cholesterol levels to predict fatal CVD event (Braun et al., 2013). Measures of age, gender, blood pressure, smoking, and FBG or total cholesterol were used in the ESC-SCORE model. The models were cross-validated with another data set and the area under the receiver-operating characteristic curve (AUC) and integrated discrimination improvement (IDI) values were used to compare accuracy. The

researchers found that not only did both low and high glucose levels generate a more accurate prediction than did cholesterol, but in a joint model (i.e., both cholesterol and glucose) only the coefficient for glucose was statistically significant (Braun, 2013).

Another study among African Americans looked at the role of fasting blood glucose and the different components of serum cholesterol in predicting the reactivity to stress associated with CHD development. Patients were exposed to a prescribed stressful situation and measurements were taken prior to, during, and while recovering from the stressful period. The analysis revealed a negative association between HDL and high blood pressure, and that FBG was better than cholesterol in predicting cardiovascular stress (Clark et al., 2015). The authors stated that excess insulin was associated with high glucose concentration and was the mediator of the stress response (Clark et al., 2015).

Glucose Lowering in CVD Treatment

Anti-diabetic drugs failed to show efficacy in cardiovascular treatment because they were producing insulin, which deleterious to the vascular system. Several clinical studies that sought to reduce CVD risk by aggressive glucose lowering failed or worse-had to be aborted due to increased morbidity and mortality in the treatment group (Ginsberg, 2011; Nordestgaard et al., 2013; Reaven et al., 2009; The ACCORD Study Group, 2011; Zoungas et al., 2009). What these trials had in common was the use of either insulin supplementation or drugs that stimulated the pancreas to produce more insulin. Not only did that approach fail to lower the HbA1C, but the CVD outcomes were poor. Recent research have indicated that while insulin may lower the glucose level, the

hormone itself has many deleterious effects on the vascular system (Faerch et al., 2012; Laakso, 2015; Sah et al., 2016).

As further proof that glucose lowering was not the source of the problem, trials that used non-insulin elevating intervention to lower glucose levels obtain greater success in reducing the participants CVD risk (Kishore et al., 2012; Xu & Rajaratnam, 2017). In fact, the use of metformin, a drug that lowers glucose while at the same time increasing insulin sensitivity was shown to confer significant CVD benefits to diabetics (Roumie et al., 2012; Skov et al., 2014). These studies confirm that the focus on glucose level is not misplaced and that association with CVD events is an important one. The strength of the glucose level association with risk of CVD emphasizes the need for its inclusion not just in CVD treatment but in CVD screening and risk assessment as well.

Summary

The incidence of cardiovascular disease began rising after WWII and rapidly became an epidemic in the middle of the last century. Despite a slight decline in the rising rate by the end of the century, it remains the deadliest of all the chronic diseases and the number cause of death in most developed and developing countries. The Framingham Heart Study (FHS), initiated by the US Government in response to the epidemic of heart disease, was a several decades long prospective study of over 5,000 adults living in Framingham, MA. The FHS successfully identified several of the major risk factors and encouraged the idea of disease prevention by reducing the incidence and severity of these risk factors in at risk individuals.

The FHS also developed and risk assessment model, the ubiquitous Framingham Risk Score (FRS), based on the risk factors of CHD and used to predict the probability of developing the disease over a given time period. The FRS, though useful and effective is limited in its accuracy and as such has undergone several rounds of optimization. The inclusion of diabetic status as a risk factor in the FRS formula has been challenged over the years and is now considered a CHD equivalent. However, using this risk factor dichotomously results in a loss of very important risk information and limits the performance of the FRS formula. Moreover, because CHD is a cardiometabolic disease, the continuous variable 'glucose levels' inclusion in the FRS model increases the accuracies and thereby improves its performance.

Chapter 3: Research Methodology

Introduction

Overview

This chapter includes an outline of the methods I used to answer the three research questions formulated in the first chapter of the dissertation. Each of the questions is examined in detail and the research design used to generate an answer is explained. The chapter begins with a description of the comprehensive research design, followed by details on the secondary data set chosen, including how the subjects were selected, how sample data were originally obtained, and how the data were ultimately handled in preparation of the statistical analysis. Also included in this chapter are the definitions of the outcome variable and all the explanatory variables featured in the analyses, along with their operationalization for inclusion in the model. The chapter ends with a presentation of the three research questions, each followed by the specifics of the statistical test used and determination of the statistical significance for each.

Study Objective

The purpose of this study was to investigate the effect that the inclusion of FBG levels, as a risk factor variable, has on the accuracy of the FRS formula in predicting 10-year CHD event. The second chapter detailed the history of the FRS formula as documented in the literature and the many studies that focused on performance optimization for this formula. One aspect of the FRS performance optimization that has been overlooked is the inclusion of glucose as a continuous variable, and it is this gap

that I sought to fill. MLR analysis was used to generate formula pairs (for FBG level and DM2 status) and the performance in representing the data compared.

Before the multivariate formulas was generated for performance comparison, the relative risk (as odds ratio or OR) of each of the two factors (FBG and DM2), for the CHD event outcome, was determined using univariate analysis, along with their respective confidence intervals. Following the univariate OR comparison, the multivariate formulas, containing the various cofactors (AGE, SEX, HDL, SMOKING, BMI and either FBG or DM2) was generated and the ORs of each compared the appropriate goodness-of-fit tests were used to compare the performance. Finally, the interaction between the each of the covariates Sys BP, BMI, and HDL and FBG was also assessed. The effect of the potential moderators (SBP, BMI, and HDL) was evaluated by the creation of three interaction variables (HDLxFBG, SBPxFBG, and BMIxFBG), that were included in the MLR formula. An evaluation of OR and significance of the interaction terms and the performance of the overall performance model was evaluated.

Research Questions

RQ1: Is the relative risk of glucose level higher than that of diabetic status for the development of CHD in the FHS data set**?**

$H_0$1: There is no difference between the relative risk of glucose level and that of diabetic status for the development of CHD in the FHS data set.

H_a1: The relative risk of glucose level is higher than that of diabetic status for the development CHD in the FHS data set.

RQ2: Is the measure of accuracy higher for the glucose level formula than that for

diabetic status formula in the FRS-CHD 10-year risk prediction model?

$H_0$2: There is no difference between the measure of accuracy for glucose level and diabetic status versions of the FRS-CHD 10-year risk prediction formula.

H_a2: The measure of accuracy for glucose level formula is higher than that for diabetic status version of the FRS-CHD 10-year risk prediction formula.

RQ3: Is the measure of accuracy for the glucose level formula independent of age, BMI, cholesterol level, or blood pressure level in the CHD 10-year CHD prediction model?

$H_0$3: The measure of accuracy for glucose level is independent of age, BMI, cholesterol level, or blood pressure level in the 10-year CHD prediction.

H_a3: The measure of accuracy for glucose level is dependent on age, BMI, cholesterol level, or blood pressure level in the CHD 10-year CHD prediction.

Research Study Design

Choosing the FHS Data Set

The data set used in this study was the third generation of FHS, a longitudinal study originating several decades ago to collect information in understanding the factors that predispose individuals to heart disease. The FHS was a prospective cohort study, where the exposure or risk factors levels are measured before the disease develops (D'Agostino et al., 2013). All the subjects were followed over a decade and information on their risk factors and their disease outcome (whether they developed heart disease or not) were recorded . This type of study, though more expensive and time consuming,

allows for greater accuracy concerning the risk of exposure and it is ideal for investigating causal relationship (Kannel & Boston, 1990).

The data set is made up of participants from the third generation of the FHS cohort (FHS-GenIII), combined with the OMNI2 and the new offspring spouses (NOS). The FHS III data accounts for 4,578 adults, who are the grandchildren of the FHS I, along with an additional 101 of their non-FHS parents who had not been included in any of the previous FHS cohorts (Splanksy, 2007). The OMNI2 are the children of the original OMNI cohort, introduced to include a more racially and ethnically diverse sample set, but they numbered only 405 (Tsao & Vasan, 2015a). The NOS (New Off-Spring) cohort, also added to increase genetic diversity, was made up of the spouses of the OMNI2 and consisted of only 101 subjects (Govindaraju et al., 2008). The sample size of the raw data was 4,578 that was reduced to about 60% (2,670) after the filtering process detailed in data handling section of Chapter 4 of this dissertation.

In the original FHS cohort, several risk factors were measured and recorded for each of the subjects and analysis of the data uncover the risk factors that were linked to CVD. This led to the development of risk prediction strategies and to the prominence of primary prevention effort in response to the rise of CHD and other CVDs (Wilson et al., 1998). The FHS data have been used extensively over 5 decades to provide additional insight into the association between a variety of risk factors and the probability of developing CVD . Despite its limitations, the large, randomly selected data set is still relevant to predictive studies today. The logistic regression model generated from the

FHS, the FRS remains the most widely used CVD prediction formula (Garg et al., 2017), making it ideal for studying the role of major risk factors such as FBG.

Experimental Validity

External Validation. External validity is a measure of the generalizability of the research findings beyond the current data set (Stoltzfus, 2011). The risk formula developed by the FHS team faced issues of validity from the beginning. The first concerns were who were the truly at-risk individuals, on which aspect of CVD should the prediction focus, and how would this outcome be defined. The decision was made to focus on the prediction of CHD, the most common CVD and with the highest clinical relevance specific to death and disability (Mahmood et al., 2014). For original FRS formula generation, the development of the first hard CHD (originally called atherosclerotic heart disease or ASHD), defined as coronary death or myocardial infraction, were the events was chosen as the outcome of interest (D'Agostino et al., 2013). With the endpoint decided, the issue was now who was considered at risk and who should be included in the study and how they would be selected. For the sake of simplicity, the first cohort was made up of individuals who were at the time free of any CVD and had not had a previous CVD event (D'Agostino et al., 2013).

The remaining issues of validity included what risk factors should be measured and what should be the cut-off for follow-up time. The decision about the time was easier, and a period of 10-year was decided as adequate for CHD risk prediction, but to ensure all events were captured, 12 years of data was collected on each subject (D'Agostino et al., 2008). Uncertainty about what factors to include and what to exclude,

the researchers decided that data on all factors making up a regular clinical exam (gender, age, height, and weight) along with suspected predisposing factors (BP, HDL, smoking status, and diabetes) would be collected (Dawber et al., 1957). The debate over which risk factors should be included continues to be part of the optimization research, but focus has always been on keeping the formula simple and convenient yet effective as possible. As a result, more definitive factors that were difficult to assess, were excluded.

An additionally important decision was to the selection of the mathematical model that would serve as the best risk estimation function for the data set. The ideal model should not only adequately estimate the risk for the period in question, but it should also provide some measure of the relative risk for each of the risk factor. The first model was based on the discriminant analysis, but was replaced by the more robust and more flexible logistic regression model as soon as its existence was known to the FRS team (D'Agostino et al., 2013). Over the years, other analyses like the Cox proportional hazard regression and accelerated failure models were considered (see D'Agostino et al., 2013). However, the absolute risk at t=10 years provided by the simpler logistic regression analysis became the preferred measure on which to make important treatment decisions.

Once I selected the model it became essential that its suitability and performance be evaluated. At the top of the list of requirements were powers of discrimination and calibration, the former reflecting the ability to separate between cases and noncases, and the later measuring the accuracy of estimating an absolute risk probability (see Dawber et al., 1957). The C-statistics or AUC (area under the curve) was the measurement of choice

as it describes the sensitivity and the specificity of a function and indicates the ability of the function to rank an individual's risk (see Bertoluci & Rocha, 2017). For calibration evaluation, the FHS researchers first used a specifically designed chi-squared test combined with Kaplan Meier (K-M) estimates to quantify the difference between the predicted probability and those reflected in the data (D'Agostino et al., 2013). The K-M test was suitable for the time-to-event risk estimation models used in the beginning but it was replaced by the Hosmer-Lemeshow test which more accurately evaluates the goodness-of-fit for logistic regression models (Giancristofaro & Salmaso, 2003).

Internal validation. Internal validation refers to the replicability of the test findings and reflects the veracity of the inferred results and a minimization of systematic error (Eichler et al., 2007; Kones, 2011). Internal validation should ideally be done on a different data set to avoid overfitting, but suitable alternatives include bootstrapping, cross validation, or simply split-sample validation (Giancristofaro & Salmaso, 2003). The 10-fold split was the first validation technique for the FRS formula, where the sample was split into deciles and nine models (developed on different sample sets) were developed and each tested on the unused 10th sample (Steyerberg et al., 2010). The C-statistics and calibration plot comparisons are performed for each of the nine models to select the one that best described the data. Fortunately, the calibration models were not significantly different from each other or the one generated on the entire data set, making this method futile (Steyerberg et al., 2010).

Despite the congruency in the internal validation procedure, the researchers (citation) knew that the best way to validate a multivariate risk model is with a

completely independent data set. However in a 1999 study, the NLBI had determined that the FRS performed relatively well on the data from other prospective studies and should be used with confidence in other populations across the United States (Kannel et al., 1999). It had also been verified as suitable for other populations across Europe, Asia, and regions of the Mediterranean, subsequent to adequate recalibration, but that would later be challenged by European experts in the field of cardiology (D'Agostino et al., 2001a). Given the racial bias of the original cohort, the FRS's performance on other populations of has always been questioned. Researchers discovered that after a recalibration process the FRS was able to distinguish the high-risk individuals from those at low risk in non-Framingham populations in predicting hard CHD (O'Donnell & Elosua, 2008). This recalibration, which included adjusting the intercepts for the formula and replacing the disease incidence with that of the current population, was found to worked well with non-Framingham populations (D'Agostino et al., 2008). This method is the basis of many recalibration methods used by those who use the FRS in communities outside the United States.

Secondary Data Set

FHS Data Set: Cohort III, OMNI & NOS

The data used in this study was obtained from the FHS group, and made up of the third generation of the original cohort that started in 1948 (Lansing, 1961). I requested permission from the institutional review board (IRB) to contact the FHS researchers and obtain a copy of the relevant data file. The FHS is a longitudinal, retrospective cohort and is thus suitable for study risk factor/outcome associations. The FHS has been the credited

with discovering all the major, now established, and minor risk factors of cardiovascular disease (Mahmood et al., 2014). This data set was used to develop the renowned formula used effectively to predict the likelihood of developing cardiovascular disease over a 10-year period (Eichler et al., 2007). FHS data has been the basis for 1,000s of published research articles on the subject related to heart disease and other related illnesses.

The long natural history of the CVD necessitated a prospective, observation-type, epidemiological cohort study. In the first cohort, the need was to explore the relationship between the disease and the predisposing factors, while at the same time collect important data on the disease's prevalence, incidence and prognosis (Dawber et al., 1951). Influenced by the enthusiastic physicians of the area and the previous success of the 6-year TB study there, participants were recruited through a random sampling of the residents from the now infamous town of Framingham, MA (Dawber & Lansing, 1966; Mahmood et al., 2014). To ensure that all the possible subject's data were accounted for, an extended follow-up period of data collection was adopted (D'Agostino et al., 2008). The study continues today, some 70 years after it began, with the second and third generation of the FHS cohorts.

Sample Selection

Initial recruitment for the FHS began with letters being sent out to two of every three (randomly selected) of the 10,000 families living in Framingham, MA, inviting anyone between 30 and 59 years of age to participate (Dawber et al., 1959; Tsao & Vasan, 2015). This group was believed to be representative of the adult U.S. population and the study was intended to run for only 20 years and to offer the benefits of a clinical

exam to all. News of the study spread, mostly by word of mouth, from first few recruits who were honored to be involved, and soon 100s of participants turned into several 1,000s (Tsao & Vasan, 2015a). Of the 6,507 contacts made, 2,336 men and 2,873 women responded to make up the original cohort 5,209 (Long & Fox, 2016; Tsao & Vasan, 2015). The data from this original cohort was intended for use in the analysis for the research, but due to concerns about the age (over 60 years) of the data set, the data from the third generation (most recent) was chosen instead.

The community-based structure of the FHS study enabled the identification of aggregation of the traits among family or communal groups (Tsao & Vasan, 2015a). The detection of shared family traits led the researchers to start recruiting the second generation participants from the children and their spouses of the original husband and wife pairs in 1972 (Tsao & Vasan, 2015a). The off-springs and their spouses of the second generation (grandchildren of the original cohort participants) made up the third FHS generation enrolled in 2002 (Mahmood et al., 2014; Tsao & Vasan, 2015). These groups provided many invaluable insight including who were susceptible, the effects of a shared environments, and most importantly, the familial and multigenerational genetic factors involved (Long & Fox, 2016; Mahmood et al., 2014). The first FHS offspring (second generation) study was organized in 1971, with 5,124 participants and consisting of the children of the original cohort participants and their spouses (Tsao & Vasan, 2015a).

The smaller OMNI1 and OMNI2, started in 1994 and 2003 respectively, focused on increasing the racial and ethnic diversity of the participant pool beyond the mostly

white (of European descent) participants that made up the first three (original, second generation and third generation) cohorts (Long & Fox, 2016; Tsao & Vasan, 2015a). All the succeeding cohorts now include state of the arts health data collection methods like ECG, imaging, and genotyping and have contributed valuable temporal trend and genotypic. Data on other diseases such as cancer, COPD, diabetes, epilepsy, arthritis, Parkinson's, and Alzheimer's is also included (Tsao, 2015). The factors measured have expanded from the original dozen or so to include almost twice as many with the addition of blood tests for insulin, HbA1C, thyroid hormones, liver function tests, c-RP, and testing for specific genes, etc. Add summary and synthesis to fully conclude the section.

Sample Size Determination

Statistical power analysis is a method used to determine the sample size required to bestow sufficient power to the research result (Rudestam & Newton, 2014). The statistical power reflects the probability of the Type II error (β) for the analysis, which is the failure to reject the null hypothesis when it is in fact false. This power is equal to 1 minus probability of the Type II error (Power = 1- β), thus as the probability of error increases, the power of the study or the significance of the results decreases. It is for this reason that the power of a model is inversely proportional to the number of factors, and formulas with the lower number of variables are preferred (Sperandei, 2014).

The statistical power not only depends on the level of Type II error, but on the effect size and the sample size as well. The effect size is a measure of the treatment effect and is usually calculated by dividing the difference of the means by the standard deviation ($(M_1 – M_2)/SD$), and the sample size is simply the number of participants

(Fuller, 2009). There are many formulas (some more complicated than others) that use the effect size, the sample size, and the probability of Type II error to calculate the statistical power. In other cases, a reverse calculation is performed, and the appropriate sample size is determined from the desired statistical power, the estimated effect size, and the predetermined probability of Type II error (Pearl, Glymour, & Jewell, 2016).

Calculating the sample size for a multivariate logistic regression is challenging. First, there is the issue of the nonlinear correlation that is represented by the odds ratio, which describes the relationship between the independent variable and the dependent variable. Second, there isn't one independent but several independent variables, with non-zero interaction between them, and each having a different probability value in the population. Choosing from among the many possible coefficients can be difficult, however, simulation using the G*Power application has simplified the process (Faul, Erdfelder, Lang, & Buchner, 2007). This app allows you to choose among several different types of test, including F tests, t tests, chi-squared tests, and z test, etc., and the input parameters include, but are not limited to, "Tails", "Odds ratio", "Power", and "R^2". I used generic values of **1.3** for the "Odds Ratio", **0.05** for the "error", and **0.8** for the "Power" and a value of **962** was returned for the "Total Sample Size". For the data set in question, with over 5000 participants, the sample size is more than adequate for statistically significant effect size determination.

Data Collection

Traditionally, the FHS data were collected by the biennial follow-up exams and consisted of face-to-face survey questions and clinical exams. The information gathered

comprised physical and medical data, including blood pressure, blood glucose, height and weight (Long & Fox, 2016). For the original cohort, those determined to have a history of CVD were excluded from the study, but for subsequent cohorts, their data were included and their information collected. Originally the researchers wanted to focus only on those who were free of any cardiovascular complications at baseline so as not to confound the data (D'Agostino et al., 2013; Kannel et al., 1999). But they later realized that a history of CVD was a strong risk factor for a future CVD event, and this eventually became one of the factors recorded for each subject in subsequent cohort (Kannel et al., 1999).

Each of the biannual exam began with a written informed consent form along with a questionnaire on the patient's medical and family history (see Appendix A). The data collection included but was not limited to information on: family history, medical history, current symptoms, medication, personal and dietary habits, weight, biometrics, cardiac exam, x-ray, ECG, blood (Hb, cholesterol, uric acid, glucose, syphilis) and a urinalysis (Dawber et al., 1951). A complete physical, targeted at the cardiovascular health, as well as collection of blood and urine samples followed the written portion of the exam. These exams were mostly conducted at the designated sites, otherwise staff members visited home-bound participants. Included in the medical consent was permission for the FHS group to access the diagnostic and chart records from clinics and doctors' offices and hospitals, medical examiners' reports, physician's files, and death certificates (Friedman, Hyg, Kannel, Dawber, & McNamara, 1967).

For each case, chart data were eventually compared to the original clinical findings on attributes associated with the development of CVD, paying attention to the

most common CVD subtype- CHD (Tsao & Vasan, 2015a). The frequency of the exams fostered relationships between the participants and the researchers and helped retention, while the comprehensive nature of the follow-up exams minimized the possibility of collection bias (Friedman et al., 1967). The dedication of the participants and the perceived value of the 'free' medical exams resulted in high retention rate, contributing to the exceptional quality of the FHS data and setting a high standard that continued with subsequent cohorts. Of the 5209 participants in recruited between 1948 and 1952, event data on 4,439 participants were collected over three exam cycles from 1956 to 1968 (Sorlie & Coady, 2004). That is a retention rate of 85%, which high by most standards.

Data Handling

The first step in handling any data set is the cleaning process, which entails detecting and removing or replacing all corrupt, inaccurate, or incomplete records. The FHS data set was assessed for missing data, incorrect data and outliers, and each variable was evaluated to ensure that no more than 10% of the data is missing, as missing data could the produce incorrect analysis (Field, 2013). In the case of missing or incorrect data, the use of mean or weighted values can be generated to compensate, depending on the nature of the variable being measured, but there was no need for this. Box or Whisker plots with the interquartile range determination was used to detect for any outlier values that may have required the removal of all corresponding for offending subject (Field, 2013). However, the potential outliers were tested and determined to be legitimate measures, and no exclusion of participants data was needed.

The data was received in the statistical analytical system (SAS) form, which was downloaded from the NIH website following the granting of permission (password) to access. The SAS zip files (FHSIII, OMNI2 and NOS2) were saved on my hard drive and unzipped for use. The relevant SAS files (CHD event, lab tests, general health screen, etc.) were opened in SPSS and working version will be created for selecting of the appropriate variables. The files were combined into one, utilizing the unique patient unique identification number, and assessed for errors then filtered, as described in chapter four. All variables will be checked to ensure that the SPSS type, length and description in SPSS matched that listed in the data dictionary provided, and data set was filtered to remove those with a history of heart disease. Following the filter, the variables were subjected to comprehensive descriptive and correlation analyses before the inferential analyses were conducted (Lani, 2015).

Ethical Considerations

Permission was obtained from the Institutional Review Board (IRB #: 10-18-18-0031692) of Walden University to request access to the FHS data set, and since it is a secondary analysis an exempt status requested. The data set from the FHS is distributed minus any personal information or any unique identifiers, ensuring participants' confidentiality. This guarantees the individuals' privacy and compliance with the Health Insurance Portability and Accountability act (HIPPA). The FHS data is available for research and analysis purposes and has been used for 100s of invaluable studies over the last several decades, with no history violation of rights.

Operational Definition of the Variables

Dependent Variable: Disease Outcome

The FHS was the first epidemiological study of its kind and it led the way in identifying CVD risk factors. The risk factors identified were then made the subject of many clinical trials that revealed the efficacy of risk factor reduction in CVD prevention strategies (Nieto, 1999a). The researchers realized that CHD was the most common CVD event, with the most lethal consequences, so they decided to focus on the risk factors for CHD in the risk assessment formula (Truett et al., 1967). However, in 2008, the FHS team decided that a more global formula was also needed and developed a formula, with a comprehensive FRS to include predictions of CVD that included cerebrovascular disease, intermittent claudication and congestive heart failure (Kones, 2011). The more commonly used formula however, estimates the 10-year absolute probability of developing the risk of hard CHD, where coronary death or myocardial infraction was the outcome (D'Agostino et al., 2013). The formula returns three categories of risk: high risk for a 10-year risk score >20%; moderate risk for a score between below 20% but above 10%; and low risk for a score below 10% (Pedersen, 2002). This analysis used CHD as determined by the FHS team and recorded as the dichotomous nominal '1' or a CHD event and '0' for no-CHD event.

