Association of Smoking and Coronary Heart Disease Among Smokers Enrolled in The Framingham Heart Study (1956-1968)

METHODS

Study population

The Framingham Heart Study (FHS) began in 1948 with 5,209 initially enrolled participants from Framingham, Massachusetts. Individuals were examined biennially since the beginning of the study. At each biennial examination, participants were interviewed, and a detailed medical history was taken along with an assessment of cardiovascular risk factors. For the current analyses, we include clinical data collected from 1956-1968 and encompass three examination periods, about 6 years apart (n=4,434). All participants repeatedly provided written informed consent to participate in the study.

Exclusion criteria

Participants who indicated smoking 0 cigarettes per day at the baseline period were considered non-smokers and excluded from the study (n=2,253). Those with missing smoking data at baseline or during the first follow-up period were excluded from the analysis (n=298). Lastly, participants with Coronary Heart Disease (CHD) at baseline or the first follow-up period were excluded (n=140), leaving a total of 1,743 to be included in our study.

Exposure assessment

The exposure period was defined as starting from the baseline period (period 1) to the completion of the first follow-up period (period 2), approximately 6 years apart, **Figure 1**. Participants were asked to report the number of daily cigarettes smoked at each study period. Using a cutoff of 20 cigarettes per day, participants were categorized as light (≤20) and heavy smokers (>20) at periods 1 and period 2. We looked at the impact of the following exposure groups on CHD; a) light vs. heavy smoking assessed at period 2, b) change in smoking intensity over periods 1 and 2 (six years), i.e., 1) Light-Light, 2) Light-Heavy, 3) Heavy-Light, and 4) Heavy-Heavy, and c) effect modification by smoking status at period 1 on the association of light vs. heavy smoking at period 2.

Covariates

A direct acyclic graph (DAG) was created to identify potential time-invariant confounders from the following covariates available in the dataset: age, sex, education, and body mass index (BMI), In addition, unmeasured confounders' activity, alcohol consumption, and diet were identified from the literature search, **Figure 2**. Demographic variables were defined as: patient age at period 1 (<55 years, 55 years or more), sex (male, female), and education (no high school, high school, some college, college degree). Clinical characteristics included: BMI at period 1 (Underweight, <18.5; Normal, 18.5–24.9; Overweight, 25–29.9; Obese, 30+), prevalent hypertension at period 2, prevalent diabetes mellitus at period 2 and total cholesterol at the period 1 (normal, <200 mg/dL; borderline high to high, <200 mg/dL).

Follow-up

The follow-up began at the end of period 2, at which the participants were categorized into the four exposure groups, **Figure 1**. The follow-up would end when there is an occurrence of any

coronary heart disease or loss of follow-up. Each participant was followed for 24 years for the outcome assessment.

Outcome assessment

The study outcome was the incidence of any CHD during the follow-up period. The outcome included the following illnesses that comprise CHDs: Angina Pectoris, Myocardial infarction (Hospitalized and silent or unrecognized), Coronary Insufficiency (Unstable Angina), or Fatal Coronary Heart Disease.

Effect modification

In addition to looking at the association of light vs. heavy smoking at period 2 on CHD and the change in intensity of smoking from period 1 to 2, we also assessed whether smoking history as determined by smoking status at period 1 was an effect modifier of association of light vs. heavy smoking at period 2 on CHD. The impact of light vs. heavy smokers on CHD was determined in the overall population at period 2 and then separately for those who were light or heavy smokers at period 1.

Statistical Analyses

For continuous variables, results are presented as median (interquartile range, IQR), overall, and stratified by four smoking exposure groups from periods 1 and 2. For categorical variables, frequencies and column percent are presented. The primary outcome of interest was any incidence of CHD following period 2. The relationship between smoking status at period 2 and CHD was assessed using the log-binomial model in SAS using PROC GENMOD (ink= log, dist = bin), and risk ratios (RR) and 95% confidence intervals (95% CI) were estimated. Model 1 was the unadjusted model. Model 2 was adjusted for categorical variables age, sex, education, and BMI. To assess the potential effect measure modification by smoking status at period 1, separate models were run for light and heavy smokers at period 1.

