
UNIVERSITY OF NORTH TEXAS
HEALTH SCIENCE CENTER AT FORT WORTH
1970

Fisher, Bettina L., Tracking of Cholesterol Among Individuals With and Without Diagnosed Cardiovascular Disease. Master of Public Health (Epidemiology), May 2003, 46 pp., 6 tables, 5 figures, bibliography, 33 titles.

Cardiovascular disease is a major public health problem among the elderly in the United States, and cholesterol is the number one risk factor for coronary heart disease. Tracking is a method of analysis used to identify at-risk subjects at an early age in order to institute preventive measures before physical implications of the disease arise. The purpose of this study is to determine the stability and predictability of total serum cholesterol, low-density lipoprotein (LDL), and high-density lipoprotein (HDL) values for people with and without diagnosed cardiovascular disease through tracking. Data was obtained from the Baltimore Longitudinal Study on Aging (BLSA) and comprised men, 45 years of age and older who had at least two measurements. The average length of time subjects were in the study was 21 years, and the average number of repeated measurements was seven. The dataset was divided into two subsets – one for subjects entering the study with diagnosed cardiovascular disease, and a second for subjects entering the study without diagnosed cardiovascular disease. Tracking coefficients were measured using the linear regression model and the linear mixed effects model. There was a high degree of tracking for HDL using the linear regression model and the linear mixed effects model (overall dataset: 0.9283, 0.8216 respectively). Comparing tracking coefficients for cholesterol among subjects with and without diagnosed cardiovascular disease (linear mixed effects: 0.6469 and 0.7668; linear regression model: 0.5408 and 0.6022) reveals

that the former subset has less stable cholesterol values than the latter subset. The linear mixed effects model was the best fit for this data, because it corrects for the variation in the BLSA in interperiod and repeated measurements between subjects.

TRACKING OF CHOLESTEROL AMONG INDIVIDUALS WITH AND WITHOUT DIAGNOSED CARDIOVASCULAR DISEASE

Bettina L. Fisher, B.S.

APPROVED:

Major Professor

Committee Member

Committee Member

Department Chair

Dean, School of Public Health

TRACKING OF CHOLESTEROL AMONG INDIVIDUALS WITH AND WITHOUT DIAGNOSED CARDIOVASCULAR DISEASE

THESIS

Presented to the School of Public Health

University of North Texas
Health Science Center at Fort Worth

in Partial Fulfillment of the Requirements

for the Degree of

Master of Public Health

By

Bettina L. Fisher, B.S.

Fort Worth, Texas

May 2003

ACKNOWLEDGEMENTS

I would like to thank the National Institute on Aging (NIA) and the faculty and staff of the University of North Texas Health Science Center School of Public Health, along with family and dear friends for their encouragement, direction and support throughout this study.

I truly appreciate Larry Brant, Ph.D., and Shan Sheng, Ph.D., of the Statistics and Experimental Design Section at the NIA who introduced me to tracking, acted as mentors throughout this study and provided the dataset from the BLSA. For the faculty and staff of the School of Public Health, I am truly grateful. My advisor, Antonio Rene, Ph.D., has been instrumental in encouraging me to exceed minimum expectations. I also owe a debt of gratitude to Thomas Moorman, Ed.D., whose door was always open. Family is where the greater part of my support lies. I am grateful for the love and Christian guidance of my parents as well as my fiancé, John Drake, whose editing skills and devotion to my success contributed to the completion of this study.

TABLE OF CONTENTS

LIST OF TABLES AND CHARTS

CHAPTER I

INTRODUCTION

Cardiovascular disease is a major public health problem, especially in the aging population. Current estimates from the American Heart Association show that 61.8 million people in America have one or more forms of cardiovascular disease. Of these, 102.3 million people have total blood cholesterol levels of at least 200 mg/dL and 41.3 million have levels of 240 mg/dL or higher (Health Net, 2003). According to Health Net (2003) elevated cholesterol is the number one risk factor for coronary heart disease. Several studies have shown that a reduction in total cholesterol or an increase in high-density lipoprotein (HDL) cholesterol predicts a reduction in cardiovascular disease, therefore identification of subjects on the high end of the total cholesterol distribution (lower part of HDL cholesterol) is of clinical importance in preventive medicine. Since it is known that prevention efforts among younger populations can curb cardiovascular disease, tracking of cholesterol is an effective method to predict future values by earlier measurements. Many investigators have studied longitudinal cholesterol measurements, developing a variety of models to explain these changes over time. Tracking is used as a simple description of the way individual anthropometric and physiologic variables change over time, and it can be a useful epidemiological and clinical tool. For example, early identification of individuals at high risk for cardiovascular disease could permit early, possibly non-pharmacologic, intervention. This will allow at-risk subjects to be

identified at an early age in order to begin preventive measures before physical implications of the disease arise.

Tracking can be defined as the stability of a certain risk factor over time, or the predictability of a risk factor later in life based on the measurement of that risk factor early in life (Twisk, 1994). The importance of tracking lies in the assumption that individuals maintain a certain high value of a risk factor for disease throughout life. Traditional tracking analyses involve calculating the correlation coefficient between two measurements of the same risk factor over time or calculating the percentage of subjects who maintain their relative position within a distribution of values at two measurements in time. However, this method does not allow all longitudinal data to be used, and tracking is assessed without adjusting for potential confounding factors. Therefore, this study will use the linear mixed effects model to overcome these limitations. Tracking is measured via a tracking coefficient, which ranges from –1 to +1. The closer a tracking coefficient is to either extreme, the more likely that risk factor is to remain steady over time. If a tracking coefficient equals 0, the risk factor under analysis does not track efficiently over time.

This study also seeks to analyze methods to perform tracking analyses on unbalanced data sets. Previous studies have only explored tracking using balanced data sets, which allows statistical tests that are not plausible on data sets with large variations between interval visits and observations. The linear regression model is used because of its ability include all measurements throughout time. However, this model does not account for the correlation between repeated measures. Therefore, the linear mixed

effects model is also used to assess tracking. The linear mixed effects model takes into consideration the correlation between repeated measures and the variance in number of observations as well as time between visits.

The purpose of this study is to use the Linear Mixed Effects Model to determine whether total serum cholesterol, low-density lipoprotein (LDL) cholesterol, and high-density lipoprotein (HDL) cholesterol track differently among individuals with cardiovascular disease than among individuals without cardiovascular disease. There are four objectives to be carried out in this analysis:

1. to determine if there are a significant number of subjects in the less favorable categories of total cholesterol, LDL, and HDL as identified by the National Cholesterol Education Program (NCEP).
2. to determine the predictability and stability of total serum cholesterol as well as LDL and HDL cholesterol in the sample population via tracking coefficients.
3. to compare tracking coefficients from the linear mixed effects model and the linear regression model
4. to assess the difference between tracking among individuals with diagnosed cardiovascular disease and individuals without diagnosed cardiovascular disease.

CHAPTER II

LITERATURE REVIEW

Background

The American Heart Association describes cholesterol as a soft waxy substance used to form cell membranes, sex hormones, and assist in the production of bile salts (2003). Because cholesterol cannot dissolve in blood, it must be transported to and from cells by carrier molecules made of a protein called apoprotein. When apoprotein is joined with cholesterol, it forms a compound called lipoprotein. The two types of lipoprotein of greatest concern are low-density lipoprotein (LDL) and high-density lipoprotein (HDL). After a cell has used enough cholesterol to meet its chemical needs, it reduces the number of LDL receptors preventing LDL cholesterol from entering cells and creating an excess of LDL to accumulate in the blood. LDLs begin to deposit cholesterol on artery walls, forming thick plaques preventing the normal flow of blood, which may cause severe cardiovascular disorders (American Heart Association, 2003). The Third Report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults shows that an LDL level of less than 100 mg/dL is the optimal level, and the high risk level of LDL is greater than 160 mg/dL (1999). In contrast, the body makes HDL cholesterol for its protection. It acts to remove this excess cholesterol and transport it to the liver for disposal (American Heart Association, 2003). Higher values of HDL are favored. HDL values of less than

40 mg/dL are considered at-risk by NCEP standards, and values over 60 mg/dL are optimal (NCEP, 1999). Despite elevated levels of LDL cholesterol being of great concern to cardiovascular disease, it has been recommended to use total cholesterol/HDL (TC/HDL) ratios because it is an accurate and simple measure of cardiovascular disease risk (Maffetone, 2003). Because total cholesterol is the addition of both LDL and HDL, its value may be skewed. For example, high total cholesterol may actually be the result of a favorably high HDL and borderline LDL. In this case, lack of the TC/HDL ratio results in a false positive. Desirable TC/HDL ratio is less than 4.5; values greater than 6.0 indicate a high risk for cardiovascular disease (Health Net, 2003; Maffetone, 2003; NCEP, 2003). A ratio of 3.5 is the optimum ratio according to the American Heart Association (American Heart Association, 2003).

