**Logistic Regression Modeling of Association between Systolic Blood Pressure and Coronary Heart Disease in the Framingham Study**

**Background**

Systolic blood pressure (SBP) has previously shown a strong association with risk of coronary heart disease (CHD)[[1]](#footnote-1). I investigated whether SBP aids in prediction of CHD, adjusting for sex, age, serum total cholesterol (STC), heart rate, and body mass index (BMI), potential confounders of the effect of SBP on CHD.

**Methods**

The dataset is from 1,150 patients from the Framingham Heart Study cohort [[2]](#footnote-2), excluding 17 patients with missing information on the outcome or covariates of interest, leaving 1,133 patients.

**Results**

The final model predicts CHD using SBP, adjusting for sex, age, and BMI, with the odds ratios summarized in Table 3. Those with SBP > 160 mmHg had 2.26 times the odds of CHD compared to those with SBP < 120 mm Hg, adjusting for the other covariates (95% CI: (1.38, 3.73)). There appears to be evidence of confounding by sex, age, and BMI, (>10% overall OR difference when these covariates were added to the model). There is no evidence of effect modification, as the interaction terms did not significantly aid in predicting CHD at the .05 level. There were no covariates that showed significant departure from linearity. Lastly, the Hosmer-Lemeshow goodness-of-fit test (p = 0.23) and area under ROC curve (0.69) indicate that the model is well calibrated and has adequate discrimination.

**Conclusions**

Systolic blood pressure, adjusting for sex, age, and BMI, aids in predicting CHD.

**Statistical Methods**

All analyses were conducted using STATA statistical software, version 13.1. The groupings for categorical variables and binary variables are described in Table 1. The characteristics of the categorical SBP groups were summarized and compared using χ2 tests for categorical and binary variables and 2-sample t tests (2-sided) for continuous variables, significant at p < 0.05. The odds ratios and their 95% confidence intervals comparing SBP categories on binary CHD outcome were computed using logistic regression and Wald tests. Separate analyses are presented for comparisons of baseline SBP category (<120 mmHg) to other SBP categories.

Forward selection logistic regression model building was used to test the inclusion of sex, age, STC, heart rate, and BMI to adjust for the main effect of SBP in CHD, with the elimination criteria of p > 0.10. Generalized additive models were used to determine the inclusion of non-linear terms. Inclusion in the final model required either biological rationale, statistical significance of the covariate’s odds ratio using a Wald test (p < 0.05), or evidence of confounding

From prior literature and biological rationale, potential confounding variables included sex (binary; male or female), age (binary; <50 or ≥50 years), serum total cholesterol (categorical; <200 mg/dL, 200-239 mg/dL, ≥240 mg/dL), body mass index (continuous; kg/m^2), and heart rate (continuous; beats/min). To analyze evidence of confounding, I looked at whether inclusion or exclusion of the covariate changed the adjusted odds ratio for the SBP categories variables by more than 10%. To assess potential effect modification, I conducted separate analyses for interactions believed to be significant: SBP category and BMI; SBP category and STC category; sex and age category; sex and BMI; age category and BMI; age category and STC category; and BMI and STC category.

**Table 1: Description of Model Variables**

|  |  |  |
| --- | --- | --- |
| **Characteristic** | **Units** | **Categories** |
| Coronary heart disease (CHD): Angina pectoris, myocardial infarction (hospitalized and silent or unrecognized), coronary insufficiency (unstable angina), or fatal coronary heart disease |  |  |
| Systolic blood pressure (SBP)[[3]](#footnote-3) (mean of last 2 of 3 measurements) | mm Hg | 0 = SBP < 120  1 = 120 ≤ SBP < 140  2 = 140 ≤ SBP < 160  3 = SBP ≥ 160 |
| Participant sex |  | 0 = Men  1 = Women |
| Age category[[4]](#footnote-4) | years | 0 = Age < 50  1 = Age ≥ 50 |
| Serum total cholesterol (STC) category[[5]](#footnote-5) | mg/dL | 0 = STC < 200  1 = 200 < STC < 240  2 = STC > 240 |
| Body mass index (BMI)6 | kg/m^2 |  |
| Heart rate (ventricular rate) | beats/min |  |

