

Research

Open Access

A longitudinal analysis of the risk factors for diabetes and coronary heart disease in the Framingham Offspring Study

Alok Bhargava*

Address: Department of Economics, University of Houston, Texas, USA

Email: Alok Bhargava* - bhargava@uh.edu

* Corresponding author

Published: 14 April 2003

Received: 17 March 2003

Population Health Metrics 2003, 1:3

Accepted: 14 April 2003

This article is available from: <http://www.pophealthmetrics.com/content/1/1/3>

© 2003 Bhargava; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Background: The recent trends in sedentary life-styles and weight gain are likely to contribute to chronic conditions such as hypertension, diabetes, and cardiovascular diseases. The temporal sequence and pathways underlying these conditions can be modeled using the knowledge from the biomedical and social sciences.

Methods: The Framingham Offspring Study in the U.S. collected information on 5124 subjects at baseline, and 8, 12, 16, and 20 years after the baseline. Dynamic random effects models were estimated for the subjects' weight, LDL and HDL cholesterol, and blood pressure using 4 time observations. Logistic and probit models were estimated for the probability of diabetes and coronary heart disease (CHD) events.

Results: The subjects' age, physical activity, alcohol consumption, and cigarettes smoked were important predictors of the risk factors. Moreover, weight and height were found to differentially affect the probabilities of diabetes and CHD events; body weight was positively associated with the risk of diabetes while taller individuals had lower risk of CHD events.

Conclusion: The results showed the importance of joint modeling of body weight, LDL and HDL cholesterol, and blood pressure that are risk factors for diabetes and CHD events. Lower body weight and LDL concentrations and higher HDL levels achieved via physical exercise are likely to reduce diabetes and CHD events.

Background

The prevalence of chronic conditions such as hypertension, non-insulin dependent diabetes mellitus, and coronary heart disease (CHD) in developed countries demand substantial medical resources [1]. Moreover, these conditions are becoming increasingly common among the well-off groups in middle and low-income countries [2]. While the quality of life and health of individuals is adversely affected by chronic conditions, there is a concomitant loss in work productivity [3,4]. This is especially important for middle and low-income countries where skilled labor is

relatively scarce and the treatment of chronic conditions may entail a reduction in health care resources available for diseases afflicting the poor [5]. A preventive approach to chronic diseases is therefore appealing.

Population surveys covering a large number of individuals such as NHANES in the U.S. [6] can provide useful insights into risk factors for chronic diseases. However, the gradual evolution of the multiple risk factors and the onset of chronic diseases cannot be addressed using cross-sectional data. By contrast, longitudinal studies such as

the Framingham Offspring study (FOS) [7] spanning over decades, can provide insights into the evolution of the risk factors and their partial and/or joint effects on chronic conditions such as diabetes and CHD.

There is a growing interest among policy makers in identifying preventive strategies for tackling chronic conditions. For example, the Women's Health Initiative is an ongoing longitudinal study in the U.S. covering 48,000 women for investigating the risk factors of breast cancer [8]. Detailed analyses of existing longitudinal data sets can provide useful insights into the effects of factors such as smoking, physical activity, and alcohol consumption on the incidence of diabetes and CHD. Because these conditions develop gradually over time, it is important to analyze their effects on intermediate risk factors such as blood pressure, and LDL and HDL cholesterol. Application of longitudinal econometric models incorporating the inter-dependence between the multiple risk factors can provide further insights.

The purpose of this paper is to analyze 4 time observations available on over 5,000 subjects in the FOS and develop empirical models for the subjects' weight, HDL and LDL cholesterol, and systolic and diastolic blood pressure that are potential risk factors for diabetes and CHD. Models were also developed for the probability of diabetes and CHD events in the 20-year period. A comprehensive analysis of the multiple risk factors in the FOS using alternative statistical techniques can facilitate an assessment of the likely effects of behavioral modifications for reducing the incidence of diabetes and CHD.

Methods

Study sample

From 1971, a cohort of 5124 men and women who were the children or spouses of the subjects in the original Framingham Heart study were recruited for the FOS [7,9]. The subjects were examined at the baseline (Exam 1) and in Exams 2, 3 and 4 that took place, respectively, 8, 12 and 16 years after Exam 1. The diabetes, CHD, and survival status was again assessed 20 years after Exam 1.

Variables measured longitudinally in Exams 1–4 in the Framingham Offspring public-use files

The subjects' age, sex, alcohol intake, and the number of cigarettes smoked per day were investigated in Exams 1–4. Systolic and diastolic blood pressure was measured in the left arm after the subjects had been sitting still for at least 5 minutes. Height was measured in inches in Exams 1 and 3 and weight was measured in pounds in all 4 exams using a standard beam balance. The surveys investigated physical activity patterns in Exams 2 and 4. An index of physical activity was constructed on the basis of the reported hours per day of sedentary, slight, moderate and heavy activities.

An index of alcohol consumption was constructed using the reported intakes of beer, wine, and other alcoholic beverages.