Human plasma cholesterol is contained primarily in apoprotein B-based LDL. Walzem, Watkins, Frankel, Hansen, and German found that atherosclerotic progression can be attenuated either by decreasing LDL or by increasing antioxidant protection of older, more susceptible LDL (1995). As cells become cholesterol replete and cease to accept additional LDL cholesterol, the level of plasma LDL elevates, which also increases the length of time that LDL particles spend in the blood stream (Walzem et al., 1995; American Heart Association, 2003). The hypothesis, supported by this study, is that increased time spent in circulation (i.e. through aging) of lipoprotein particles will alter their susceptibility to oxidative modification. Oxidative modification of LDL particles enhances their atherogenicity causing atherosclerotic cardiovascular disease (Walzem et al., 1995).

Wissler and Robert reported from the Eighth Münster International Arteriosclerosis Symposium that, in general, total cholesterol values increase with age in both men and women, and most of this increase is linked to increases in LDL cholesterol. Subsequently, cholesterol clearance from the blood stream is decreased in older individuals (1996). Among the elderly, cardiovascular disease is the most prevalent cause of morbidity and mortality in men and women. The incidence and prevalence are highest in people older than 65 years, and most coronary events occur in this age group (Wissler & Robert, 1996; American Heart Association, 2003). The value of detection and management of elevated serum cholesterol for this age group needs to be carefully assessed. In reviewing the effects of cholesterol in the elderly population, Grundy et al., has found that the public's response to the National Cholesterol Education Program (NCEP) has been to modify its eating habits to lower total cholesterol levels. Medical professionals have increased the frequency of testing for high levels and intensified the treatment of various cholesterol disorders (1999).

The results of a study by Zuliani, et al. suggest that among the elderly, severe disability is strongly associated with low HDL levels. Low HDL levels may be a marker for ongoing disability in basic activities of daily living (1999). Using multiple logistic regression in a sample of 344 institutionalized patients over 65 years of age, severe disability was found to be significantly associated with HDL and total cholesterol with odds ratios over 2.0 for both variables (Zuliani et al., 1999). Corti et al., (as cited in Grundy et al., 1999) reported that even in patients older than 75 years, HDL levels and TC/HDL ratios retained predictive power of later values in life from early cholesterol

values. Combining HDL cholesterol with total cholesterol may help identify high-risk, older patients for cholesterol-lowering therapy.

Determining the stability of cholesterol to predict the risk of cardiovascular disease may also be used as a tool to assess the predictability of diabetes. Cardiovascular disease and type 2 diabetes share a common physiologic antecedent. Although patients with type 2 diabetes are affected by microvascular complications, cardiovascular disease complications exert the greatest toll. James Meigs reports that cardiovascular disease is more than twice as common among patients with diabetes as it is among those who do not have diabetes, and it is a major cause of morbidity and death in patients with type 2 diabetes (2002). Reaven (as cited in Meigs, 2002) proposed that obesity, hyperglycemia, hypertension and dyslipidemia are at least partially caused by insulin resistance and are recognizable as a syndromic precursor to type 2 diabetes and cardiovascular disease. For example, data from the San Antonio Heart Study revealed a greater prevalence of obesity, hypertension, low HDL, hyperglycemia, and hyperinsulimia among subjects who initially did not have diabetes but developed type 2 diabetes over eight years of follow-up compared with subjects who did not develop diabetes. In addition, Burke et al., agrees that there are a number of modifiable factors such as diet, exercise and smoking cessation, to reduce the risk of cardiovascular disease. However, biological factors, such as higher levels of HDL cholesterol and the absence of diabetes, are strongly associated with continued health and reduced risk of cardiovascular disease (2001). Likewise, a follow-up study from San Antonio found that prediabetic subjects with insulin resistance had more cardiovascular disease risk factors than insulin-sensitive prediabetic subjects

(Meigs, 2002). This supports the joint pathogenesis of type 2 diabetes and cardiovascular disease. Meigs makes note of new data that suggests type 2 diabetes can be prevented (2002); therefore, tracking of cholesterol may be an important step in developing primary prevention efforts for type 2 diabetes.

Solfrizzi et al., also analyzed the relationship between lipoprotein(a) cholesterol levels and type 2 diabetes through the Italian Longitudinal Study on Aging, but only among individuals age 65 and over (2002). Lipoprotein(a), as described by Mercola is a compound that consists of LDL plus a protein called apoprotein(a) whose exact function is not known (2003). The interaction between low lipoprotein(a) values and high LDL cholesterol values did not significantly increase the risk of coronary heart disease. Also, the interaction between lipoprotein(a) and LDL cholesterol in the absence of type 2 diabetes did increase the risk of coronary heart disease risk by 2.75 but only with borderline significance (95% CI 7.7 – 0.99). Maggi et al., supported these results in an analysis of the Italian Longitudinal Study on Aging that found total and LDL cholesterol in women were significantly lower in the highest insulin quartile, while no significant differences were seen in non-diabetic women or men (2001). However, Solfrizzi et al., found the most notable interaction was the combined effect of elevated lipoprotein(a) values and elevated LDL cholesterol values in the presence of type 2 diabetes which yielded a significant increase in the risk of coronary heart disease by 6.65 (2002). This study suggests that even in the elderly population, lipoprotein levels may contribute to cardiovascular diseases through synergistic mechanisms with high serum LDL cholesterol and type 2 diabetes. The increase with age in LDL cholesterol levels is often

associated with a triglyceride increase and insulin resistance (Wissler & Robert, 1996). Overall, the relationship between cholesterol and insulin in the elderly population is still unclear; therefore, tracking of cholesterol may be an important method to compare the stability of cholesterol and predictability of diabetes among an elderly population.

History of Tracking

In Western populations, the level of serum cholesterol increases with age. Various studies have proven this, however few have considered whether subjects consistently remain in the same percentage group relative to their peers or if early values of cholesterol are sufficient to predict later values. A longitudinal study by Clark, Allen and Wilson on students of the United States Military Academy at West Point has shown that both average serum-cholesterol and LDL levels increased throughout the study (1967). In contrast, average HDL levels followed a downward trend. Similar results are presented in a later study by Berns, DeVries, and Katan that shows an increase in cholesterol among the elderly population (1988). The average increase in total cholesterol was 7% in women and 14% in men. Berns et al. ventured further to determine the stability of cholesterol by performing a correlation between the level of cholesterol at baseline and at follow-up measurement (1988). Studies such as these may have prompted further studies on the stability and predictability of variables as well as the various methods of tracking used today.

Longitudinal studies are often used in epidemiologic analysis to study the development of chronic diseases. Tracking is a statistical measure that may significantly contribute to this type of study. There is no single definition of tracking but there are two

themes that are carried in all methods of tracking. The first is the longitudinal stability of a certain variable, and the second is the predictability of future values by early measurements (Twisk, Kemper, & Mellenbergh, 1994; Foulkes & Davis, 1981; Wilsgaard et al., 2001; Twisk et al., 1997a). A survey conducted by Foulkes and Davis in 1981 of several U.S. epidemiologists revealed these two themes to hold true for most epidemiologists in the United States. The first relates to the consistency of ranking over time within the same percentile group among a group of peers (Foulkes & Davis, 1981). The research question is commonly whether persons with extreme values at times t_1, t_2, ...t_k also be extreme at time t_{k+1}. The second is determining how predictive early measurements are for values later in life. However, the problem here is that serial observations on the same individual at different times are assumed to be correlated.

A sample of N subjects from a given population is needed to evaluate tracking with measurements at different times. A measure of Y(i,j) for subject i, where i = 1,2,....,N, at time points t(j), where j = 1,2,...., T determines if Y(t) tracks for any two time points (t_1 and t_2) within the time period (Twisk et al., 1994). The degree of tracking is expressed in a tracking coefficient. In order to determine the reliability of an early measurement, $Y(t_1)$, to predict the value of a measurement later in life, $Y(t_2)$, the magnitude of the prediction is expressed in a predictive value or risk measure, which is related to the tracking coefficient (Twisk et al., 1994).

Special precautions should be taken when performing any method of tracking. The measurement Y(t) can be either ordinal or dichotomous. However, tracking between $Y(t_1)$and $Y(t_2)$ can be influenced by bias of the observers. This may cause measurement

errors of Y(t) which may lead to a miscalculated tracking coefficient (Twisk et al., 1994). In most studies there are differences in the length of the inter-period between the related measurements or in the ages of the subjects at which the initial measurement is carried out, therefore, it is difficult to compare the magnitudes of tracking indices or predictability measures (Twisk et al., 1994). The result of tracking depends heavily on the period under consideration and the method of tracking used. There are many differences in the methodology used to assess tracking, however they may be categorized into two approaches: nonparametric and parametric approaches (Twisk et al., 1994).

Nonparametric Approaches

The Spearman's rank coefficient is a nonparametric method commonly used to calculate tracking between two measurements (T=2). If these measurements are bivariate normally distributed, the value of Spearman's rank correlation coefficient is similar to the value of Pearson's correlation coefficient (Twisk et al., 1994). The correlation coefficient becomes the tracking coefficient, which ranges from -1 to +1. A coefficient of 0 would denote no tracking, and thus the variable is said to have no stability over the investigated time period and T = 1 has no predictability of T = 2. The optimal tracking coefficient is either –1 or +1, which is interpreted as the variable of interest being consistently stable compared to the population in the specified population, as well as, measurement 1 having perfect predictability of measurement 2.