**Table 2: Characteristics of Participants in the Framingham Study Sample by Systolic Blood Pressure (SBP) Category**

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Characteristic** | **SBP < 120 mmHg** | | **120 ≤ SBP < 140 mmHg** | | | **140 ≤ SBP < 160 mmHg** | | | **SBP ≥ 160 mmHg** | | |
|  | **Mean (SD)** | **No (%)** | **Mean (SD)** | **No (%)** | **P-value****[[6]](#footnote-6)** | **Mean (SD)** | **No (%)** | **P-value6** | **Mean (SD)** | **No (%)** | **P-value6** |
| Total |  | n=343 |  | n=460 |  |  | n=210 |  |  | n=137 |  |
| Participant sex |  |  |  |  | 0.02 |  |  | 0.03 |  |  | 0.22 |
| *Men* |  | *133 (39)* |  | *217 (47)* |  |  | *101 (48)* |  |  | *45 (33)* |  |
| *Women* |  | *210 (61)* |  | *243 (53)* |  |  | *109 (52)* |  |  | *92 (67)* |  |
| Age |  |  |  |  | <0.01 |  |  | <0.01 |  |  | <0.01 |
| *Age < 50* |  | *252 (73)* |  | *243 (53)* |  |  | *75 (36)* |  |  | *18 (13)* |  |
| *Age ≥ 50* |  | *91 (27)* |  | *217 (47)* |  |  | *135 (64)* |  |  | *119 (87)* |  |
| STC |  |  |  |  | <0.01 |  |  | <0.01 |  |  | <0.01 |
| *STC < 200* |  | *96 (28)* |  | *78 (17)* |  |  | *31 (15)* |  |  | *18 (13)* |  |
| *200< STC <240* |  | *125 (37)* |  | *152 (33)* |  |  | *65 (32)* |  |  | *44 (32)* |  |
| *STC > 240* |  | *119 (35)* |  | *225 (50)* |  |  | *110 (53)* |  |  | *74 (54)* |  |
| BMI[[7]](#footnote-7) | 24.23 (3.51) |  | 25.67 (3.82) |  | <0.01 | 26.86 (3.96) |  | <0.01 | 28.37 (5.47) |  | <0.01 |
| Heart rate | 74.40 (11.31) |  | 75.22 (11.28) |  | 0.31 | 77.36 (12.48) |  | <0.01 | 81.18 (14.13) |  | <0.01 |

Abbreviations: SD, standard deviation; CHD, coronary heart disease; STC, serum total cholesterol; BMI, body mass index

**Table 3: Summary of Logistic Regression Analysis for Covariates Predicting CHD**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Characteristic** | **Crude OR** | **95% CI** | **Age, Sex-Adjusted OR[[8]](#footnote-8)** | **95% CI** | **Fully Adjusted OR****[[9]](#footnote-9)** | **95% CI** |
| SBP < 120 | 1.00 |  | 1.00 |  | 1.00 |  |
| 120 ≤ SBP < 140 | 1.89 | (1.33, 2.69) | 1.57 | (1.09, 2.26) | 1.46 | (1.01, 2.11) |
| 140 ≤ SBP < 160 | 2.48 | (1.65, 3.73) | 1.85 | (1.21, 2.83) | 1.60 | (1.04, 2.48) |
| SBP ≥ 160 | 3.96 | (2.54, 6.19) | 2.85 | (1.76, 4.60) | 2.26 | (1.38, 3.73) |

**Figure 1: Logistic Regression Model Equation**

**Selected Outputs:**

***Final model, Hosmer-Lemeshow goodness-of-fit test, area under the ROC curve:***

The final model, testing of good calibration, and testing of good discrimination are presented below, with rationale behind the model-building in outputs further below.

****

****

****

***Model-building, looking at including variables that significantly aid in predicting CHD:***

From prior knowledge, I chose the appropriate categorical and continuous versions of the covariates of interest. I also wanted to ensure that I adjusted for age and sex in assessing the relationship of systolic blood pressure in predicting CHD. Thus, I performed a forward selection approach to include additional variables significantly adding to the prediction of CHD in the model, on top age category and sex. Heart rate was insignificant at the 0.10 level in adding to the model prediction using the forward selection approach (p = 0.33), so that covariate was not included in the model.

**

Because both BMI and STC category were significant when added to the model (BMI, p < 0.01; STC, p = 0.01, < 0.01 for each indicator variable), I added one and then the other. All of the covariates were significant at the p = 0.10 level, although I next assessed evidence of confounding, effect modification using interaction terms, and potential nonlinear terms, to determine addition or deletion of final terms to the model.



