Final Project Report: Smoking in the Framingham Heart Study

Team members: Gaeun Kim, Zhaoyu Liu, Youn Kyeong Chang, Lynette Pan

OBJECTIVE: The goal of this longitudinal data analysis is to study the relationship between current smoking status in the Framingham Heart study and age as well as the number of cigarettes smoked per day and age with the impact of smoking on certain health outcomes including systolic blood pressure, diastolic blood pressure and serum total cholesterol.

STUDY DESIGN: The Framingham Heart Study is a long term prospective study of the etiology of cardiovascular disease among participants living in Framingham, Massachusetts. This study began in 1948 and 5,209 subjects were initially enrolled. This analysis was conducted using a subset of Framingham study data which includes laboratory, clinic, questionnaire, and adjudicated event data on 4,434 participants. Participant clinic data was collected during three examination periods, approximately 6 years apart.

METHODS: To explore the relationships between variables of interest, they were visualized with and without modifier. Baseline values were retained as part of the outcome vector and no assumptions were made about the group differences at baseline. Missing values were assumed as missing at random (MAR). Confounders were identified both by literature and detecting significant change in OR/mean estimates for main explanatory variable in the model when potential confounder was introduced. OR/mean estimates was obtained from the mixed effect model for dichotomous/count outcome and generalized least square model for continuous outcome. Effect modifiers were inferred graphically and potentially considered as a covariate in the model with interaction term. Random intercept models were fitted under MAR assumption due to the unbalanced data and difference at baseline among individuals. In linear models, the significance of fixed effects was tested using bootstrap and random effects were tested using RLRsim function[1] or the graph at baseline. Based on the AIC score, the final model was selected among candidate models. For comparing nested model and checking overdispersion in the poisson model, Chi-square deviance test was used.

RESULTS: Among 4434 patients in the study, there are 447 patients with only one observation, 781 patients with two observations and 3206 patients have all three observations.

(1)-1. Current smoking status(cursmoke) for period 1 was completely obtained but there were 504 missing values for period 2 and 1171 missing for period 3. The average of age across periods is 54.8 (SD: 9.6) ranging from 32 to 81. Also, the probability of current smoking by categorized age was ranged from 0.13 to 0.60 with mean 0.41 (SD: 0.19). Figure 1 shows that propensity of cursmoke decreases as getting older and this relationship is modified by sex. Other than sex, attained education(educ)[2] and prevalent hypertensive(prevhyp) were identified as modifiers and high and low density lipoprotein cholesterol(hdlc and ldlc) as confounders. Among model candidates with various combinations of variables, age, educ, use of anti-hypertensive medication at exam(bpmeds), bmi, heart rate(heartrte), hdlc and ldlc were adjusted, supported by the data with AIC 2964.4. Given that the variance of random effects is almost zero in this model, estimates could be interpreted in population level. OR of cursmoke across age was modified by sex, decreasing by 5% (p-value: 3.38e-10) as one year older in male group and decreasing by 8% (p-value: 5.4e-25) in female group, holding other variables fixed. The OR of cursmoke was also decreased by 14% (p-value: < 2e-16) with one unit of BMI increase but increased about 2% (p-value: 4.08e-05) by one unit of heartrte increases after adjusting other variables.

(1)-2. non-smokers were filtered out to assess the relationship between number of cigarettes smoked per day(cigpday) and age. Among 2294 individuals with average number of examination cycle (period) 1.789 (SD: 0.783). The age of participants ranged from 32 to 80 and mean age was 52 (SD: 8.752). The average of cigpday across periods was 19.247 (SD: 11.611). Based on the final model (AIC: 7870.3), on a population level, compared to a group of subjects who are a year younger, we can expect the mean number of cigarettes smoked per day to be 1.019*exp(-0.0004549*heartrte) times higher (p-value: 0.273), holding sex, heartrter and hdlc constant. Although relationship of unadjusted cigpday and age seems to be modified by sex (Figure 1), after adjusting for other factors, gender is not a significant effect modifier at the significancelevel of 5%. Instead, cigpday and age were modified by heartrte (p-value: 0.0372).

(2)-1. The mean of systolic blood pressure (sysbp) is 133 (SD: 22.4) at first period, ranging from 83.5 to 295. Figure 2 shows that the SBP of non-current smoker is higher than current smoker, without adjustment. Age, diastolic blood pressure (diabp), bmi and prevhyp were identified as confounders. According to the result, the average SBP of current smoker is 9.766 (95% CI is [6.521, 12.714]) higher than that of non-current smoker, adjusting for other confounders at 0.05

significance level. In terms of effect modification, there is evidence that the proportion of current smokers declines by age[3]. All covariates are significantly related to sysbp.

(2)-2. The mean of diastolic blood pressure is 83 (SD: 12.1) at first period, ranging from 48 to 142.5. In figure 2, the diabp of non-smoker is slightly higher than smoker. Based on the OR method and literature, the confounders are sysbp, bmi and prevhyp. From the result, the average diabp of current smoker is 7.977 (95% CI is [10.303, 5.930]) lower than that of non-current smoker, adjusting for other confounders at 0.05 significance level. There is an significant interaction between current smoke and sysbp (p-value of 1.18e-24). All the covariates show significant relations with diabp.