In each exam, blood was drawn after a 12-hour fast for determination of plasma glucose; non-insulin dependent diabetes mellitus was defined as glucose greater than 140 mg per deciliter of blood or if the subjects were taking prescription medication. HDL cholesterol was measured after precipitation of the plasma with heparin-manganese. LDL cholesterol was determined according the techniques described in Lipid Research Clinic Program [10]. A CHD event was defined as the occurrence of angina pectoris, myocardial infarction, coronary insufficiency, or coronary death.

The analytical framework

The risk factors for diabetes and CHD events such as weight, HDL and LDL, and blood pressure respond gradually over time to dietary intakes, life-style, smoking, etc. Of these, changes in body weight resulting from energy imbalance are apparent to the subjects themselves. While the nutrient composition of the diet was not measured in the FOS, one would expect body weight to be a predictor of LDL because fat intakes have increased in the observation period [2]. Moreover, body size is a predictor of energy needs. Thus, one would expect height and weight to be predictors of nutrient intakes, LDL, and other risk factors. Because height is a good approximation for skeletal size, height is an important predictor of weight [11–13]; height also reflects nutrition in the early years that is influenced by socioeconomic factors.

In models for the risk factors, it may be inappropriate to combine height and weight as the BMI [14]. From the standpoint of CHD events, it may be more risky for shorter individuals to gain weight than for taller individuals because coronary artery diameter is likely to be higher in taller subjects [15]. By contrast, the risk of diabetes may be less dependent on height; persisting energy imbalances may lead to similar outcomes in terms of the development of insulin resistance. It is therefore desirable to include height and weight in the empirical models and test the null hypothesis that these variables can be combined as the BMI [13].

Because specification of empirical models is limited by data availability, it was necessary to make simplifying assumptions in analyzing the risk factors. The empirical model for the data from the FOS is outlined in the Figure 1. The subjects' age, sex, and height were fixed characteristics. However, due to a paucity of socioeconomic variables such as education and income, the box containing the background characteristics also included the number of cigarettes smoked, alcohol intake, and physical activity.

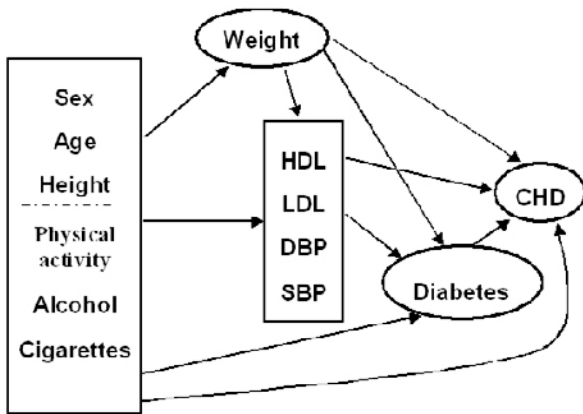


Figure 1
Empirical model for risk factors for diabetes and coronary heart disease. A graphical representation of the empirical model for risk factors for diabetes and coronary heart disease in the Framingham Offspring Study.

Because weight changes can influence risk factors such as blood pressure and LDL cholesterol, weight of the n subjects was first explained by a dynamic formulation allowing the current weight to depend on previous measurement ($i = 1, 2, \dots, n; t = 2, 3, 4$):

$$\ln (WT)_{it} = a_0 + a_1 (Sex)_i + a_2 \ln (Age)_i + a_3 [\ln (Age)]_i^2 + a_4 \ln (Height)_i \quad (1)$$

$$+ a_5 (Alcohol \text{ index})_{it} + a_6 (Cigarettes)_{it} + a_7 (Physical \text{ activity})_{it}$$

$$+ a_8 \ln (WT)_{it-1} + u_{1it}$$

Here, \ln represents natural logarithm. Subjects' weight, age, and height were transformed into natural logarithms, partly to reduce heteroscedasticity [16]. The coefficients of the explanatory variables in logarithms were thus the "elasticities" (percentage change in the dependent variable resulting from a 1% change in the independent variables). Because the model in equation (1) was dynamic, the long run impact of an explanatory variable on weight was the estimated coefficient in equation (1) divided by $(1-a_8)$. Indicator variables for the observations from Exams 3 and 4 were included in the model to allow different time means in all 4 examinations. The u_{1it} in equation (1) were random error terms that could be decomposed in a random effects fashion as:

$$u_{1it} = \delta_i + v_{1it} \quad (2)$$

where, δ_i were subject specific random effects that were assumed to be normally distributed with zero mean and a constant variance, and v_{1it} were normally distributed error terms with zero mean and constant variance [17].

The models for HDL and LDL cholesterol and systolic and diastolic blood pressure were also dynamic and, in addition to the explanatory variables in equation (1), contained weight as an explanatory variable. For example, the dynamic random effects model for HDL could be written as ($i = 1, 2, \dots, n; t = 2, 3, 4$):

$$\ln (HDL)_{it} = b_0 + b_1 (Sex)_i + b_2 \ln (Age)_i + b_3 [\ln (Age)]_i^2 + b_4 \ln (Height)_i \quad (3)$$

$$+ b_5 (Alcohol \text{ index})_{it} + b_6 (Cigarettes)_{it} + b_7 (Physical \text{ activity})_{it}$$

$$+ b_8 \ln (WT)_{it} + b_9 \ln (HDL)_{it-1} + u_{2it}$$

The error term u_{2it} in equation (3) also had a random effects structure as in equation (2). Moreover, the random effects affecting u_{2it} could be correlated with those influencing weight. Such problems can be addressed using the econometric techniques for "endogeneity" briefly outlined in the next section.