Another nonparametric method of tracking when there are two measurements is performed by dividing the population into percentile groups (quartiles, quintiles, deciles…) according to the initial measurement. Next, the percent of subjects who stayed

in the same upper or lower percentile group at one or more follow-up measurements is calculated. If this percent is more than the expected percent the population is said to track for that particular variable in the specified time period (Twisk et al., 1994). The expected percent is calculated by how the subjects are divided into each percentile group (more than 25 percent for quartiles, more than 20 percent for quintiles, more than 10 percent for deciles...). The problem with this approach of calculating tracking coefficients is that the magnitude of the coefficients depends on the grouping of the data.

Two other methods of nonparametric tracking exist that are based on the division into percentile groups. Nishio, et al. developed a method to calculate a tracking coefficient when T = 2. The method gives different weights to movements to and from different percentile groups thereby calculating the tracking index (TI):

$$TI = \frac{T(s)}{T(h)} \qquad \text{Eq. 1}$$

T(h) equals the tracking of a hypothetical group whose values change randomly between the percentile groups during the measurement period, and T(s) equals the tracking of the study group:

$$T(s) = \frac{(2x + y - z)}{(x + y + z)} \qquad \text{Eq. 2}$$

where, x equals the number of subjects who remained in the same percentile group; y equals the number of subjects who moved to a neighboring percentile group; and z equals the number of subjects who moved to a remote percentile group (Nishio et al., 1987). In this approach by Nishio et al., not only subjects who stay in the same percentile group

during follow-up but also the subjects who moved to a neighboring percentile group positively influence the tracking coefficient TI (1987).

Twisk identifies another method of tracking called Cohen's kappa (κ) that is calculated in longitudinal studies where $T \geq 2$ (1994). For each individual, the number of times that the individual is in each of the particular percentile groups is counted and compared with the value that is expected if the individuals are randomly assigned to the different percentile groups at each measurement (Twisk et al., 1994). Kappa ranges from 0.0 to 1.0, and if $\kappa > 0.75$, then the variable tracks well. If $\kappa < 0.40$, then the variable tracks poorly, and if κ has a value in between these two, then there is moderate tracking for the variable of interest. The major advantage of kappa is that the index is very easy to compute, but one of the problems of kappa is that all movements between two percentile groups are weighted equally, irrespective of the length of the movement (Twisk, et al., 1994).

Parametric Approaches

Since the late 1970s statisticians have used the correlation coefficient to measure tracking. When $T = 2$, this is the most used parametric approach of tracking (Twisk et al., 1994). However, correlation is only a suitable measure of tracking when $Y(t_1)$ and $Y(t_2)$ are bivariate normally distributed. Also, when there are more than two longitudinal measurements, correlation coefficients do not employ all available data. Beckett et al. and Guo et al. (as cited in Twisk et al., 1994) have overcome this problem using the general linear model as a measure of tracking.

$$Y_i = X_i\beta + \varepsilon_i \qquad \text{Eq. 3}$$

In this equation, Y_i equals the vector of observed values of individual i; X_i equals the design matrix for individual i; β equals the vector of regression parameters; and ε_i equals the vector of measurement errors. The estimation of the regression parameters includes an estimation of the correlation matrix between pairs of measurements (Twisk et al., 1994). These estimated correlations are interpreted as tracking coefficients. One of the assumptions with this method is that all longitudinal measurements are multivariate, normally distributed.

Another parametric method of tracking, outlined by Foulkes and Davis, is based on the probability that a follow-up measurement will be greater than a predetermined cut-point, given that the initial measurement is greater than the predetermined cut-point (1981). If this probability is high, the variable is said to 'track'. This method of tracking is related to both correlation and probability, which is an index of tracking called gamma (γ) (Foulkes & Davis, 1981). Their approach is based on a statistical model in which the longitudinal values of an individual change as a certain function of time. For each individual the original values are replaced by the predicted values, based on this function of time. Through this method Foulkes and Davis estimate the probability that the true serial values of two randomly selected individuals are in the same order at all study times (1981). The coefficient γ can take values between 0.0 and 1.0. A value of 0.0 means that every individual curve crosses every other individual curve at least one time; a value of 1.0 indicates that none of the individual curves cross; and a value > 0.5 indicates tracking, because two individual curves chosen at random would be more likely to have curves that

do not cross (Foulkes & Davis, 1981). A problem with γ, like most parametric tracking methods, is that it highly depends on the length of the observed time period.

However, this method does not provide a basis for prediction of future values, which Ware and Wu (1981) address by using growth techniques and individual design matrices. In a comparison between growth-curve analysis and linear regression, it was found that growth-curve analysis was especially useful when each individual has a different design matrix (Ware & Wu, 1981). Their analysis was compared to that of McMahan who developed a tracking coefficient, tau (τ). Tau is calculated under the assumption that Y(t) are multivariate normally distributed with common covariance matrix (McMahan, 1981). It is the average value of the ½ T(T-1) Pearson correlation coefficient, where T is the number of times a value is measured. A population tracks for a certain variable if, for each individual growth curve, the relative deviation from the population mean growth curve remains unchanged over time. When measurements and thus the tracking coefficients are biased by measurement error, the actual data can be replaced by predicted data based on a function of time. τ ranges from -1 to +1. If τ has the value of +1.0, there is perfect tracking for that variable. If the coefficient is 0.0, there is no tracking for that variable (McMahan, 1981). Negative values indicate a reversal of the values between two observed time points.

Twisk, in the International Journal of Sports Medicine, describes the use of generalizing estimating equations (GEE) to analyze epidemiological longitudinal data (1997). The tracking coefficient of this model is the standardized regression coefficient of β_1 that represents the magnitude of the relationship between the outcome variable (Y_{it})

and the different predictor variables (X_{ijt}). The analysis of these two is not only analyzed under correction of time (t), but also under correction of time-dependent covariates (Z_{ikt}) and time-independent covariates (G_{im}) (Twisk, 1997). With GEE, a pooled analysis of cross-sectional (between-subjects) and longitudinal (within-subjects) relationships is carried out simultaneously into one tracking coefficient. GEE has many advantages and seems to provide the most cohesive tracking coefficient for analysis. However, it does not include random effects that may control for the variation in measurements between subjects. This may be corrected by using a linear mixed effects model.

Morrell, Brant, Pearson, Verbeke, and Fleg notice the problem epidemiologic investigators have about correcting risk estimates for the bias due to measurement error in the risk factor and have performed a simulation study to prove that the linear mixed effects model is a method for determining the true level of risk at baseline (2003). The underestimation of the true association between a risk factor and a disease outcome has been termed regression dilution bias. This may occur because an individual's measurement of a risk factor may differ from its true level due to random variability in the measurement process or due to real but short-term biological variability. Because of regression toward the mean, observed low values are probably lower than the true value of the risk factor, while observed high values are probably higher than the true value (Morrell et al., 2003). One method Morrell et al. (2003) suggests to protect against regression dilution bias is to use a study where replicated measurements are collected near baseline. The data could then be modeled using the mixed effects model to obtain a baseline predicted value. Zhang and Davidian describe the linear mixed effects model as

incorporating random effects to accommodate among-subject variation (2001). The fundamental assumption of this model is that within-subject errors and random effects are normally distributed (Zhang & Davidian, 2001). The linear mixed effects approach is flexible in dealing with data sets that are unbalanced in the number of repeated measurements. Studies have been conducted on applying this model to the problem of measurement error in epidemiologic analysis. This is a beneficial quality of the linear mixed effects model used to assess tracking, because measurement error may create bias in a tracking coefficient.

Factors Influencing Tracking of Cholesterol

Another study by Twisk et al., in 1996 revealed that total cholesterol was positively influenced by body fatness, measured by sum of skin folds, and the amount of daily physical activity at the initial measurement. However, HDL was not found to be influenced by any biological, psychological or lifestyle variable. Adding the initial values of these parameters to the tracking model did not have a significant influence on the magnitude of the tracking coefficient (Twisk et al., 1996). Also, diet may affect the tracking of cholesterol if it is not correctly considered in the analysis. An excess of saturated fat uptake may cause an increase in blood cholesterol (Health Net, 2003). Diet is an important measure, though difficult to measure, in analyzing cholesterol in a longitudinal analysis.

Matthews et al. discovered that some natural and physical characteristics may influence the level of lipids (2001). Based on results from the Seasonal Variation of Blood Cholesterol Study (SEASON), seasonal variation in physical activity was found to

coincide with seasonal changes in blood lipid levels (Matthews et al., 2001). As seasons change, there is a variation in the amount of physical activity that men and women undergo, thus affecting the biological stability of blood cholesterol, lipid levels and body fatness. Excess weight gain tends to lower HDL levels and raise LDL levels (Mattthews et al., 2001; Health Net, 2003).