Independent Variables: The Risk Factors

The FHS is credited with uncovering the multifactorial nature of CVD and with the recognition of interactions between the risk factors, leading to the creation of the multivariate risk prediction formula (Kannel & Boston, 1990). The risk factors -gender,

age, systolic blood pressure (SBP), HDL cholesterol (HDL), body mass index (BMI), smoking behavior (SMKG) and 'diabetes status' (DM2)- all comprise the current algorithm that is used as the Framingham Risk Score (FRS) for estimating a patient's CHD. There have been changes in the way a few of the variables are measured and utilized in the formula, namely blood glucose, which although measured for determining diabetic status it was not included in the FRS formula; and HDL values replacing total cholesterol by the second cohort as technological advancement facilitated it determination, and only systolic BP (SBP) as being a better predictor of CHD than diastolic BP (DBP). The following section will take an in-depth look at the factors in the FRS formula and the role they each play in predicting CHD.

Cohorts	Original Cohort 1948–1953		Offspring Cohort 1971–1975		Third Generation 2002–2005	
Variables	Men, $N = 2336$	Women, $N = 2873$	Men, $N = 2483$	Women, $N = 2641$	Men, $N = 1912$	Women, $N = 218$
Age (years)	44 (9)	44 (9)	37 (11)	36 (10)	40 (9)	40 (9)
Current smoking (%)	78	41	45	44	19	16
Systolic BP (mmHg)	136 (19)	135 (24)	126 (16)	118 (16)	121 (13)	113 (14)
Diastolic BP (mmHg)	86 (12)	84 (13)	82 (11)	76 (10)	78 (9)	73 (9)
Hypertensive medication (%)	0	0	4	3	10	7
Hypertension (%)	45	39	26	13	13	8
BMI (kg/m^2)	25.8 (3.5)	25.4 (4.7)	26.4 (3.7)	24.0 (4.6)	27.9 (4.7)	26.0 (6.1)
BMI $\geq 30\ kg/m^2$ (%)	12	15	15	10	26	21
Blood glucose (mg/dl)	82 (24)	82 (20)	106 (16)	99 (15)	99 (18)	92 (18)
Total cholesterol (mg/dl)	221 (43)	221 (46)	201 (40)	192 (39)	193 (37)	185 (34)
HDL-C (mg/dl)			44 (12)	56 (15)	47 (12)	61 (16)
Lipid lowering medication (%)			1	0.3	11	4
Prevalent CVD (%)	4	2	3	1	2	1

The Risk Factors Measured in FHS Cohort 1, 2, and 3. (Govindaraju et al., 2008).

Cholesterol. Cholesterol is a fatty substance transported in the blood as small bundles of proteins and fat molecules, aka lipoproteins. The two major types of lipoproteins are high-density lipoproteins (HDL) and low-density lipoproteins (LDL) and are named based on the ratio of lipid to protein. HDL, the larger, heavier molecule, has a higher protein content, while LDL is smaller and lighter with more fat than protein

(Chowdhury et al., 2014). Cholesterol is carried to the tissues as LDL, which why it is found deposited on the arterial walls and is the major component of plaque (National Heart Lung & Blood Institute, 2013). High levels of LDL are associated with an increased risk of CHD, heart attack and stroke (Srikanthan et al., 2016). Cholesterol is removed from the arteries and tissues by HDL molecules and taken to the liver for expulsion (National Heart Lung & Blood Institute, 2013). High levels of HDL are associated with a reduced risk of CHD, making it the "good/healthy" cholesterol.

Blood Pressure. Blood pressure is a measure of the force exerted on the walls of the arteries as by the blood flowing through it. The pumping of the heart pumps forces blood to move through the blood vessels delivering oxygen and nutrients to the cells and tissues of the body. Damaged arteries, narrowed by plaque deposits, require extra force or higher pressure to overcome the increased resistance. This persistent high blood pressure or hypertension is overworks the heart muscles, and together with the damaged arteries, if left untreated, leads to heart disease (Franklin & Wong, 2013). High blood pressure is referred to as the "silent killer" as it develops slowly and can go undetected for many years. Fortunately, the FHS discovered that systolic blood pressure (SBP) is more strongly associated with CVD than is diastolic blood pressure (DBP), and the FRS only includes the former (Kannel, Gordon, et al., 1971). Hypertension, now defined as an SBP above 140 mmHg, and is associated with an increased risk of CHD (Pedersen, 2002).

Diabetic Status. Diabetes Mellitus II (DM2) or hyperglycemia describes the condition of persistently high concentration glucose in the blood. Normally, ingested carbohydrates are converted to glucose and transported into the cells by insulin to meet

the cells energy need. However, when the supply of glucose supply is excessive, the cells become unresponsive to insulin and this excess glucose remains dissolved in the blood (Sah et al., 2016). This increased glucose causes an upsurge in the pancreas' insulin production (Faerch et al., 2012), and it is this elevated levels of glucose and insulin is known to damage the walls of the arteries and increase the risk of heart disease (Faerch et al., 2012; Laakso, 2015; Tostes & et al., 2009). For first FHS cohort, diabetic status was made by either of three determinations: (i) the use of insulin; or (ii) two separate incidents of a glucose level >150mg/dL; or (iii) an abnormal glucose tolerance test (Qazi & Malik, 2013a). It was the FHS researchers who first reported the link between diabetes and CVD (Kannel & McGee, 1979a), a link that was confirmed by many subsequent studies (Carter et al., 2016; Cockram et al., 2001; Ray et al., 2014; Zhang et al., 2014). Although elevated glycated hemoglobin (HbA1c) is becoming more popular in diagnosing DM2, in many cases a FBG above 120 mg/dL is also used.

Glucose Level.

In the original FHS study, glucose level was used only for the determination of diabetic status. Fortunately, the raw values of fasting blood glucose (FBG) have always been recorded for all cohorts and are available for use in the current analysis. High FBG is now recognized as being a strong risk factor for CVD as well as other diseases such as diseases of the kidney, the eye, the brain, and the central and peripheral nervous (Ismail-Beigi, Moghissi, Kosiborod, & Inzucchi, 2008; Laakso, 2015). However, the only established glucose level threshold is that used for the diagnosis of diabetes, and glucose levels in the diabetic range are now accepted as a risk of microvascular damage that leads

to other disease (Sarwar et al., 2010). But the threshold for diabetes diagnosis was not based on the risk of CHD and that threshold may not relevant to the risk of a CHD outcome. In fact, research now shows that the link between glucose and atherosclerosis and cardiovascular disease may extends below the diabetes threshold (Braun et al., 2013; Desouza, Raghavan, & Fonseca, 2010; Valentino et al., 2015). Low levels, intermediate (impaired glucose tolerance) levels, and low-normal (pre-diabetes) levels of glucose have all been shown to be associated with a risk of developing CHD (Coutinho et al., 1999; Desouza et al., 2010; Faerch, Vistisen, Borup, Marit, & Jørgensen, 2014; Park et al., 2013; Valentino et al., 2015). For this reason, this study contends that FBG, a continuous and modifiable CHD risk factor, can improve the efficacy of the CHD risk prediction.

Body Mass Index. An overweight person is one whose weight is above what is considered normal or healthy. Body mass index or BMI is a measure of overweight and is calculated by dividing the weight in kilograms (Kg) by the square of the height in meters (m^2). i.e. BMI = Kg/ m^2 (Herman & Rothberg, 2015). For adults, a BMI below 25 is normal, above 25 is overweight, and above 29 is obese (American Diabetes Association, 2015). A person increasing BMI is an indication of an increasing dysfunction in their body's ability to metabolize glucose (Srikanthan et al., 2016). The high BMI in overweight and obese persons is usually associated other CHD risk factors such as high blood cholesterol, high blood pressure, and hyperglycemia, making an elevated BMI strong risk factor CHD and heart attack (Grundy et al., 2002).

Smoking. Cigarettes or tobacco smoking or exposure to secondhand smoke subjects the body to toxic gaseous molecules that are harmful to its tissues and organs.

These toxins cause damage the blood vessels, mainly to the lungs and heart first, which initiates a buildup of plaque and the formation of blood clots in the arteries (Huxley et al., 2011). Smoking is also associated with a lower level of the healthy HDL, exacerbating the development of heart disease risks. Smoking as a CHD risk is dependent on the how long the individual has smoked and how many packs are smoked per day (Mannan, Stevenson, Peeters, Walls, & McNeil, 2010). Ideally, smoking should be used as a continuous risk factor (number of cigarettes or number of packs per day, or a similar measure), and some optimized versions of the FRS do so. However, the current FRS formula used for purposed of CHD risk screening uses as a binary (yes/no) factor of current smoking status and will be similarly used in this study.

Age. As the body ages so does the blood vessels and older blood vessels are less supple and more susceptible to damage. Both the buildup of plaque in response to any damage and the loss in flexibility makes blood flow more difficult, resulting in the need for increase in blood pressure from the heart (Rodondi et al., 2012). Over time, these conditions combined with an extended exposure to other harmful influences makes heart disease more common as subjects get older. Persons in an older age group are more likely to die from heart disease than those in a younger age group, in fact over 80% of the people who die from CHD are over 65 years and (Mozaffarian, Benjamin, Arnett, et al., 2015). Older people are not only more likely to develop heart disease they are also more likely to die from it, making age a major CHD risk factor (Tsao, 2015). The age of the participants, as recorded in years, at the original exam was used in the MLR formula.

Gender. Being female is somehow heart protective, relative to make, and women are less likely to develop CHD. Not only is there is a lower prevalence of the heart disease among women but and women also tend to develop CHD at a later age than men (Franklin & Wong, 2013). Unfortunately, there are gender specific factors such as preeclampsia during pregnancy or menopause, that do increase women's predisposition to CHD risk factors such as atherosclerosis and hypertension (Hosseini, 2015). Additionally, the risk posed by cigarette smoking is more harmful to women and furthermore women have greater incidences of metabolic dysfunction (obesity, prediabetes, diabetes) than do men (Huxley et al., 2011; Moore et al., 2017). Moreover, the major CVD risk factors affect the female body differently than it does the male, for example diabetes raises the risk of CVD for both genders, but it does so to a greater extent for women than it does for men (Hosseini, 2015; Qazi & Malik, 2013b). Diabetic women tend to have a greater burden of risk as well as a higher risk of developing CHD, making gender another important non-modifiable risk factor (Amsterdam, 2011; Booth et al., 2006). All these factors combined, makes gender or sex a major risk factor for CHD and in some instances, separate gender-specific formulas are used for screening purposes (Çevik et al., 2015). Here gender was included as a nominal dichotomous variable along with the other risk factors in the multivariate formula.

Data Analysis

This study sought to use statistical analysis of the FHS data set to answer the following three research questions:

Research Question 1:

Is the relative risk of 'glucose level' higher than that of 'diabetic status' for the development of CHD in the FHS data set?

Null Hypothesis. There is no difference between the relative risk of 'glucose level' and that of 'diabetic status' for the development of CHD in the FHS data set.

Alternative Hypothesis. The relative risk of 'glucose level' is higher than that of 'diabetic status' for the development CHD in the FHS data set.

Statistical Analysis. This first question compared the strength of the association between the two predictors, FBG and DM2, and the outcome. Given the binary outcome variable (10-year CHD event), a logistic regression is most suitable for this analysis. Here two different univariate logistic regression formulas were generated, one for 'diabetic status' (DM2) and one for 'glucose level' (FBG), and the odds ratio (OR) for each was recorded. The OR is a measure of the change in odds for the outcome (CHD) corresponding with a unit change in the predictor, and an OR > 1 implies that the presence of predictor or exposure increases the odds or the probability of the outcome event (Tripepi, Jager, Dekker, Wanner, & Zoccali, 2007). This ORs from each of the formulas, which indicated the crude or unadjusted measure of association for the outcome, was used to compare the relative strength of two independent variables and to make a determination of which was the stronger (Field, 2013).

Logistic regression (LR) analysis generates a regression coefficient for the predictor variable, and the OR is calculated by finding the exponential or inverse-logit function of that regression coefficient (Sperandei, 2014). In LR, the categorical nature of

the outcome violates the regular assumption of linearity, but the relationship between the predictor variables and the logit of the outcome variable is assumed to be linear (Lani, 2015). The statistical significance of the OR is based on both the *p*-value and the 95% confidence interval (CI). A α level of 0.05 was used and an OR with a *p*-value less than 0.05 was considered statistically significant. The CI reflects the range within which the OR value can be found 95% of the times the test is run on this data set, (Giancristofaro & Salmaso, 2003; Sullivan, n.d.) The precision with which the OR was estimated from the width of the CI and a ranges that did not cross "1" (a pair of positive numbers of a pair of negative numbers) was considered statistically significant (Field, 2013; Sullivan, n.d.).

Research Question 2:

Is the measure of accuracy higher for the 'glucose level' formula than that for 'diabetic status' formula in the FRS-CHD 10-year risk prediction model?

Null Hypothesis. There is no difference between the measure of accuracy for 'glucose level' and 'diabetic status' versions of the FRS-CHD 10-year risk prediction formula.

Alternative Hypothesis. The measure of accuracy for 'glucose level' formula is higher than that for 'diabetic status' version of the FRS-CHD 10-year risk prediction formula.

Statistical Analysis. This second question dealt with the relative performance of the two versions of the Multivariate Logistic regression (MLR) model. An MLR is used to describe a data set containing several predictor variables and one dichotomous outcome. Its ability to determine strength of association and to predict risk while

controlling for confounding effects makes it the analysis of choice for medical research (Stoltzfus, 2011). MLR analysis is common with longitudinal data, which is collected on the various risk factors believed to be associated with a disease or medical event. Here the independent or predictor variables can be categorical, or continuous, but the outcome must be binomial (Stoltzfus, 2011). For this research question, two MLRs were generated, one with FBG and another with DM2, and both included the other independent variables including- age, gender, HDL, SBP, smoking status, and BMI. The dependent variable was the presence or absence (yes/no) of a 10-year CHD event outcome.

For assessing the performance of screening models, there are several tests that can used, depending on whether the goal is prediction, selection, or causal modeling (Steyerberg et al., 2010). In this study the goodness-of-fit or modeling power was assessed using three parameters: the pseudo-R^2 values, the Hosmer-Lemeshow test, and the maximum likelihood function in the form of chi-square values. The pseudo-R^2 measures the amount of variance in the dependent variable explained by the independent variables and provides a measure of the model's discriminatory capability (Field, 2013). The maximum likelihood function estimates the how likely is the inclusion of the variable (or variables) in the model, relative to not including it, to make the correct outcome prediction (Hu, 2007). The Hosmer-Lemeshow (H-L) chi-squared test is another goodness-of-fit test that measures the model's calibration, giving an average of fit for the sub-groups within the data set when divided into deciles of predicted risk values (Demler, Paynter, & Cook, 2015; Stoltzfus, 2011). Unlike the chi-square values for the maximum likelihood, a lower H-L chi-square (and larger *p*-value) was considered a superior fit, as

the null hypothesis assumes that the model is a good fit of the data (Giancristofaro & Salmaso, 2003).

Assumption/Rationale. As a regression analysis, the MLR does not depends on the ordinary least square algorithms required in general linear models. A linear relationship between the independent variable (IV) and outcome is not required, neither is a normal distribution of the residuals (homoscedasticity), nor that the dependent variable should be an interval or ratio scale (Lani, 2015). The assumptions for the MLR instead include: independence of little or no multicollinearity among the variables and linearity between the variable log odds measure for the outcome (Stoltzfus, 2011). Additionally, the MLR should be generated from a large enough sample size, with a minimum of 10 cases for each IV, to avoid overfitting of the data (Sperandei, 2014; Stoltzfus, 2011). The maximum number of variables in the main multivariate formula was seven (when gender, and BMI were included), requiring a minimum of 70 cases, which was far exceeded by the sample size of over 2,600 after filtering. Pearson correlation coefficient (for continuous IVs) and the Chi-Square Cramer's V values (dichotomous IVs and continuous IVs as groupings) were used to analyze the strength of the association among the variables. The pairs that were deemed strongly correlated were used to create interaction terms, which were then tested for significance by including in the MLR formula.

Statistical Significance.

Determining the statistical significance of a logistic regression is relatively straightforward, as the model results are generated with a corresponding *p*-value and CI for each of the predictors in the equation. If the *p*-value is less than the predetermined α-

value of 0.05 then the term, and by extension the association between the predictor and the outcome, is considered significant (Giancristofaro & Salmaso, 2003). The size or strength of the association is based from the odds ratio, which is generated with a 95% CI. If the *p*-value is less than 0.05 and the OR is outside of one, statistical significance is accepted and the null hypothesis is rejected (Cook & Sheikh, 2000; Sullivan, n.d.).

Since this research question focuses on the model's performance or how well the model fits the data, it is important to evaluate the goodness-of-fit results. The goodness-of-fit test tells the extent to which the predicted probabilities deviate from that recorded in the data (Demler et al., 2015). For logistic regression, pseudo-R^2 can be used as a goodness-of-fit test, that reflects the model's predictive power, as a measure of variation accounted for by the variable included. There is no established threshold of significance for this measure, but when used to compare models, the greater the R^2 value the better the fit of the model (Steyerberg et al., 2010). The maximum likelihood measure does not test the model independently, but compares the fit of models generated from the same data set and each is reported with a corresponding *p*-value (Burnham & Anderson, 2004). For the H-L test, unlike in the logistic regression test on which it is based, the alternative hypothesis assumes a poor match between the observed and the predicted, and therefore a $\chi 2 < 20$ and ap > 0.05 is considered statistically significant (Demler et al., 2015)..

Research Question 3:

Is the measure of accuracy for the 'glucose level' formula independent of 'age', 'BMI', 'cholesterol level' or 'blood pressure level' in the CHD 10-year CHD prediction model?

Null Hypothesis. The measure of accuracy for 'glucose level' is independent of 'age', 'BMI', 'cholesterol level' or 'blood pressure level' in the 10-year CHD prediction.

Alternative Hypothesis. The measure of accuracy for 'glucose level' is dependent on 'age', 'BMI', 'cholesterol level' or 'blood pressure level' in the CHD 10-year CHD prediction.

Statistical Analysis. This third question evaluates extent to which the interaction between the independent variables moderates the association between 'glucose level' and the 10-year CHD outcome. The variables chosen for interaction testing were considered factors involved metabolism and are thus connected in the cardiometabolic hypothesis of CVD development. For the moderation analysis, the FBG model in addition to the respective interaction items (a product of the moderator and the predictor) (Field, 2013). Since the FBG itself was determined to have an insignificant relationship with the outcome, the FBG-group variable was created, and this was used to generate the interaction terms. Since the correlation coefficient between these variables were all below 0.350, there was no need for centering of the variables.

Statistical Significance. To determine the statistical significance of the interaction term, the properties of its coefficient was evaluated (Newsom, 2016). Like with the other ORs, a *p*-value that is less than α and a 95% CI that does not include "1" indicate statistical significance for the interaction terms. The statistical significance of the overall model, as well as the *p*-value and CI for each of the interaction and non-interaction variables of the model was evaluated for comparing the accuracy of the models- FBG with interaction terms, and FBG model without interaction terms

(Newsom, 2016). The relevant performance analyses will also be compared for the new 'interaction' model against that for FBG model without interaction terms to determine if the interaction terms provide any improvement to the model.

Summary

The goal of this study was to measure the effect of fasting blood glucose (FBG) levels on the performance of the Framingham Risk Score (FRS). The FRS was designed from the FHS data, from the first cohort of which originating several decades ago, and the more commonly used current version includes the major CHD risk factors; age, gender, HDL cholesterol, systolic blood pressure, smoking behavior and diabetic status. The 10-year FRS score is a measure of an individual's probability of succumbing to a CHD event over the next 10-years. Limits to the generalizability and discriminatory power of the FRS has led to customization and optimization of the formula over the past four decades, with goal of improving its performance and widening is applicability in a diversity of populations.

In this study, multivariate logistic regression (MLR) model was the statistical analysis of choice, as it was the method used to first develop the FRS model. Using the data from the third generation (most recent) of the FHS cohort as secondary data, univariate and multivariate regression analyses was performed to evaluate and compare the predictive power of the 'diabetic status' (DM2) model with that of the FBG model. The strength of the relationship between the respective variables and the outcome, the predictive power of the two models, and the effect of interactions (in the FBG model) between the independent variables were all assessed based on the *p*-values and the CI of

the odds ratio for each. The goodness-of-fit parameters were employed included the odds ratio or OR (assessed the unadjusted strength of the association), the pseudo-R^2 (how well the variance in the outcome is accounted for by the predictive factors), the Chi-square values (the extent to which the model predictions matches the recorded data), and the Hosmer-Lemeshow test (the average of the fit of model for each decile of the outcome variable). The next chapter will provide specific details of the steps involved in the statistical analysis performed to answer the three research questions and whether the null hypothesis was rejected or not.

Statistical Analyses

Research Question	**Statistical Test**	**Assumptions**	**Interpretation**	**Significance**
RQ1: Strength of Association *Glucose vs. Diabetes*	Univariate Logistic Regression (ULR)	Outcome matches test Correct labeling of outcome Large sample size	Estimated from OR $OR_{DM2} > OR_{FBG}$?	$p < \alpha$
RQ2: Model Comparison *Glucose model vs. Diabetes model*	Multivariate Logistic Regression (MLR)	Outcome matches test Correct labeling of outcome Only meaningful variables No/little collinearity Large sample size	$R^2_{DM2} > R^2_{FBG}$	$p < \alpha$
			DM2 χ^2 > FBG χ^2	
			Hosmer-Lemeshow Test	$p > \alpha$*
RQ3: Model Comparison *'Glucose' x 'Moderators' model*	Multiple Logistic Regression w/ Interaction	Same as MLR Minimal collinearity	Coefficient of interaction term *OR from* MLR_{G+Ms}	p-value $< \alpha$

Chapter 4: Results

Overview

The objective of this quantitative analysis was to assess the contribution of the variable FBG level compared to that of DM2 diagnosis as a predictor for (CHD in the FHS cohort dataset. The variables were compared unadjusted, adjusted and then possible interactions tested, in the three research questions examined. The research questions and the related null and alternative hypotheses are as follows:

RQ1: Is the relative risk for glucose level higher than that for diabetic status in the CHD 10-year risk prediction?

$H_0$1: There is no difference between the relative risk for glucose level and that for diabetic status' in the CHD 10-year risk prediction.

H_a1: The relative risk for glucose level is higher than that for diabetic status in the CHD 10-year risk prediction.

RQ2: Is the measure of accuracy for glucose level higher than that for diabetic status in the CHD 10-year risk prediction formula?

$H_0$2: There is no difference between the measure of accuracy for glucose level and diabetic status in the CHD 10-year risk prediction formula.

H_a2: The measure of accuracy for glucose level is higher than that for diabetic status in the CHD 10-year risk prediction formula.

RQ3: Is the measure of accuracy for glucose level independent of age, BMI, cholesterol level or blood pressure level in the CHD 10-year risk prediction formula?

$H_0$3: The measure of accuracy for glucose level is independent of age, BMI, cholesterol level or blood pressure level in the CHD 10-year risk prediction formula.

H_a3: The measure of accuracy for glucose level is dependent on age, BMI', cholesterol level or blood pressure level in the CHD 10-year risk prediction formula.

This chapter outlines the data collection (obtaining access to the secondary data) process, how the data was treated and analyzed, and the results of the statistical analyses. The data I analyzed is from the third generation of the original cohort (GenIII), combined with data from the OMNI2 and the NOS cohorts. The reporting of results begins with the descriptive analyses performed on the dependent variable as well as all the independent variables included in the study. This is followed by the results of each of the logistic regression analysis conducted to answer the specific research questions. Finally, the chapter summary highlights the overall findings as it relates to the objective of the study.

Data Collection

The data analysis for this study was conducted using secondary data obtained from the FHS, via the BioLINCC division of the NHBLI. The data set from the third cohort (FHSIII), as this was the most recent, were collected from the grandchildren of the participants of the original cohort (FHSI) beginning in 2001. The recruitment and response rates were described in detail in chapter three. Following IRB approval received from Walden University (Appendix B), I was granted access to the data from the BioLINC website according the signed agreement (Appendix C) for third party researchers. The data was provided as large zip file that was downloaded to my laptop

and opened to separate folders containing the data, the documentation, and the data dictionary. The data set folders contained two folders each, one with the data in CSV (Excel files) and the other with a SAS duplicate of all the files. The CSV files were opened in the SPSS application, subjected to several data screening and data cleaning steps before being used for the data analyses described below.

Preanalysis Data Cleaning

I combined several pertinent files in SPSS to create a new file that contained only the variables of interest. This involved duplicating the main file (CHD outcome, CHD date, etc.), deleting unwanted variables (CVD, CHF outcomes, etc.) and then merging into one file (based on the ID numbers) that included the predictor variables (age, sex, HDL, glucose, smoking status, BP, DM status and CHD outcome). Some of the variables were renamed (e.g. age @ exam 1 to simply AGE) for clarity, others required reformatted (e.g. CDH update status to 10-year update status) or used to create new variables (age groups, HDL groups, etc.). This file was then filtered twice, first to remove all those who had a history of CHD and then to remove those whose CHD event happened after more than 10 years. The filtered file was saved as the working file and used to perform the descriptive and inferential analyses.