The binary logistic or probit models for whether the subjects' developed diabetes or had a CHD event in the 20-year period, explained by the characteristics in Exam 1, were, respectively:

$$(Diabetes)_i = c_0 + c_1 (Sex)_i + c_2 (Age)_i + c_3 (Height)_i + c_4 (Alcohol \text{ index})_{i1} \quad (4)$$

$$+ c_5 (Cigarettes)_{i1} + c_6 (Physical \text{ activity})_{i1} + c_7 (HDL)_{i1} + c_8 (LDL)_{i1}$$

$$+ c_9 (SBP)_{i1} + c_{10} (DBP)_{i1} + c_{11} (WT)_{i1} + u_{3i}$$

and

$$(CHD)_i = d_0 + d_1 (Sex)_i + d_2 (Age)_i + d_3 (Height)_i + d_4 (Alcohol \text{ index})_{i1} \quad (5)$$

$$+ d_5 (Cigarettes)_{i1} + d_6 (Physical \text{ activity})_{i1} + d_7 (HDL)_{i1} + d_8 (LDL)_{i1}$$

$$+ d_9 (SBP)_{i1} + d_{10} (DBP)_{i1} + d_{11} (WT)_{i1} + d_{12} (Diabetes)_{i1} + u_{4i}$$

Here, the variables Diabetes or CHD were latent variables assuming values 0 or 1 depending on if the subject reported to have had diabetes or a CHD event, respectively. Explanatory variables in equations (4) and (5) were the

Table 1: Sample means and standard deviations of selected variables for the subjects in the 4 exams in the Framingham offspring study¹

	Exam 1 N = 5120	Exam 2 n = 3861	Exam = 3 n = 3871	Exam 4 n = 4017
Age, y	36.3 ± 10.4			
Sex, 0–1 (Female = 1)	0.52 ± 0.50			
Height, inches	66.21 ± 3.76			
Weight, lb	158.34 ± 34.28	161.73 ± 34.23	165.78 ± 34.84	168.67 ± 35.31
BMI, Kg/m ²	25.39 ± 4.42	25.82 ± 4.41	26.28 ± 4.58	26.91 ± 4.74
Systolic blood pressure, mm Hg	121.59 ± 15.96	122.34 ± 16.74	123.64 ± 16.78	129.97 ± 18.42
Diastolic blood pressure, mm Hg	78.32 ± 11.25	79.35 ± 11.08	79.42 ± 9.97	77.71 ± 10.72
HDL, mg/dL	50.52 ± 14.68	48.31 ± 13.46	50.88 ± 14.86	49.48 ± 14.83
LDL, mg/dL	124.86 ± 35.43	130.59 ± 35.19	133.88 ± 36.55	-
Cigarettes smoked per day ² , n	13.95 ± 14.74	8.49 ± 13.76	6.83 ± 12.88	5.49 ± 11.55
Physical activity score, n	-	3.91 ± 3.61	-	5.75 ± 4.20
Alcohol index, n	3.71 ± 5.09	3.77 ± 5.29	3.34 ± 4.84	2.85 ± 4.35
Proportion with diabetes	0.016 ± 0.13	0.026 ± 0.16	0.037 ± 0.19	0.051 ± 0.22
Proportion with coronary heart disease	0.013 ± 0.11	0.032 ± 0.18	0.047 ± 0.21	0.065 ± 0.25
Age diabetes diagnosed ³ , y	52.13 ± 10.20			
Age coronary heart disease diagnosed ³ , y	54.42 ± 8.76			

¹Values are means ± standard deviations. ²Only the smokers were included in these calculations. ³Only the 6.7% of the subjects with diabetes and 9.3% subjects with coronary heart disease were included in these calculations.

measurements taken at Exam 1. The analyses were also done separately for men and women and the null hypothesis of constancy of the model parameters in the two groups was tested using likelihood ratio tests. The estimation methods assumed the errors u_{3i} and u_{4i} to be drawings from a logistic distribution or from a normal distribution for the probit models. For probit analysis using current levels of explanatory variable in the 4 time periods, random effects models were estimated under the assumptions that u_{3i} and u_{4i} were normally distributed with a random effects structure as in equation (2). Finally, Cox proportional hazard models were estimated for the age at first CHD event.

Statistical methods

Because only 4 time observations were available, estimation of the dynamic models given in equation (1) was based on the assumptions that the number of subjects (n) was large but the number of time periods was fixed. Thus, initial observations on the dependent variables (e.g. WT_{i1} in equation (1)) were treated as "endogenous" variables (correlated with the errors, [18]). The errors on equations (1) were assumed independent across subjects but correlated over time with a positive definite variance-covariance matrix. The random effects decomposition in equation (2) was a special case of this formulation.