***Testing for confounding:***

I tested for confounding by whether the addition of the variable changed the odds ratio from the previous model by more than 10%: |(OR SBP\_simple – OR SBP\_complex)| / (OR SBP\_simple) \* 100%. As the age, sex, and STC adjusted ORs appear to not significantly change (i.e., less than 10%) from the age and sex-adjusted ORs, it appears that STC is not a confounder in our model. Thus, I did not adjust for STC in the model and dropped that covariate from the model. There appeared to be evidence of confounding effects of sex, age, and BMI on SBP, as overall they changed by greater than 10% from their simpler reference model. Thus, I only adjusted for age, sex, and BMI as covariates, in addition to SBP, in order to predict SBP in the final model.

**Table 3: Summary of Logistic Regression Analysis for Covariates Predicting CHD**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Characteristic** | **Crude OR** | **Age, Sex-Adjusted OR** | **% Change from Crude OR** | **Age, Sex, BMI-Adjusted OR** | **% Change from Age, Sex-Adjusted OR** | **Age, Sex, STC Adjusted OR** | **% Change from Age, Sex-Adjusted OR** |
| SBP < 120 | 1.00 | 1.00 | 0% | 1.00 | 0% | 1.00 | 0% |
| 120 ≤ SBP < 140 | 1.89 | 1.57 | 17% | 1.46 | 7% | 1.49 | 5% |
| 140 ≤ SBP < 160 | 2.48 | 1.85 | 25% | 1.60 | 14% | 1.76 | 5% |
| SBP ≥ 160 | 3.96 | 2.85 | 28% | 2.26 | 21% | 2.80 | 2% |

****

****

****

******

***Testing for effect modification:***

Next, I looked at potential effect modification through identifying interaction terms of interest from prior knowledge. I computed all of these interactions terms of interest and ran a logistic regression with these terms in the model, to see if any of these interaction terms were statistically significant at the .05 level. None of the interaction terms of interest were significant at the .05 level. Thus, it appeared that there was no evidence of effect modification between the selected covariate interactions.

**

***Assessing need for non-linear terms using generalized additive models (GAMs):***

To assess whether any of the continuous or categorical covariates of interest had significant departure from linearity, I performed generalized additive models (GAMs), allowed the covariates to have multiple degrees of freedom to fit the data (heart rate, 4; sysbpcat, 3; bmi, 4). None of the covariates showed significant departure from linearity, as the p-values for the gain were not significant (heart rate, p = 0.1625; sysbpcat, p = 0.6936; bmi, p = 0.3147). Thus, I did not add any non-linear terms to the final logistic regression model.

****

****

1. Stokes III J et al. Blood pressure as a Risk Factor for Cardiovascular Disease- The Framingham Study- 30 Years of Follow-Up. Hypertension. 1989;13(5):I13-18. [↑](#footnote-ref-1)
2. <https://biolincc.nhlbi.nih.gov/teaching/> [↑](#footnote-ref-2)
3. Systolic blood pressure is categorized according to National Heart, Lung, and Blood Institute definitions: <http://www.nhlbi.nih.gov/health/health-topics/topics/hbp/> [↑](#footnote-ref-3)
4. Age is categorized according to prior evidence of distinct binary age categories in evaluating systolic blood pressure and hypertension: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2805932/> [↑](#footnote-ref-4)
5. Serum total cholesterol is categorized according to Mayo Clinic recommendations, with below 200, 200-239, and 240 and above as separate categories: <http://www.mayoclinic.org/diseases-conditions/high-blood-cholesterol/in-depth/cholesterol-levels/art-20048245> [↑](#footnote-ref-5)
6. P values were calculated using χ2 tests for categorical variables and 2-sample t tests for continuous variables [↑](#footnote-ref-6)
7. Weight (kg)/height (m)2. [↑](#footnote-ref-7)
8. Age and sex-adjusted OR from logistic regression model. Covariates included sex (binary; male or female) and age (binary; <50 or ≥50 years). [↑](#footnote-ref-8)
9. Fully adjusted OR from logistic regression model. Covariates included sex (binary; male or female), age (binary; <50 or ≥50 years), serum total cholesterol (categorical; <200 mg/dL, 200-239 mg/dL, ≥240 mg/dL), body mass index (continuous; kg/m^2). [↑](#footnote-ref-9)