Results of Data Analysis

Descriptive Analysis

Valid Number of Sample Size. After filtering, a total of 2,677 participants were included (from the original 4578 sample size) in the data set, with a valid *N* of

2,649 after accounting for the missing values from some (BP, FBG, HDL and DM) of the variables (See Table 1).

Table 1

Valid Number of Variables

	N	
	Valid	Missing
Sex of participant	2,677	0
Age (years)	2,677	0
Diastolic blood pressure	2,675	2
Fasting blood glucose (FBG)	2,652	25
HDL cholesterol	2,676	1
Diabetes Mellitus Status (DM2)	2,674	3
CHD status	2,677	0
Valid N (listwise)	2,649	

Measures of central tendency (Continuous Variables). The average age of the sample of participants was 40 years and 6 months, with the youngest being 19 and the oldest being 82 years old. As the histogram of the age distribution shows (Figure 4), there is a right shift, as there are fewer persons at the extreme end (elderly) than are at the left (<25 years). Despite this skewness, the trend line of frequency distribution indicates that the assumption of normalcy is met. Similarly, with the distribution of FBG levels, some very high values cause right skewness, but overall the distribution is still considered normal (Figure 4).

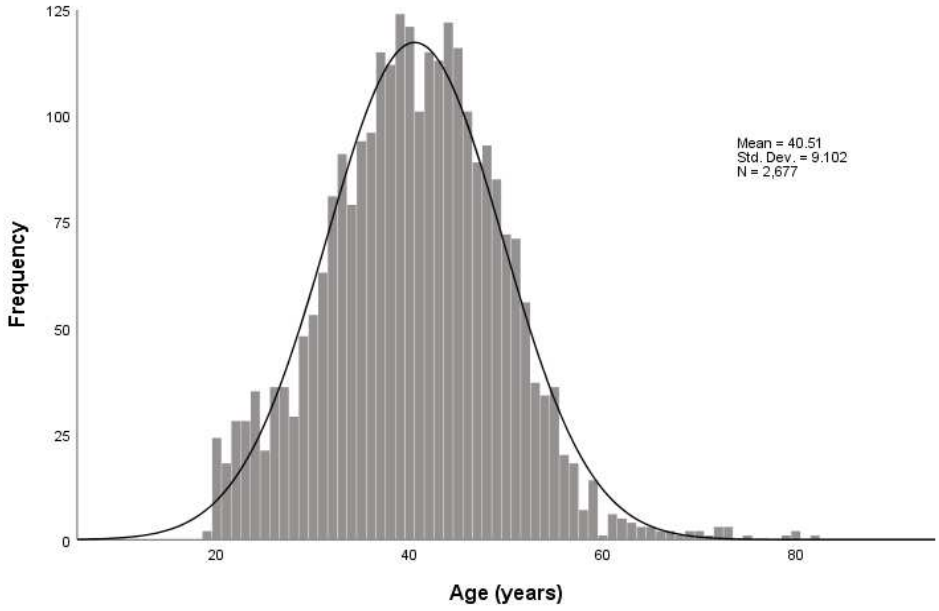


Figure 4. Histogram of AGE distribution.

The mean value of FBG was 94.43 but the median was only 92.00, confirming the presence of very high FBG values causing the shift in the mean to the right of the median (See Figure 5). The relatively large SE also suggests that the sample mean may be greater than the true population mean, but still within range. As can be seen in Table 2, the other continuous variables, BMI and SysBP, were also normally distributed, with respective means 26.7 and 116.8 (mmHg). The BMI was calculated based on weight and height measures, and like all other measurement, were based at the initiation exam (Exam 1) and are regarded as baseline measurements.

Table 2

Measures of Central Tendency of CHD Risk Factors

	Mean	S.E. Mean	Median	Std. Dev	Min	Max
Age (years)	40.51	.176	41.00	9.102	19	82
Systolic Blood Pressure	116.80	.278	115.00	14.403	81	192
Body mass index	26.684	.105	25.720	5.440	16	59
Fasting blood glucose	94.43	.322	92.00	16.588	64	357
HDL Cholesterol	54.87	.316	53.00	16.352	19	206

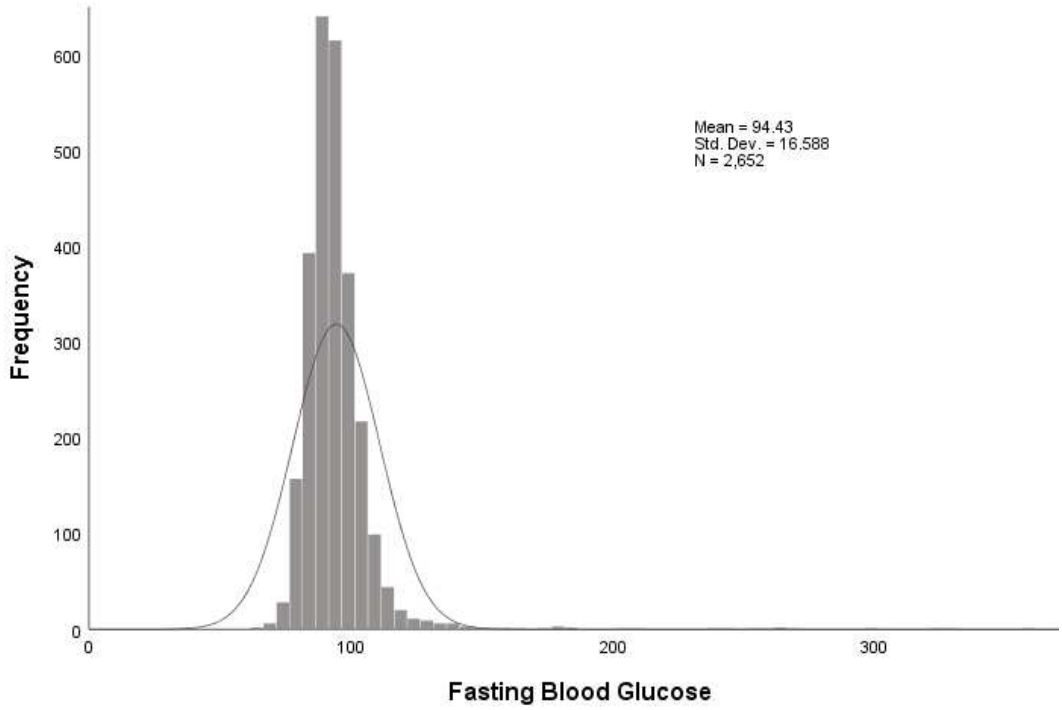


Figure 5. Histogram of fasting blood glucose (FBG) levels.

Distribution of CHD risk factors. In the case of the nonmodifiable risk factors, Age and Sex, there was an equal representation of the major groups. There were 271 more females than males, but the difference accounted for only a 10% difference (See

Table 3). And all the participants were adults (at least 19 years old), but the two largest groups were those between 26 and 40 (44.0%) and those between 41 and 55 (46.4%). Those younger than 25 and those older than 56 years made up a little less than 10% (combined) of the total sample. This distribution of age is consistent with the fact that most of the participants that make up this sample were from the third generation of the original FHS participants. The rest of the sample was made up from the second generation of the OMNI sample group and the new offspring group, both used to provide diversity.

Table 3

Distribution of Risk Factors

	Frequency	Percent	Cumulative Percent
Sex of participant			
Male	1204	45.0	45.0
Female	1473	55.0	100.0
Age Groups			
< 25	156	5.8	5.8
26 - 40	1178	44.0	49.8
41 - 55	1241	46.4	96.2
56 - 70	90	3.4	99.6
> 70	12	.4	100.0
HDL Chol Group			
< 41 (Risk)	504	18.8	18.8
41-60 (Boderline)	1301	48.6	67.4
> 60 (Optimal)	872	32.6	100.0
Systolic Blood Pressure Group			
<91 (Low)	23	.9	.9
91-120 (Normal)	1728	64.5	65.4
121-140 (Prehypertensive)	748	27.9	93.4
>140 (HyperTension)	178	6.6	100.0
Smoking Status			
Non-Smoker	2334	87.2	87.3
Smoker	341	12.7	100.0
FBG Groups			
<81 Hypo	143	5.3	5.3
81 - 100 Normal	2017	75.3	80.7
101 - 125 PreDM2	443	16.5	97.2
126 - 150 Boderline DM2	27	1.0	98.2
151 - 200 High DM2	9	.3	98.6
>200 Very High DM2	13	.5	99.1
Missing	25	.9	100.0
BMI Groups			
<18 Underweight	19	.7	.7
18 - 24.9 Normal	1117	41.7	42.4
25 - 39.9 Overweight	1443	53.9	96.3
>= 40 Morbid Obesity	98	3.7	100.0

Bivariate Analysis

Pearson correlation of risk factor. The correlation between the risk factors measured as continuous variables (age, HDL, BMI, FBG and SysBP) were statistically significant, with *p* values less than 0.000. The Pearson correlation values were all positive, except for HDL, as was expected, which was negatively correlated with BMI (-0.321), FBG (-0.200), and SysBP (-0.128). Since the strongest correlation was between BMI and SysBP (0.384) all the correlations were determined relatively weak or negligible**.**

Table 4

Pearson Correlation Between Variables

	Age	HDL cholesterol	Body mass index	Blood glucose	Systolic blood pressure
Age					
Pearson Correlation	1	.068	.161	.234	.323
p-value		.000	.000	.000	.000
HDL cholesterol					
Pearson Correlation	.068	1	-.352	-.200	-.128
p-value	.000		.000	.000	.000
Body mass index					
Pearson Correlation	.161	-.352	1	.341	.384
p-value	.000	.000		.000	.000
Blood glucose					
Pearson Correlation	.234	-.200	.341	1	.309
p-value	.000	.000	.000		.000
Systolic Blood Pressure					
Pearson Correlation	.323	-.128	.384	.309	1
p-value	.000	.000	.000	.000	

Association between risk factor groups across gender. When comparing the distribution of risk factors across genders, the differences in age group and smoking status were not statistically significant and the differences among the other variables were. The 17% (200) excess of females (1,204 males and 1,473 females) was reflected in all of the age groups. Similarly, despite there being more male smokers than female smokers, the difference in male (177) smokers compared female smokers (164) was not statistically significant ($p = 0.215$). Advanced age and smoking are the strongest, modifiable risk factors for CHD, and both they were both found to be equally distributed across the genders in this data set.

All the other risk factors demonstrated a statistically significant difference in association across the genders, with all p-values less than or equal to 0.001. However, for the effect size of the gender difference for diabetic status was considered weak (Cramer's V = 0.062), while the effect size for HDL, BP, and BMI and FBG were moderate in effect, with HDL being the largest (Cramer's V values of 0.431, 0.207, 0.312, 0.431, and 0.265 respectively). The effect sizes of these association did not warrant the centering of the variables prior to the interaction study, but it did justify investigating the interactions.

Table 5

Chi-square Analysis of Association

	Male	Female	χ^2	*p*-value	Cramer's V
Age			2.494	0.646	0.031
<25	67	89			
26- 40	545	633			
41 - 55	546	695			
56 - 70	39	51			
>70	7	5			
Diabetic Status			10.226	0.001	0.062
Non-Diabetic	1160	1447			
Diabetic	43	24			
Smoking Status			1.539	0.215	0.024
Non-Smoker	1039	1295			
Smoker	164	177			
HDL Chol			497.074	0.000	0.431
<41 (Risk)	396	108			
41 – 60 (Normal)	650	651			
>60 (Optimal)	158	714			
Sys Blood Press			114.982	0.000	0.207
<90 (Low)	2	21			
91-120 (Normal	660	1068			
121-140 (Prehypertensive)	446	302			
>40 (Hypertension)	96	82			

	Male	Female	χ^2	*p*-value	Cramer's V
BMI			260.973	0.000	0.312
<18 (Underweight)	2	17			
18 – 24.9 (Normal)	315	802			
25 – 39.9 (Overweight)	855	588			
>= 40 (Morbid Obesity)	32	66			
Blood Glucose			188.646	0.000	0.265
< Hypoglycemic	15	128			
81- 100 (Normal)	842	1175			
101 – 125 (PreDM2)	306	137			
126 – 150 (Borderline DM2)	19	8			
151 – 200 (High DM2)	5	4			
> 200 (Very High DM2)	6	7			

Distribution of Risk Factors Across Diabetic Status

The other set of bivariate analyses conducted, were the distribution of the risk factors across the diabetic status groups (DM2 and Non-DM2). The chi-square analyses (Table 6) indicated that there was a statistically significant difference between the two groups for all the risk factors, with the exception of Smoking status ($p = 0.201$). In the case of Age, HDL, SysBP, and BMI, even though the differences were statistically significant, the value of the Cramer's V were all below 0.200, with the largest effect seen across the age groups (Cramer's V=0.198). The percent of diabetics increased with age, with an average of 0.6% for those younger than 40, increasing to 3.7% between the ages 40 to 55, to 12.5% for those over 56, and to 57% for those over 70 years (See

Table 6). Similar trends occur in the risk factors that are also negatively predictive, such as SysBP and BMI, with the percentage of diabetics increasing as the value of the risk factor increases, e.g. from Normal BMI to Overweight BMI or from Normal SysBP to Hypertensive SysBP. The converse happens with HDL, which is the only factor that is positively predictive (of a reduced risk) of CHD. As the HDL value goes up from one group to the next, the percentage diabetics in the group goes down.

Table 6

Association of Risk Factor Groups Across Diabetic Status

	Diabetic	Non-Diabetic	χ^2	*p*-value	Cramer's V
Age			105.124	0.000	0.198
<25	1	155			
26- 40	7	1170			
41 - 55	45	1195			
56 - 70	10	80			
>70	4	7			
Smoking Status			1.636	0.201	0.025
Non-Smoker	55	2276			
Smoker	12	329			
HDL Chol			30.842	0.000	0.107
<41 (Risk)	30	474			
41 – 60 (Normal)	25	1274			
>60 (Optimal)	12	859			
Sys Blood Press			51.338	0.000	0.139
<90 (Low)	0	23			

	Diabetic	Non-Diabetic	χ^2	*p*-value	Cramer's V
91-120 (Normal	19	1708			
121-140 (PreHyperT)	33	713			
>140 (Hpertension)	15	163			
BMI			93.053	0.000	0.187
<18 (Underweight)	0	19			
18 – 24.9 (Normal)	8	1107			
25 – 39.9 (Overweight)	43	1399			
>= 40 (Morbid Obesity)	16	82			
Blood Glucose			1957.410	0.000	0.856
< Hypoglycemic	1	141			
81- 100 (Normal)	2	2015			
101 – 125 (PreDiabetic)	14	429			
126 – 150 (Borderline Diabetic)	27	0			
151 – 200 (High Diabetic	9	0			
> 200 (Very High Diabetic)	13	0			

Expectedly, the association between fasting blood glucose (FBG) group and DM2 was statistically significant with a very large effect size (Cramer's V=0.856). As can be seen from the Table 6, all those with very high *blood* glucose levels (above 126 mg/dL) belonged to the diabetic group. But surprisingly, all of the very high glucose levels were with the younger diabetics, while many the older diabetics have blood glucose levels closer to that of the Non-Diabetic (See Figure 6). Also, of note is the association between those who those who had a CHD event and those who didn't, relative to increasing age and blood glucose level. Most of the CHD events occurred

with those whose blood glucose levels were within the normal range. According to Figure 6, the majority of diabetics were 40 years and older, but as can be seen in Figure 7, those for whom a CHD event was recorded, the ages were both above and below 40 years, but mostly in the normal FBG region. These findings would have an impact on the inferential analysis and is discussed further in chapter 5.

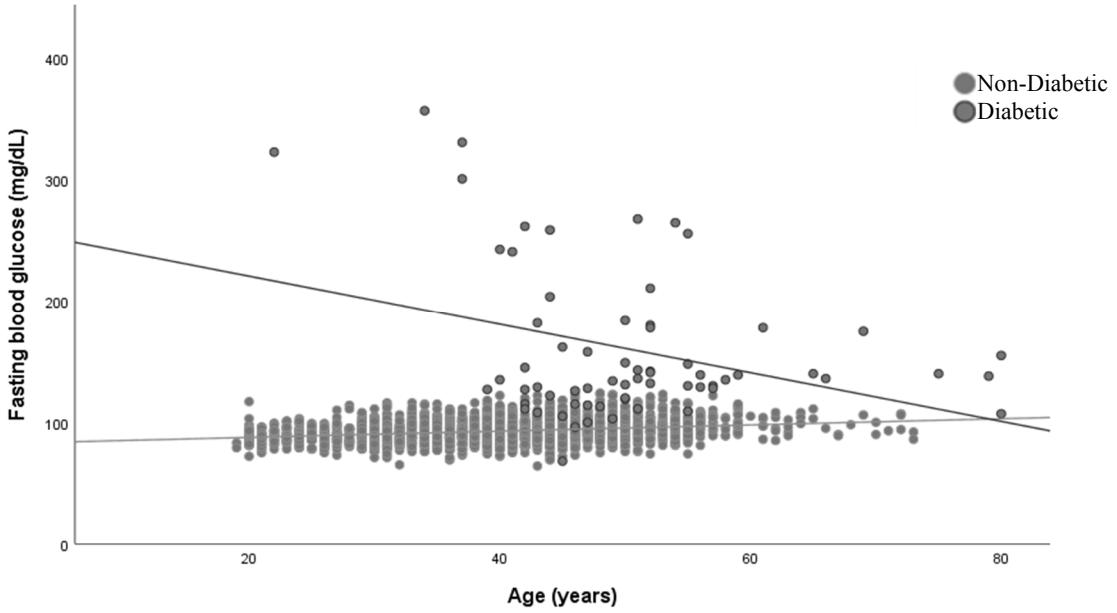


Figure 6. Scatterplot of fasting blood glucose (FBG) vs. AGE by diabetic status (DM2).

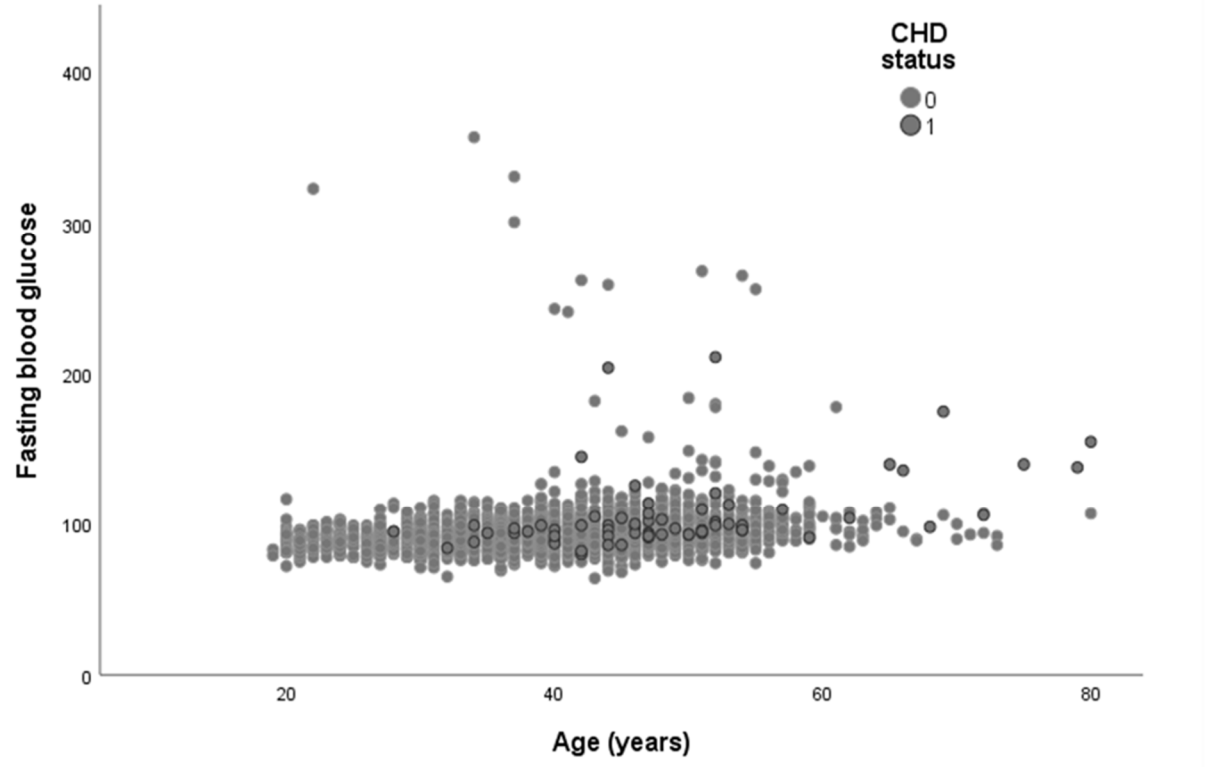


Figure 7. Scatterplot of fasting blood glucose (FBG) vs AGE by CHD status.

Research Question 1

Research Question 1: Is the relative risk of 'glucose level' higher than that of 'diabetic status' for the development of CHD in the FHS data set?

Null Hypothesis. The relative risk of 'glucose level' is less than that of 'diabetic status' for the development of CHD in the FHS data set.

Alternative Hypothesis. The relative risk of 'glucose level' is higher than that of 'diabetic status' for the development CHD in the FHS data set.

Two univariate binary logistic regression analyses were conducted to answer this first research question. The dichotomous outcome variable was the occurrence of a CHD event over the 10-year period. In this unadjusted model, Diabetic Status (DM2) was predictive of CHD as the model was statistically significant, with the $\chi^2 = 24.652$, $p = .000$ and the pseudo-$R^2 = 0.043$ (See Table 7). In a similar unadjusted model, fasting blood glucose (FBG) was also statistically significant, with the $\chi^2 = 14.649$, $p = .000$ and the pseudo-$R^2 = 0.025$. The larger pseudo-R^2 of the DM2 model would imply that it is a superior representation of the variance in the CHD outcome than the FBG model.

Table 7

Comparison of the Unadjusted Models of the Odds for CHD

	χ^2	Pseudo- R^2	*p*-value
DM2 Model	24.652	0.043	0.000
FBG Model	14.649	0.025	0.000

With respect to predicting CHD, the odds ratio (OR) for glucose level was 1.015, 95% CI [1.009, 1.022], suggesting that for each unit level increase in FBG the likelihood of CHD goes up by 1.5%. The OR for DM2 was 8.483, 95% CI [4229,

17.015], implying that on average, a diagnosis of diabetes increases the likelihood of CHD by 848%. According to the results from the two models, in the absence of adjusting for any other risk factor, DM2 status is shown to have a higher OR and is therefore a better predictor of CHD. The relative risk of an increasing FBG level, for one each unit change, is lower than that for change in DM2 Status, from non-diabetic to diabetic. The results of the analysis indicate that the null hypothesis cannot be rejected and that the relative predictive value of FBG is less (based on OR measures) than that of DM2 for the probability of CHD, in this FHS data set.

Table 8

Two Models of Odds of CHD Relative to Diabetic Status or Fasting Blood Glucose

							95% C.I.	
	B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
DM2 Model	2.138	.355	36.249	1	.000	8.483	4.229	17.015
FBG Model	.015	.003	21.060	1	.000	1.015	1.009	1.022

Research Question 2

Research Question 2: Is the measure of accuracy higher for the 'glucose level' formula than that for 'diabetic status' formula in the FRS-CHD 10-year risk prediction model?

Null Hypothesis. There is no difference between the measure of accuracy for 'glucose level' and 'diabetic status' versions of the FRS-CHD 10-year risk prediction formula.

Alternative Hypothesis. The measure of accuracy for 'glucose level' formula is higher than that for 'diabetic status' version of the FRS-CHD 10-year risk prediction formula.

For answering this second research question, a multivariate logistic regression was used. The variables SEX, AGE, HDL, SysBP, Smoking Status, and BMI were entered as the cofactors, but SEX proved to be a statistically insignificant predictor and was eliminated from subsequent formulations. The adjusted DM2 model (Table 9), was statistically significant, with $\chi^2 = 100.242$, $p = .000$ and a pseudo-$R^2 = .171$, as was the adjusted FBG model (Table 10), with $\chi^2 = 92.098$, $p = .000$, and the pseudo-$R^2 = .163$. The adjusted OR for DM2 (Table 9), dropped to 2.295 (from 8.483), 95% CI [1.035, 5.087] but it remained statistically significant ($p = .041$). However, in the case of FBG (Table 10), the OR decreased to 1.008 (from 1.015) and was no longer a statistically significant predictor, as $p = .079$.