In order to determine the effects of smoking on cholesterol, Blanco-Cedres et al. conducted a study that connected smoking to a higher risk of cardiovascular disease at every level of total serum cholesterol (2002). For all total cholesterol categories, the relative risks for smoking ranged from 1.57 to 2.78. Yuan et al. (as cited in Blanco-Cedres et al., 2002), conducted a prospective study of male residents of Shanghai, China with average cholesterol levels less than 180 mg/dl and found significant relative risks of 2.1 for cardiovascular disease. The relationship between cholesterol and cigarette smoking is more evident when investigating LDL and HDL cholesterol levels. Valkonen and Kuusi found during a clinical trial that the exposure of non-smoking subjects to secondhand smoke breaks down the serum antioxidant defense, leading to accelerated lipid peroxidation, LDL modification, and accumulation of LDL cholesterol in human macrophages (1998). This occurs when free radicals from secondhand smoke are trapped by serum aqueous and lipophilic antioxidants. A failure of the antioxidant barrier occurs, LDL oxidation takes place, and finally LDL cholesterol accumulates in macrophages (Valkonen & Kuusi, 1998). This accumulation of cholesterol in the macrophages is the key event in atherosclerosis. Neufeld, Mietus-Snyder, Beiser, Baker, and Newburger found similar results when analyzing HDL cholesterol levels via a cross-sectional, pilot-

scale study (1997). Average HDL cholesterol levels were significantly lower in hyperlipidemic children from households with smokers compared with those from nonsmoking households (Valkonen & Kuusi, 1998; Health Net, 2003).

All of these factors influence the severity of cholesterol among most age groups. Studies suggest that there may be an interaction among body mass index (BMI), physical activity, and smoking with serum cholesterol levels. Lussier-Cacan, Bolduc, Xhignesse, Niyonsenga, and Sing have measured the interaction between these variables and concluded that the interactions between these variables are the primary causes of variation in diseases with multifactorial etiology such as cardiovascular disease (2002).

Previous Tracking Results

Earlier studies of tracking have examined a wide range of cardiovascular disease risk factors including BMI, blood pressure, cholesterol, and triglycerides. Although it is difficult to compare tracking coefficients between studies, it is beneficial to future studies to note the methods used, study sample and subsequent results.

In a study conducted by Wilsgaard et al. two methods of tracking were performed in a population-based cohort study over a period of sixteen years (2001). The first was the linear regression model with β_1 as the tracking coefficient, which is interpreted as the prediction of the dependent variable's initial value when the dependent variable changes at time t_2 and t_3. The second method presented the number of subjects who remained in the same sextile group throughout the different examinations as well as the number of subjects who changed sextiles after the initial examination. For all risk factors, a significantly higher proportion of participants changed fewer than three sextiles,

indicating the presence of tracking. To assess the stability of each risk factor over time, and to use all available data, Wilsgaard, et al. calculated the proportions of participants who remained in the same sextile between two examinations. For men, 36.2%, 39.9%, and 27.1% remained in the same sextile for high-density lipoprotein cholesterol, total cholesterol and triglycerides, respectively. For women, 36.3%, 40.7%, and 28.0% remained in the same sextile for high-density lipoprotein cholesterol, total cholesterol and triglycerides, respectively. Also, higher values of HDL cholesterol were associated with decreased odds for tracking of BMI and triglycerides (Wilsgaard et al, 2001).

In an observational longitudinal study by Twisk, Kemper, Mellenberg, and Mechelen, tracking was assessed via six repeated measurements on 181 subjects over a period of fifteen years (1997a). Baseline measurements for all subjects were taken at age 13. Tracking was assessed for hypercholesteremia, which is characterized by high total serum cholesterol, low HDL and high TC/HDL ratio. Subjects were divided into risk groups according to standard risk values defined by the Report of the National Cholesterol Education Program. The risk groups of total cholesterol, HDL and the TC/HDL ratio consisted of eleven to thirteen percent of the total sample population. An odds ratio (OR) was calculated as the tracking coefficient for subjects who were at risk for hypercholesteremia from the generalized estimating equations. For HDL the OR was 14.4, and for the TC/HDL ratio the OR was 25.6. This indicates that the subjects with TC/HDL values $\geq$ 4.0 at the initial measurement at the age of 13 years had a 25 times higher risk of developing TC/HDL values $\geq$ 5.5 later in life compared to the subjects with TC/HDL values $<$ 4.0 at the initial measurement (Twisk et al., 1997a). These results can

have important implications in the planning of primary prevention procedures regarding hypercholesteremia.

In the Amsterdam Growth and Health Study, Twisk, Kemper, Mechelen, and Post studied biologic and lifestyle risk factors for coronary heart disease (1997b). This study began in 1977 and concluded in 1991with six repeated measurements and a total of 181 subjects. Because initial biologic age is more important than calendar age when considering cardiovascular disease, biologic age was calculated by measuring the skeletal age from radiographs of the left had, according to the Tanner-Whitehouse 2 method. In order to determine if it is worthwhile to screen subjects at an early age for cardiovascular disease risk factors, ORs were calculated using GEE. The highest odds ratios were found for the lipoproteins and for a measure of weight, the sum of four skin folds. The OR for total cholesterol was 10.4 and the OR for HDL was 14.1. The highest OR of all factors analyzed was the TC/HDL ratio, which was 22.9. All ORs for the biologic risk factors reached statistical significance based on the 95% confidence interval. Overall, tracking for cholesterol and lipoproteins was significantly high indicating relatively high stability during adolescence and young adulthood and relatively good predictability of measurements early in life (Twisk et al., 1997b).

CHAPTER III

METHODS

Data

Data to conduct analyses were obtained from the Baltimore Longitudinal Study of Aging (BLSA), which was initiated in 1958 to trace effects of aging in humans. Subjects are well-educated, community-dwelling volunteers of middle- to upper-class socioeconomic status. Each volunteer visits the Gerontology Research Center once every two years for two and one-half days of testing (Shock et al., 1984).

Inclusion/Exclusion Criteria

Only identification numbers under 5,000 were included, limiting the analysis to men. The addition of women in 1978 to BLSA did not warrant sufficient information to provide a comparison. This analysis was restricted to subjects with an initial age of 45 years or older. By focusing the tracking analysis on a specific age group, discrepancies over the effects of cholesterol at younger ages as opposed to older ages could be eliminated. In order to calculate a tracking coefficient, at least two measurements were needed. From the sample of men with initial age of 45 years and older, only individuals with two or more visits were included.

Variables

The dependent variables were total cholesterol, LDL, and HDL (starting from second measurement), and the independent variable was the initial value of each dependent variable. Time-independent covariates were initial age and current smoking status, while the time-dependent covariate was time itself.

Design

This investigation used an analytic and descriptive component. The analytic component was the assessment of tracking using both the linear regression model and a linear mixed effects model. The formulas for these two models are presented below.

The formula for the linear regression model is

$$Y_{it} = \beta_0 + \beta_1 Y_{it1} + \beta_2 t + \sum_{j=1}^{J} \beta_{3j} X_{itj} + \sum_{k=1}^{K} \beta_{4k} Z_{ik} + \varepsilon_{it} \qquad \text{Eq. 4}$$

where Y_{it} is the observation of individual i from t_2 to t_m, β_0 is the intercept, β_1 is the regression coefficient used as tracking coefficient, Yit_1 is the initial observation of individual i, β_2 is the regression coefficient for time, β_{3j} is the regression coefficient of the time-dependent covariate, X_{itj} is the time-dependent covariate j for individual i, β_{4k} is the regression coefficient of the time-independent covariate, Z_{ik} is the time-independent covariate k of individual i, and ε_{it} is the measurement error for individual i.

The formula used for the linear mixed effects model is

$$Y_{it} = (\beta_0 + b_{i0}) + \beta_1 Y_{it1} + (\beta_2 + b_{i2})t + \sum_{j=1}^{J} \beta_{3j} X_{itj} + \sum_{k=1}^{K} \beta_{4k} Z_{ik} + \varepsilon_{it} \qquad \text{Eq. 5}$$

where b_{i0} and b_{i2} are the random effect coefficients for intercept and time respectively. Random effects account for the variability among persons and the interaction of persons with the independent variables.

The descriptive component measured the number and percentage of subjects in each category for total, LDL, and HDL cholesterol. These cut-points were obtained from

the Third Report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (1999). (See Table 1).

Table 1.
NCEP Classification of Total, LDL, HDL Cholesterol

Total Cholesterol	
< 200	Desirable
200-239	Borderline high
≥ 240	High
LDL Cholesterol	
<100	Optimal
100-129	Near optimal / above optimal
130-159	Borderline high
160-189	High
≥ 190	Very high
HDL Cholesterol	
< 40	Low
≥ 60	High

Analysis

The statistical package used for all data analysis was SAS. SAS and Microsoft Excel were used to create graphs and charts. Only men with an initial age ranging from 45 to 95 years were selected from the BLSA. Subjects must have had at least two visits to be included in the study population. Tracking coefficients were calculated over an average period of 21 years with a mean of 7 repeated measurements per subject. There was a maximum of 180 subjects and 1,465 observations analyzed. Analyses controlled for current smokers by including this variable as a covariate in both models. Random effects were placed on both intercept and time.