The joint determination of the 4 observations in the dynamic models for weight (and HDL, LDL, and systolic and diastolic blood pressure) implied that econometric techniques for estimating simultaneous equations models were useful. Details of the maximum likelihood estima-

tion method are presented in [18]. The profile log-likelihood functions were optimized using a numerical scheme E04 JBF from [19]; asymptotic standard errors of the parameters were obtained by approximating second derivatives of the function at the maximum. Possible correlation between the random effects δ_i and the mean over time of the subject's body weight was tested using a likelihood ratio statistic that was distributed, for large n , as a Chi-square variable with 4 degrees of freedom. The restrictions for combining height and weight as the BMI in equation (3), for example, were:

$$b_4 + 2b_8 = 0 \quad (6)$$

These were tested by a likelihood ratio test that was distributed for large n as a Chi-square variable with 1 degree of freedom.

For the descriptive statistics, the package SPSS [20] was used. Binary logistic models for diabetes and CHD events were estimated using SPSS; Cox proportional hazard models were also estimated using SPSS. Probit models were estimated using the packages LIMDEP [21] and STATA [22].

Results

Descriptive statistics

The sample means and standard deviations of selected variables from the 4 exams in the FOS are presented in Table 1. The subjects' average age was 36 years and 52% were women. There was an increase in the body weight over time. This was also true for systolic blood pressure and for

Table 2: Maximum likelihood estimates of dynamic random effects model for Weight, HDL and LDL of the subjects in the Framingham offspring study in the 4 exams explained by demographic, behavioral and anthropometric variables^{1,2}

Independent variable	Dependent variable					
	Weight n = 2481		HDL n = 2481		LDL ³ n = 2503	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
Constant	1.303*	0.136	3.6750*	0.032	2.106*	0.451
Sex	-0.050*	0.008	0.100*	0.012	-0.023	0.023
Age ²	0.240*	0.018	0.035*	0.004	-0.203*	0.060
Age-squared ²	-0.033*	0.002	-0.005*	0.001	0.053*	0.004
Physical activity score	-0.001*	0.0006	0.002*	0.001	0.001	0.002
Alcohol index	0.0007*	0.0002	0.007*	0.001	-0.001	0.001
Cigarettes smoked	-0.0007*	0.0001	-0.002*	0.0002	0.001*	0.0003
Height ²	0.796*	0.097	0.747*	0.087	-1.291*	0.174
Weight ²	-	-	-0.469*	0.030	0.461*	0.055
Lagged dependent variable ²	0.495*	0.053	0.453*	0.017	0.305*	0.114
Indicator variable for Exam 3	0.012*	0.002	0.077*	0.005	0.005	0.007
Indicator variable for Exam 4	0.014*	0.003	0.032*	0.004	-	-
Chi-square statistic ⁴ , df = 4	-	-	20.2*	-	26.7*	-
Chi-square statistic ⁵ , df = 1	-	-	31.3*	-	57.3*	-

¹ Values are slope coefficients and standard errors; see equations (1) and (3) in the text for the explanation of the models. ² The dependent and the marked independent variables were in natural logarithms. ³ Only 3 time observations were available on LDL. ⁴ Chi-square test for the exogeneity of the mean over time of body weight. ⁵ Chi-square test for combining height and weight as the BMI. * P < 0.05.

LDL cholesterol. Table 1 also reports the mean number of cigarettes smoked per day calculated for smokers; there was a decline over time in the cigarettes smoked. For the subjects that were diagnosed with diabetes, the mean age at the time of the diagnosis was 52 years; the first CHD event was diagnosed at the average age of 54 years.

Results from estimating dynamic random effects models for body weight, HDL, LDL, and systolic and diastolic blood pressure

Table 2 presents the maximum likelihood estimates of dynamic random effects models for weight, and HDL and LDL cholesterol; the results for diastolic and systolic blood pressure are presented in Table 3. In all cases, the dependent variables were transformed into natural logarithms; the independent variables age, height, and weight were also transformed into logarithms.

Body weight

The results for body weight in the second column of Table 2 showed that men were significantly heavier than women. Both age and age-squared were significant predictors of weight thereby showing an increase in weight with age, though at a declining rate. From the estimated parameters, the turning point of weight with respect to age occurred at approximately 38 years. However, this estimate was subject to considerable estimation error and may have also been influenced by attrition in the sample.

The coefficient of physical activity score was negative but was not statistically significant. This could be because physical activity was measured only in Exams 2 and 4. The alcohol index was positively associated with weight, whereas cigarettes smoked were negatively associated; both these coefficients were statistically significant at the 5% level. Height was a significant predictor of weight, though the coefficient estimate 0.796 was significantly lower than the value 2; the data did not indicate a preference for modeling the BMI. Coefficient of the lagged dependent variable was estimated to be approximately 0.5 and was significant. Thus, the long run effects of an independent variable on weight were twice the magnitude of the corresponding short run coefficients in Table 2 (i.e. the coefficients divided by 1- the coefficient of the lagged dependent variable). Coefficients of the indicator variables for Exams 3 and 4 were positive and statistically significant.