Table 9

Odds of Diabetic Status for CHD Adjusted for Age, Smoking Status, HDL, and SysBP

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. Lower	95% C.I. Upper
Age	.089	.015	36.985	1	.000	1.093	1.062	1.125
Smoking Status	.921	.301	9.351	1	.002	2.512	1.392	4.533
HDL cholesterol	-.026	.009	8.506	1	.004	.974	.957	.991
Systolic blood	.021	.008	7.395	1	.007	1.022	1.006	1.037
DM2	**.831**	**.406**	**4.182**	**1**	**.041**	**2.295**	**1.035**	**5.087**
Constant	-9.029	1.088	68.808	1	.000	.000		

Table 10

Odds of FBG for CHD adjusted for Age, Smoking Status, HDL, and SysBP

	B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. Lower	95% C.I. Upper
Age	.093	.015	41.080	1	.000	1.098	1.067	1.129
Smoking Status	.884	.307	8.279	1	.004	2.422	1.326	4.423
HDL cholesterol	-.028	.009	8.973	1	.003	.972	.955	.990
Systolic blood	.018	.008	5.119	1	.024	1.018	1.002	1.035
FBG	**.008**	**.004**	**3.077**	**1**	**.079**	**1.008**	**.999**	**1.017**
Constant	-9.504	1.154	67.788	1	.000	.000		

To evaluate and compare the performance of the two models relative to each other, various performance parameters were used. In the case of the predictive power (the Log Likelihood and the Wald statistics), and for the discriminating power (pseudo R-squared) the values for the DM2 model were greater than these of FBG model (See Table 11) suggesting a superior DM2 performance. For the goodness-of-fit measure, as determined by the Hosmer-Lemeshow values, the FBG model had the lower chi-square and the higher *p*-value indicating the FBG superiority over the DM2. These seemingly contradiction in performance measures and the lack of statistical significance of the actual FBG variable in its model, were considered evidence that the FBG model was not an improvement to the DM2 model. Consequently, the null hypothesis was rejected as there was no significant difference between the measure of accuracy for ‘glucose level’ and ‘diabetic status’ versions of the FRS-CHD 10-year risk prediction formula.

Table 11

Comparison of Model's Performance

	DM2 Model		FBG Model	
	χ^2	*p*-value	χ^2	*p*-value
Overall Model				
Log Likelihood	547.800	.000	32.995	.000
Wald	891.077	.000	71.000	.000
Goodness-of-fit				
Hosmer-Lemeshow	8.115	.422	.524	.589
	Coefficient of Determination			
Pseudo-R-square				
Cox & Snell	.037		.034	
Nagelkerke	.171		.163	

Research Question 3:

Is the measure of accuracy for the 'glucose level' formula independent of 'BMI', 'HDL' or 'SysBP' in the CHD 10-year CHD prediction model?

Null Hypothesis. The measure of accuracy for 'glucose level' is independent of 'BMI', 'HDL' or 'SysBP' in the 10-year CHD prediction.

Alternative Hypothesis. The measure of accuracy for 'glucose level' is dependent on 'BMI', 'HDL' or 'SysBP in the CHD 10-year CHD prediction.

In answering this third research question, instead of using the fasting blood glucose level as a continuous variable it was used as an ordinal variable. The sample was divided into blood glucose level group (FBG group) variable, which was first run in an unadjusted logistic regression model and determined to be statistically significant.

When the FBG group variable was entered in the model, the model was also statistically significant, as $\chi^2 = 100.781$, $p = 0.000$ and the pseudo R-square = 0.172 (Table 12). All the variables, including the FBG groups were shown to be statistically predictive on a CHD event, with the FBG being the strongest predictor (0R = 1.360, p = 0.01, 95% CI [1.076, 1.720].

The introduction of interaction the terms, FBGxHDL, FBGxBMI, and FBGxSysBP (created from the metabolically linked factors), did produce a statistically significant overall model, with $\chi^2 = 102.407$, $p = 0.000$ and the pseudo R-square = 0.175 (Table 13). However, the only variables that remained statistically significant in this new 'interaction' model were Age, and Smoking status. All the other variables, including the interaction variables along with HDL and SysBP, were no longer statistically significant in this model.

Table 12

Odds of FBG Groups for CHD Adjusted for Age, Smoking Status, HDL, and SysBP

							95% C.I.	
	B	**S.E.**	**Wald**	**df**	**Sig.**	**Exp(B)**	**Lower**	**Upper**
Age (years)	.085	.014	35.940	1	.000	1.088	1.059	1.119
Smoking Status	.951	.300	10.016	1	.002	2.588	1.436	4.664
HDL cholesterol	-.025	.009	8.020	1	.005	.975	.958	.992
Sys blood pressure	.021	.008	7.025	1	.008	1.021	1.005	1.037
FBG Groups	.308	.120	6.610	1	.010	1.360	1.076	1.720
Constant	-9.514	1.083	77.211	1	.000	.000		

Table 13

Odds FBG Groups and Interaction Variables for CHD Adjusted for Age, Smoking Status, HDL, BMI, and SysBP

							95% C.I.	
	B	S.E.	Wald	df	Sig.	Exp(B)	Lower	Upper
Age (years)	.085	.014	35.802	1	.000	1.089	1.059	1.119
Smoking Status	.955	.301	10.074	1	.002	2.600	1.441	4.689
HDL cholesterol	-.036	.023	2.479	1	.115	.964	.921	1.009
Sys blood pressure	-.002	.020	.009	1	.925	.998	.959	1.039
FBG Groups	-.831	1.060	.614	1	.433	.436	.055	3.481
FBG_SysBP	.009	.007	1.494	1	.222	1.009	.995	1.023
FBG_HDL	.004	.008	.199	1	.655	1.004	.988	1.020
FBG_BMI	-.004	.008	.326	1	.568	.996	.981	1.011
Constant	-6.157	2.921	4.441	1	.035	.002		

The comparison of the measures of performance statistics of the two FBG Group model (with and without the interaction variables) demonstrated very little difference. As shown in Table 14, the Log Likelihood value did go down from 547.368 to 545.376 with the addition of the interaction terms (representing a decrease in performance), the Hosmer-Lemeshow value also went up from 9.319 to 10.881 (representing a slight increase in performance). Collectively, the very small changes in these values and the contradiction from one performance maker to another renders the results insignificant. Together with the results of the logistic regression model, it can be concluded that there is no significant interaction between FBG levels (when taken as

groups) and the metabolic terms of HDL, BMI and SysBP. The null hypothesis cannot be rejected, and it is assumed that the measure of accuracy of the prediction of the CHD, in the FRS-CHD 10-year risk prediction formula, by FBG level groups is not influenced the interaction terms.

Table 14

Comparing the Performance of the FBG Group Models

	Non-Interaction Model		Interaction Model	
	χ^2	*p*-value	χ^2	*p*-value
Overall Model				
Log Likelihood	547.368	0.000	545.376	0.000
Wald	891.474	0.000	891.276	0.000
Goodness-of-fit				
Hosmer-Lemeshow	9.319	0.316	10.881	0.209
	Coefficient of Determination			
Pseudo R-square				
Cox & Snell	0.037		0.038	
Nagelkerke	0.172		0.175	

Summary

For this research, the outcome variable was the binary measure of the presence or absence of a CHD event within a 10-year period. The dichotomous nature of the outcome warranted the use of a binary logistic regression analysis to answer the research question. For the first question, where the relative risk of a diagnosis of Diabetes was compared with the relative risk of a measure of *Fasting* Blood glucose level, a univariate binary logistic regression test was used. The results indicated that

both variables were statistically significant predictors, but the OR for Diabetes was 8.483 and that for glucose was 1.015. This suggests that the odds of a CHD event for a Diabetic is approximately 8.5 times that of the odds for a non-diabetic, while the odds of a CHD event go up by 0.015 for each unit increase in blood glucose level. While the glucose would appear to give a more precise measure of risk, the results of the χ^2-value (24.652 vs 14.649) and the pseudo R-square (0.043 vs 0.025) would indicate that a diagnosis of diabetes (in the absence of other risk factors) is a superior predictor.

For the second question, the other risk factors (Age, Sex, SysBP, HDL and Smoking Status) were included in the model and each of the variables (glucose/FBG and diabetes/DM2) tested separately. Here both models were statistically significant, but the OR for diabetes dropped considerably (OR = 2.295, 95% CI [1.035, 5.087] with adjustment, and glucose level was rendered statistically insignificant in the formula. The performance measures also confirmed the superiority of the Diabetes model and the null hypothesis was not rejected as 'Diabetic' FRS-CHD 10-year risk prediction formula as there was no difference between the measure of accuracy for 'glucose level' and 'diabetic status' versions of the FRS-CHD 10-year risk prediction formula.

The third research question investigate the impact of any possible interaction between glucose and the metabolic risk factors (HDL, BMI and SysBP). For this analysis, the glucose level was transformed into an ordinal variable of groups of glucose levels. This variable was statistically significant predictive of CHD, with a OR=1.36, $p = 0.010$ and 95% CI [1.076, 1.720]. The addition of the interaction terms did not have any significant effect on the performance, but with the exception of Age

and Smoking Status, all the other risk factors, along with the interaction term, were not statistically significant for the outcome (p-values > 0.05). Based on these results the null hypothesis was not rejected, and the interaction terms were considered to not influence the relationship between glucose (when used as FBG levels group) and the CHD event outcome in the FRS-CHD 10-year risk prediction formula.

Chapter 5: Discussion, Recommendations

Introduction

Purpose of the Study

CVDs continue to be the leading cause of death in developed and developing countries around the world . Despite the efforts in the field of medicine and public health, these diseases continue to threaten the lives of many, and causing serious debilitation in others, while costing several billions of dollars in health care cost annually . Screening for CVDs has proven to be an effective means of curbing the spread and slowing the continued growth among the many affected communities around the world. Identifying those at risk and exposing them to health care intervention methods has reduced the incidence as well as the devastating effect of all the major CVDs.

The FRS is still the most common screening tool used by health care providers to identify those at risk and allocate resources to help prevent CVD or minimize its effects on affected individuals . The risk factors included in the FRS have proven to be very strong predictors and their combination in the multivariate FRS makes this model most effective. The FRS was developed over 50 years ago and its usefulness effectiveness notwithstanding, it has undergone optimization and customization that has made it more specific the various people groups for whom it is used. The optimization methods have focused on either the inclusion of additional risk factors or altering the measuring of current risk factors based on novel information from research in the fields of biochemistry and cardiology .

This present study deals with the optimization of the FRS formula based on the use of FBG levels as a risk factor in the place of a diagnosis of DM2. Other derivations of the FRS that were developed in for other countries, with different populations, have seen an increase in the sensitivity of the screening formula when similar changes were made (see citation). The objective of this study was to carry out this substitution and compare the models using various performance evaluation methods. Additionally, the interaction between glucose and the metabolically significant risk factors were assessed for the FBG model. Pairs of the univariate-, multivariate-, and the interaction-models were all generated for the both FBG and the DM2, and the respective odds ratios, the pseudo-R^2 values, and the model-fit parameters (Log Likelihood, Wald, and Hosmer-Lemeshow statistics) were compared for each pair.

The data set used for this study was a combination of the Gen III (third generation of the Framingham cohort), together with the OMNI II and the NOS cohort. This is prospective data, collected based on a series of exams over many years, beginning in 2001 and continuing today . Before beginning the analysis, the data was streamlined to remove those with a history of heart disease, leaving only those whose heart disease developed within 10-years period following the original exam when the baseline data was collected. The risk factors glucose level and DM2 diagnosis along with smoking status (dichotomous), HDL-cholesterol (continuous), systolic blood pressure (continuous), gender (nominal) and age (continuous) were included. The outcome variable was the dichotomous measure of coronary heart disease (CHD =1, no CHD = 0) and a series of logistic regression analyses was used to carry out the

inferential statistics to answer the research questions. Before conducting the inferential statistics, descriptive statistics on the predictor variables was carried out, the results of which are presented in Chapter 4 of this dissertation.

Results and Findings

RQ1: Is the relative risk for glucose level higher than that for diabetic status in the CHD 10-year risk prediction?

$H_0$1: There is no difference between the relative risk for glucose level and that for diabetic status in the CHD 10-year risk prediction.

H_a1: The relative risk for glucose level is higher than that for diabetic status in the CHD 10-year risk prediction.

RQ2: Is the measure of accuracy for glucose level higher than that for diabetic status in the CHD 10-year risk prediction formula?

$H_0$2: There is no difference between the measure of accuracy for glucose level and diabetic status in the CHD 10-year risk prediction formula.

H_a2: The measure of accuracy for glucose level is higher than that for diabetic status in the CHD 10-year risk prediction formula.

RQ3: Is the measure of accuracy for glucose level independent of age, BMI, cholesterol level, or blood pressure level in the CHD 10-year risk prediction formula?

$H_0$3: The measure of accuracy for glucose level is independent of age, BMI, cholesterol level, or blood pressure level in the CHD 10-year risk prediction formula.

H_a3: The measure of accuracy for glucose level is dependent on age, BMI, cholesterol level, or blood pressure level in the CHD 10-year risk prediction formula?

Descriptive Statistics: How This Data Set Compares with Others

Comparing Nonmodifiable Risk Factors. The descriptive analysis conducted on this data set revealed that the distribution of the gender reflects that of most other data sets used for this type of study, as well as that of the distribution in the general population (Center for Health Statistics, 2017; D'Agostino, Grundy, Sullivan, & Wilson, 2001b; Jahangiry et al., 2017). In the current study, there was almost an equal balance, with slightly more female (55%) than males, similar to the distribution in the original FHS cohort, and in the subsequent (FHSII) cohort (Lloyd-Jones et al., 2002; Tsao & Vasan, 2015b). This distribution is also reflective of the distribution of the sexes in the general population according to the U. S. Census Bureau (U. S. Department of Commerce, n.d.).

Unlike the distribution of the genders, the racial distribution in this data set does not reflect that of the general population but is considered like the original FHS cohort (mostly White), given that most of the subjects were descendants of the original FHS participants. This current data set is made up of three groups, the FHS III, the OMNI II, and the NOS cohorts. The FHS III data accounts for 4,578 adults, who are the grandchildren of the FHS I, along with an additional 101 of their non-FHS parents who had not been included in any of the previous FHS cohorts (Splanksy, 2007). The OMNI II are the children of the original OMNI cohort, introduced to include a more racially

and ethnically diverse sample set, but they numbered only 405 (Tsao & Vasan, 2015a). The NOS cohort also added to increase genetic diversity, was made up of the spouses of the OMNI2 and consisted of only 101 subjects (Govindaraju et al., 2008). Unfortunately, the ethnically diverse participants made up no more 10% of the total sample size, in its original iteration. However, that percentage was reduced with each subsequent generation, and race data was only collected for the OMIN2 participants, so race was not included in any of these analyses.

The age values followed a normal distribution of adults (over 19 years), with an average of the sample being slightly above 40 years, and 93.8% of the subject between 26 and 55 years old. In the original cohort, the age range (30 to 62) was narrower that the current 20 to 92 and this could have impacted the results of the data analysis. Age is not only a major CHD risk factor; it influences the prevalence of other risk factors as well and thus the overall reliability of the model. Other risk model optimization studies used data sets with older subjects and shorter age ranges, e.g. 45 to 64 (Conroy et al., 2003), or 35 to 74 (Hippisley-Cox et al., 2007), or 30 to 74 (D'Agostino et al., 2001b). A 10-year CHD prediction formula is less accurate for younger adults than it is for older adults and this is one of the reasons why formula optimization is preferred with a age distribution with smaller standard deviation (Booth et al., 2006).

Comparing Modifiable Risk Factors. The results of the prevalence evaluation of the various risk factors among the participants in this data set revealed that some of the prevalences were similar to the national prevalence but most were lower. For example, the prevalence of lower HDL (CHD protectant) values was 18.8%, like the

national average of 17.1% (Zwald, 2017), but the prevalence of hypertension (>140 m) was 6.6%, compared to a 29% national average (Paulose-Ram, 2017). In this data set, only 1.3% of the subject were reported as diabetic as baseline, which is much lower than the national average of 9.4% (or even 7.4% for Whites) (Benjamin et al., 2017). And while it is not known how many patients developed DM2 during the 10-year period, those who were classified as prediabetic made up 16.5%, almost half of what is the average percentage nationally- 33.9% (Benjamin et al., 2017). The prevalence of overweight/obesity is also a lot less 57.6% compared to the 65.7% reported by the CDC at the time of the initial exam (Health E Stats, 2003-4). Based on these analyses, the current data set was made up of subjects with lower prevalence of the major risk factors.

The descriptive analyses revealed that the people making up had a lower risk of heart disease than the general population average and this was reflected in the CHD outcome. After 10 years, only 2.6% had a CHD event, 4.7% CVD in general, which is much less than in previous decades, 18.09% for men and 10.08% women for the (D'Agostino et al., 2008), or the 15.9% men and 6.9% women for CVD for the two previous FHS cohorts (Ford, 2014). The rate of CHD across the U.S. population has gone from 10.3% (2001-2), down to 8.0% (2011-12) (Yoon, 2016), and currently is at a national average of 3.9%, but ranges from WV's 7.4% to HI's 2.5% (CDC BRFSS, 2017). This low CHD prevalence in this data set countered the effectiveness of the logistic regression modelling and evaluation, which works best with not so rare

outcomes (Giancristofaro & Salmaso, 2003; Stoltzfus, 2011) and made the formula development more challenging.

Bivariate Analysis: Gender Differences

Historically, the distribution of heart disease and its risk factors have always differed across the genders (Amsterdam, 2011; Hosseini, 2015; Tunstall-Pedoe, Woodward, Tavendale, Brook, & Mccluskey, n.d.). In the present study also, there were more almost twice as many male diabetics (n=43) as female diabetics (n=23) contrasted with the more than four times as many women (n=714) than men (n=158) with optimal HDL measures. These findings were also congruent with others, including that the fact that the overweight/obese men (887 or 73.7%) outnumbered the overweight/obese women (654 or 44.4%). The association between gender and the various groups of systolic blood pressure was also significant, (p = .000) with 45% of men being hypertensive compared to 22% of the women. These findings are consistent with previous research where men had higher prevalence of CVD risk factors (D'Agostino et al., 2013, 2008; Fawwad et al., 2016). A related study that used electronic health records (EHR) to compare the performance of three CVD risk score functions (FRS, ASCVD and QRISK), found that 82% of the men were overweight or obese compared to 63% of the women; 14% of men were diabetic compared with 9% of women; and ultimately 8% of the men and 4% of the women had a CVD event during the 10-year follow-up (Pike et al., 2016). These gender differences in the current study were ultimately reflected in the CHD rate outcome as 3.25% of men and 1.31% of women had a 10-year CHD event. Interestingly in this data set, when the covariates

were adjusted for in the formula, gender was not a statistically significant predictor. This lack of effectiveness of gender in the CHD outcome has been reported in previous research (Jahangiry et al., 2017; Kozakova et al., 2017).

Bivariate Analysis: Diabetics versus Nondiabetics

The bivariate analysis of this data set confirmed the association between DM2 and the other cardiometabolic risk factors. Increasing BMI, Sys BP, and age, and decreasing HDL were all associated with an increased probability of a DM2 diagnosis (see Table 6). The older the subjects the higher the percentage of diabetes, going from 3.76% of those 41 - 55, to 12.5% for 56 - 70, and to 57.2% of those over 70 years old. Similarly, only 1.1% of the nonhypertensive (Sys BP < 120) were diabetic, but 9.2% of the hypertensive were diabetic; only 0.7% of the normal/underweight (BMI < 25) were diabetic but 3.98% of the overweight or obese were. Only 1.3% of those with optimal HDL levels were diabetic, but of those with low (not-optimal) HDL, 3.15% were diabetic. These results are similar to those found in other DM2-CVD risk factor correlation studies (Bragg et al., 2014; Haregu et al., 2016; Qazi & Malik, 2013a; Selvin et al., 2010). The Chi-square analysis of all these association, proved they were all statistically significant (p = .000), and based on the Cramer's V, relatively strong, measuring 0.107, 0.139 and 0.187 for HDL, SBP and BMI respectively. Despite these values, however, in the multivariate formula, BMI was not predictive, and the interaction variables for each of HDL and SBP were not statistically significant.

Smoking is traditionally the strongest CVD risk factor (CDC, 2014; Mannan et al., 2010), yet it had no significant association with the second strongest risk factor-

DM2 ($p = 0.201$). In this analysis, smokers were no more likely to be diagnosed as diabetic than nonsmokers. This is probably an indication that though both smoking and DM2 increases the risk of developing CVD, the mechanism by which these two factors cause damage to the cardiovascular system are different. Smoking introduces toxins and external sources of oxidative stress that ultimately damages the endothelial and this precipitates atherosclerosis (CDC, 2014; Huxley et al., 2011). On the other hand, as a disorder of metabolism, DM2-driven CVD is mediated by the internally produced biochemicals, namely excessive insulin, that eventually generates ROS and the associated tissue damage then leads to atherosclerosis and (Faerch et al., 2012; Kumar et al., 2019; Laakso, 2015). These differences may be a possible explanation for absence of association between the variables at baseline.

The etiology of CVD as a cardiometabolic disease that has been postulated in many prospective epidemiological studies, though not clearly defined, is increasing being supported by the clinical and genetic research (Holmes, Pulit, & Lindgren, 2017; Nichols et al., 2012; Wilson & Meigs, 2008). One such study is the comprehensive meta review of over 100,00 cases that used gene analysis to research the causal role of central/general obesity in several CVD outcomes (Dale et al., 2017. The results suggest that patients with a genetic disposition to metabolic disorder, of which central or abdominal obesity is symptomatic, are more susceptible to both DM2 and CVD (Dale et al., 2017). Another genetic study of over 26,000 Finnish subjects verified the causal role of metabolic disruption and rising insulin as the source of heart damage and that hyperinsulinemia was absent in the obese patients who did not develop CVD (Tikkanen

et al., 2016). Another study, conducted with 1.3 million obese or overweight (BMI > 25), nondiabetic adults on the distribution of four metabolic risk factors -elevated blood pressure, low HDL, elevated triglycerides and prediabetes- found the number and prevalence of the CVD increased significantly with each age group indicating a progression of the risk factors culminating in CVD (Assmann et al., 2008). This may also explain, the correlation between BMI in a univariate model, but no statistical significance in the multivariate model, when adjustments are made for other more strongly predictive risk factors.

Answer to Research Questions

RQ1: Is the relative risk for glucose level higher than that for diabetic status in the CHD 10-year risk prediction?

$H_0$1: There is no difference between the relative risk for glucose level and that for diabetic status in the CHD 10-year risk prediction.

H_a1: The relative risk for glucose level is higher than that for diabetic status in the CHD 10-year risk prediction.

Comparison of Unadjusted Odds Ratio. The results of the univariate analyses do indicate that a DM2 diagnosis is stronger predictor of CHD risk than is a measure of fasting glucose level. However, it is difficult to compare these results, as one variable is dichotomous, and the other is continuous. The odds ratio indicate that a diabetic person has an almost 850% (*OR* = 8.483) increased risk of CHD compared to a non-diabetic. On the other hand, every unit increase in FBG confers an increased risk of 1.5% (*OR* = 1.015) of CHD, but it also means that a 10 unit increase in FBG corresponds to a 16%

(*OR* = 1.161), and a 20-unit FBG increase confers a 35% (*OR* = 1.347) increased CHD risk. Despite the large discrepancy in the odds for DM2 and FBG, they were both predictive, but the difference in odds is mainly due to the way the variables are measured as both are statistically significant (p-value < 0.001).

This minimal effect in the predictive value of FBG may be further explained by a number of reasons, including (a) the formula used for evaluating; (b) the cohort used to generate the formula; and (c) the method of comparison. Firstly, previous studies that have evaluated the performance of glucose included models that were based on different (optimized) variations of the FRS than the one used in this study (Collins & Altman, 2009; Rücker et al., 2016), or other studies used hazard ratio (HR) measures instead of *OR* for comparison of the variables (Demler et al., 2015; Kadowaki et al., 2008; Sarwar et al., 2010), or still others compared classification categories for CHD risk outcomes and not the goodness of fit parameters (Conroy et al., 2003; Garg et al., 2017). Secondly, the Framingham cohort is known to be made up of predominantly White subjects and as have been confirmed by many studies race dose play a role risk of heart disease (see Nichols et al., 2012; Pike et al., 2016; Sarwar et al., 2010; Singhal, Tien, & Hsia, 2016). Two such study in particular found that glucose level was a strong predictor of CVD risk than blood pressure and cholesterol values, for African Americans (Carter et al., 2016). These factors may have impacted the performance of the model in this study and thus render the model optimization futile.

Other explanations for the apparent underperformance of the glucose level as a CHD risk predictor may lie in the unique nature of the relationship between the

variables. The coefficient and the corresponding odds ratio for each independent variable approximates the relationship. One of the assumptions of the logistic regression is that the predictor has a linear relationship with the log of the outcome, else the strength and significance of the relationship is underestimated (Lani, 2015; Sperandei, 2014). For example, a few studies have reported that the (log odds) relationship between glucose level and heart disease risk follows a J-shaped curve, representing that the glucose level associated with a reduced risk is only a narrow range, and an increased CVD risk for levels below and above that range (Mongraw-Chaffin et al., 2019; Park et al., 2013; Selvin et al., 2010). In the present study, when the FBG was turned into a categorical variable (FBG groups) the odds ratio varied considerably among the groups, but the relationship was not linear. Other studies have reported that the strength of the relationship between glucose and CHD decrease with the age of the subjects, as other CHD risk factors become more prominent (Booth et al., 2006; Conroy et al., 2003). This combined with the fact that CHD prevalence in the population increases with age, are the main reasons why the hazard ratio is the preferred method of risk evaluation. With the current study, the only age range within which glucose remained statistically significant in the model was between 56 and 70 years. Using ORs as was done in here, may complicate and even preclude the relationship necessary for formula optimization.