(2)-3. Among the 4419 subjects who have serum totchol records, the propensity of being a smoker at the first period was 0.49. The average totchol at first period was 236.9 (SD: 44.6) mg/dL and ranged from 107 to 696 mg/dL, indicating that individuals have diverse baseline measures. Figure 2 shows that there seems to be no relationship between totchol and cursmoke without controlling for other variables. Sex, weight, bmi, age and sysbp were adjusted in the model[4,5] because of being confounders. Based on the resulting model (AIC:112160.4) with confounders adjusted, there is a significant difference in totchol between smokers and nonsmokers where gender acts as an effect modifier (p-value: 0.0042). Male current smoker on average would have 6.6 (95% CI of [3.847, 9.249]) higher in totchol compared to that of male non current smokers, whereas female current smoker on average would have 1.47 (95% CI of [-4.5, 7.93]) higher totchol compared to female non-current smokers after adjusted for age, bmi and sysbp.

CONCLUSION: In this sample of participants, odds of cursmoke was decreased with age and its effect was modified by sex, adjusting for educ, bpmeds, bmi, heartrte, hdlc and ldlc. The mean cigpday had a positive relationship with age, adjusting for sex, heartrter and hdlc and was modified by heartrte. As for the impact of smoking on certain health outcome, there is a positive relationship between both sysbp and totchol and current smoking status, adjusting for corresponding confounders in each model, but diabp and current smoking status was negatively related, adjusting sysbp, bmi and prevhyp. Specifically, the impact of current smoking status is modified by gender where for men, current smoker would have greater effect in increasing the expected totchol.

Appendix

Reference

1. https://cran.r-project.org/web/packages/RLRsim/RLRsim.pdf

2. Int J Epidemiol. 2008 Jun; 37(3): 615–624. Educational attainment and cigarette smoking: a causal association

3. BMC Public Health. 2005; 5: 57. Gender and age differences among current smokers in a general population survey.

4. https://medbroadcast.com/condition/getcondition/high-cholesterol

5. https://www.healthline.com/health/serum-cholesterol#prevention

Figure 1. Relationship between the categorized age and the outcome by sex

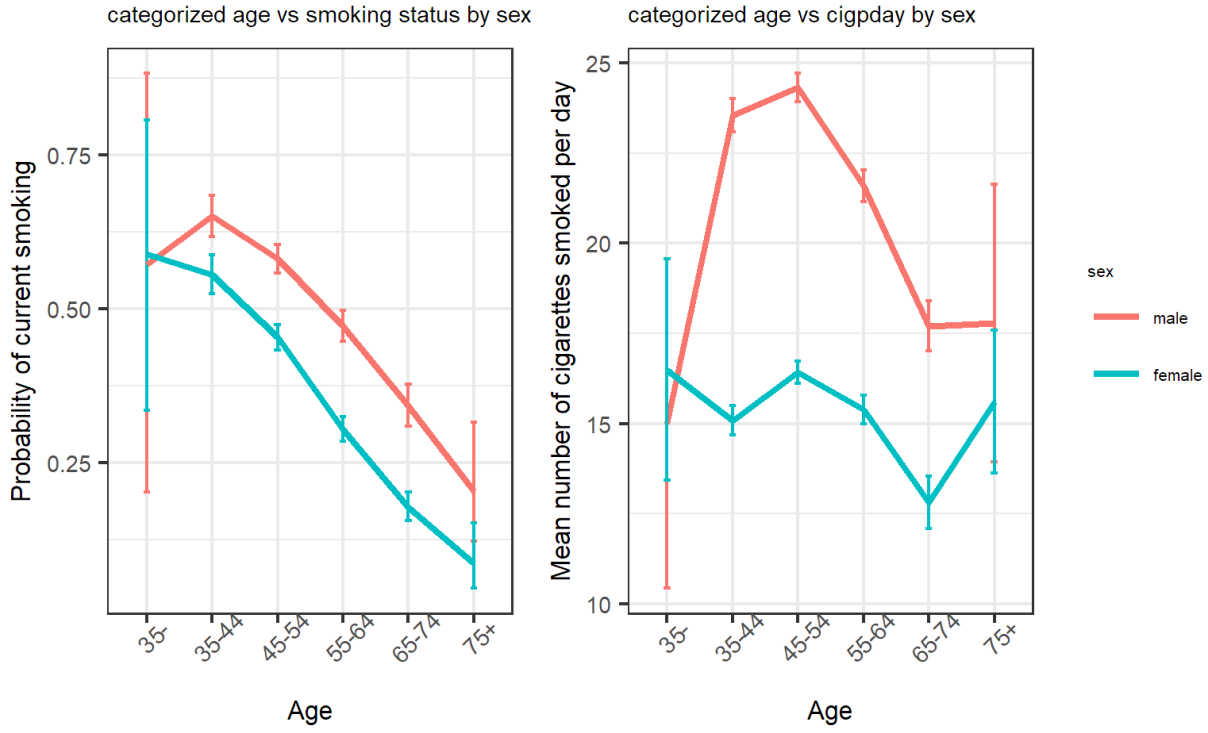

Figure 2. Relationship between current smoking stutus and the outcome without adjustment