HDL cholesterol

The results for HDL cholesterol are in the third column of Table 2. Men had significant lower concentrations of HDL than women. There was an increase in HDL with age though at a declining rate. The physical activity score and alcohol index were significant predictors of HDL with positive coefficients; the number of cigarettes smoked was negatively associated with HDL. The coefficient of height was positive and that of weight was negative in the model for HDL. However, the likelihood ratio test for combining height and weight as the BMI rejected the restrictions in

Table 3: Maximum likelihood estimates of dynamic random effects model for Systolic and Diastolic Blood Pressure of the subjects in the Framingham offspring study in the 4 exams explained by demographic, behavioral and anthropometric variables^{1,2}

Independent variable	Dependent variable			
	Systolic Blood Pressure n = 2481		Diastolic Blood Pressure n = 2481	
	Coefficient	SE	Coefficient	SE
Constant	3.791*	0.156	1.585*	0.158
Sex	-0.001	0.006	0.003	0.005
Age ²	-0.707*	0.029	0.425*	0.025
Age-squared ²	0.122*	0.003	-0.054*	0.003
Physical activity score	0.0007	0.0005	0.001	0.0006
Alcohol index	0.002*	0.0003	0.002*	0.0003
Cigarettes smoked	0.0000	0.0001	-0.0001	0.0001
Weight ²	0.262*	0.016	0.317*	0.018
Height ²	-0.451*	0.049	-0.373*	0.058
Lagged dependent variable	0.223*	0.046	0.180*	0.054
Indicator variable for Exam 3	0.005*	0.002	-0.005*	0.002
Indicator variable for Exam 4	0.026*	0.002	-0.009*	0.003
Chi-square statistic ³ , df = 4	30.8*		35.5*	
Chi-square statistic ⁴ , df = 1	35.1*		43.9*	

¹ Values are slope coefficients and standard errors; see equations (1)-(3) in the text for the explanation of the models.² The dependent and the marked independent variables were in natural logarithms. ³ Chi-square test for the exogeneity of the mean over time of body weight. ⁴ Chi-square test for combining height and weight as the BMI. * P < 0.05.

equation (6). This may have been partly due to the relatively large samples used in the estimation. The likelihood ratio test for exogeneity of the mean over time of body weight in the model for HDL also rejected the null hypothesis. Thus, the factors affecting body weight appeared to be correlated with the unobserved random effects affecting HDL. The results in Table 2 took account of these correlations in the estimation.

LDL cholesterol

The results for LDL cholesterol reported in the last column of Table 2 were based on the observations in Exams 1, 2 and 3; only the indicator variable for Exam 3 was included in this model. Sex differences in LDL were not statistically significant. The relationship between age and LDL was again a quadratic one though the turning point occurred at age 6.8 years indicating that, for the age range in the sample, there was a steady increase in LDL with time. Both the physical activity score and the index of alcohol intake were insignificant predictors of LDL. However, the number of cigarettes smoked was positively associated with LDL. Height was negatively associated whereas weight was positively associated with LDL. Although the likelihood ratio test again rejected the combination of height and weight as the BMI, one can broadly interpret the results as implying that subjects with higher BMI had lower HDL and higher LDL concentrations. Coefficient of the lagged dependent variable was statistically significant and was smaller than in the model for HDL presumably

due to greater changes in LDL in response to the dietary factors [23].

Systolic and diastolic blood pressure

Table 3 presents the maximum likelihood estimates of the dynamic random effects models for the systolic and diastolic blood pressure. There were no significant sex differences in the results from the two models. The quadratic terms in age were significant in both models though with opposite signs. The coefficients of age variables implied that systolic blood pressure declined until the age of approximately 18 years and thereafter increased. By contrast, diastolic blood pressure increased until the age 51 years and then showed a decline. These non-linear estimates were indicative of the complex time profiles of blood pressure [24].

Results from binary logistic and probit regressions for diabetes in the 20-year period

The results from estimating binary logistic and probit models for whether the subjects had diabetes during the 20-year period are in Table 4. To circumvent the problems due to censoring, the risk factors included were the measurements taken at Exam 1. Table 4 also reports the confidence intervals for the odds ratio using the parameter estimates from the logistic regression, and "marginal effects" estimated from the probit model. Chi-square statistics for combining height and weight as BMI are also reported.

Table 4: Maximum likelihood estimates of binary logistic and probit regression models for diabetes for subjects in the Framingham offspring study in the 20-year observation period predicted by the demographic, behavioral and anthropometric variables measured at Exam 1¹