This study population was further divided by those individuals entering the study with diagnosed cardiovascular disease and those entering the study without diagnosed cardiovascular disease. The resulting three datasets will henceforth be called Data A, Data B, and Data C. Data A is defined as the initial dataset including all subjects. Data B is a subset of Data A which includes all subjects in Data A that entered the study with diagnosed cardiovascular disease. Data C is also a subset of Data A which includes all subjects in Data A that entered the study without diagnosed cardiovascular disease.

CHAPTER IV

RESULTS

Tracking analysis was conducted on a study population obtained from the BLSA. Subjects included in the analysis had entered the study at the age of 45 or older and had been to the Gerontology Research Center for at least two visits. Subjects averaged seven visits every two years. Table 2 shows the average and range of values for time spent in the study and age of subjects. The initial age of subjects ranges from 45 years old to 84 years old. The range of each subject's age at their most recent visit ranges form 53 to 80. As shown in Table 3, Data A contains 180 subjects and 1,465 observations. From the 180 subjects under analysis, 34 entered Data B resulting in 300 observations and 146 entered Data C resulting in 1,165 observations.

Table 2.
Descriptives of time and age for Data A

	Range	Average
# of Visits	2 – 26	6.9
Years between visits	1 – 25	2.3
Length in study	4 – 38	21
Age	45 – 95	69
Initial	45 – 84	59
Final	53 – 95	80

Table 3.
Descriptives for Data A, Data B, Data C

	Data A	Data B	Data C
Subjects	180	34	146
Observations	1465	300	1165
Max observations per subject	22	18	22

The National Cholesterol Education Program (NCEP) has specified categories for cholesterol, LDL, and HDL as specified in Table 1. The percentage of subjects in each of these categories was calculated and can be observed in Figures 1, 2, and 3. Based on categories outlined by the NCEP, this study population has approximately equal distribution among the three categories of desirable, borderline high and high for cholesterol. (See Figure 1).

Figure 1.
Cholesterol Categories from Data A based on NCEP standards

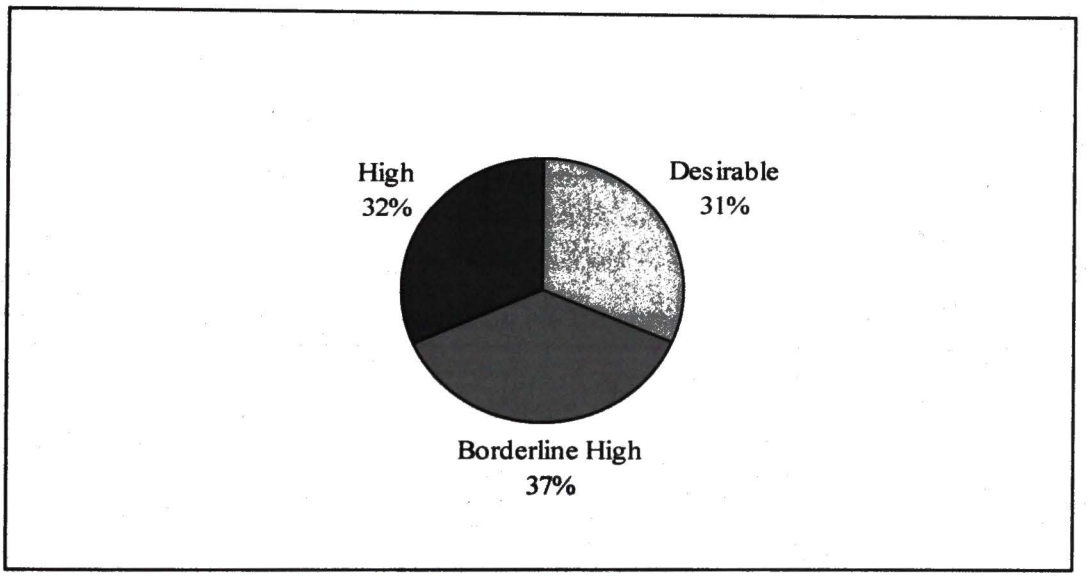


Figure 2 depicts the percentages of subjects in the NCEP categories for LDL. Twenty-eight percent of the study population has optimal levels of LDL. Thirty-five percent are near optimal values, 24% are borderline high and 11% of the study participants have high values for LDL. Only 2% of the study participants have LDL values greater than or equal to 190 mg/dL. Higher HDL values are more desirable. The majority of the participants lie in the normal range for HDL, 54%, however a significantly greater

number, 34%, of participants are in the low HDL category than the 9% in the high HDL category.

Figure 2.
LDL Categories from Data A based on NCEP standards

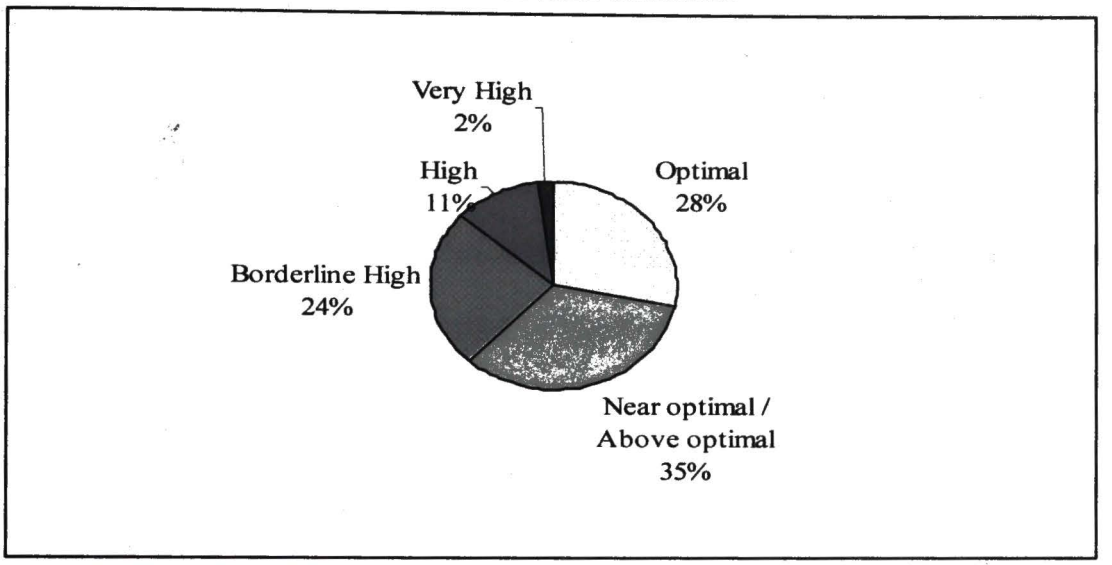
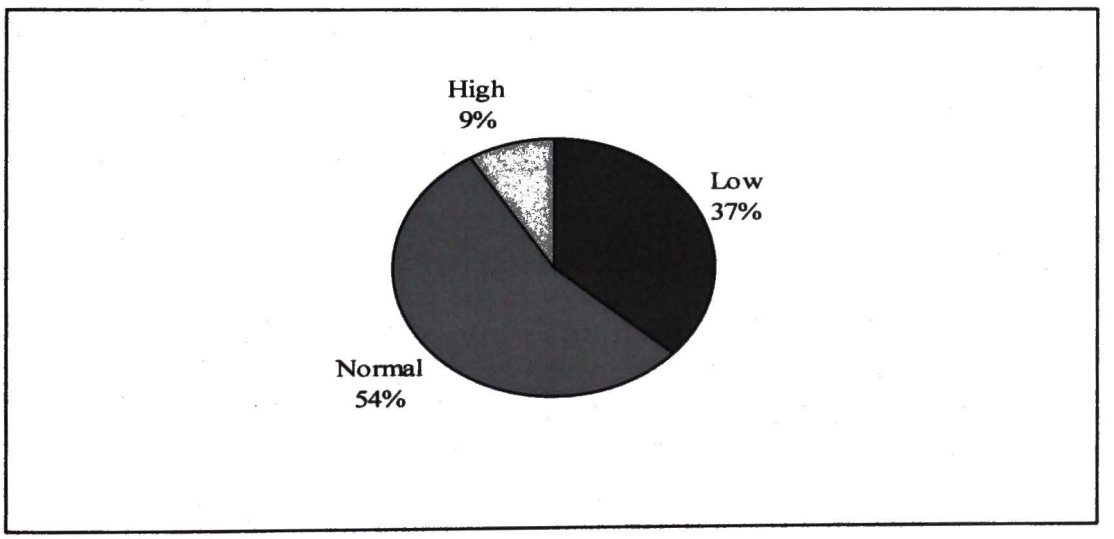
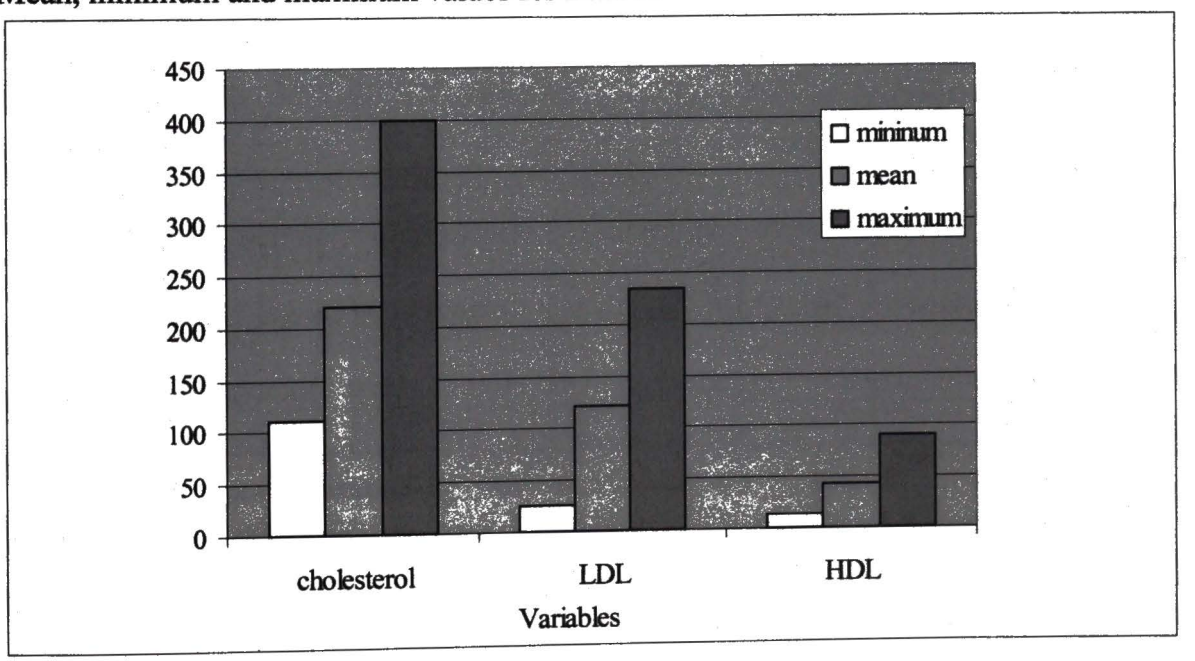