RQ2: Is the measure of accuracy for ‘glucose level’ higher than that for ‘diabetic status’ in the CHD 10-year risk prediction formula?

H_{02}: There is no difference between the measure of accuracy for 'glucose level' and 'diabetic status' in the CHD 10-year risk prediction formula.

H_{A2}: The measure of accuracy for 'glucose level' is higher than that for 'diabetic status' in the CHD 10-year risk prediction formula.

Comparison of Adjusted Odds Ratios

Although both factors (glucose level and diabetic status) were significant in their respective univariate models, only DM2 remained statistically significant after adjusting for the other cofactors. In comparing the two models, the various parameters indicated that the both models performed similarly, with the glucose model being slightly weaker model than that for DM2 model. Looking at the chi-square (measure of the association between the observed and the expected probabilities) values for the multivariate models, shows that the DM2 model is slightly superior fit to the overall data (χ^2: 547.800_{DM2} vs 532.995_{FBG}); and similarly the pseudo R-square values indicating that the DM2 model has marginally better predictive power (0.171_{DM2} vs 0.163_{FBG}). Based on these two parameters, there is no advantage gained by using blood glucose level instead of diabetic status in the FRS model when screening for CHD risk in asymptomatic clients.

The Hosmer-Lemeshow (H-L) test tells a similar story about the similarity of the models, with one interesting caveat. It should be noted that with the H-L test, which gives an average of the chi-square for deciles of the outcome probability, unlike the Omnibus test, a lower chi-square and higher *p*-values are preferred (Giancristofaro & Salmaso, 2003). In the H-L analysis of the two models, the FBG model was slightly

superior to the diabetic status model, as indicated by the lower chi-square values (H-L: 6.524_{FBG} vs 8.115_{DM2}) and the higher *p*-values (0.589_{FBG} vs 0.422_{DM2}). This points to the superiority of the glucose model as a better fit (predicted versus observed) for the data when divided into sections (Moons et al., 2012.; Peng, Lee, & Ingersoll, 2002). The H-L results lend further support to the recommendation that model performance evaluation is best estimated only after dividing the data into various strata.

RQ3: Is the measure of accuracy for 'glucose level' independent of 'age', 'BMI', 'cholesterol level' or 'blood pressure level' in the CHD 10-year risk prediction formula?

H_{03}: The measure of accuracy for 'glucose level' is independent of 'age', 'BMI', 'cholesterol level' or 'blood pressure level' in the CHD 10-year risk prediction formula.

H_{A3}: The measure of accuracy for 'glucose level' is dependent on 'age', 'BMI', 'cholesterol level' or 'blood pressure level' in the CHD 10-year risk prediction formula.

Assessing the Interaction Variables

Testing the interaction variables in the given formula proved futile, as the FBG variable itself was not statistically significant in the multivariate model. Consequently, different interaction terms were created with the 'FBG-groups' variable, which was statistically significant, but again the interaction variables were not statistically significant. Though only the age and smoking status variables remained significant in this new interaction-included model, the addition of the interaction terms had very little effect on the model's performance. This is probably because, even in its most effective

version, the model only accounted for 17.2% of the variability in the outcome, and the low prevalence of the disease conferred an already high accuracy (~97.4%) to the null (no variable) model leaving very little room for improvement. The area under curve (AUC) values, which assesses the classification power of the models, were 0.816 for DM2 and 0.813 for glucose; with the DM2 model have a slightly higher specificity and sensitivity, at the 20%-CHD risk cut off point, 14.3% vs 11.7% and 1.0% vs. 0.9% respectively. The UC values are similar to those found in previous studies, higher than some, 0.69 – 0.71 (Pike et al., 2016), 0.809-0.834 (Pandya, Weinstein, & Gaziano, 2011) and lower than others, 0.811 -0.819 (Günaydın et al., 2016). However, based on the confidence intervals for the present AUC values (overlapping CIs), the differences between the two models were not considered significant.

Limitations of the Findings

This study presented insight into the performance of the FRS formula, but there are some specific limitations to the generalizability of the findings. Firstly, the main limitation of these findings relates to the uniqueness of all the FHS cohorts, which is specific to a particular population, namely White American of European descent, connected to the Framingham, MA region. As other studies have proven, FRS findings are sometimes not applicable to the other population that include other ethnicities, and require recalibration prior to application (Elosua, 2014; Hemann et al., 2007; Rücker et al., 2016). Furthermore, this FRS III cohort is the most recent version and was chosen for that reason, but it should be noted that there were differences between this and the FRS I from which the formula was initially developed. Most importantly, the diabetes

diagnosis has changed since the 1950s, and was mentioned before, many more diabetes were previously diagnosed and receiving treatment than was the case with the participants of the FRS I.

Another important limitation relates to the FRS formula used here was restricted to very specific risk factors, namely Age, Sex, Sys BP, HDL, DM2 and FBG and did not include others. This model did not account for factors, such as family history of heart disease, presence of other diseases (chronic kidney or liver disease or inflammatory disease), level of physical activity, etc., all of which have been proven to impact CHD risk levels (Kones, 2011; Ray et al., 2014; Sarwar et al., 2010; Tsao, 2015; Tsao & Vasan, 2015b; Wilson et al., 1998; Woodward et al., 2007). The presence of these attenuating factors for some participants, but not included in the analysis, may have weakened the measure of the association between the outcome and the risk factors being assessed.

A most important issue in this model is that it did not include relevant information on patient medication, especially glucose lowering medications for the diabetics, including sulfonylureas, biguanides, alpha glucosides inhibitors or exogenous insulin. Like with other medications, this was not included for the sake parsimony of the model, but it should be noted that doing so limits the model's accuracy. In this study, almost 25% of the diabetics had FBG levels below the threshold (see Table 6), which could be the results of medication, or lifestyle, or both. The lowered FBG for these diabetics would have a major effect on the relationship between the measured

FBG levels and the CHD outcome, thereby reducing the predictive power of dysglycemia on a CHD event, reducing the reliability of the results.

There are several other limitations related to all disease risk score studies in general that are also worth mentioning. Any disease risk score, based on cross-sectional data, can only be considered as risk estimation and cannot be used for causality determination (Dahlöf, 2010; Nieto, 1999b). Also, all risk formula should be carefully validated by applying and testing them to a data set other than the one used to develop it. Relying of statistical significance or internal calibration is not enough, for this type of one-time measurement of risk factors that are known to fluctuate and interact with each other in unpredictable ways (Collins & Altman, 2009). Finally, an issue that is common to risk estimation studies, is the imprecision and low sensitivity of the formulas generated. These models tend to have very low sensitivity and specificity values and should only be used for general classification of asymptomatic patients. This is advised to reduce the reported cases of underestimation of high-risk patients going undetected and therefore untreated (Collins & Altman, 2012; Rücker et al., 2016; Steyerberg et al., 2010; Tsao & Vasan, 2015b). The CVD risk prediction models have been very effective in predicting risk probability and stratifying individuals into risk groups, but because of their many limitations then cannot be used in place of a thorough medical examination.

Recommendations for Future Research

The unexpected findings in this study point to the need for additional research before it can be utilized for making clinical and resource allocation decisions related to

heart health. The next step in this type of study would be to conduct similar performance estimation with a cohort from a different, more diverse population. For example, the results may be different in African Americans populations, given that glucose level have be shown to have on the strong predictive association with CVD risk and that there is a higher the prevalence of CHD as well (Carter et al., 2016; Clark et al., 2015; Marshall, 2005). The prevalence of the different types of CVD as well as the prevalence of the risk factors vary among the different population types and these prevalence difference have been shown to impact the performance of the risk models (Al-Nooh, Abdulabbas Abdulla Alajmi, & Wood, 2014; Nichols et al., 2012; Yosaputra et al., 2010).

Another area of future research is the model evaluation generated with different stratifications of the same data set. There are several ways that the data could be stratified, including by gender, age groups, diabetic status, smoking status, etc. Other risk assessment models have been known to have variation in their performance for different strata of the same population (Mendis, 2010; Steyerberg et al., 2010). In this data set, as in others reported, the prevalence of CHD is twice as high as it is for women, consequently some forms of the risk models are sex-specific (Amsterdam, 2011; Jahangiry et al., 2017; Wilson et al., 1998). Separating the data set into different groups or sections could potentially reveal any nuances in the performance of the models across the groups that may be affecting overall performance.

Another important recommendation for future research is the use of other measures of metabolic dysfunction that may prove to be more effective risk factors,

namely glycated hemoglobin. A supplementary analysis of this data set was conducted, replacing glucose with insulin or HbA1C values. Insulin levels were not statistically significant but HbA1C was found to be a statistically significant predictor, even having a stronger predictive relationship (with an OR = 1.378, p =0.039), than all the other risk factors, except for smoking. A previous study found HbA1C to be a stronger predictor of CHD risk than total cholesterol or total cholesterol/HDL ratio, with OR of 1.23, 1.047 and 1.073 respectively (Faeh et al., 2013). Several other studies also determined that HbA1C has a greater association with CVD risks than does glucose (Ahn, 2017; Carson et al., 2015; Sarwar et al., 2010; Selvin et al., 2010). Others have suggested that HBA1C levels could replace glucose in screening for both DM2 and CVD (Danesh, 2014), or it could even be used a single parameter in determining CVD risk for non-diabetics (Garg et al., 2014). All these studies point to a strong rationalization to including HbA1C in risk models, or at least testing the merit of its inclusion.

The final recommendation as it relates to CVD risk estimation is the use of a different type of regression modeling to represent the prospective data. The models that have been traditionally used are based on Logistic Regression and the Cox Proportional Hazard, both of which are parametric analyses and therefore rely on various assumptions about the underlying data. The Logistic regression assumes that there is a linear relationship between the variable and the log odds of the outcome (Lani, 2015), and the Cox Hazard assumes that the ratio of the hazard of the succumbing to the disease constant over time (D'Agostino et al., 1990). Both of these assumptions may

not be true in the case of CVD and its many risk factors, making models based on them unreliable.

Non-parametric modeling are free of assumptions and are able to account for the structure in the data and generate more accurate representation of the data (Austin, Tu, & Alter, 2003). Methods such as decision/classification tree analysis; structural equation model (SEM); non-parametric path analysis; as well as data mining and big data analysis techniques are transforming medical disease modeling (Dunson, Xing, & Associate, 2012)(Kennedy, Wiitala, Hayward, & Sussman, n.d.) (Dunson et al., 2012). These novel techniques are outperforming older modeling approaches with increased accuracy and improved reliability as well as providing causal inferences from prospective data. These types of models allowing improved patient classification and clinical treatment decision making (Vistisen et al., 2016). The combination of large quantities of patient data, both epidemiological and HER, and the sophisticated computer application tools may make traditional modeling obsolete.

General Conclusion

This study was the first of its kind to assess the performance of the Framingham Risk Score model on the third generation of FHS cohort (FHS III). The present analysis confirmed that both glucose and diabetic status had a significant association with the outcome of a CHD event. The findings, however, indicate that there was no statistically significant difference and no improvement in performance of the FRS model when DM2 replaces FBG levels. The study did also corroborate the predictive value of the major risk factors including smoking, diabetic status, age, HDL and systolic blood

pressure. Specific to this data set, factors of gender and BMI were not statistically significant to the outcome, however, it should be noted that this data set, like all other FHS cohorts, is made up of unique population of relatively heart-healthy participants. As such further research into the accuracy and reliability of the results, along with responsible validation studies, are needed before generalizable application of these findings is to be considered.

References

Ahn, C. H. (2017). Hemoglobin Glycation Index Is Associated With Cardiovascular Diseases in People With Impaired Glucose Metabolism. *Journal of Clinical Endocrinology Metabolism*, *102*, 2905–2913. Retrieved from https://watermark.silverchair.com/jc.2017-00191.pdf?token=AQECAHi208BE49Ooan9kkhW_Ercy7Dm3ZL_9Cf3qfKAc485ysgAAAlcwggJTBgkqhkiG9w0BBwagggJEMIICQAIBADCCAjkGCSqGSIb3DQEHATAeBglghkgBZQMEAS4wEQQMl0tEUxpp2mqywGCeAgEQgIICCqCih7K3kz6Uor5rzcFqk7gKUHUY31i2xEQle_y-t

Al-Nooh, A. A., Abdulabbas Abdulla Alajmi, A., & Wood, D. (2014). The Prevalence of Cardiovascular Disease Risk Factors among Employees in the Kingdom of Bahrain between October 2010 and March 2011: A Cross-Sectional Study from a Workplace Health Campaign. *Cardiology Research and Practice*, *2014*, 1–9. https://doi.org/10.1155/2014/832421

Alatawi, Y. M., Kavookjian, J., Ekong, G., & Alrayees, M. M. (2016). The association between health beliefs and medication adherence among patients with type 2 diabetes. *Research in Social and Administrative Pharmacy*, *12*(2016), 914–925. https://doi.org/10.1016/j.sapharm.2015.11.006

American Diabetes Association. (2009). Diagnosis and Classification of Diabetes Mellitus. *Diabetes Care*, *33*(1), S63–S69. https://doi.org/10.2337/dc10-S062

American Diabetes Association. (2015). Classification and diagnosis of diabetes. *Diabetes Care*, *38*(1), S8–S16. https://doi.org/10.2337/dc15-S005

Amsterdam, E. A. (2011). Primary Prevention of Coronary Heart Disease in Men and Women: Does 1 Size Fit All? Yes! *Clinical Cardiology*, *34*(11), 658–662. https://doi.org/10.1002/clc.20993

Anderson, K. M., Wilson, P. W. F., Odell, P. M., & Kannel, W. B. (1991). An updated coronary risk profile : a statement for health professionals. *Circulation*, *83*(1), 356–362. https://doi.org/10.1161/01.CIR.83.1.356

Anderson, Keaven M, Wilson, P. W. F., Odell, P. M., & Kannel, W. B. (1991). An Updated Coronary Risk Profile A Statement for Health Professionals. *AHA Medical/Sciientific Statement*, *83*, 356–362. Retrieved from http://circ.ahajournals.org/content/circulationaha/83/1/356.full.pdf

Assmann, G., Schulte, H., & Seedorf, U. (2008). Cardiovascular risk assessment in the metabolic syndrome: results from the Prospective Cardiovascular Munster (PROCAM) Study. *International Journal of Obesity*, *32*(S2), S11–S16. https://doi.org/10.1038/ijo.2008.29

Association, A. H., Center for Disease Control and Prevention, & Health, N. I. of. (2015). Heart Disease Facts American Heart Association. Retrieved July 8, 2017, from http://www.theheartfoundation.org/heart-disease-facts/heart-disease-statistics/

Austin, P. C., Tu, J. V, & Alter, D. A. (2003). Comparing hierarchical modeling with traditional logistic regression analysis among patients hospitalized with acute myocardial infarction: Should we be analyzing cardiovascular outcomes data differently? *American Heart Journal*. https://doi.org/10.1067/mhj.2003.23

Avitabile, N. A., Banka, A., & Fonseca, V. A. (2012). Glucose control and cardiovascular outcomes in individuals with diabetes mellitus: lessons learned from the megatrials. *Heart Failure Clinics*, *8*(4), 513–522. https://doi.org/10.1016/j.hfc.2012.06.009

Benjamin, E J, Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S., Deo, R., … Jimenez, M. C. (2017). *Heart Disease and Stroke Statistics 2017 At-a-Glance*. *American Heart Association & American Stroke Association*. Retrieved from https://www.heart.org/idc/groups/ahamah-public/@wcm/@sop/@smd/documents/downloadable/ucm_491265.pdf

Benjamin, Emelia J., Blaha, M. J., Chiuve, S. E., Cushman, M., Das, S. R., Deo, R., … Muntner, P. (2017). Heart Disease and Stroke Statistics—2017 Update: A Report From the American Heart Association. *Circulation*, *135*(10). Retrieved from http://circ.ahajournals.org/content/135/10/e146

Bennet, A., Angelantonio, E. D., Erqou, S., Eurujstittir, G., Sigurdsson, G., Woodward, M., … Gudnason, V. (2008). Lipoprotein(a) Levels and Risk of Future Coronary Heart Disease<subtitle>Large-Scale Prospective Data</subtitle>. *Archives of Internal Medicine*, *168*(6), 598. https://doi.org/10.1001/archinte.168.6.598

Bennet, A., Di Angelantonio, E., Erqou, S., Eiriksdottir, G., Sigurdsson, G., Woodward, M., … Gudnason, V. (2008). Lipoprotein(a) Levels and Risk of Future Coronary Heart Disease. *Archives of Internal Medicine*, *168*(6), 598–608. https://doi.org/10.1001/archinte.168.6.598

Bertoluci, M. C., & Rocha, V. Z. (2017). Cardiovascular risk assessment in patients

with diabetes. *Diabetology & Metabolic Syndrome*, *9*(25), 1–13. Retrieved from https://dmsjournal.biomedcentral.com/track/pdf/10.1186/s13098-017-0225-1?site=dmsjournal.biomedcentral.com

Bitton, A., & Gaziano, T. (2010). The Framingham Heart Study's Impact on Global Risk Assessment. *Progress in Cardiovascular Diseases*, *53*(1), 68–78. https://doi.org/10.1016/j.pcad.2010.04.001

Bonow, R. O. (2002). Primary Prevention of Cardiovascular Disease: A Call to Action. *Circulation*, *106*(25), 3140–3141. https://doi.org/10.1161/01.CIR.0000048067.86569.E1

Booth, G. L., Kapral, M. K., Fung, K., & Tu, J. V. (2006). Relation between age and cardiovascular disease in men and women with diabetes compared with non-diabetic people: a population-based retrospective cohort study. *Www.Thelancet.Com*, *368*. Retrieved from https://pdfs.semanticscholar.org/7bab/289ad870459c5f902478ecefd30f09b27a8d.pdf

Bragg, F., Li, L., Smith, M., Guo, Y., Chen, Y., Millwood, I., … Chen, Z. (2014). Research: Epidemiology Associations of blood glucose and prevalent diabetes with risk of cardiovascular disease in 500 000 adult Chinese: the China Kadoorie Biobank. *Diabet. Med*, *31*, 540–551. https://doi.org/10.1111/dme.12392

Braun, J., Bopp, M., & Faeh, D. (2013). Blood glucose may be an alternative to cholesterol in CVD risk prediction charts. *Cardiovascular Diabetology*, *12*(24), 1. https://doi.org/10.1186/1475-2840-12-24

Brindle, P, Smith, G. D., Upton, M., McConnachie, A., Hart, C., & Watt, G. (2005). The accuracy of the Framingham risk-score in different socioeconomic groups: a prospective study. *British Journal of General Practice*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1570792/pdf/bjpg55-838.pdf

Brindle, PM, McConnachie, A., Upton, M., Carole, L., Smith, G., & Graham, C. (2005). The accuracy of the Framingham risk-score in different socioeconomic groups: a prospective study. *British Journal of General Practice*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1570792/pdf/bjpg55-838.pdf

Brunzell, J. D., Davidson, M., & Fuberg, C. D. (2009). Cardiometabolic Risk FAQS. *Health Dialog*. Retrieved from https://www.amerihealth.com/pdfs/providers/resources/connections/cardiometabolic_ah.pdf

Burnham, K. P., & Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, *33*(2), 261–304. https://doi.org/10.1177/0049124104268644

Carson, A. P., Steffes, M. W., Carr, J. J., Kim, Y., Gross, M. D., Carnethon, M. R., … Lewis, C. E. (2015). Hemoglobin A 1c and the Progression of Coronary Artery Calcification Among Adults Without Diabetes Diabetes. *Diabetes Care*, *38*. https://doi.org/10.2337/dc14-0360

Carter, A. N., Ralston, P. A., Young-Clark, I., & Ilich, J. Z. (2016). Diabetic indicators are the strongest predictors for cardiovascular disease risk in African American adults. *Am J Cardiovasc Dis*, *6*(3), 129–137. Retrieved from www.AJCD.us

CDC. (2014). *SMOKING AND CARDIOVASCULAR DISEASE*. Retrieved from www.smokefree.gov

Center for Disease Control & Prevention. (2017). National Diabetes Statistics Report, 2017 Estimates of Diabetes and Its Burden in the United States Background. *National Center for Chronic Disease Prevention and Health Promotion* . Retrieved from https://www.cdc.gov/diabetes/pdfs/data/statistics/national-diabetes-statistics-report.pdf

Center for Health Statistics, N. (2017). *Table A-1. Selected circulatory diseases among adults aged 18 and over, by selected characteristics: United States, 2017*. Retrieved from https://www.cdc.gov/nchs/data/series/sr

Centers for Disease Control and Prevention. (2013). Vital Signs: Avoidable Deaths from Heart Disease, Stroke, and Hypertensive Disease — United States, 2001–2010. Retrieved July 24, 2019, from https://www.cdc.gov/mmwr/preview/mmwrhtml/mm6235a4.htm?s_cid=mm6235a4_w

Centers for Disease Control and Prevention. (2017). Long-term Trends in Diabetes. *CDC's Division of Diabetes Translation*. Retrieved from https://www.cdc.gov/diabetes/statistics/slides/long_term_trends.pdf

Çevik, A. B., Özcan, Ş., & Satman, İ. (2015). Sensitivity of FRAMINGHAM, PROCAM and SCORE models in Turkish people with Type 2 diabetes: comparison of three cardiovascular risk calculations. *Contemporary Nurse*, *50*. https://doi.org/10.1080/10376178.2015.1111153

Chi, C., Snaith, J., & Gunton, J. E. (2017). Diabetes Medications and Cardiovascular Outcomes in Type 2 Diabetes. *Heart, Lung and Circulation*, *26*. https://doi.org/10.1016/j.hlc.2017.02.030

Chowdhury, R., Warnakula, S., Kunutsor, S., Crowe, F., Ward, H. A., Johnson, L., … Di Angelantonio, E. (2014). Association of dietary, circulating, and supplement fatty acids with coronary risk. *Annals of Internal Medicine*, *16*, 398–406. Retrieved from https://www.repository.cam.ac.uk/bitstream/handle/1810/247312/Chowdhury_et_al-2014-Annals_of_Internal_Medicine.pdf?sequence=1&isAllowed=y

Clark, V. R., Perkins, P., Carson, B. L., Boyd, K., & Jefferson, T. M. (2015). FASTING SERUM GLUCOSE AND CHOLESTEROL AS PREDICTORS OF CARDIOVASCULAR REACTIVITY TO ACUTE STRESS IN A SAMPLE OF AFRICAN AMERICAN COLLEGE STUDENTS. *Ethnicity & Disease*, *25*, 175–179.

Cleeman, J. I. (2001). Executive Summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *JAMA*, *285*. Retrieved from www.nhlbi.nih.gov.