Independent variable	Dependent variable: Diabetes in the 20-year period n = 3718							
	Binary logistic model				Probit model			
	Coefficient	SE	95% CI for exp (β)		Coefficient	SE	Marginal effect	SE
Constant	-8.658*	0.749	-	-	-4.666*	0.373	-	-
Sex	0.289	0.172	0.952	1.870	0.153	0.085	0.013	0.007
Age	0.053*	0.008	1.038	1.072	0.026*	0.413	0.002*	0.0003
Physical activity score	0.010	0.019	0.973	1.048	0.004	0.010	0.0003	0.0008
Alcohol index	0.016	0.012	0.992	1.040	0.008	0.007	0.0007	0.0006
Cigarettes smoked	0.009*	0.004	1.000	1.018	0.005*	0.002	0.0004*	0.0002
Systolic Blood Pressure	-0.004	0.008	0.981	1.011	-0.0001	0.004	-0.0001	0.0003
Diastolic Blood Pressure	0.035*	0.012	1.012	1.060	0.016*	0.006	0.001*	0.0005
LDL	-0.001	0.002	0.995	1.003	-0.0005	0.001	0.00004	0.0009
HDL	-0.038*	0.006	0.951	0.975	-0.018*	0.003	-0.002*	0.0003
BMI	0.112*	0.016	1.085	1.154	0.059*	0.008	0.005*	0.007
Chi-square statistic ² , df = 1	0.005	-			1.29	-		

¹ Values are slope coefficients and standard errors; see equation (4) in the text. ² Chi-square test for combining height and weight as the BMI based on a logarithmic specification of the model. * P < 0.05.

The results for the model for diabetes using logistic and probit models were consistent across the respective models. The coefficients of sex, physical activity score, alcohol index, systolic blood pressure, and LDL concentrations were not statistically different from zero. However, cigarette smoked, diastolic blood pressure, and weight were significantly positively associated with the probability of diabetes, whereas HDL and height were negatively associated. The likelihood ratio statistic accepted the null hypothesis that height and weight could be combined as the BMI. Thus, for example, a unit increase in the BMI at Exam 1 increased the chances of getting diabetes by between 8%–15% in the 20-year period.

Results from binary logistic and probit regressions for a CHD event in the 20-year period

The results from the binary logistic and probit regressions for CHD events during the 20-year period are presented in Table 5. Women had between 43%–75% lower chances of CHD events. Coefficients of physical activity, alcohol index, and systolic blood pressure were not statistically different from zero in Table 5. However, LDL and HDL concentrations, cigarettes smoked, and diastolic blood pressure were significant predictors of CHD events.

Height and weight were not significant predictors of CHD events, though the P-value of the coefficient of height was 0.053. When height and weight were combined as BMI, the likelihood ratio test rejected the restrictions implied by the BMI transformation in the logistic regression, and

the coefficient of BMI was not significantly different from zero. When weight was dropped from the model for CHD, height was significantly negatively associated with the probability of CHD in both the logistic and the probit models. By contrast, when height was dropped from the model, weight was not a significant predictor of CHD. These results indicated that diameter of coronary arteries was likely to be influenced by height and hence taller subjects had lower chances of CHD events. By contrast, the significance of BMI in the model for diabetes showed that, irrespective of height, being over-weight increased the chances of diabetes.

The coefficient of the variable for diabetes was positive and was a statistically significant predictor of CHD events; subjects with diabetes in Exam 1 had between 70%–534% higher chances of a CHD event. Lastly, the random effects probit models were estimated for diabetes and CHD events using current values of the explanatory variables in the 4 exams. Including the random effects, however, often led to boundary solutions using the algorithms in the software packages [21,22]. These results could be due to the serial correlation in the errors affecting longitudinal probit models. Moreover, Cox proportional hazard models were estimated for the age at which the subjects first had the CHD event using the explanatory variables measured at Exam 1. The predictive power of such models was poor in comparison with the results for the binary logistic and probit models indicating the uncertainties in predicting subjects' ages at the time of the first CHD event.

Table 5: Maximum likelihood estimates of binary logistic and probit regression models for a CHD event for subjects in the Framingham offspring study in the 20-year observation period predicted by the demographic, behavioral and anthropometric variables and diabetes in Exam I¹

Independent variable	Dependent variable: CHD event in the 20-year period n = 3718							
	Binary logistic model				Probit model			
	Coefficient	SE	95% CI for exp (β)		Coefficient	SE	Marginal effect	SE
Constant	-3.531	1.902	-	-	-2.072*	0.996	-	-
Sex	-0.979*	0.209	0.249	0.566	-0.507*	0.107	-0.051*	0.011
Age	0.080*	0.008	1.067	1.100	0.041*	0.004	0.004*	0.0004
Physical activity score	-0.021	0.017	0.947	1.013	-0.010	0.009	-0.001	0.001
Alcohol index	-0.004	0.011	0.974	1.019	0.002	0.006	-0.0002	0.0006
Cigarettes smoked	0.015*	0.004	1.008	1.023	0.009*	0.002	0.0009*	0.0002
Systolic Blood Pressure	-0.008	0.007	0.978	1.005	-0.003	0.004	-0.0003	0.0004
Diastolic Blood Pressure	0.033*	0.010	1.013	1.055	0.017*	0.005	0.002*	0.0006
LDL	0.007*	0.002	1.004	1.011	0.004*	0.001	0.0004*	0.0001
HDL	-0.020*	0.005	0.969	0.991	-0.009*	0.002	-0.0009*	0.0003
Weight	0.000	0.006	0.988	1.012	0.0005	0.003	0.0001	0.003
Height	-2.077	1.075	0.015	1.031	-1.075	0.564	-0.109	0.057
Diabetes in Exam I	1.195*	0.333	1.721	6.343	0.695*	0.196	0.070*	0.020
Chi-square statistic ² , df = 1	4.65*	-	-	-	2.46	-	-	-

¹ Values are slope coefficients and standard errors; see equation (5) in the text. ² Chi-square test for combining height and weight as the BMI based on a logarithmic specification of the model. * P < 0.05.