Figure 3.
HDL Categories from Data A based on NCEP standards



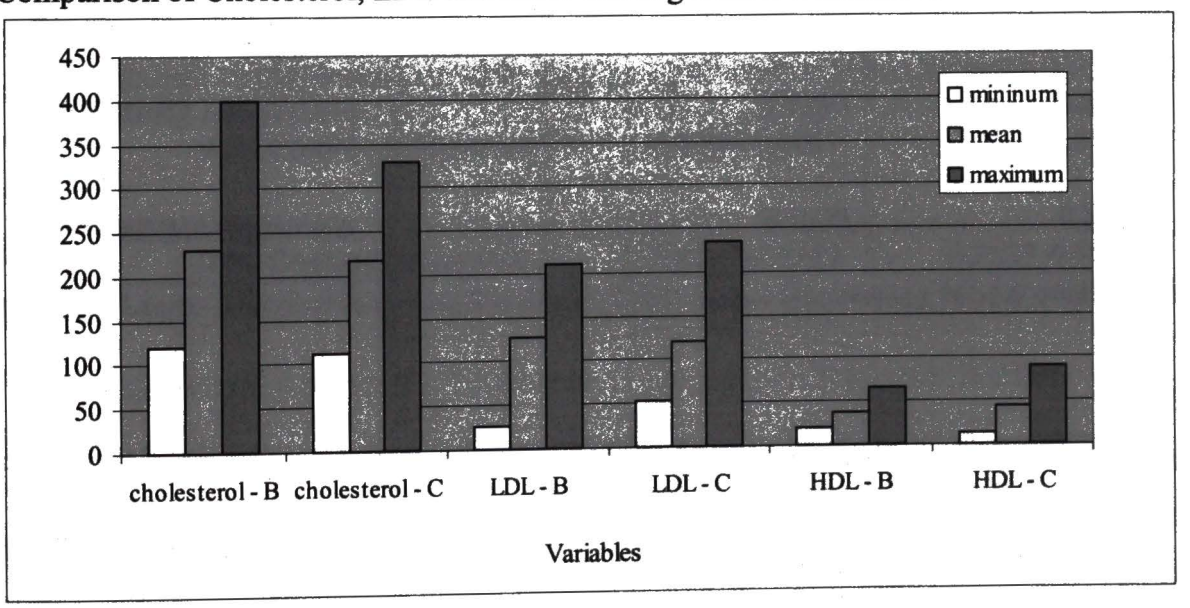
The study population has cholesterol values that range from "desirable" to "high" with values of 111 mg/dL to 399 mg/dL respectively. However, the average cholesterol value for all observations is 221 mg/dL, which is in the "borderline high" classification (see Figure 4). Both LDL values and HDL values follow the same pattern. LDL values for Data A range from 26 mg/dL to 234 mg/dL, which corresponds to the "optimal" and "very high" categories respectively based on NCEP classification. The mean LDL value of 122 mg/dL is of "near optimal" classification. HDL values less than 40 mg/dL are considered low and values greater than 60 mg/dL are high according to Table 1. Higher HDL values are more desirable. Data A resulted in HDL values ranging from 14 to 90 mg/dL with an average of 44 mg/dL. It is concluded that elevated cholesterol is common among this population.

Figure 4.
Mean, minimum and maximum values for Data A



Based on mean values, subjects entering the study with diagnosed cardiovascular disease had higher values for cholesterol, LDL and lower values for HDL (see Figure 5). Average cholesterol for Data B is 230 mg/dL as compared to 218 mg/dL for Data C. Mean values for LDL between the two groups are close with values of 128 mg/dL for Data B and 121 mg/dL for Data C. However, both LDL for Data C and Data A have values of 121 mg/dL, which is less than that of LDL of Data B. HDL values for Data A and Data C are 44 mg/dL and 45 mg/dL respectively. These values are more desirable than the HDL value of 38 mg/dL for Data B. Based on these average results, subjects entering the study diagnosed with cardiovascular disease (Data B) collectively had higher values of cholesterol, LDL and lower HDL values than Data C. This result was expected, however, it did not address the stability of these variables over time.

Figure 5.
Comparison of Cholesterol, LDL and HDL Among Data B and Data C



Tracking coefficients are measured from 0 to +1 if the variable is increasing over time, and are measured from -1 to 0 if the variable is decreasing over time. For this study, cholesterol, HDL, and LDL increase as people age, therefore, all coefficients range from 0 to +1. The closer a coefficient is to –1 or +1, the more stable the variable is over time and the higher the probability of predicting future values from present measurements. For Data A, as shown in Table 4, HDL is the variable that is the most stable over time. However, its coefficient for the linear mixed effects model is of borderline significance. The coefficients for all variables are above 0.5, which denotes some stability over time.

Table 4.
Tracking coefficients for Data A

Variable	LME*	LRM*
Cholesterol	**0.7408**	**0.5908**
95% Confidence Interval	0.6749-0.8067	0.5513-0.6303
High-Density Lipoprotein	**0.9283**	**0.8216**
95% Confidence Interval	0.8499-1.0067	0.7589-0.8842
Low-Density Lipoprotein	**0.9130**	**0.8262**
95% Confidence Interval	0.8247-1.0013	0.7523-0.9000

*LME = Linear Mixed Effects coefficients; LRM = Linear Regression Model coefficients

Data B, in Table 5, has the fewest number of subjects and observations among HDL and LDL, which may account for its lack of significance. With a coefficient of 0.91, HDL is the most stable variable though not statistically significant. A coefficient was not able to be calculated for HDL using the linear mixed effects model, because the convergence

criteria were not met. There were too many likelihood evaluations for this analysis. Subjects entering the study without diagnosed cardiovascular disease (Data C) are significantly stable for cholesterol and HDL and of borderline significance for LDL (see Table 6). Cholesterol has significantly strong stability coefficients derived from both the linear mixed effects model and the linear regression model of 0.7668 and 0.6022 respectively.

Table 5.
Tracking coefficients for Data B

Variable	LME*	LRM*
Cholesterol	**0.6469**	**0.5408**
95% Confidence Interval	0.4797-0.8140	0.4407-0.6409
High-Density Lipoprotein	***	**0.9058**
95% Confidence Interval		0.5071-1.3045
Low-Density Lipoprotein	**0.8687**	**0.7311**
95% Confidence Interval	0.6650-1.0724	0.5139-0.9482

*** Because of lack of information for this subset of the data, the model did not converge

The linear mixed effects model accounts for variance in number of repeated measurements that is not accounted for in the linear regression model. In Data A, Data B and Data C, the linear mixed effects model has higher coefficients than the linear regression model. Subjects in Data C have the most stable cholesterol values, with a linear mixed effects coefficient of 0.7668 and a linear regression model coefficient of 0.6022, as compared to subjects in Data A and Data B.