Coch, R. W., & Green, J. B. (2016). Current cardiovascular outcomes trials in type 2 diabetes: Perspectives and insight. https://doi.org/10.1016/j.numecd.2016.06.004

Cockram, C., Albert, G., Allgot, B., Al Nakhi, A., Aschner, P., Dwyer, T., & Haffner, S. (2001). *Diabetes and Cardiovascular Disease: Time to Act*. *International*

Diabetes Federation (Vol. 44). https://doi.org/10.2337/diabetes.44.4.369

Collins, G. S., & Altman, D. G. (2009). An independent external validation and evaluation of QRISK cardiovascular risk prediction: a prospective open cohort study professor of statistics in medicine. *BMJ*. https://doi.org/10.1136/bmj.b2584

Collins, G. S., & Altman, D. G. (2012). Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. *BMJ*, *344*(4181), 1–12. https://doi.org/10.1136/bmj.e4181

Conroy, R. M., Pyörälä, K., Fitzgerald, A. P., Sans, S., Menotti, A., De Backer, G., … Graham, I. M. (2003). Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *European Heart Journal*, *24*, 987–1003. https://doi.org/10.1016/S0195-668X(03)00114-3

Cook, A., & Sheikh, A. (2000). Descriptive statistics (Part 2): Interpreting study results. https://doi.org/10.1038/pcrj.2000.6

Coutinho, M., Gerstein, H. C., Wang, Y., & Yusuf, S. (1999). The Relationship Between Glucose and Incident Cardiovascular Events. *DIABETES CARE*, *22*(2). Retrieved from http://care.diabetesjournals.org/content/diacare/22/2/233.full.pdf

D'Agostino, R. B., Grundy, S., Sullivan, L. M., & Wilson, P. (2001a). Validation of the Framingham Coronary Heart Disease Prediction Scores. *JAMA*, *286*(2), 180. https://doi.org/10.1001/jama.286.2.180

D'Agostino, R. B., Grundy, S., Sullivan, L. M., & Wilson, P. (2001b). Validation of the Framingham Coronary Heart Disease Prediction Scores. *JAMA*, *286*(2), 180. https://doi.org/10.1001/jama.286.2.180

D'Agostino, R. B., Lee, M. L., Belanger, A. J., Cupples, L. A., Anderson, K. M., & Kannel, W. B. (1990). Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. *Statistics in Medicine*, *9*(12), 1501–1515. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/2281238

D'Agostino, R. B., Pencina, M. J., Massaro, J. M., & Coady, S. (2013). Cardiovascular Disease Risk Assessment: Insights from Framingham. *Global Heart of Elsevier Ltd.*, *8*(11), 11–23. https://doi.org/10.1016/j.gheart.2013.01.001

D'Agostino, R. B., Russell, M. W., Huse, D. M., Ellison, R. C., Silbershatz, H., Wilson, P. W., & Hartz, S. C. (2000). Primary and subsequent coronary risk appraisal: new results from the Framingham study. *American Heart Journal*, *139*(2 Pt 1), 272–281. https://doi.org/10.1016/S0002-8703(00)90236-9

D'Agostino, R. B., Vasan, R. S., Pencina, M. J., Wolf, P. A., Cobain, M., Massaro, J. M., & Kannel, W. B. (2008). General cardiovascular risk profile for use in primary care: The Framingham heart study. *Circulation*, *117*(6), 743–753. https://doi.org/10.1161/CIRCULATIONAHA.107.699579

Dahlöf, B. (2010). Cardiovascular Disease Risk Factors: Epidemiology and Risk Assessment. *American Journal of Cardiology*, *105*, 3A-9A. https://doi.org/10.1016/j.amjcard.2009.10.007

Dale, C. E., Fatemifar, G., Palmer, T. M., White, J., Prieto-Merino, D., Zabaneh, D., … Casas, J. P. (2017). Causal Associations of Adiposity and Body Fat Distribution With Coronary Heart Disease, Stroke Subtypes, and Type 2 Diabetes Mellitus: A Mendelian Randomization Analysis. *Circulation*, *135*(24), 2373–2388.

https://doi.org/10.1161/CIRCULATIONAHA.116.026560

Dalen, J. E., Alpert, J. S., Goldberg, R. J., & Weinstein, R. S. (2014). The Epidemic of the 20 th Century: Coronary Heart Disease. *The American Journal of Medicine*, *127*, 807–812. https://doi.org/10.1016/j.amjmed.2014.04.015

Damkondwar, D., Rajiv, R., Suganeswari, G., Kulothungan, V., & Sharma, T. (2011). Assessing Framingham cardiovascular risk scores in subjects with diabetes and their correlation with diabetic retinopathy. *Indian Journal of Ophthalmology*. https://doi.org/10.4103/0301-4738.91344

Danesh, J. (2014). Glycated Hemoglobin Measurement and Prediction of Cardiovascular Disease CVD or diabetes, additional assessment of HbA 1c values in the context of CVD risk assessment provided little incremental benefit for prediction of CVD risk. HHS Public Access. *JAMA*, *311*(12), 1225–1233. https://doi.org/10.1001/jama.2014.1873

Dauriz, M., Porneala, B. C., Guo, X., Bielak, L. F., Peyser, P. A., Durant, N. H., … Meigs, J. B. (2015). Association of a 62 Variants Type 2 Diabetes Genetic Risk Score With Markers of Subclinical AtherosclerosisCLINICAL PERSPECTIVE. *Circulation: Cardiovascular Genetics*, *8*(3), 507–515. https://doi.org/10.1161/CIRCGENETICS.114.000740

Dawber, T. R., George, F. E., & Mann, V. (1957). 1. Coronary Heart Disease in the Framingham Study. *American Journal of Clinical Nutrition*, *47*, 4–24. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1550985/pdf/amjphnation01096-

0007.pdf

Dawber, T. R., Kannel, W. B., Revotskie, N., Stokes, J., Kagan, A., & Gordon, T. (1959). Some Factors Associated with the Development Of Coronary Heart Disease Six Years' Follow-Up Experience in the Framingham Heart Study. *Am J Public Health Nations Health*, *49*, 1349–1356. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1372949/pdf/amjphnation00326-0051.pdf

Dawber, T. R., & Lansing, A. I. (1966). The Framingham Study An Epidemiological Approach to Coronary Heart Disease. *Circulation*, *34*(4), 1283–1285. https://doi.org/10.1161/01.CIR.24.6.1283

Dawber, T. R., Meadors, G. F., & Moore, F. E. (1951). Epidemiological Approaches to Heart Disease: The Framingham Study*. In *Joint Sessions of the Epidemiology, Health Officers, Medical Care and Statistics Sections of the APHA* (Vol. 41, pp. 279–289). St. Louis, MO. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1525365/pdf/amjphnation00421-0020.pdf

Demler, O. V, Paynter, N. P., & Cook, N. R. (2015). Tests of calibration and goodness-of-fit in the survival setting. *Statistics in Medicine*, *34*(10), 1659–1680. https://doi.org/10.1002/sim.6428

Desouza, C., Raghavan, V. A., & Fonseca, V. A. (2010). The enigma of glucose and cardiovascular disease. *Heart*, *96*, 649–651.

Duckworth, W., Abraira, C., Moritz, T., Reda, D., Emanuele, N., Reaven, P. D., …

Huang, G. D. (2009). Glucose Control and Vascular Complications in Veterans with Type 2 Diabetes for the VADT Investigators*. *New England Journal of Medicine*, *360*(2), 129–139. https://doi.org/10.1056/NEJMoa0808431

Dunson, D. B., Xing, C., & Associate, P. (2012). Nonparametric Bayes Modeling of Multivariate Categorical Data. *J Am Stat Assoc*, *104*. https://doi.org/10.1198/jasa.2009.tm08439

Eichler, K., Puhan, M. A., Steurer, J., & Bachmann, L. M. (2007). Prediction of first coronary events with the Framingham score: A systematic review. *American Heart Journal*, *153*, 723–733. https://doi.org/10.1016/j.ahj.2007.02.027

Elosua, R. (2014). Cardiovascular Risk Functions: Usefulness and Limitations, *67*(2), 77–79. https://doi.org/10.1016/j.rec.2013.09.012

Esparaza-del Villar, O. A., Montañez-Alvarado, P., Gutiérrez-Vega, M., Carrillo-Saucedo, I. C., Gurrola-Peña, G. M., Ruvalcaba-Romero, N. A., … Ochoa-Alcaraz, S. G. (2017). Factor structure and internal reliability of an exercise health belief model scale in a Mexican population. *BMC Public Health*, *17*(229). https://doi.org/10.1186/s12889-017-4150-x

Faeh, D., Rohrmann, S., & Braun, J. (2013). Better risk assessment with glycated hemoglobin instead of cholesterol in CVD risk prediction charts. *European Journal of Epidemiology*, *28*(7), 551–555. https://doi.org/10.1007/s10654-013-9827-6

Faerch, K., Bergman, B., & Perreault, L. (2012). Does Insulin Resistance Drive the Association between Hyperglycemia and Cardiovascular Risk? *PLoS ONE*, *7*.

https://doi.org/10.1371/journal.pone.0039260

Faerch, K., Vistisen, D., Borup, N., Marit, J. &, & Jørgensen, E. (2014). Cardiovascular Risk Stratification and Management in Pre-Diabetes. https://doi.org/10.1007/s11892-014-0493-1

Farooq Al-Azzawi, O. (2015). Prevalence of prediabetes and metabolic syndrome and their association in an Iraqi sample. *IOSR Journal of Dental and Medical Sciences*, *14*, 2279–2861. https://doi.org/10.9790/0853-14951016

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). Statistics textbooks in the social. *Behavior Research Methods*, *38*. Retrieved from https://link.springer.com/content/pdf/10.3758/BF03193146.pdf

Fawwad, A., Moin, H., Siddiqui, I. A., Hydrie, M. Z. I., & Basit, A. (2016). Framingham risk score in impaired glucose tolerant population: A sub analysis of Diabetes Prevention and Awareness Program of Pakistan. *Pak J Med Sci*, *32*(5). https://doi.org/10.12669/pjms.325.10448

Feresu, S. a, Zhang, W., Puumala, S. E., Ullrich, F., & Anderson, J. R. (2008). The frequency and distribution of cardiovascular disease risk factors among Nebraska women enrolled in the WISEWOMAN screening program. *Journal of Women's Health*, *17*(4), 607–617. https://doi.org/10.1089/jwh.2007.0438

Ferreira-GonzáLez, I. (2014). The Epidemiology of Coronary Heart Disease. *Sociedad Espanola de Cardiologia*, *672*, 139–144. https://doi.org/10.1016/j.rec.2013.10.002

Field, A. (2013). *Discovering Statistics using IBM SPSS Statistics*. Sage. Retrieved from

https://books.google.com/books?hl=en&lr=&id=c0Wk9IuBmAoC&oi=fnd&pg=PP2&dq=discovering+statistics+using+ibm+spss&ots=LbBnLM3u5D&sig=uRLvAGfdnpawZrGkqotYeYL5A3s#v=onepage&q=discovering statistics using ibm spss&f=false

Ford, E. S., Giles, W. H., & Mokdad, A. H. (2004). Coronary Artery Disease Findings From the National Health and Nutrition Examination Survey III. *Journal of the American College of Cardiology*, *43*, 1791–1796. https://doi.org/10.1016/j.jacc.2003.11.061

Fox, C. S. (2010). Cardiovascular Disease Risk Factors, Type 2 Diabetes Mellitus, and the Framingham Heart Study. *Trends in Cardiovascular Medicine*, *20*(3), 90–95. https://doi.org/10.1016/j.tcm.2010.08.001

Framingham Risk Score Calculator for Coronary Heart Disease. (2018). Retrieved March 12, 2018, from https://www.thecalculator.co/health/Framingham-Risk-Score-Calculator-for-Coronary-Heart-Disease-745.html

Franklin, S. S., & Wong, N. D. (2013). Hypertension and Cardiovascular Disease: Contributions of the Framingham Heart Study. https://doi.org/10.1016/j.gheart.2012.12.004

Friedman, G. D., Hyg, S., Kannel, W. B., Dawber, T. R., & McNamara, P. M. (1967). An Evaluation of the Follow-Up Methods in the Framingham Heart Study. *American Journal of Public Health*, *57*, 1015–1024. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1227450/pdf/amjphnation00074-0089.pdf

Fuster, V., & Kelly, B. B. (2010). Committee on Preventing the Global Epidemic of Cardiovascular Disease: Meeting the Challenges in Developing Countries Board on Global Health. *The Institute of Medicine of National Academies Press*. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK45693/pdf/Bookshelf_NBK45693.pdf

Garg, D., Moorthy, Nagaraja, D., Kapoor, A., Tewari, D., Satyendra, D., & Kumar. (2014). *Hemoglobin A^sub 1c^ in Nondiabetic Patients: An Independent*. *Mayo Clinic Proceedings* (Vol. 89). Retrieved from https://search-proquest-com.ezp.waldenulibrary.org/docview/1545042480/fulltextPDF/2A9A7BF6AED84683PQ/1?accountid=14872

Garg, N., Muduli, S. K., Kapoor, A., Tewari, S., Kumar, S., Khanna, R., & Goel, P. K. (2017). Comparison of different cardiovascular risk score calculators for cardiovascular risk prediction and guideline recommended statin uses. *Indian Heart Journal*, *69*(4), 458–463. https://doi.org/10.1016/J.IHJ.2017.01.015

Giancristofaro, R. A., & Salmaso, L. (2003). MODEL PERFORMANCE ANALYSIS AND MODEL VALIDATION IN LOGISTIC REGRESSION. *Statistica*, (2). Retrieved from https://rivista-statistica.unibo.it/article/viewFile/358/351

Ginsberg, H. N. (2011). The ACCORD (Action to Control Cardiovascular Risk in Diabetes) Lipid Trial What we learn from subgroup analyses. *Diabetes Care*, *34*(2). https://doi.org/10.2337/dc11-s203

Go, A. S., Mozaffarian, D., Roger, V. L., Benjamin, E. J., Berry, J. D., Blaha, M. J., … Turner, M. B. (2014). *Heart Disease and Stroke Statistics - 2014 Update: A report*

from the American Heart Association. *Circulation* (Vol. 129). https://doi.org/10.1161/01.cir.0000441139.02102.80

Goff, D. C., Lloyd-Jones, D. M., Bennett, G., Coady, S., D'Agostino, R. B., Gibbons, R., … Wilson, P. W. F. (2014). 2013 ACC/AHA Guideline on the Assessment of Cardiovascular Risk. *Circulation*, *129*(25 SUPPL. 1). https://doi.org/10.1161/01.cir.0000437741.48606.98

Goong, H., Ryu, S., & Xu, L. (2016). A structural model of health behavior modification among patients with cardiovascular disease. *Applied Nursing Research*, *29*(2016), 70–75. https://doi.org/10.1016/j.apnr.2015.06.005

Gordon, T., & Kannel, W. B. (1982). Multiple risk functions for predicting coronary heart disease: the concept, accuracy, and application. *American Heart Journal*, *103*(6), 1031–1039. https://doi.org/10.1016/0002-8703(82)90567-1

Govindaraju, D. R., Adrienne Cupples, L., Kannel, W. B., O, C. J., Atwood, L. D., D'Agostino, R. B., … Benjamin, E. J. (2008). Genetics of the Framingham Heart Study Population. *Advanced Genetics*, *62*. https://doi.org/10.1016/S0065-2660(08)00602-0

Grimes, D. S. (2012). An epidemic of coronary heart disease. *QJM*, *105*(6), 509–518. https://doi.org/10.1093/qjmed/hcr265

Grundy, S. M., Becker, D., Clark, L. T., Cooper, R. S., Denke, M. A., James Howard, W., … Sempos, C. T. (2002). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) Final Report National

Cholesterol Education Program Expert Panel on Dete. *National Cholesterol Education Program National Heart, Lung, and Blood Institute*, *02*. Retrieved from http://circ.ahajournals.org/

Grundy, S. M., Cleeman, J. I., Merz, C. N. B., Brewer, H. B., Clark, L. T., Hunninghake, D. B., … American Heart Association, C. (2008). Implications of recent clinical trials for the National Cholesterol Education Program Adult Treatment Panel III guidelines. *Circulation*, *110*(2), 227–239. https://doi.org/10.1161/01.CIR.0000133317.49796.0E

Günaydın, Y. Z., Karagöz, A., Bektaş, O., Kaya, A., Kırış, T., Erdoğan, G., … Ayhan, E. (2016). Comparison of the Framingham risk and SCORE models in predicting the presence and severity of coronary artery disease considering SYNTAX score. *Antol Journal of Cardiology*, (16), 412–418. https://doi.org/10.5152/AnatolJCardiol.2015.6317

Haregu, T. N., Oti, S., Ngomi, N., Khayeka-Wandabwa, C., Egondi, T., & Kyobutungi, C. (2016). Interlinkage among cardio-metabolic disease markers in an urban poor setting in Nairobi, Kenya. *Global Health Action*, *9*(0), 30626. https://doi.org/10.3402/gha.v9.30626

Hemann, B. A., Bimson, W. F., & Taylor, A. J. (2007). The Framingham Risk Score: an appraisal of its benefits and limitations. *The American Heart Hospital Journal*, *5*(2), 91–96. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17478974

Herman, W. H., & Rothberg, A. E. (2015). Prevalence of Diabetes in the United States: A Glimer of Hope. *JAMA*, *314*(10), 1005.

https://doi.org/10.1001/jama.2015.10030

Hermansson, J., & Kahan, T. (2018). Systematic Review of Validity Assessments of Framingham Risk Score Results in Health Economic Modelling of Lipid-Modifying Therapies in Europe. *PharmacoEconomics*, *36*(2), 205–213. https://doi.org/10.1007/s40273-017-0578-1

Hippisley-Cox, J., Coupland, C., Vinogradova, Y., Robson, J., May, M., & Brindle, P. (2007). Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ*, *335*. https://doi.org/10.1136/bmj.39261.471806.55

Hobbs, F. D. R., & Hobbs, R. (2004). CARDIOVASCULAR DISEASE: DIFFERENT STRATEGIES FOR PRIMARY AND SECONDARY PREVENTION? *Heart*, *90*, 1217–1223. https://doi.org/10.1136/hrt.2003.027680

Holmes, M. V., Pulit, S. L., & Lindgren, C. M. (2017). Genetic and epigenetic studies of adiposity and cardiometabolic disease. *Genome Medicine*, *9*. https://doi.org/10.1186/S13073-017-0474-5

Hosseini, H. (2015). Stroke and Cardiovascular Disease among Women. *Health Stroke and Cardiovascular Disease among Women. Health*, *7*(7), 1847–1854. https://doi.org/10.4236/health.2015.714203

Hu, S. (2007). Akaike Information Criterion. Retrieved from http://www4.ncsu.edu/~shu3/Presentation/AIC.pdf

Huang, Y., Cai, X., Mai, W., Li, M., & Hu, Y. (2016). Association between prediabetes and risk of cardiovascular disease and all cause mortality: systematic review and

meta-analysis. *British Medical Journal*, *355*. https://doi.org/10.1136/bmj.i5953

Hutcheson, R., & Rocic, P. (2012). The Metabolic Syndrome, Oxidative Stress, Environment, and Cardiovascular Disease: The Great Exploration. *Experimental Diabetes Research*. https://doi.org/10.1155/2012/271028

Huxley, R. R., Woodward, M., Huxley, R. R., & Woodward, M. (2011). Cigarette smoking as a risk factor for coronary heart disease in women compared with men: a systematic review and meta-analysis of prospective cohort studies. *Www.Thelancet.Com Lancet*, *378*, 1297–1305. https://doi.org/10.1016/S0140-6736(11)60781-2

Hyman, M. (2012). *The blood sugar solution : the ultrahealthy program for losing weight, preventing disease, and feeling great now!* Little, Brown and Co.

International Diabetes Federation. (2017). *IDF Diabetes Atlas*. Retrieved from https://www.idf.org/e-library/epidemiology-research/diabetes-atlas/134-idf-diabetes-atlas-8th-edition.html

Ismail-Beigi, F., Moghissi, E., Kosiborod, M., & Inzucchi, S. E. (2008). Intensive Blood Glucose Control and Vascular Outcomes in Patients with Type 2 Diabetes. *N Engl J Med*, *358*, 2560–2572. https://doi.org/10.1056/NEJMoa0802987

Jackson, R., Brennan, S., Fielding, P., Sims, M. W., Challiss, R. J., Adlam, D., … Rainbow, R. D. (2016). Distinct and complementary roles for α and β isoenzymes of PKC in mediating vasoconstrictor responses to acutely elevated glucose. *British Journal of Pharmacology*, *173*(5), 870–887. https://doi.org/10.1111/bph.13399

Jahangiry, L., Farhangi, M. A., & Rezaei, F. (2017). Framingham risk score for

estimation of 10-years of cardiovascular diseases risk in patients with metabolic syndrome. *Journal of Health, Population, and Nutrition*, *36*(1), 36. https://doi.org/10.1186/s41043-017-0114-0

Janz, N. K., & Becker, M. H. (1984). The Health Belief Model: A Decade Later. *Health Education Quarterly*, *11*(1), 1–47. https://doi.org/10.1177/109019818401100101

Jones, D. S., & Greene, J. A. (2013). The Decline and Rise of Coronary Heart Disease. *Am J Public Health*, *103103*(7), 1207–1218. https://doi.org/10.2105/AJPH.2013.301226

Kadowaki, S., Okamura, T., Hozawa, A., Kadowaki, T., Kadota, A., Murakami, Y., … Ueshima, H. (2008). Relationship of elevated casual blood glucose level with coronary heart disease, cardiovascular disease and all-cause mortality in a representative sample of the Japanese population. NIPPON DATA80. *Diabetologia*, *51*(4), 575–582. https://doi.org/10.1007/s00125-007-0915-6

Kannel, W. B., & Boston, F. (1990). Contribution of the framingham study to preventive cardiology. *Journal of the American College of Cardiology*, *15*, 206–211. https://doi.org/10.1016/0735-1097(90)90203-2

Kannel, W. B., Castelli, W., Gordon, T., & Mcnamara, P. (1971). Serum Cholesterol, Lipoproteins, and the risk of coronary heart disease The framingham study. *Annals of Internal Medicine*, *74*(1), 1–12.

Kannel, W. B., D'Agostino, R. B., Silbershatz, H., Belanger, A. J., Wilson, P. W. F., & Levy, D. (1999). Profile for Estimating Risk of Heart Failure. *Archives of Internal Medicine*, *159*. Retrieved from

https://pdfs.semanticscholar.org/5dd2/7701eb6ce90920a642ac12122ce0b76f0b40.pdf

Kannel, W. B., Gordon, T., & Schwartz, M. J. (1971). Systolic versus diastolic blood pressure and risk of coronary heart disease: The Framingham study. *The American Journal of Cardiology*, *27*(4), 335–346. https://doi.org/10.1016/0002-9149(71)90428-0

Kannel, W. B., & McGee, D. L. (1979a). Diabetes and Cardiovascular Disease. *JAMA*, *241*(19), 2035. https://doi.org/10.1001/jama.1979.03290450033020

Kannel, W. B., & McGee, D. L. (1979b). Diabetes and cardiovascular risk factors: the Framingham study. *Circulation*, *59*(1), 8–13. https://doi.org/10.1161/01.CIR.59.1.8

Kannel, W. B., Vasan, R. S., & Levy, D. (2003). Is the relation of systolic blood pressure to risk of cardiovascular disease continuous and graded, or are there critical values? *Hypertension (Dallas, Tex. : 1979)*, *42*(4), 453–456. https://doi.org/10.1161/01.HYP.0000093382.69464.C4

Kelly, T. N., Bazzano, L. A., Fonseca, V. A., & al., et. (2009). Systematic review: glucose control and cardiovascular disease in type 2 diabetes. *Ann Intern Med*, *151*, 394–403.

Kennedy, E. H., Wiitala, W. L., Hayward, R. A., & Sussman, J. B. (n.d.). Improved cardiovascular risk prediction using nonparametric regression and electronic health record data. https://doi.org/10.1097/MLR.0b013e31827da594

Khorsandi, M., Fekrizadeh, Z., & Roozbahani, N. (2017). Investigation of the effect of

education based on the health belief model on the adoption of hypertension-controlling behaviors in the elderly. *Clinical Interventions in Aging*, *Volume 12*, 233–240. https://doi.org/10.2147/CIA.S117142

Kim, H. C. (2016). *SYMPOSIUM 04 ASSESSMENT INSTRUMENT IN DEVELOPING*.