Discussion

This paper analyzed the effects of risk factors such as smoking, weight, HDL, LDL, and blood pressure for the development of chronic condition diabetes and CHD events using data from the FOS. Because of the gradual evolution of the risk factors, dynamic random effects models were used to explain the risk factors by age, physical activity, alcohol consumption, and cigarettes smoked. An advantage of the present approach was that one can discuss pathways through which the multiple risk factors affected the diabetes and CHD outcomes [25]; alternative approaches are available in the statistical literature [26].

While the inter-relationships between behavioral and biological factors are complex, the present analysis enables the estimation of the combined effects of the risk factors on CHD events under certain assumptions. The model represented by equations (4) and (5) was "triangular" and in the calculations reported below, we ignored the endogeneity of weight and potentially small bias in the estimate of the coefficient of diabetes in the model for CHD. The total effect of an explanatory variable on CHD events was thus the coefficient reported in Table 5, plus the coefficient of diabetes (1.195 in the logistic model) multiplied by the respective coefficient of the explanatory variable in the model for diabetes (Table 4).

First, after controlling for sex, age, physical activity, smoking, blood pressure, and LDL and HDL cholesterol, alco-

hol intake was not significantly associated with the probability of diabetes and CHD events in Tables 4 and 5, respectively. This is in contrast with the beneficial effects of alcohol intake on CHD among diabetic nurses in the U.S. [27], and male British doctors [28]. An important aspect in the analyzing the effects of alcohol intake is the type of drinks consumed and if they were consumed with meals [29]. Because alcohol intake data in the FOS cannot make such distinctions, it was perhaps reasonable to expect that the analysis would not provide unambiguous evidence on this issue. Further, in the dynamic random effects models, alcohol intake was positively associated with body weight, HDL, and diastolic and systolic blood pressure. Of these 4 variables, only HDL predicted lower chances of CHD events. Thus, the analysis of the multiple risk factors in the FOS indicated an overall tendency of the beneficial and harmful effects of alcohol intake to cancel out.

Second, LDL was seen to be an important risk factor for CHD events. The average LDL concentration at the first examination was approximately 125 mg/ dL. A decrease of 35 mg in LDL to 90 mg, for example, would constitute a 28% decline. Using the estimated parameters from the logistic regression (Table 5), the effect of this decrease would be to lower the chances of CHD events by between 14%–39%. Because LDL was not significant in the model for diabetes, there were no additional effects of lowering LDL. By contrast, an increase of 15 mg/dL in HDL concen-

tration would constitute an approximate 30% increase and predict a decline in CHD events by 15%–45%. Because HDL was negatively associated with diabetes, this decrease in HDL would further lower chances of CHD events by 3%–5%. These results also show the importance of disaggregating serum cholesterol into the HDL and LDL categories for the analysis of risk factors for cardiovascular diseases [30].

Third, the effects of smoking 14 cigarettes per day at the first examination would imply a likely increase in CHD events between 11%–32%; this effect would increase to between 12%–33% by taking into account the effects of smoking on diabetes. Using the estimates from Table 5, halving the average number of cigarettes smoked would reduce CHD events between 6%–16%. Lastly, the effects of body weight on CHD events in this population appeared to operate through the decline in diabetes. Using the results in Table 4, a 12% decrease in average BMI in Exam 1 to 22 was likely to reduce the number of subjects with diabetes from approximately 350 to 110. Because diabetic subjects had approximately a 3-fold greater chance of CHD events, the 12% reduction in BMI was likely to lead to a 10% decline in CHD events. Overall, the results from FOS indicated that the importance of reducing weight, LDL cholesterol and blood pressure [31] and increasing HDL for reducing the prevalence of diabetes and CHD events in the U.S. The econometric modeling of the risk factors indicated that it is better to rely on joint rather than the partial effects of risk factors in part because the time sequence of chronic diseases such as diabetes and CHD is known. The importance of modeling multiple risk factors for various diseases has also been emphasized in recent studies [32].

Competing interests

None declared.

Authors' contributions

AB conceived of this study, carried out the analyses and wrote the paper.

Acknowledgements

This study was part of a World Health Organization project, supported by a grant from the National Institute of Aging, National Institutes of Health (P01 AG17625-02), on which the author served as a consultant. The author is responsible for the views in the paper.