Table 6.
Tracking coefficients for Data C

Variable	LME*	LRM*
Cholesterol	**0.7668**	**0.6022**
95% Confidence Interval	0.6941-0.8396	0.5587-0.6458
High-Density Lipoprotein	**0.8972**	**0.8200**
95% Confidence Interval	0.8101-0.9843	0.7554-0.8846
Low-Density Lipoprotein	**0.9419**	**0.8426**
95% Confidence Interval	0.8402-1.0435	0.7603-0.9249

CHAPTER V

DISCUSSION

The purpose of this study was to determine the stability and predictability of cholesterol, LDL, and HDL values for people with and without cardiovascular disease through tracking. Stability of these variables does not mean their values will not increase over time. Studies show that both cholesterol and LDL increase as people age. The stability under observation is the extent to which peers remain in the same percentage ranking over time. The purpose is achieved through four objectives: (1) to categorize levels of cholesterol, LDL, and HDL according to NCEP classification guidelines, and to determine where the majority of the study population lies; (2) to determine the stability of cholesterol, LDL, and HDL using tracking coefficients derived from linear regression and linear mixed effects models; (3) to assess the difference between tracking coefficients using the linear regression model and tracking coefficients using the linear mixed effects model; and (4) to compare tracking coefficients of subjects with diagnosed cardiovascular disease to tracking coefficients of subjects without diagnosed cardiovascular disease.

The NCEP standards were used to categorize the average values of cholesterol on the whole study population before dividing the population by diagnosed cardiovascular disease. This is important because it provides a comparison for the tracking and average cholesterol values of Data B and Data C. Based on NCEP standards, the distribution of cholesterol in the study population, Data A, is consistent in all three categories. For average cholesterol values, approximately one-third of the population lies in each

category: desirable, borderline, and high. Categories for the average LDL values are similar to cholesterol (see Figure 2). There are five categories for LDL as defined by NCEP, however if the top three – borderline high, high, and very high – are combined, approximately one-third of the population will exist in each category, much like the distribution of cholesterol. Twenty-eight percent have optimal levels of LDL, 35% have near or above optimal levels of LDL, and a slightly larger 37% of the population has an average borderline high to very high LDL value. Categories for HDL distribution do not follow the same pattern. Although more than half of the study participants, 54%, have normal values ranging from 40 to 59 only 9% have high HDL values. High HDL values are desirable in order to remove stored plagues in the arteries. Also, 37% have low HDL values. Having established percentages of the study population in NCEP categories, a framework is set to assess the implications of tracking.

Tracking coefficients for Data A, Data B, and Data C are described in Tables 4, 5, and 6 respectively. Standardized coefficients are most desirable, because they allow for comparison with other coefficients in similar studies. However, the mixed effects model was not standardized, and is therefore compared to un-standardized linear regression coefficients.

As shown in Table 4, cholesterol in Data A has a moderate stability as compared to HDL and LDL. The coefficients for the mixed effects model for HDL and LDL are of borderline significance, which may be due to a smaller sample size and observations for these variables. Even with borderline significance, HDL proves to be most stable over time with a coefficient of 0.9283. Individuals with optimal HDL levels early in life tend

to maintain these optimal levels when compared over time with the same group of people. The most statistically stable variable is cholesterol with linear mixed and linear regression coefficients of 0.7408 and 0.5908 respectively.

Because traditional tracking methods do not bring into account random measurement error, within-person variability can also lead to underestimation of the tracking coefficient. The linear mixed effects model controls for this random measurement error by placing random effects on intercept and time. In Datasets A, B, and C the tracking coefficients from the linear mixed effects model are higher than the tracking coefficients from the linear regression model. This is an example of the underestimation of the true value by the linear regression model. For cholesterol the coefficients from the mixed model and regression model are 0.7408 and 0.5908 respectively in Data A. Cholesterol coefficients in Data B are 0.6469 for linear mixed effects model and 0.5408 for linear regression model. In Data C, cholesterol coefficients are 0.7668 and 0.6022 for linear mixed effect model and linear regression model respectively.

In comparing tracking coefficients between Data B and Data C, Data C is found to have the most stable cholesterol values obtained from mixed effects tracking coefficients (0.7668, CI 0.6941-0.8396). Also, this coefficient is similar to the mixed effects tracking coefficient for the overall study population, Data A (0.7408, CI 0.6749-0.8067). The distribution of cholesterol is equal among all categories, as shown in Chart 1. Therefore, the stability cholesterol among individuals without diagnosed cardiovascular disease is similar to stability of cholesterol in the whole study population.

Based on these tracking coefficients, the percentage of individuals in each of the NCEP categories for cholesterol will remain stable over time. Tracking coefficients for LDL cholesterol, 0.8687 (95% CI 0.6650-1.0724) and 0.9419 (95% CI 0.8402-1.0435) are both of borderline significance in the linear mixed effects model for Data B and Data C respectively. The results are similar for the linear regression model coefficients and are statistically significant. The corresponding coefficients are 0.7311 (95% CI 0.5139-0.9482) and 0.8426 (95% CI 0.7603-0.9249) for Data B and Data C.

In conclusion, the values of cholesterol have more stability for individuals without diagnosed cardiovascular disease. This result is expected because of the interventions and medications placed on individuals entering the study who have diagnosed cardiovascular disease. Based on tracking coefficients, those individuals entering the study without diagnosed cardiovascular disease will maintain their high values in future measurements. Given the descriptives of the data, the linear mixed effects model is the best fit for the analysis of tracking.

Implications of Findings

The basic question behind tracking analysis is: Is it worthwhile to screen subjects at an early age on risk factors for a particular chronic disease such as cardiovascular disease? A "perfect" tracking method does not exist. It is best to use a method of tracking that best fits the data under analysis. It is preferable to use a tracking method that is as simple as possible. In this instance, a categorical nonparametric approach will suffice; however, tracking of the data in the BLSA is the description of the longitudinal development of certain variables and the possible prediction of future values by early

measurements. Observations per subject vary in number and interval between visits. The method of tracking had to have the possibility to use all available longitudinal data and control for confounding variables. A regression modeling technique is the most appropriate, such as the linear regression model and the linear mixed effects model.

Before beginning tracking analysis, the researcher must decide whether to use a parametric or nonparametric approach. Parametric approaches can only be used for continuous variables, while nonparametric approaches can also be used for categorical variables. In this analysis, a nonparametric approach best fits the study population. In most nonparametric approaches, the population is divided into subgroups according to some arbitrary cut points. This causes the magnitude of the analysis to depend heavily upon this arbitrary decision. For example, when comparing Foulkes and Davis's tracking coefficient γ with Cohen's κ, or any other tracking coefficient based on the division of the population into subgroups, two randomly chosen individuals can have the same coefficient but move from two different percentile groups to another (Twisk et al., 1994; Foulkes & Davis, 1981).

The advantages of using the linear mixed effects model to analyze tracking are (1) all longitudinal data are used to estimate one regression coefficient indicating the overall causal relationship; (2) all causal relationships can be estimated under correction for both time-dependent and time-independent covariates; and (3) this method may be used in longitudinal studies with both equally and unequally spaced time periods. An alternative to using the mixed method approach, as described by Twisk (1997), to assess tracking is to use generalizing estimating equations (GEE). Its advantage is its ability to analyze by

continuous and discrete outcome variables. However, its greatest disadvantage is the fact that the method does not provide any information on how well the model fits the data.

Tracking coefficients for both HDL and LDL were either not statistically significant or were of borderline significance. Twisk et al., however, doubts the importance of the significance test for tracking coefficients. The magnitude of every tracking coefficient depends on the length of the interperiod. A significant tracking coefficient calculated over an interperiod of one year does not necessarily track better than a non-significant tracking coefficient calculated over a much longer interperiod (1994). In this study, the interval between visits ranges from one year to twenty-five years with an average of seven visits per subject. Based on the reasoning of Twisk et al., the significance of HDL and LDL is increased slightly. However, Data B and Data C have fewer subjects and observations, which does contribute to the wider confidence intervals and borderline significance.

Limitations

Because the data is derived from a longitudinal study, subjects diagnosed with cardiovascular disease are given treatment and remain a part of the study. Subjects with high cholesterol values have an increased chance of being diagnosed with cardiovascular disease, and begin medical treatment for this condition. Also, these subjects may undergo a change in behavior and diet that they normally would not have experienced without participating in this longitudinal study. Similarly, subjects without diagnosed cardiovascular disease may not receive pharmacological treatment; however, being exposed to a thorough physical examination every two years may encourage subjects to

alter their diet and lifestyle in order to maintain positive physical examination results. Consequently, the results of this analysis may not be generalizable to all individuals over 45 years of age in the United States population.

Throughout the history of the BLSA, different techniques have been used to measure cholesterol levels, which may alter the validity of the results. Although repeated measures are used in the mixed effects model, which allow for correction in measurement error, they may not be as effective in allowing for measurement error between different measurement techniques. A study performed on the BLSA by Hershcopf, et al., examined data using three techniques to compensate for methodologic drift or for blocks of time during which technical errors may have occurred (1982). In order to correct this limitation, three techniques may be applied to the subsample of the BLSA data set used in this study. (1) Reanalyze a random sample of frozen and lyophilized plasma over the entire time period at the end of the study and compare to the original analysis. (2) Examine the mean age-specific cholesterol values over the time span under analysis. (3) Compute the regression of cholesterol on age for each subject and the deviation of each datum from the value predicted. These deviations are grouped by time period. If in any block of time, cholesterol values are measured incorrectly, the mean deviation for that time period will differ significantly from zero (Hershcopf, et al., 1982).