Using the stratified plug-in g-formula, we estimated the average treatment effect (i.e., marginal causal) on the entire population (ATE). The ATE is the average effect, at the population level, of moving an entire population from untreated to treated. Using the stratified plug-in g-formula, we estimated the risk of CHD had all the individuals were heavy smokers at period 1, and the risk of CHD had all the individuals were light smokers at period 1 while adjusting for all the confounders mentioned above. The causal risk ratio was estimated as $(E[Y^{a=1}=1]) / E[Y^{a=0}=1])$. The plug-in-formula is flexible as it allows modeling the relation between the outcome, exposure, and confounders. The analysis requires the following assumptions: a) exchangeability, b) positivity, c) consistency, and d) correct outcome model specification. We picked stratified plug-in g-formula over pooled as stratifying on exposure gives smaller bias due to less model specification than pooled plug-in g-formula, which gives better precision. Finally, a 95% confidence interval was calculated around the causal risk ratio using non-parametric bootstraps. Statistical computations were performed on SAS 9.4 system (SAS Institute, Cary, NC).

RESULTS

Descriptive Statistics

A total of 1,743 participants with smoking information at periods 1 and 2, with no history of CHD, were included in the analysis. The sample comprises 896 (51.1%) men and 847 (48.6%) women, **Table 1**. The median baseline age was 46 years (IQR: 41-52), with 80% of the sample starting the study when they were less than 55 years old. The median BMI of the sample was 24.6 (IQR: 22.4-27.2), with a majority having a normal weight status (n=917, 52.6%). More than half indicated finishing high school and beyond (n=1,059, 60.8%). Most of the sample was free of hypertension (n=1,329, 76.2%) and not diabetic (n=1,716) at the start of follow-up, i.e., period 2.

However, approximately 77% *n=1,337) of the sample had a total cholesterol level that was borderline high to high.

Regarding smoking behavior, the median daily cigarettes smoked were the same at Period 1 and Period 2 (Median: 20, P1 IQR: 10-20, P2 IQR: 7-25). Most of the sample did not change their smoking behavior from baseline to first follow-up, with 68% (n=1,180) categorized as Light-Light and 15% (n=257) categorized as Heavy-Heavy. Across smoking groups, there were more males in the Heavy-Light group (79%) and the Heavy-Heavy group (78%) compared to females (21% and 22%, respectively). Additionally, most participants categorized as Light-Heavy (86%) and Heavy-Heavy (89%) were less than 55 years old.

Main Outcomes

Table 2a reports the crude, non-casual adjusted, plug-in g-formula RR estimates and their 95% CI for the association of smoking behavior at period 2 and CHD. It compares the risk of CHD among heavy smokers with reference to the risk among light smokers. On crude analysis, heavy smokers were 1.17 (95% CI 0.96-1.42) times more likely to develop CHD than participants who were light smokers. When adjusted for age, sex, education, and BMI, heavy smokers had 1.06 (95% CI 0.87-1.30) times the risk of CHD compared to light smokers. The Plug-in g-formula estimate was close to what we obtained when adjusted for covariates (RR= 1.04; 95% CI 0.84-1.27).

For the change in smoking intensity over periods 1 and 2, results from crude, non-causal adjusted, and plug-in g-formula analysis are presented in **Table 2b**. First, the association of smoking behavior on CHD was determined for participants who changed from light to heavy smoking from period 1 to period 2 with reference to light smokers who did not change their behavior over the exposure period. The crude, non-casual adjusted, and g-formula RR estimates (95% CI) were 1.15 (0.87-1.52), 1.07 (0.81-1.41), and 1.13 (0.81–1.49), respectively. With reference to participants who stayed as heavy smokers from period 1 to period 2, the crude adjusted and g-formula RR estimates (95% CI) for those who changed to light smokers were 1.04 (0.72-1.52), 0.93 (0.63-1.36) and 0.98 (0.65–1.45), respectively.