Kishore, P., Kim, S. H., & Crandall, J. P. (2012). Glycemic Control and Cardiovascular Disease: What's a Doctor to Do? *Current Diabetes Reports*, *12*. https://doi.org/10.1007/s11892-012-0268-5

Kones, R. (2011). Primary prevention of coronary heart disease: integration of new data, evolving views, revised goals, and role of rosuvastatin in management. A comprehensive survey. *Drug Design, Development and Therapy*, *5*, 325–380. https://doi.org/10.2147/DDDT.S14934

Kones, R., & Rumana, U. (2014). Cardiovascular Prevention: Components, Levels, Early Origins, and Metrics. *Hospital Practice*, *42*(3), 84–95. https://doi.org/10.3810/hp.2014.08.1121

Kozakova, M., Morizzo, C., Carrubba, S. La, Fabiani, I., Della Latta, D., Jamagidze, J., … Palombo, C. (2017). Associations between common carotid artery diameter, Framingham risk score and cardiovascular events. *Nutrition, Metabolism and Cardiovascular Diseases*, *27*, 329–334. https://doi.org/10.1016/j.numecd.2017.01.001

Kumar, A., Patel, D. R., Wolski, K. E., Lincoff, A. M., Kashyap, S. R., Ruotolo, G., … Menon, V. (2019). Baseline fasting plasma insulin levels predict risk for major adverse cardiovascular events among patients with diabetes and high-risk vascular

disease: Insights from the ACCELERATE trial. *Diabetes & Vascular Disease Research*, *16*(2), 171–177. https://doi.org/10.1177/1479164119827604

Laakso, M. (2015). Is Insulin Resistance a Feature of or a Primary Risk Factor for Cardiovascular Disease ? *Curr*, *15*(105). https://doi.org/10.1007/s11892-015-0684-4

Lammertyn, L., Schutte, R., Schutte, A. E., Huisman, H. W., van Rooyen, J. M., Malan, N. T., … Malan, L. (2011). Associations of Cholesterol and Glucose with Cardiovascular Dysfunction in Black Africans: The SABPA Study. *Clinical & Experimental Hypertension*, *33*(3), 159–166. https://doi.org/10.3109/10641963.2010.531842

Lani, J. (2015). Assumptions of Logistic Regression. *Statistics Solutions*. Retrieved from http://www.statisticssolutions.com

Lansing, A. I. (1961). The Framingham Study: An Epidemiological Approach to Coronary Heart Disease. *American Heart Association*, *34*(4), 1283–1285. https://doi.org/10.1161/01.CIR.24.6.1283

Laurier, D., Chau, N. P., Cazelles, B., & Segond, P. (1994). Estimation of CHD risk in a French working population using a modified Framingham model. *Journal of Clinical Epidemiology*, *47*(12), 1353–1364. https://doi.org/10.1016/0895-4356(94)90079-5

Levy, D., Walmsley, P., & Levenstein, M. (1996). Principal results of the Hypertension and Lipid Trial (HALT): A multicenter study of doxazosin in patients with hypertension. Retrieved from https://ac-els-cdn-

com.ezp.waldenulibrary.org/S0002870396901817/1-s2.0-S0002870396901817-main.pdf?_tid=b2453ffa-a625-4441-9aee-29424cd7121f&acdnat=1532011667_787454d277692198bf0774706cbb4ef2

Lloyd-Jones, D., Adams, R. J., Brown, T. M., Carnethon, M., Dai, S., De Simone, G., … Wylie-Rosett, J. (2010). Executive summary: Heart disease and stroke statistics-2010 update: A report from the american heart association. *Circulation*, *121*(7). https://doi.org/10.1161/CIRCULATIONAHA.109.192667

Lloyd-Jones, D. M., Larson, M. G., Leip, E. P., Beiser, A., D, R. B., Kannel, W. B., … Levy, D. (2002). Lifetime Risk for Developing Congestive Heart Failure The Framingham Heart Study. https://doi.org/10.1161/01.CIR.0000039105.49749.6F

Long, M. T., & Fox, C. S. (2016). The Framingham Heart Study — 67 years of discovery in metabolic disease. *Nature Publishing Group*, *12*, 177–187. https://doi.org/10.1038/nrendo.2015.226

Mahmood, S. S., Levy, D., Vasan, R. S., & Wang, T. J. (2014). The Framingham Heart Study and the Epidemiology of Cardiovascular Diseases: A Historical Perspective. *Lancet*, *383*(9921), 999–1008. https://doi.org/10.1016/S0140-6736(13)61752-3

Mannan, H., Stevenson, C., Peeters, A., Walls, H., & McNeil, J. (2010). Framingham risk prediction equations for incidence of cardiovascular disease using detailed measures for smoking. *Heart International*, *5*(2), e11. https://doi.org/10.4081/hi.2010.e11

Marshall, M. C. (2005). Diabetes in African Americans. *Postgrad Med J*, *81*, 734–740. https://doi.org/10.1136/pgmj.2004.028274

Matfin, G. (2010). The metabolic syndrome: what's in a name? *Therapeutic Advances in Endocrinology and Metabolism*, *1*. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3475284/pdf/10.1177_2042018810374290.pdf

Mayfield, J. A. (1988). Diagnosis and Classification of Diabetes Mellitus: New Criteria. *American Family Physician*, *58*(6), 1355–1362. Retrieved from http://www.aafp.org/afp/1998/1015/p1355.html

Mccormack, J. P., Levine, M., & Rangno, R. E. (1997). Primary prevention of heart disease and stroke: a simplified approach to estimating risk of events and making drug treatment decisions. Retrieved from http://www.cmaj.ca/content/cmaj/157/4/422.full.pdf

Mendis, S. (2010). The Contribution of the Framingham Heart Study to the Prevention of Cardiovascular Disease: A Global Perspective. *Progress in Cardiovascular Diseases*, *53*, 10–14. https://doi.org/10.1016/j.pcad.2010.01.001

Menke, A., Casagrande, S., Geiss, L., & Cowie, C. C. (2015). Prevalence of and Trends in Diabetes Among Adults in the United States, 1988-2012. *JAMA*, *314*(10), 1021. https://doi.org/10.1001/jama.2015.10029

Mensah, G., Roth, G., Sampson, U., Moran, A., Feigin, V., Forouzanfar, M., … Murray, C. (2015). Mortality from cardiovascular diseases in sub-Saharan Africa, 1990?2013: a systematic analysis of data from the Global Burden of Disease Study 2013: cardiovascular topic. *Cardiovascular Journal Of Africa*, *26*(2), S6–S10. https://doi.org/10.5830/CVJA-2015-036

Miller, M. E., Byington, R. P., Goff, vid C., Forest, W., Thomas Bigger, J., York, N., … Friedewald, W. T. (2008). Effects of Intensive Glucose Lowering in Type 2 Diabetes. *New England Journal of Medicine*, *35824358*, 2545–2559. https://doi.org/10.1056/NEJMoa0802743

Mongraw-Chaffin, M., Bertoni, A. G., Golden, S. H., Mathioudakis, N., Sears, D. D., Szklo, M., & Anderson, C. A. M. (2019). Association of Low Fasting Glucose and HbA1c With Cardiovascular Disease and Mortality: The MESA Study. *Journal of the Endocrine Society*, *3*(5), 892–901. https://doi.org/10.1210/js.2019-00033

Moons, K. G. M., Kengne, A. P., Grobbee, D. E., Royston, P., Vergouwe, Y., Altman, D. G., & Woodward, M. (n.d.). Risk prediction models: II. External validation, model updating, and impact assessment. https://doi.org/10.1136/heartjnl-2011-301247

Moons, K. G. M., Kengne, A. P., Grobbee, D. E., Royston, P., Vergouwe, Y., Altman, D. G., & Woodward, M. (2012). Risk prediction models: II. External validation, model updating, and impact assessment. *British Medical Journal*. https://doi.org/10.1136/heartjnl-2011-301247

Moore, J. X., Chaudhary, N., & Akinyemiju, T. (2017). Metabolic Syndrome Prevalence by Race/Ethnicity and Sex in the United States, National Health and Nutrition Examination Survey, 1988–2012. *Centers for Disease Control and Prevention*, *14*(E24), 1–16. https://doi.org/10.5888/pcd14.160287

Mozaffarian, D., Benjamin, E., Go, A., & et al. (2015). Heart Disease and Stroke Statistics – At-a-Glance. Retrieved June 18, 2017, from

https://www.heart.org/idc/groups/ahamah-public/@wcm/@sop/@smd/documents/downloadable/ucm_470704.pdf

Mozaffarian, D., Benjamin, E. J., Arnett, D. K., Blaha, M. J., Cushman, M., de Ferranti, S., … Fullerton, H. J. (2015). Older Americans & Cardiovascular Disease. Retrieved May 16, 2018, from http://www.heart.org/idc/groups/heart-public/@wcm/@sop/@smd/documents/downloadable/ucm_472923.pdf

National Heart, L. and B. I., & Boston University. (2017). Research Milestones | Framingham Heart Study. Retrieved May 7, 2017, from https://www.framinghamheartstudy.org/about-fhs/research-milestones.php

National Heart Lung & Blood Institute. (2013). Managing Blood Cholesterol in Adults Systematic Evidence Review From the Cholesterol Expert Panel, 2013. *US Dept of Health & Human Services*. Retrieved from https://www.nhlbi.nih.gov/sites/default/files/media/docs/cholesterol-in-adults.pdf

Newsom. (2016). Interactions with Logistic Regression. *Categorical Data Analysis*. Retrieved from http://web.pdx.edu/~newsomj/cdaclass/ho_interactions.pdf

Nguyen, Q. N., Pham, S. T., Do, L. D., Nguyen, V. L., Wall, S., Weinehall, L., … Byass, P. (2012). Cardiovascular Disease Risk Factor Patterns and Their Implications for Intervention Strategies in Vietnam. *International Journal of Hypertension*, *11*. https://doi.org/10.1155/2012/560397

NHLBI, & NIH. (2002). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) Final Report. *Circulation*,

106(25).

Nichols, G. A., Horberg, M., Koebnick, C., Young, D. R., Waitzfelder, B., Sherwood, N. E., … Ferrara, A. (2012). Cardiometabolic Risk Factors Among 1.3 Million Adults With Overweight or Obesity, but Not Diabetes, in 10 Geographically Diverse Regions of the. *United States Prev Chronic Dis*, *14*. https://doi.org/10.5888/pcd14.160438

Nieto, F. J. (1999a). *Cardiovascular Disease and Risk Factor Epidemiology: A Look Back at the Epidemic of the 20th Century*. *American Journal of Public Health* (Vol. 89). Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1508603/pdf/amjph00003-0014.pdf

Nieto, F. J. (1999b). Cardiovascular Disease and Risk Factor Epidemiology: A Look Back at the Epidemic of the 20th Century. *Am J Public Health*, *89*, 292–294. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1508603/pdf/amjph00003-0014.pdf

Nolan, C. J., Damm, P., & Prentki, M. (2011). Seminar Type 2 diabetes across generations : from pathophysiology to prevention and management. *The Lancet*, *6736*(11). Retrieved from https://www.researchgate.net/profile/Marc_Prentki/publication/288258284_Type_2_diabetes_across_generations_from_pathophysiology_to_prevention_and_management/links/568ab4e508aebccc4e1a16c4/Type-2-diabetes-across-generations-

from-pathophysiology-to-preventi

Nordestgaard, B. G., Chapman, M. J., Humphries, S. E., Ginsberg, H. N., Masana, L., Descamps, O. S., … Tybjaerg-Hansen, A. (2013). Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: Consensus Statement of the European Atherosclerosis Society. *European Heart Journal*, *34*(45), 3478–3490. https://doi.org/10.1093/eurheartj/eht273

O'Donnell, C., & Elosua, R. (2008). Cardiovascular Risk Factors. Insights From Framingham Heart Study. *Rev Esp Cardiol*, *61*(3), 299–310. Retrieved from http://appswl.elsevier.es/watermark/ctl_servlet?_f=10&pident_articulo=13117552&pident_usuario=0&pcontactid=&pident_revista=255&ty=142&accion=L&origen=cardio&web=www.revespcardiol.org&lan=en&fichero=255v61n03a13117552pdf001.pdf&anuncioPdf=ERROR_publi_pdf

Ockene, I., Daley, J., & Tran, J. (2014). Cardiovascular Disease: How it Became a Problem, Risk Factors, and the Role of Public Policy. In *2014 massachusetts family impact seminar*. Retrieved from https://www.purdue.edu/hhs/hdfs/fii/wp-content/uploads/2015/07/s_mafis05c01.pdf

Pandya, A., Weinstein, M. C., & Gaziano, T. A. (2011). A Comparative Assessment of Non-Laboratory-Based versus Commonly Used Laboratory-Based Cardiovascular Disease Risk Scores in the NHANES III Population. *PLOS*, *6*. https://doi.org/10.1371/journal.pone.0020416

Park, C., Guallar, E., Linton, J. A., Lee, D., Jang, Y., Koog Son, D., … Samet, J. M.

(2013). Fasting Glucose Level and the Risk of Incident Atherosclerotic Cardiovascular Diseases. *Epidemiology/Health Services Research*, *36*. https://doi.org/10.2337/dc12-1577

Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal inference in statistics : a primer*. Retrieved from https://www.wiley.com/en-us/Causal+Inference+in+Statistics%3A+A+Primer-p-9781119186847

Pedersen, K. (2002). Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *National Cholesterol Education Program National Heart, Lung, and Blood Institute*, *02–5235*(December). https://doi.org/http://www.royalcommission.vic.gov.au/finaldocuments/summary/PF/VBRC_Summary_PF.pdf

Pencina, M. J., D'Agostino, R. B., Larson, M. G., Massaro, J. M., Vasan, R. S., & Vasan, R. S. (2009). Predicting the 30-year risk of cardiovascular disease: the framingham heart study. *Circulation*, *119*(24), 3078–3084. https://doi.org/10.1161/CIRCULATIONAHA.108.816694

Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). *An Introduction to Logistic Regression Analysis and Reporting*. Retrieved from https://datajobs.com/data-science-repo/Logistic-Regression-[Peng-et-al].pdf

Perreault, L., Temprosa, M., Mather, K. J., Horton, E., Kitabchi, A., Larkin, M., … Group, D. P. P. R. (2014). Regression From Prediabetes to Normal Glucose Regulation Is Associated With Reduction in Cardiovascular Risk: Results From the Diabetes Prevention Program Outcomes Study. *Diabetes Care*, *37*.

https://doi.org/10.2337/dc14-0656

Piepoli, M. F., Hoes, A. W., Agewal, S., Albus, C., Brotons, C., Catapano, A. L., … Hall, M. S. (2016). 2016 European Guidelines on cardiovascular disease prevention in clinical practice. *European Heart Journal*, *37*. https://doi.org/10.1093/eurheartj/ehw106

Pike, M. M., Decker, P. A., Larson, N. B., St Sauver, J. L., Takahashi, P. Y., Roger, V. L., … Author, C. (2016). Improvement in Cardiovascular Risk Prediction with Electronic Health Records HHS Public Access. *J Cardiovasc Transl Res. J Cardiovasc Transl Res*, *9*(3), 214–222. https://doi.org/10.1007/s12265-016-9687-z

Qazi, M. U., & Malik, S. (2013a). Diabetes and Cardiovascular Disease: Insights from the Framingham Heart Study. *Global Heart*, *8*, 43–48. https://doi.org/10.1016/j.gheart.2012.12.008

Qazi, M. U., & Malik, S. (2013b). Diabetes and Cardiovascular Disease: Insights from the Framingham Heart Study. *Global Heart*, *8*(1), 43–48. https://doi.org/10.1016/J.GHEART.2012.12.008

Raggi, P., Shaw, L. J., Berman, D. S., & Callister, T. Q. (2004). Prognostic Value of Coronary Artery Calcium Screening in Subjects With and Without Diabetes. *J Am Coll Cardiol*, *43*, 1663–1669. https://doi.org/10.1016/j.jacc.2003.09.068

Ray, K. K., Kastelein, J. J. P., Matthijs Boekholdt, S., Nicholls, S. J., Khaw, K.-T., Ballantyne, C. M., … Lü Scher, T. F. (2014). The ACC/AHA 2013 guideline on the treatment of blood cholesterol to reduce atherosclerotic cardiovascular disease risk in adults. *European Heart Journal*, *35*, 960–968.

https://doi.org/10.1093/eurheartj/ehu107

Reaven, P. D., Moritz, T. E., Schwenke, D. C., Anderson, R. J., Criqui, M., Detrano, R., … Duckworth, W. (2009). Intensive Glucose-Lowering Therapy Reduces Cardiovascular Disease Events in Veterans Affairs Diabetes Trial Participants With Lower Calcified Coronary Atherosclerosis. *Diabetes*, *58*. https://doi.org/10.2337/db09

Reddy, K. S., & Satija, A. (2010). The Framingham Heart Study: Impact on the Prevention and Control of Cardiovascular Diseases in India. *Progress in Cardiovascular Diseases*, *53*, 21–27. https://doi.org/10.1016/j.pcad.2010.02.011

Roberts, C. K., Hevener, A. L., & Barnard, R. J. (2014). Metabolic Syndrome and Insulin Resistance: Underlying Causes and Modification by Exercise Training. *Comparative Physiology*, *3*. https://doi.org/10.1002/cphy.c110062

Rodondi, N., Locatelli, I., Aujesky, D., Butler, J., Vittinghoff, E., Simonsick, E., … Health ABC Study, for the H. A. (2012). Framingham risk score and alternatives for prediction of coronary heart disease in older adults. *PloS One*, *7*(3), e34287. https://doi.org/10.1371/journal.pone.0034287

Roger, V. L., Go, A. S., Lloyd-Jones, D. M., Benjamin, E. J., Berry, J. D., Borden, W. B., … Turner, M. B. (2012). Heart disease and stroke statistics-2012 update: A report from the American heart association. *Circulation*, *125*(1). https://doi.org/10.1161/CIR.0b013e31823ac046

Root, M., Hu, G., & Duncan, A. (2014). Adding multiple risk factors improves Framingham coronary heart disease risk scores. *Vascular Health and Risk*

Management, *10*, 557. https://doi.org/10.2147/VHRM.S69672

Rosen G. (1958). *A History of Public Health*. Retrieved from http://booksite.elsevier.com/samplechapters/9780123708908/Sample_Chapters/02~Chapter_1.pdf

Rosenstock, I. M. (1974). The Health Belief Model and Preventive Health Behavior. *Health Education & Behavior*, *2*(4), 354–386. https://doi.org/10.1177/109019817400200405

Roumie, C. L., Hung, A. M., Greevy, R. A., Grijalva, C. G., Liu, X., Murff, H. J., … Griffin, M. R. (2012). Comparative effectiveness of sulfonylurea and metformin monotherapy on risk of cardiovascular events in type 2 diabetes mellitus. *Annals of Internal Medicine*, *157*. https://doi.org/10.7326/0003-4819-157-9-201211060-00003

Rücker, V., Keil, U., Fitzgerald, A. P., Malzahn, U., Prugger, C., Ertl, G., … Neuhauser, H. (2016). Predicting 10-Year Risk of Fatal Cardiovascular Disease in Germany: An Update Based on the SCORE-Deutschland Risk Charts. *PLOS*. https://doi.org/10.1371/journal.pone.0162188

Rudestam, K. E., & Newton, R. R. (2014). *Surviving your dissertation : a comprehensive guide to content and process* (3 rd). Sage. Retrieved from https://us.sagepub.com/en-us/nam/surviving-your-dissertation/book239510

Sah, S. P., Singh, B., Choudhary, S., & Kumar, A. (2016). Animal models of insulin resistance: A review. *Pharmacological Reports*, *68*(6), 1165–1177. https://doi.org/10.1016/J.PHAREP.2016.07.010

Saisho, Y. (2014). Glycemic variability and oxidative stress: A link between diabetes and cardiovascular disease? *International Journal of Molecular Sciences*, *15*(10), 18381–18406. https://doi.org/10.3390/ijms151018381

Sarwar, N., Gao, P., Seshasai, S. R. K., Gobin, R., Kaptoge, S., Di Angelantonio, E., … Danesh, J. (2010). Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *Lancet (London, England)*, *375*(9733), 2215–2222. https://doi.org/10.1016/S0140-6736(10)60484-9

Scheerbaum, M., Langenbach, C., Scheerbaum, P., Heidemann, F., Rieß, H. C., Heigel, H., … Behrendt, C.-A. (2017). Prevalence of cardiovascular risk factors among 28,000 employees. *Vasa*, *46*(3), 203–210. https://doi.org/10.1024/0301-1526/a000611

Schlendorf, K. H., Nasir, K., & Blumenthal, R. S. (2009). Limitations of the Framingham risk score are now much clearer. *Preventive Medicine*, *48*(2), 115–116. https://doi.org/10.1016/J.YPMED.2008.12.002

Selvin, E., Steffes, M. W., Zhu, H., Matsushita, K., Wagenknecht, L., Pankow, J., … Brancati, F. L. (2010). Glycated Hemoglobin, Diabetes, and Cardiovascular Risk in Nondiabetic Adults. *New England Journal of Medicine*, *362*(9), 800–811. https://doi.org/10.1056/nejmoa0908359

Shaw, J. E., Sicree, R. A., & Zimmet, P. Z. (2010). Global estimates of the prevalence of diabetes for 2010 and 2030. *Diabetes Research and Clinical Practice*, *87*(1), 4–14. https://doi.org/10.1016/j.diabres.2009.10.007

Shaw, J. E., Zimmet, P. Z., McCarty, D., & de Courten, M. (2000). Type 2 Diabetes Worldwide According to the New Classification and Criteria. *Diabetes Care*, (23), 5–10. Retrieved from http://journal.diabetes.org/diabetescare/FullText/Supplements/DiabetesCare/Supplement400/B5.asp

Singhal, A., Tien, Y. Y., & Hsia, R. Y. (2016). Racial-ethnic disparities in opioid prescriptions at emergency department visits for conditions commonly associated with prescription drug abuse. *PLoS ONE*, *11*(8), 1–14. https://doi.org/10.1371/journal.pone.0159224

Skov, V., Cangemi, C., Gram, J., Christensen, M. M., Grodum, E., Sørensen, D., … Rasmussen, L. M. (2014). Metformin, but not rosiglitazone, attenuates the increasing plasma levels of a new cardiovascular marker, fibulin-1, in patients with type 2 diabetes. *Diabetes Care*, *37*(3), 760–766. https://doi.org/10.2337/dc13-1022

Sorlie, P., & Coady, S. (2004). Framingham Heart Study Longitudinal Data Documentation. *National Heart, Lung, and Blood Institute* . Retrieved from https://www.stat.ncsu.edu/courses/sibs/datasets/framingham/Framingham.Longitudinal.Data.Documentation.pdf

Sperandei, S. (2014). Understanding logistic regression analysis Lessons in biostatistics. *Biochemia Medica*, *24*(1), 12–18. https://doi.org/10.11613/BM.2014.003

Srikanthan, K., Feyh, A., Visweshwar, H., Shapiro, J., & Sodhi, K. (2016). Systematic

Review of Metabolic Syndrome Biomarkers: A Panel for Early Detection, Management, and Risk Stratification in the West Virginian Population. *International Journal of Medical Sciences*, *13*. https://doi.org/10.7150/ijms.13800

Stehbens, W. E. (1995). The Hypothetical Epidemic of Coronary Heart Disease and Atherosclerosis. *Medical Hypotheses*, *45*, 449–454. Retrieved from http://ac.els-cdn.com.ezp.waldenulibrary.org/0306987795902198/1-s2.0-0306987795902198-main.pdf?_tid=db4959ce-7635-11e7-8323-00000aacb35d&acdnat=1501536115_53984a7d8696e876c56340829e6877a5

Steyerberg, E. W., Vickers, A. J., Cook, N. R., Gerds, T., Gonen, M., Obuchowski, N., … Kattan, M. W. (2010). Assessing the performance of prediction models: a framework for some traditional and novel measures. *Epidemiology*, *21*(1), 128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2

Stoltzfus, J. C. (2011). Logistic Regression: A Brief Primer. *ACADEMIC EMERGENCY MEDICINE*, *18*, 1099–1104. https://doi.org/10.1111/j.1553-2712.2011.01185.x

Stone, N. J., Jennifer Robinson, C., Chair Alice Lichtenstein, V. H., Chair Noel Bairey Merz, V. C., Donald Lloyd-Jones, F. M., Conrad Blum, F. B., … Win-Kuang Shen, F. (2013). 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults. *Circulation*. Retrieved from http://circ.ahajournals.org/content/circulationaha/early/2013/11/11/01.cir.0000437738.63853.7a.full.pdf

Sullivan, L. M. (n.d.). Confidence Intervals. Retrieved May 17, 2018, from http://sphweb.bumc.bu.edu/otlt/MPH-Modules/BS/BS704_Confidence_Intervals/BS704_Confidence_Intervals_print.html

Tabák, A. G., Herder, C., Rathmann, W., Brunner, E. J., & Kivimäki, M. (2012). Prediabetes: A high-risk state for developing diabetes, *379*(9833). https://doi.org/10.1016/S0140-6736(12)60283-9

The ACCORD Study Group. (2011). Action to Control Cardiovascular Risk in Diabetes (ACCORD) Trial. *Diabetes Care*, *34*. Retrieved from https://ac-els-cdn-com.ezp.waldenulibrary.org/S0002914907004389/1-s2.0-S0002914907004389-main.pdf?_tid=7b3a4a7a-a7cc-4d5d-a765-79f232131fee&acdnat=1524246204_35592b1a0a2de40c684f8b814dbcf5f3

The Emerging Risk Factors Collaboration. (2010). Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies. *The Lancet*, *375*(9733), 2215–2222. https://doi.org/10.1016/S0140-6736(10)60484-9

Tikkanen, E., Pirinen, M., Sarin, A.-P., Havulinna, A. S., Männistö, S., Saltevo, J., … Ripatti, S. (2016). Genetic support for the causal role of insulin in coronary heart disease. *Diabetologia*, *59*(11), 2369–2377. https://doi.org/10.1007/s00125-016-4081-6

Tomasik, T., Krzysztoń, J., Dubas-Jakóbczyk, K., Kijowska, V., & Windak, A. (2015). The Systematic Coronary Risk Evaluation (SCORE) for the prevention of

cardiovascular diseases . Protocol of a systematic review. *Family Medicine Topics*, *XVII*(2), 72–79.