Future Study

Using a mixed model approach is a new method to assess tracking; therefore, many areas have not been explored using this model. Future analysis should be

conducted using data from the BLSA or a similar longitudinal dataset in order to produce results comparable to these.

In addition to analyzing total serum cholesterol, LDL cholesterol, and HDL cholesterol, the TC/HDL ratio would be a beneficial factor to analyze. This ratio is a better measure for cardiovascular disease risk, because it weighs the amount of HDL in the total cholesterol value. Predicting the stability of this ratio will define the stability of a specified risk for cardiovascular disease. Within the two days of testing during each subject's visit to the Gerontology Research Center, data is collected on diet, exercise, and specific medications taken. Access to these variables was not available for this analysis, however, future analysis will benefit from controlling for these variables. All three alter total cholesterol as well as LDL and HDL levels.

For future studies, analysis should be conducted on the linear mixed effects model to learn how to standardize its coefficients. Most tracking coefficients are standardized to allow comparability between studies; however, there is currently no method to standardize linear mixed effects coefficients. Also, the BLSA lacks a sufficient number of repeated measurements to have a significant population of women or racial minorities. Similar tracking analysis using the linear mixed effects model on a more diverse population is needed. There are differences among races and between genders on cholesterol levels as well as risk for cardiovascular disease. It is necessary to determine how these groups track, both separately and within a study group representative of the general population.

CHAPTER VI

REFERENCES

American Heart Association (2003). About cholesterol Retrieved on March 1, 2003 from http://www.americanheart.org/presenter.jhtml

Berns, M.A., DeVries, J.H., & Katan, M.B. (1988). Determinants of the increase of serum cholesterol with age: A Longitudinal Study. International Journal of Epidemiology, 17(4), 789-796.

Blanco-Cedres, L., Daviglus, M.L., Garside, D.B., Liu, K., Pirzada, A., Stamler, J., & Greenland, P. (2002). Relation of cigarette smoking to 25-year mortality in middle-aged men with low baseline serum cholesterol. American Journal of Epidemiology, 155(4), 354-360.

Burke, G.L., Arnold, A.M., Bild, D.E., Cushman, M., Fried, L.P., Newman, A., Nunn, C., & Robbins, J. (2001). Factors associated with healthy aging: The Cardiovascular Health Study. Journal of the American Geriatric Society, 49, 254-262.

Clark, D.A., Allen, M.F., & Wilson, F.H. (1967). Longitudinal study of serum lipids. The American Journal of Clinical Nutrition, 20(7), 743-752.

Foulkes, M.A. & Davis, C.E. (1981). An index of tracking for longitudinal data. Biometrics, 37, 439-446.

Grundy, S.M., Cleeman, J.I., Rifkind, B.M., & Kuller, L.H. (1999). Cholesterol lowering in the elderly population. Arch Internal Medicine, 159, 1670-1678.

Health Net. What you should know about cholesterol. Brochure. Retrieved March 24, 2003 from www.healthnet.com/general/wellsite/healthlibrary/pdf/aboutcholesterol.pdf

Hershcopf, R.J., Elahi, D., Andres, R., Baldwin, H.L., Raizes, G.S., Schocken, D.D., & Tobin, J.D. (1982). Longitudinal changes in serum cholesterol in man: An epidemiologic search for an etiology. Journal of Chronic Disease, 35, 101-114.

Lussier-Cacan, S., Bolduc, A., Xhignesse, M., Niyonsenga, T., & Sing, C.F. (2002). Impact of alcohol intake on measures of lipid metabolism depends on context defined by gender, body mass index, cigarette smoking, and apolipoprotein E genotype. Arteriosclerosis, Thrombosis, and Vascular Biology, 22(5), 824-831.

Maffetone, P. Clarifying cholesterol. MAF BioNutritionals Pursuing Longevity Naturally. Retreived March 24, 2003 from http://mafgroup.securedata.net/news_pv/pv_123101.shtml

Maggi, S., Minicuci, N., Harris, T., Motta, Luciano, Baldereschi, M., DiCarlo, A., Inzitari, D., & Brepaldi, G. (2001). High plasma insulin and lipids profile in older individuals: The Italian Longitudinal Study on Aging. Journal of Gerontology, 56A(4), M236-M242.

Matthews, C.E., Freedson, P.S., Hebert, J.R., Stanek, E.J., Merriam, P.A., Rosal, M.C., Ebbeling, C.B., & Ockene, I.S. (2001). Seasonal variation in household occupational and leisure time physical activity: Longitudinal analyses from the Seasonal Variation of Blood Cholesterol Study. American Journal of Epidemiology, 153(2), 172-183.

McMahan, C.A. (1981). An index of tracking. Biometrics, 37, 447-455.

Meigs, J.B. (2002). Epidemiology of the metabolic syndrome, 2002. American Journal of Managed Care, 8(11), S283-S292.

Mercola, J. (2003). Lipoprotein (a) increases heart disease risk. Retrieved March 20, 2003 from http://www.mercola.com/2000/sept/17/lipoprotein_a.htm

Morrell, C.H., Brant, L.J., Pearson, J.D., Verbeke, G.N., & Fleg, J.L. (2003). Applying linear mixed effects models to the problem of measurement error in epidemiologic studies. Communications in Statistics: Simulation and Computation, 32(2).

Neufeld, E.J., Mietus-Snyder, M., Beiser, A.S., Baker, A.L., newburger, J.W. (1997). Passive cigarette smoking and reduced HDL cholesterol levels in children with high-risk lipid profiles. Circulation, 96, 1403-1407.

Nishio T., Mori, C., Haneda, N., Haneda N., Watanabe K., Kishida K., Saito M., & Kajino Y. (1987). Quantification of blood pressure tracking of children by tracking index: The Shimane Heart Study. Japanese Circulation Journal, 51, 1404-1408.

Shock, N.W., Greulick, R.C, Andres, R., Arenberg, D., Costa, P.T., Lakatta, E.G., & Tobin, J.D. (1984). Normal Human Aging: The Baltimore Longitudinal Study of Aging. Baltimore, MD: National Institutes of Health.

Solfrizzi, V., Panza, F., Colacicco, A.M., Capurso, C., D'Introno, A., Torres, F., Baldassarre, G., & Capurso, A. (2002). Relation of lipoprotein(a) as coronary risk factor to type 2 diabetes mellitus and low-density lipoprotein cholesterol in patients $\geq$ 65 years of age (The Italian Longitudinal Study on Aging). American Journal of Cardiology, 89, 825-829.

Twisk, J.R. (1997). Different statistical models to analyze epidemiological observational longitudinal data: An example from the Amsterdam Growth and Health Study. International Journal of Sports Medicine, 18(3), S216-S224.

Twisk, J.R., Kemper, H.G., Mechelen, W., & Post, G.B. (1997b). Tracking of risk factors for coronary heart disease over a 14-year period: A comparison between lifestyle and biologic risk factors with data from the Amsterdam Growth and Health Study. American Journal of Epidemiology, 145(10), 888-898.

Twisk, J.R., Kemper, H.G., & Mellenbergh, G.J. (1994). Mathematical and analytical aspects of tracking. Epidemiologic Reviews, 16(2), 165-183.

Twisk, J.R., Kemper, H.G., Mellenbergh, G.J., & Mechelen, W. (1997a). A new approach to tracking of subjects at risk for hypercholesteremia over a period of 15 years: The Amsterdam Growth and Health Study. European Journal of Epidemiology, 13, 293-300.

Twisk, J.R., Kemper, H.G., Mellenbergh, G.J., & Mechelen, W. (1996). Factors influencing tracking of cholesterol and high-density lipoprotein: The Amsterdam Growth and Health Study. Preventive Medicine, 25(3), 355-364.

Valkonen, M., Kuusi, T. (1998). Passive smoking induces atherogenic changes in low-density lipoprotein. Circulation, 97, 2012-2016.

Walzen, R.L., Watkins, S., Frankel, E.N., Hansen, R.J., & German, J.B. (1995). Older plasma lipoproteins are more susceptible to oxidation: A linking mechanism for the lipid and oxidation theories of atherosclerotic cardiovascular disease. Applied Biological Sciences, 92, 7460-7464.

Ware, J.H. & Wu, M.C. (1981). Tracking: Prediction of future values from serial measurements. Biometrics, 37, 427-437.

Wilsgaard, T., Jacobsen, B.K., Schirmer, H., Thune, I., Løchen, M., Njølstad, I., Arnesen, E. (2001). Tracking of cardiovascular risk factors: The Tromsø Study, 1979-1995. American Journal of Epidemiology, 154(4), 418-426.

Wissler, R.W., & Robert, L. (1996). Aging and cardiovascular disease: A summary of the Eighth Münster International Arteriosclerosis Symposium. Circulation, 93, 1608-1612.

Zhang, D. & Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. Biometrics, 57, 795-802.

Zuliani, G., Romagnoni, F., Bollini, C., Leoci, V., Soattin, L., & Fellin, R. (1999). Low levels of high-density lipoprotein cholesterol are a marker of disability in the elderly. Gerontology, 45, 317-322.