Effect Modification

To assess whether smoking status at period 1 was an effect modifier of the association of light vs. heavy smoking at period 2 on CHD, crude, non-causal adjusted, and plug-in g-formula risk ratio estimates were calculated. Using the plug-in g-formula analysis results, the overall risk of heavy smoking compared to light smoking at period 2 was 1.04 (95% CI 0.84-1.27), **Table 2a**. When stratified by the effect modifier smoking status at period 1, the risk of CHD for heavy smoking compared to light smoking was higher for those who were light smokers (RR=1.13, 95% CI 0.81–1.49) compared to those who were heavy smokers at period 1 (RR=1.02; 95% CI 0.70 – 1.57), indicating that smoking at baseline period was an effect modifier of association of smoking at period 2 and CHD. In the group of light smokers at period 1, heavy smoking at period 2 was associated with an increase of CHD compared to light smoking, while no such effect was noted in those who were heavy smokers at period 1.

Sensitivity Analysis

Finally, we repeated the analyses presented in Tables 2a and 2b after excluding underweight participants at baseline. This was done to exclude the possibility of any pre-clinical illness which could impact survival. The results were largely similar for most analyses.

CONCLUSION

Overall, these results did not find an adverse effect of heavy smoking on the incidence of coronary artery disease after accounting for potential confounders through a traditional conditional

model and using the causal mechanism technique of plug-in g-formula. When the results were stratified by history of smoking, an adverse effect of heavy vs. light smoking was noted in those with a history of light smoking. No such effect was noted among those who were already heavy smokers. In the context of change in smoking behavior, compared to those who remained light smokers over time, those who changed to heavy smoking reported an increased risk of CHD. No benefit was noted for those who switched from heavy to light smoking compared to those who remained heavy smokers.

There are some limitations to our study. The population of Framingham is predominantly white, so the results of this study may not be generalized to other populations. Also, there was no information on unmeasured confounders such as alcohol consumption, dietary habits, and physical activity. The prospective nature of this study, with follow-up data available for almost 24 years, is a major strength of this study. Future studies should address these limitations and include other sources of confounding. The importance of this study lies in guiding us toward a better understanding of the change in smoking behavior over time and its impact on CHD so that appropriate preventive measures may be taken to reduce the burden of this commonly occurring outcome.

Figure 1: Study design

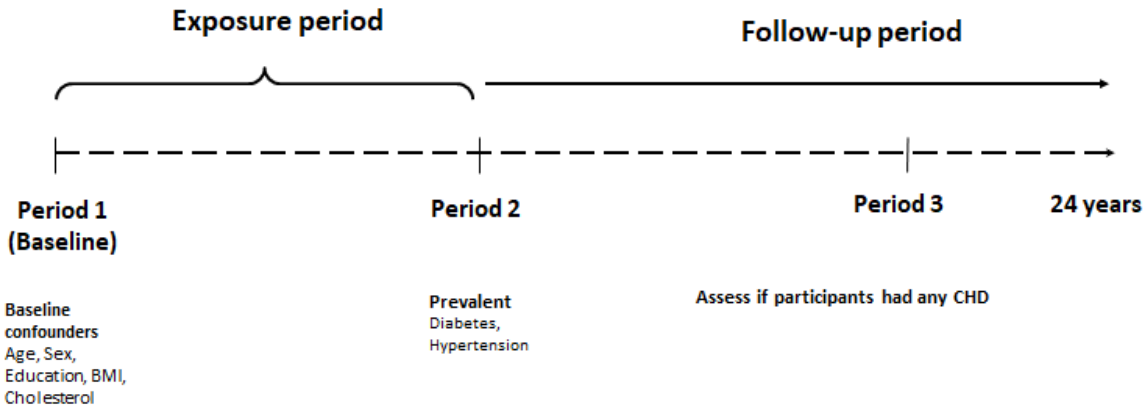

Figure 2.