Tostes, R. C., & et al. (2009). New Evidence Of How High Glucose Damages Blood Vessels Could Lead To New Treatments -- ScienceDaily. Retrieved March 14, 2018, from https://www.sciencedaily.com/releases/2009/05/090511140951.htm

Tripepi, G., Jager, K., Dekker, F., Wanner, C., & Zoccali, C. (2007). Measures of effect: Relative risks, odds ratios, risk difference, and 'number needed to treat.' *Kidney International*, *72*. Retrieved from https://ac.els-cdn.com/S0085253815527464/1-s2.0-S0085253815527464-main.pdf?_tid=fe39c451-5d68-4956-969c-c0e5df7cc02c&acdnat=1523638947_a681703d312535e63571691d9236413e

Truett, J., Cornfield, J., & Kannel, W. B. (1967). A MULTIVARIATE ANALYSIS OF THE RISK OF CORONARY HEART DISEASE IN FRAMINGHAM. *I. Chron. Dis*, *20*, 511–524. Retrieved from https://waldenu-illiad-oclc-org.ezp.waldenulibrary.org/illiad/illiad.dll?Action=10&Form=75&Value=29609

Tsao, C. W. (2015). The Framingham Heart Study: past, present and future. *International Journal of Epidemiology*, *44*(6), 1763–1768. https://doi.org/10.1093/ije/dyv336

Tsao, C. W., & Vasan, R. S. (2015a). Cohort Profile: The Framingham Heart Study (FHS): overview of milestones in cardiovascular epidemiology. *International Journal of Epidemiology*, *44*(6), 1800–1813. https://doi.org/10.1093/ije/dyv337

Tsao, C. W., & Vasan, R. S. (2015b). Cohort Profile: The Framingham Heart Study

(FHS): overview of milestones in cardiovascular epidemiology. *International Journal of Epidemiology*, *44*(6), 1800–1813. https://doi.org/10.1093/ije/dyv337

Tunstall-Pedoe, H. (2011). Cardiovascular Risk and Risk Scores: ASSIGN, Framingham, QRISK and others: how to choose. *BMJ*, *97*. Retrieved from http://www.assign-score.com/downloads/Heart2011.pdf

Tunstall-Pedoe, H., Woodward, M., Tavendale, R., Brook, R. A., & Mccluskey, M. K. (n.d.). Comparison of the prediction by 27 different factors of coronary heart disease and death in men and women of the Scottish heart health study: cohort study. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2127508/pdf/9314758.pdf

U. S. Department of Commerce. (n.d.). U.S. Census Bureau QuickFacts: UNITED STATES. Retrieved April 23, 2019, from https://www.census.gov/quickfacts/fact/table/US/LFE046217

U.S. Department of Health and Human Services. (2005). Theory At a Glance: A Guide For Health Promotion Practice. *NIH Cancer Control and Population Science*. Retrieved from http://www.sbccimplementationkits.org/demandrmnch/wp-content/uploads/2014/02/Theory-at-a-Glance-A-Guide-For-Health-Promotion-Practice.pdf

U.S. Department of Health and Human Services. (2018). Heart Disease and Stroke | Healthy People 2020. Retrieved March 11, 2018, from https://www.healthypeople.gov/2020/topics-objectives/topic/heart-disease-and-stroke

Valentino, G., Kramer, V., Orellana, L., Bustamante, M. J., Casasbellas, C., Adasme, M., … Acevedo, M. (2015). Impaired Fasting Glucose in Nondiabetic Range: Is It a Marker of Cardiovascular Risk Factor Clustering? *Disease Markers*. https://doi.org/10.1155/2015/804739

Van Staa, T.-P., Gulliford, M., Ng, E. S.-W., Goldacre, B., & Smeeth, L. (2014). Prediction of Cardiovascular Risk Using Framingham, ASSIGN and QRISK2: How Well Do They Predict Individual Rather than Population Risk? *PLOS ONE*, *9*(10). https://doi.org/10.1371/journal.pone.0106455

Vistisen, D., Andersen, G. S., Hansen, C. S., Hulman, A., Henriksen, J. E., Bech-Nielsen, H., & Jorgensen, M. E. (2016). Prediction of First Cardiovascular Disease Event in Type 1 Diabetes Mellitus. *Circulation*, *133*(2016), 1058–1066. https://doi.org/10.1161/CIRCULATIONAHA.115.018844

Wareham, N. (1998). The changing classification and diagnosis of diabetes. *British Medical Journal*, *317*, 359–360. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1113665/pdf/359.pdf

Westendorp, R. G. J., Assendelft, W. J. J., Den Elzen, W. P. J., De Craen, A. J. M., Le Cessie, S., & Gussekloo, J. (2009). Use of Framingham risk score and new biomarkers to predict cardiovascular mortality in older people: population based observational cohort study Wouter de Ruijter, general practitioner and clinical researcher. *BMJ*. https://doi.org/10.1136/bmj.a3083

Wilson, P. W. F., D'Agostino, R. B., Levy, D., Belanger, A. M., Silbershatz, H., & Kannel, W. B. (1998). Prediction of Coronary Heart Disease Using Risk Factor

Categories. *Circulation*, *97*(18). Retrieved from http://circ.ahajournals.org/content/97/18/1837.long

Wilson, P. W. F., & Meigs, J. B. (2008). Cardiometabolic risk: a Framingham perspective. *International Journal of Obesity*, *32*, 17–20. https://doi.org/10.1038/ijo.2008.30

Woodruff, V. D. (2012). Physiologic capacity as a predictor of postoperative complications and associated costs in three types of oncological surgeries. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, *73*(4-B), 2170. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2012-99200-108&lang=es&site=ehost-live

Woodward, M., Brindle, P., & Tunstall-Pedoe, H. (2007). Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC) for the SIGN group on risk estimation*. *Heart*, *93*, 172–176. https://doi.org/10.1136/hrt.2006.108167

World Health Organization. (2016). WHO | Cardiovascular diseases (CVDs). Retrieved May 7, 2017, from http://www.who.int/mediacentre/factsheets/fs317/en/

Xu, J., & Rajaratnam, R. (2017). Cardiovascular safety of non-insulin pharmacotherapy for type 2 diabetes. *Cardiovascular Diabetology*, *2017*, 16–18. Retrieved from http://download.springer.com/static/pdf/170/art%253A10.1186%252Fs12933-017-0499-5.pdf?originUrl=http%3A%2F%2Fcardiab.biomedcentral.com%2Farticle%2F10.1

186%2Fs12933-017-0499-5&token2=exp=1494363182~acl=%2Fstatic%2Fpdf%2F170%2Fart%25253A10.1186%25252Fs12933

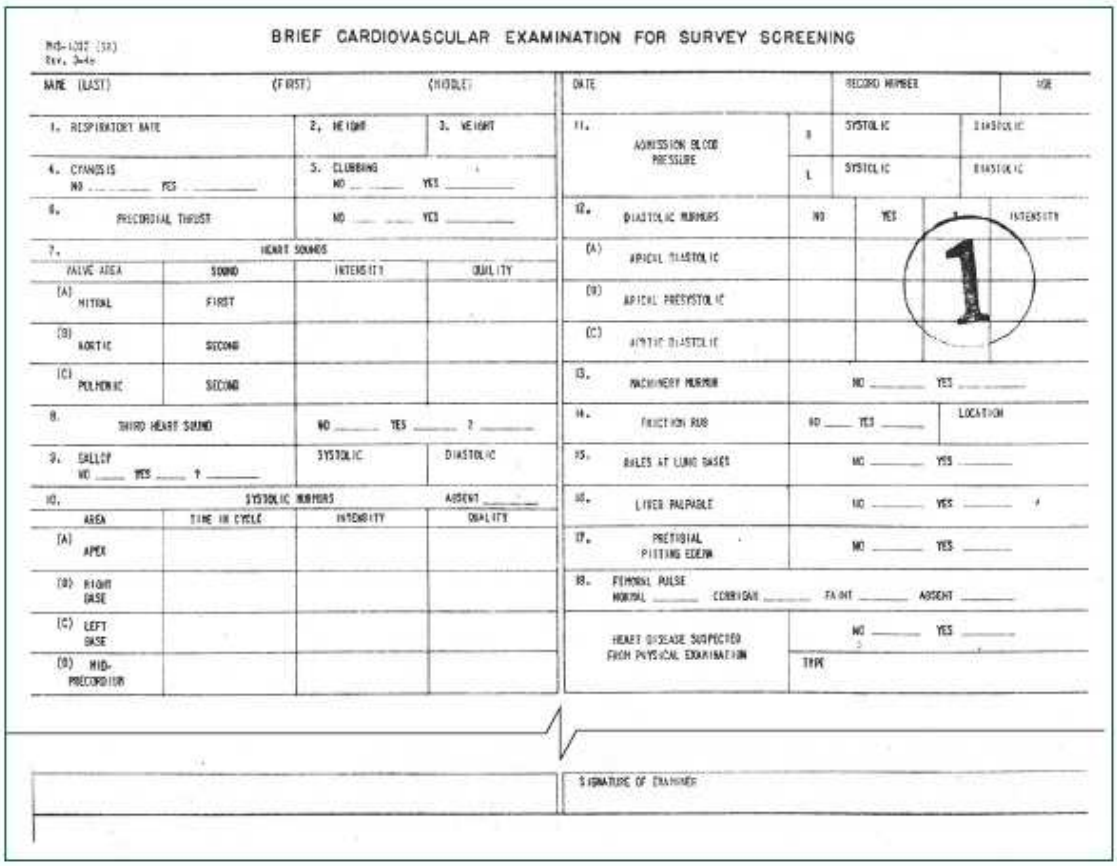
Yeung, D. C.-S., Yuan, X., Hui, S. S.-C., & Feresu, S. A. (2016). Determinants of moderate to vigorous physical activity and obesity in children: a structural equation modeling analysis. *World Journal of Pediatrics : WJP*, *12*(2), 170–176. https://doi.org/10.1007/s12519-015-0057-8

Yosaputra, C., Kholinne, E., & Susanto Taufik, E. (2010). High Framingham risk score decreases quality of life in adults UNIVERSA MEDICINA. *Univ Med*, *29*(11111), 27–33. Retrieved from http://www.univmed.org/wp-content/uploads/2011/02/koas.pdf

Zhang, X. G., Zhang, Y. Q., Zhao, D. K., Wu, J. X., Zhao, X. M., Jiao, B., & Chen, X. F. (2014). Relationship between blood glucose fluctuation and macrovascular endothelial dysfunction in type 2 diabetic patients with coronary heart disease. *European Review for Medical and Pharmacological Sciences*, *18*, 3593–3600. Retrieved from http://www.europeanreview.org/wp/wp-content/uploads/3593-3600.pdf

Zoungas, S., de Galan, B. E., Ninomiya, T., Grobbee, D., Hamet, P., Heller, S., … Travert, F. (2009). Combined effects of routine blood pressure lowering and intensive glucose control on macrovascular and microvascular outcomes in patients with type 2 diabetes: New results from the ADVANCE trial. *Diabetes Care*, *32*(11), 2068–2074. https://doi.org/10.2337/dc09-0959

Appendix A: Original FHS Questionnaire



BRIEF CARDIOVASCULAR EXAMINATION FOR SURVEY SCREENING

NAME (LAST) (FIRST) (MIDDLE) | DATE | RECORD NUMBER | AGE

1. RESPIRATORY RATE | 2. HEIGHT | 3. WEIGHT

4. CYANOSIS NO _____ YES _____ | 5. CLUBBING NO _____ YES _____

6. PRECORDIAL THRUST NO _____ YES _____

7. HEART SOUNDS

VALVE AREA	SOUND	INTENSITY	QUALITY
(A) MITRAL	FIRST		
(B) AORTIC	SECOND		
(C) PULMONIC	SECOND		

8. THIRD HEART SOUND NO _____ YES _____ ? _____

9. GALLOP NO _____ YES _____ ? _____ | SYSTOLIC | DIASTOLIC

10. SYSTOLIC MURMURS ABSENT _____

AREA	TIME IN CYCLE	INTENSITY	QUALITY
(A) APEX			
(B) RIGHT BASE			
(C) LEFT BASE			
(D) MID-PRECORDIUM			

11. ADMISSION BLOOD PRESSURE R SYSTOLIC DIASTOLIC L SYSTOLIC DIASTOLIC

12. DIASTOLIC MURMURS NO YES INTENSITY

(A) APICAL DIASTOLIC

(B) APICAL PRESYSTOLIC

(C) AORTIC DIASTOLIC

13. MACHINERY MURMUR NO _____ YES _____

14. FRICTION RUB NO _____ YES _____ LOCATION

15. RALES AT LUNG BASES NO _____ YES _____

16. LIVER PALPABLE NO _____ YES _____

17. PRETIBIAL PITTING EDEMA NO _____ YES _____

18. FEMORAL PULSE NORMAL _____ CORRIGAN _____ FAINT _____ ABSENT _____

HEART DISEASE SUSPECTED FROM PHYSICAL EXAMINATION NO _____ YES _____ TYPE

SIGNATURE OF EXAMINER

Appendix B: Signed NHLBI/BIOlincc Agreement Form

NHLBI Research Materials Distribution Agreement (RMDA)

Introduction and Definitions

The National Heart, Lung, and Blood Institute (NHLBI), the RECIPIENT Organization (RECIPIENT) and the Principal Investigator (PI) hereby enter into this Research Materials Distribution Agreement (RMDA) as of the effective date specified on the final signature page.

The Research Materials and Research Plan covered by this RMDA are:

- Name of Clinical Study: FRAMCOHORT, GEN3, FRAMOFFSPRING
- Title of Research Plan: The role of blood glucose level on the performance of the Framingham CHD Risk Score
- Research Materials Requested: Data
- Research Plan includes a Commercial Purpose: No
- Name of Principal Investigator (PI): Dr. Shinga A. Feresu
- Email of Principal Investigator (PI): shingairai.feresu@mail.waldenu.edu
- Name of Other Approved Users at PI's Institution: Uohna J. Thiessen

The Research Materials are provided through the Biologic Specimen and Data Repository Information Coordinating Center. The Center was established by the NHLBI to develop and maintain the infrastructure necessary to facilitate and maximize access to Research Materials from NHLBI-sponsored studies in accordance with NHLBI approved procedures.

The Research Materials were collected as part of the above clinical study, hereafter referred to as "STUDY". They constitute a unique scientific resource and the NHLBI is committed to making them available in a timely manner, on appropriate terms and conditions, to the largest possible number of qualified investigators who wish to analyze the materials in a secondary study designed to enhance the public health benefit of the original work. The RECIPIENT and PI acknowledge responsibility for ensuring the review of and agreement to the terms within this RMDA and the appropriate research use of the Research Materials, subject to applicable laws and regulations.

The RECIPIENT and PI acknowledge that other researchers are entitled to access to the Research Materials on the same terms as RECIPIENT so that duplication of research may occur. RECIPIENT and PI also recognize that the STUDY Investigators have made a substantial long-term contribution in establishing the Research Materials and the NHLBI encourages appropriate collaborative relationships by outside investigators with the STUDY Investigators and proper acknowledgement of their contributions.

The NHLBI believes that the confidentiality and privacy of the STUDY participants can best be assured by requiring all who are interested in accessing the Research Materials to acknowledge their review of this RMDA and agree to adhere to its provisions. Violation of its confidentiality provisions could lead to legal action on the part of STUDY participants, their families, or the U.S. Government.

Note: RECIPIENT requests access to NHLBI Research Materials for its PI at its sole risk.

For the purpose of this Agreement

"RECIPIENT" is any organization that is seeking access to STUDY Research Materials, and may be a: Public/State Controlled Institution of Higher Education; Private Institution of Higher Education; Nonprofit organization with 501(c)(3) IRS Status (Other than Institution of Higher Education); Nonprofit Organization without 501(c)(3) IRS Status (Other than Institution of Higher Education); Small Business; For-Profit Organization (Other than Small Business); State Government; Government of a U.S. Territory or Possession; Non-domestic (non-U.S.) Entity (Foreign Organization); or Eligible Agency of the U.S. Government.

"Principal Investigator (PI)" is an individual judged by the RECIPIENT to have the appropriate level of authority and responsibility to lead the scientific investigation proposed in the Research Plan using the requested materials, oversee the supporting staff who are provided access to the Research Materials and contribute to the analytic effort and public disclosure of STUDY results, and assume responsibility for all team members' compliance with the terms and conditions of this RMDA.

"APPROVED USERS" are all individuals specifically identified in the Research Plan, including the PI. Only individuals listed in the Research Plan may have access to the Research Materials.

"Research Plan" is a description of the proposed research that includes the identities of the investigators participating in the research effort. The Research Plan must include the project title, the RECIPIENT's name, the PI's name, the name of other APPROVED USERS, and the proposed research protocol with the research objectives and design. For plans including biospecimens, the biospecimen material type, number, minimum volume, and required characteristics needed to meet the objectives of the protocol must also be included.

"Research Materials" are the requested materials covered by this RMDA and may include STUDY data, defined as clinical or epidemiologic subject data, and/or STUDY biospecimens. STUDY biospecimens may have associated characterization data. Characterization data serve to describe STUDY biospecimens only and are not considered to be STUDY data; they are exempt from STUDY data requirements that may be described elsewhere in this RMDA.

"STUDY" is the clinical study that collected the Research Materials described in this RMDA.

"STUDY Investigator" is a research investigator with a current or previous grant, contract or consulting agreement with the NHLBI, or one of its contractors, to work on the STUDY.

Terms of Access

1. **Research Use**

   The RECIPIENT and APPROVED USERS agree that they will use the Research Materials solely in connection with the research project described in the Research Plan named in this RMDA. Substantive modifications to the research project will require submission of a revised RMDA.

2. **Institutional and Approved User Responsibilities**

   RECIPIENT and APPROVED USERS acknowledge that RECIPIENT's Institutional Review Board (IRB) has reviewed the Research Plan and either approved it or determined that it is exempt from review. Access to Research Materials from some STUDIES requires IRB approval and/or compliance with other limitations, and RECIPIENT agrees to abide by all such conditions and limitations on the Research Materials. RECIPIENT certifies that its IRB is operating under an Office of Human Research Protections (OHRP) - approved Assurance and in accordance with Department of Health and Human Services regulations at 45 CFR Part 46. RECIPIENT and APPROVED USERS agree to comply fully with all such conditions.

   RECIPIENT and APPROVED USERS agree to report promptly to the NHLBI any proposed change in the Research Plan and any unanticipated problems involving risks to subjects or others. Changes to the Research Plan include changes in the APPROVED USERS list. This RDMA is made in addition to, and does not supersede, any of RECIPIENT's institutional policies or any local, State, and/or Federal laws and regulations that provide additional protections for human subjects.

   Evidence of local IRB review and/or approval (where appropriate) from an expedited or convened review to conduct the Research Plan with the requested STUDY data must be included in a supplemental Adobe PDF document that will be uploaded during the application process and attached to the RMDA form.

3. **Public Posting of Approved User's Research Use Statement**

   The RECIPIENT and PI agree that information about the proposed research use can be posted on a public web site that describes the project(s) included in the Research Plan. The information will include the PI's name, RECIPIENT institution, project title, and a brief summary of the research. In addition, citations resulting from the use of Research Materials may be posted on the Biologic Specimen and Data Repository Information Coordinating Center Website.

4. **Non-Identification**

   The PI agrees not to use the Research Materials, either alone or in concert with any other information, to identify or contact individual STUDY subjects without specific approval to contact STUDY subjects obtained from the IRB(s) responsible for the STUDY.

5. **Non-Transferability of Research Materials**

   The RECIPIENT and PI agree to retain control over the Research Materials, and further agree not to release or distribute Research Materials in any form to any entity or individual unless required by NHLBI policies. The RECIPIENT and PI agree to store Research Material data on a computer with adequate security controls (see Section 6), and to maintain appropriate control over the Research Materials at all times. Research Materials data containing individual-level information, in whole or in part, may not be sold to any entity or individual at any point in time for any purpose.

   The PI agrees that if his or her relationship with the RECIPIENT terminates and a relationship with a different RECIPIENT is established during the period of the RMDA, a new RMDA from the second RECIPIENT will be submitted and approved before the PI resumes use of the Research Materials. Any versions of Research Material data stored at the first RECIPIENT will be destroyed and their destruction documented. However, if advance written notice and approval by the NHLBI Program Office is obtained to transfer responsibility for the approved Research Plan to a different PI with a relationship with the first RECIPIENT, the Research Material data may not need to be destroyed.

6. **Security of Research Materials**

   The RECIPIENT and PI agree to store Research Material data on a computer with security controls adequate to protect sensitive or identifiable information, to ensure that only approved, supervised persons have access to the data, and to maintain appropriate control over the Research Materials at all times. Hard copies of any Research Material must similarly be stored under conditions sufficiently secure to avoid inappropriate access, and shredded prior to discarding.

   This RMDA will be in effect for a period of three (3) years from its effective date for the requested STUDY data set. At the end of the three (3) year period, the RECIPIENT and PI agree to destroy all copies of the STUDY data, and all derivatives that contain individual-level information. Characterization data associated with the STUDY biospecimens are exempt from this requirement.

   An extension of this RMDA may be permitted by the NHLBI upon submission by the PI and RECIPIENT of evidence of IRB approval for the extended period.

7. **Intellectual Property**

   By requesting access to the STUDY Research Materials, the REQUESTER and APPROVED USERS acknowledge the intent of the NHLBI to see that anyone authorized for research access through the attached Research Plan, follow the intellectual property principles within the NIH GWAS Policy for Data Sharing (https://grants.nih.gov/grants/guide/notice-files/NOT-OD-07-088.html) as summarized below:

   - Achieving maximum public benefit is the ultimate goal of Research Material distribution through the NHLBI Biological and Data Repository Information Coordinating Center. The NIH believes that Research Materials, such as these covered by this RMDA, should be considered as pre-competitive, and urges APPROVED USERS to avoid making IP claims derived directly from the STUDY Research Materials. However, the NIH also recognizes the importance of the subsequent development of IP on downstream discoveries, especially in therapeutics, which will be necessary to support full investment in products to benefit the public.
   - It is expected that these NHLBI-provided data, and conclusions derived there from, will remain freely available, without requirement for licensing. The NIH encourages broad use of shared Research Materials coupled with a responsible approach to management of intellectual property derived from downstream discoveries in a manner consistent with the NIH's Best Practices for the Licensing of Genomic Inventions (https://www.ott.nih.gov/sites/default/files/documents/pdfs/70fr18413.pdf) and the NIH Research Tools Policy (https://grants.nih.gov/grants/intell-property_64FR72090.pdf).

8. **Acknowledgement of BioLINCC Research Resources**

   RECIPIENT agrees to acknowledge the contribution of the STUDY in all oral and written presentations, disclosures, or publications resulting from any analyses conducted on the STUDY Research Materials.

   If the Research Plan involves collaboration with STUDY Investigators, then the APPROVED USERS will comply with all policies established by the STUDY's publications committee. In addition, the APPROVED USERS will acknowledge the source of the data by including language similar to the following either in the acknowledgment or in the text of the manuscript: "This manuscript was prepared using FRAMCOHORT, GEN3, FRAMOFFSPRING Research Materials obtained from the NHLBI".

   If the Research Plan does not involve collaboration with STUDY Investigators or the STUDY has ended, the RECIPIENT will acknowledge the source of the data by including language similar to the following either in the acknowledgment or in the text of the manuscript: "This Manuscript was prepared using FRAMCOHORT, GEN3, FRAMOFFSPRING Research Materials obtained from the NHLBI Biologic Specimen and Data Repository Information Coordinating Center and does not necessarily reflect the opinions or views of the FRAMCOHORT, GEN3, FRAMOFFSPRING or the NHLBI." Manuscripts and abstracts resulting from the Research Plan should not use the name of the STUDY in the title of the manuscript/abstract unless the title clearly denotes the source of the Research Materials as being from the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (e.g., "...An investigation using the "). The purpose is to delineate manuscripts from the Research PI and APPROVED USERS from manuscripts from the STUDY and STUDY Investigators.

   The RECIPIENT and PI agree to ensure that all APPROVED USERS will not include in any manuscripts derived from Research Materials any case studies that describe the characteristics of individual participants, or a small number or groups of participants.

9. **Research Use Reporting**

   Prompt publication or other public disclosure of the results of the Research Plan is encouraged.

   When requested by the NHLBI, the APPROVED USERS agree to provide general comments regarding topics such as the effectiveness of the NHLBI Biological Specimen and Data Repository Information Coordinating Center Research Material access process (ease of access and use; appropriateness of STUDY data format; challenges in following the policies; suggestions for improving research material access; or the program in general).

10. **Non-Endorsement, Indemnification**

    The RECIPIENT and PI acknowledge that although all reasonable efforts have been taken to ensure the accuracy and reliability of Research Materials, the NHLBI, and STUDY Investigators do not and cannot warrant the results that may be obtained by using any Research Materials included therein. The NHLBI and all contributors to these Research Materials disclaim all warranties as to performance or fitness of the Research Materials for any particular purpose.

    No indemnification for any loss, claim, damage or liability is intended or provided by any party under this Agreement. Each party shall be liable for any loss, claim, damage, or liability that said party incurs as a result of its activities under this Agreement, except that the NIH, as an agency of the United States, assumes liability only to the extent provided under the Federal Tort Claims Act, 28 U.S.C. 2671 et seq.

11. **Termination and Violations**

    The NHLBI may terminate this Agreement if RECIPIENT or APPROVED USERS are in default of any of its conditions and such default has not been remedied within 30 days after the date of written notice of such default by an authorized representative of the NHLBI. Past violations will be taken into consideration by the NHLBI for future requests from the RECIPIENT and APPROVED USERS to access NHLBI Research Materials.

12. **Amendments**

    Amendments to this Agreement must be made in writing and signed by authorized representatives of all parties.