References

1. Posner BM, Cupples LA, Gagnon D, Wilson PWF, Chetwynd K and Felix D **Healthy people 2000: The rationale and potential efficacy of preventive nutrition in heart disease: The Framingham Offspring-Spouse study** *Arch Intern Med* 1993, **153**:1549-1556
2. Popkin B and Doak C **The obesity epidemic is a worldwide phenomenon** *Nutr Rev* 1998, **56**:106-114
3. Murray CJL and Lopez A **The global burden of disease** Cambridge, Harvard University Press 1996,
4. Bhargava A, Jamison DT, Lau LJ and Murray CJL **Modeling the effects of health on economic growth** *J Health Econ* 2001, **20**:423-440
5. Bhargava A **World Health Report 2000: Data limitations, statistical methods and policy analysis** *Lancet* 2001, **358**:1097-1098
6. **NHANES National Health and Nutrition Examination Survey III Atlanta, Center for Disease Control** 1994,
7. Wilson PWF, Garrison RJ, Castelli WP, Feinleib M, McNamara PM and Kannel WB **Prevalence of coronary heart disease in the Framingham Offspring study: Role of lipoprotein cholesterols** *Am J Cardiol* 1980, **46**:649-654
8. Women's Health Initiative Study Group **Design of the Women's Health Initiative Clinical Trial and Observational Study** *Controlled Clinical Trials* 1998, **19**:61-109
9. Wilson PWF, Anderson KM and Castelli WP **Twelve-year incidence of coronary heart disease in middle-aged adults during the era of hypertensive therapy: The Framingham Offspring study** *Am J Med* 1991, **90**:11-16
10. Lipid Research Clinic Program **Manual of laboratory operation volume 1 DHEW publication number**. Bethesda, National Institutes of Health 1975, **1**:75-628
11. Ehrenberg ASC **The elements of lawlike relationships** *J R Statist Soc A* 1968, **131**:280-302
12. Cole TJ **Weight-stature indices to measure underweight, overweight and obesity** In *Anthropometric assessment of nutritional status* (Edited by: Himes JH) New York: Wiley-Liss 1991, 83-111
13. Bhargava A **Modelling the health of Filipino children** *J R Statist Soc A* 1994, **157**:417-432
14. Kronmal RA **Spurious correlation and the fallacy of the ratio standard revisited** *J R Statist Soc A* 1993, **156**:379-392
15. Palmer JR, Rosenberg L and Shapiro S **Stature and the risk of myocardial infarction in women** *Am J Epidemiol* 1990, **132**:27-32
16. Nelson M, Black AE, Morris JA and Cole TJ **Between-and-within subject variation in nutrient intake from infancy to old age: Estimating the number of days to rank dietary intakes with desired precision** *Am J Clin Nutr* 1989, **50**:155-167
17. Laird N and Ware J **Random effects models for longitudinal data** *Biometrics* 1982, **38**:963-974
18. Bhargava A and Sargan JD **Estimating dynamic random effects models from panel data covering short time periods** *Econometrica* 1983, **51**:1635-1660
19. Numerical Algorithm Group **Numerical Algorithm Group Mark 13** Oxford, Oxford University 1991,
20. **SPSS SPSS for Windows version 10** Chicago, SPSS 1999,
21. **LIMDEP Econometric Software Inc** New York 1995,
22. **STATA Statistics, Graphics and data management** College Station, Stata Corporation 2001,
23. Keys A **Serum cholesterol response to dietary cholesterol** *Am J Clin Nutr* 1994, **40**:351-359
24. Flack JM, Neaton J, Grimm R, Shih J, Cutler J, Ensrud K and MacMahon S **Blood pressure and mortality among men with prior myocardial infarction** *Circulation* 1995, **92**:2437-2445
25. Neaton JD and Wentworth D **Serum cholesterol, blood pressure, cigarette smoking, and death from coronary heart disease. Overall findings and differences by age for 316,099 white men. Multiple Risk Factor Intervention Trial Research Group** *Arch Intern Med* 1992, **152**:56-64
26. Robins JM **Marginal structural models versus nested structural models as tools for causal inference** In *Statistical models in epidemiology* (Edited by: Halloran E) New York, Springer Verlag 1999,
27. Solomon C, Hu F, Stampher M, Colditz G, Speizer F, Rimm E, Willett W and Manson J **Moderate alcohol consumption and the risk of coronary heart disease among women with Type 2 diabetes mellitus** *Circulation* 2000, **102**:494-499
28. Doll R, Peto R, Hall E, Wheatley K and Gray R **Mortality in relation to consumption of alcohol: 13 years' observations on male British doctors** *BMJ* 1994, **309**:911-918
29. Vogel R **Alcohol, heart disease, and mortality: A review** *Rev Cardiovasc Med* 2002, **3**:7-13
30. Stamler J, Daviglius ML, Garside DB, Dyer AR, Greenland P and Neaton JD **Relationship of baseline serum cholesterol levels in 3 large cohorts of younger men to long-term coronary, cardiovascular, and all-cause mortality and to longevity** *JAMA* 2000, **284**(3):365-367
31. Bhargava A and Guthrie J **Unhealthy eating habits, physical exercise and macronutrient intakes are predictors of anthro-**

pometric indicators in the Women's Health Trial: Feasibility Study in Minority Populations *B J Nutr* 2002, **88**:719-728

32. Ezzati M, Hoorn SV, Rodgers A, Lopez AD, Mathers CD, Murray CJL and the Comparative Risk Assessment Collaborative Group **Potential health gains from reducing multiple selected major risk factors: Global and regional estimates** *Lancet*

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