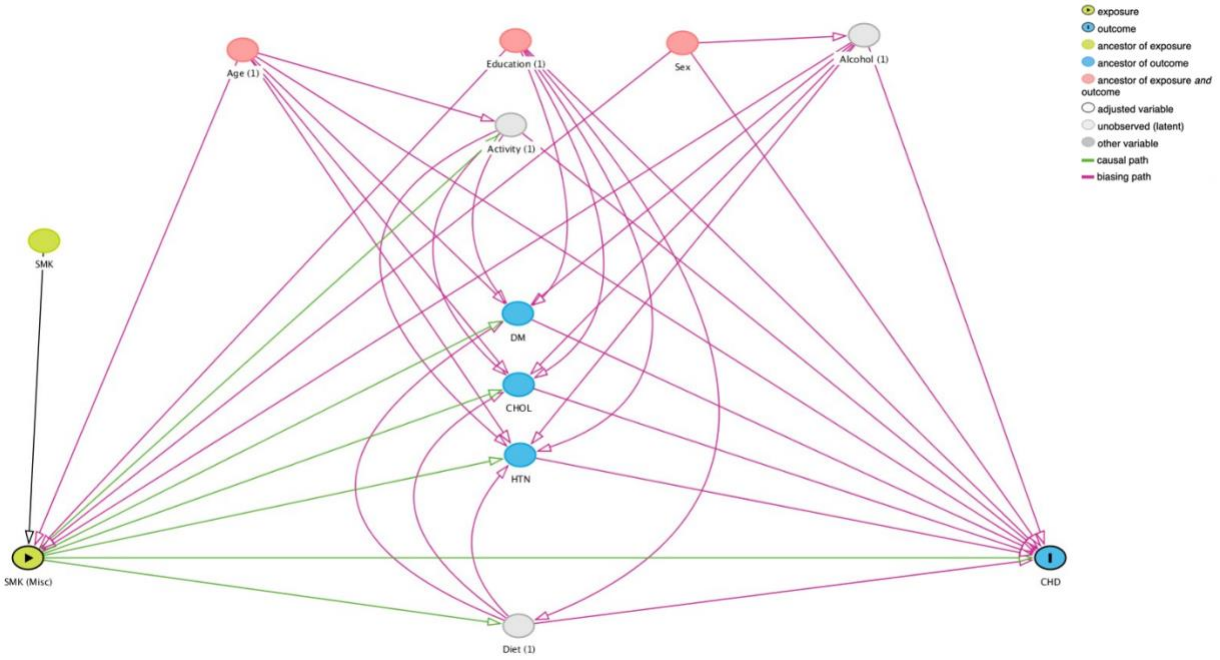
exposure
outcome
ancestor of exposure
ancestor of outcome
ancestor of exposure and outcome
adjusted variable
unobserved (latent)
other variable
causal path
biasing path
Age (1)
Education (1)
Sex
Alcohol (1)
Activity (1)
SMK
DM
CHOL
HTN
SMK (Misc)
CHD
Diet (1)

Table 1. Baseline demographic characteristics by smoking status pairs from Period 1 (P1) and Period 2 (P2) (n=1,743)					
Characteristics	**Total (n=1,743)**	**P1 Light[1] - P2 Light (n=1,180)**	**P1 Light[1] - P2 Heavy[3] (n=196)**	**P1 Heavy[3] - P2 Light[2] (n=110)**	**P1 Heavy[3] - P2 Heavy[3] (n=257)**
	Median (IQR[4])				
Cigarettes smoked per day at P1	20 (10-20)	15 (8-20)	20 (20-20)	30 (30-35)	30 (30-40)
Cigarettes smoked per day at P2	20 (7-25)	13.5 (3-20)	30 (30-35)	20 (0-20)	40 (30-40)
Age, *years*	46 (41-52)	46 (41-53)	44.5 (40-50)	48 (42-54)	44 (40-50)
Body Mass Index (BMI) (n=1,739)	24.6 (22.4-27.2)	24.2 (22.2-26.9)	24.8 (22.7-27.0)	25.8 (23.7-28.6)	25.6 (23.1-27.9)
	n (column %)				
Sex					
Male	896 (51.1)	491 (41.6)	117 (59.7)	87 (79.1)	201 (78.2)
Female	847 (48.6)	689 (58.4)	79 (40.3)	23 (20.9)	56 (21.8)
Age, *years*					
<55	1403 (80.5)	919 (77.9)	169 (86.2)	86 (78.2)	229 (89.1)
≥55	340 (19.5)	261 (22.1)	27 (13.8)	24 (21.8)	28 (10.9)
Education					
No high school	647 (37.1)	438 (37.1)	69 (35.2)	43 (39.1)	97 (37.7)
High School degree	594 (34.1)	401 (34.0)	78 (39.8)	24 (21.8)	91 (35.4)
Some college	259 (14.9)	187 (15.9)	24 (12.2)	19 (17.3)	29 (11.3)
College degree	206 (11.8)	128 (10.9)	24 (12.2)	17 (15.5)	37 (14.4)
Missing	37 (2.1)	26 (2.2)	1 (0.5)	7 (6.4)	3 (1.2)
Body Mass Index (BMI)					
Underweight (<18.5)	33 (1.9)	27 (2.3)	2 (1.0)	0 (0.0)	4 (1.6)
Normal (18.5 – 24.9)	917 (52.6)	664 (56.3)	99 (50.5)	45 (40.9)	109 (42.4)
Overweight (25 – 29.9)	635 (36.4)	397 (33.6)	77 (39.3)	50 (45.5)	111 (43.2)
Obese (30+)	154 (8.8)	88 (7.5)	18 (9.2)	15 (13.6)	33 (12.8)
Missing	4 (0.23)	4 (0.34)	0 (0.0)	0 (0.0)	0 (0.0)
Prevalent Hypertension[5]	414 (23.8)	272 (23.1)	39 (19.9)	32 (29.1)	71 (27.6)
Prevalent Diabetes[6]	27 (1.6)	20 (1.7)	4 (2.0)	2 (1.8)	1 (0.4)
Total cholesterol (mg/dL)[7]					
Borderline high to high (>200)	1337 (76.7)	890 (75.4)	157 (80.1)	85 (77.3)	205 (79.8)
Normal (<200)	381 (21.9)	277 (23.5)	34 (17.4)	23 (20.9)	47 (18.3)
Missing	25 (1.4)	13 (1.1)	5 (2.6)	2 (1.8)	5 (2.0)

[1]P1 Light: 1-20 cigarettes per day
[2]P2 Light: 0-20 cigarettes per day
[3]P1 and P2 Heavy: >20 cigarettes per day
[4]IQR: Interquartile range
[5]Prevalent hypertension: The participant was defined as hypertensive if treated or if Period 2 exam at which mean systolic was >=140 mmHg or mean Diastolic >=90 mmHg
[6]Prevalent diabetes: Diabetic according to criteria of the first exam treated or the first exam with casual glucose of 200 mg/dL or more
[7]Total cholesterol (mg/dL): Serum Total Cholesterol (milligrams/deciliters)

Table 2a. Crude, non-causal adjusted, plug-in g-formula risk ratio (RR) estimates, and their 95% Confidence Intervals (CI) for the association of smoking behavior at Period 2 (P2) on any Coronary Heart Disease (CHD)

	N	Events (%)	Crude estimate (95% CI)	Non-causal adjusted estimate[1] (95% CI)	Plug-in g-formula estimate[1] (95% CI)
P2 Smoking status group					
P2 Heavy	453	111 (24.5)	1.17 (0.96 – 1.42)	1.06 (0.87 – 1.30)	1.04 (0.84 – 1.27)
P2 Light	1,290	270 (20.9)	Ref.	Ref.	Ref.

[1]Adjusting for sex, age, education, and BMI.

Table 2b. Non-causal adjusted, plug-in g-formula risk ratio (RR) estimates, and their 95% Confidence Intervals (CI) for the association of smoking behavior from Period 1 (P1) to Period 2 (P2) on any Coronary Heart Disease (CHD) by smoking status					
	N	Events (%)	Crude estimate (95% CI)	Non-causal adjusted estimate[1] (95% CI)	Plug-in g-formula estimate[1] (95% CI)
	Light Smokers P1				
P1 Light - P2 Heavy	196	46 (23.5)	1.15 (0.87 – 1.52)	1.07 (0.81 – 1.41)	1.13 (0.81 – 1.49)
P1 Light - P2 Light	1,180	241 (20.4)	Ref.	Ref.	Ref.
	Heavy Smokers P1				
P1 Heavy - P2 Light	110	29 (26.4)	1.04 (0.72 – 1.52)	0.93 (0.63 – 1.36)	0.98 (0.65– 1.45)
P1 Heavy - P2 Heavy	257	65 (25.3)	Ref.	Ref.	Ref.
	Heavy Smokers P1 (to assess effect modification)				
P1 Heavy - P2 Heavy	257	65 (25.3)	0.96 (0.66 – 1.40)	1.08 (0.74 -1.57)	1.02 (0.70 – 1.57)
P1 Heavy - P2 Light	110	29 (26.4)	Ref.	Ref.	Ref.

[1]Adjusting for sex, age, education, and BMI.

Table 3a. Crude, non-causal adjusted, plug-in g-formula risk ratio (RR) estimates, and their 95% Confidence Intervals (CI) for the association of smoking behavior at Period 2 (P2) on any Coronary Heart Disease (CHD): **Excluding underweight participants.**

	N	Events (%)	Crude estimate (95% CI)	Non-causal adjusted estimate (95% CI)	Plug-in g-formula estimate (95% CI)
P2 Smoking status group					
P2 Heavy	447	110 (24.6)	1.16 (0.96 – 1.41)	1.07 (0.88 – 1.30)	1.06 (0.84– 1.30)
P2 Light	1,263	268 (21.2)	Ref.	Ref.	Ref.

Table 3b. Crude, Non-causal adjusted, plug-in g-formula risk ratio (RR) estimates, and their 95% Confidence Intervals (CI) for the association of smoking behavior from Period 1 (P1) to Period 2 (P2) on any Coronary Heart Disease (CHD) by smoking status: **Excluding underweight participants.**

	N	Events (%)	Crude estimate (95% CI)	Non-causal adjusted estimate (95% CI)	Plug-in g-formula estimate (95% CI)
	Light Smokers P1				
P1 Light - P2 Heavy	194	45 (23.2)	1.12 (0.85 – 1.48)	1.05 (0.79 – 1.39)	1.13 (0.81– 1.51)
P1 Light - P2 Light	1,153	239 (20.7)	Ref.	Ref.	Ref.
	Heavy Smokers P1				
P1 Heavy - P2 Light	110	29 (26.4)	1.03 (0.70 – 1.49)	0.96 (0.71 – 1.32)	0.98 (0.64– 1.50)
P1 Heavy - P2 Heavy	253	65 (25.7)	Ref.	Ref.	Ref.
	Heavy Smokers P1 (to assess effect modification)				
P1 Heavy - P2 Heavy	253	65 (25.7)	0.97 (0.67 – 1.42)	1.04 (0.76 – 1.41)	1.02 (0.67 – 1.57)
P1 Heavy - P2 Light	110	29 (26.4)	Ref.	Ref.	Ref.