STAT 341 Honors Option: Biomarkers and Mortality

Boston Lee

December 19, 2019

Introduction

The purpose of this project is to discover whether or not there is a relationship between inflammatory biomarkers (CRP and fibrinogen) and cardiovascular death, among the elderly. C-reactive protein is a protein whose blood concentration increases drastically in people with inflammatory conditions. Fibrinogen is another protein, which is also a blood clotting factor. Involvement in clotting means that Fibrinogen levels rise in conditions that cause tissue damage and inflammation. This project aims to discover whether the relationship between inflammatory biomarkers and cardiovascular death exists individually, in combination, or as a function of confounding. This will be discussed later in the **Methods** section.

The data for this project was acquired from the Cardiovascular Health Study, a government-sponsored cohort study, which included adults age 65 and older across 4 sites.

Descriptive Statistics

The two tables below provide summary statistics about the relevant variables from the dataset. The variables from the raw data excluded from the tables were patient ID, site number, time to death, and indication of overall survival.

Table 1: Table of continuous variables included in the study

	Died of CD		Did not die of CD	
	Mean	Standard Dev.	Mean	Standard Dev.
Age	75.669	6.224	72.429	5.340
BMI	142.705	22.786	135.548	21.428
Systolic Blood Pressure (mmHg)	0.966	0.222	1.075	0.164
Anke-Arm Ratio	206.671	37.736	212.273	39.169
Cholesterol (mg/dl)	5.115	7.695	3.436	5.963
C-Reactive Protein (mg/l)	341.890	75.590	320.761	65.701
Fibrinogen (mg/dl)	1500.365	796.860	2477.735	575.991

From the tables above, we can see that the population who died from cardiovascular death had higher levels on average of C-reactive protein, and lower levels of fibrinogen, than the population who did not die of cardiovascular disease.

Furthermore, those who died from cardiovascular disease had approximately half the proportion of estrogen users and a higher average BMI.

Table 2: Table of binary variables included in the study

	Died of CD		Did not die of CD	
	No.	%	No.	%
Sex (Male = 1)	1735	40.613	272	57.749
African-American	643	15.051	53	11.253
Smoking status (Y = 1)	497	11.634	73	15.499
Estrogen Use	312	7.303	13	2.760
Previous Disease	838	19.616	245	52.017
Type-2 Diabetes	616	14.419	130	27.601

Methods

To model the relationships between all of the provided variables, a directed acyclic graph was constructed, as seen in Figure 1.

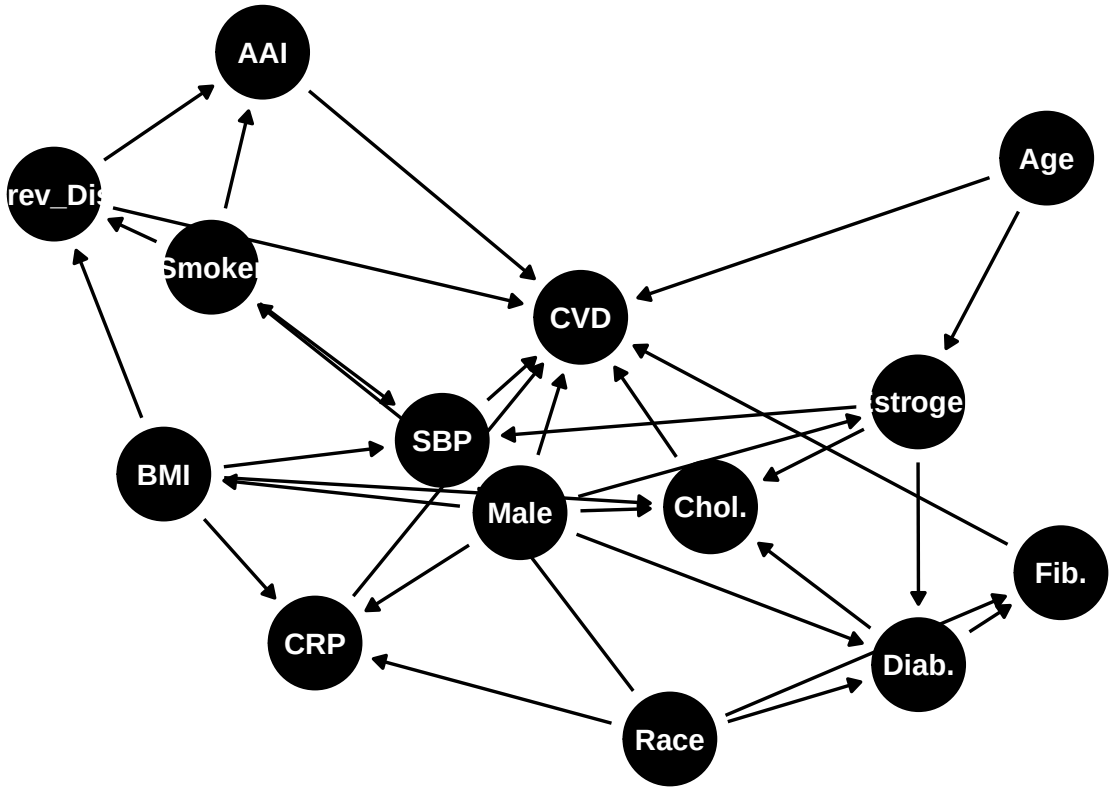

Figure 1: DAG illustrating a scientific scheme of the relationship between the given varaibles

Based on this conceptual scheme of the relationships between variables, six variables were identified as confounding variables, and thus will be adjusted for in all three of the models. Those variables were the following: age, gender, race, use of hormone replacement therapy, type-II diabetes, and BMI.

To test the relationship between the inflammatory biomarkers (CRP and fibrinogen) and the outcome of cardiovascular death, three logistic regression models were created. Two of these models included CRP and Fibrinogen separately, along with confounding variables. The final model included both CRP and Fibrinogen, along with the confounding variables. The coefficients for the biomarkers were be examined for significance

in each of the three models, to see if a significant relationship persisted while the variables were alone, and while they were included in the same model.

To understand the significance of the variables in each model, we should examine the coefficients for that variable in the model. To establish a relationship between both inflammatory biomarkers and cardiovascular death, both biomarkers should be significant alone in a model, and that significance should persist when they are in a model together. That is, one biomarker's significance should not depend on the presence of the other.

Results

The table below contains point estimates and confidence intervals for the exponentiated coefficients of CRP and Fibrinogen. That is, the estimated odds ratio for the relevant coefficients ($exp(\hat{\beta}_i)$). Addressing each value in turn will show what the three models demonstrate about the relationship between the variables.

Table 3: Confidence interval of the expoentiated coefficients for CRP and Fibrinogen in each model

	CRP	Fibrinogen
CRP-only model	1.025 (1.014, 1.038)	–
Fibrinogen-only model	–	1.004 (1.002, 1.005)
Both biomarker model	1.012 (0.997, 1.025)	1.003 (1.002 1.005)

In the model only including C-Reactive protein and the control variables, the estimate of the odds ratio with respect to CRP was 1.02 (95% CI: 1.014, 1.038). This means it was estimated that, on average, a one mg/l difference in blood C-reactive protein was associated with a multiplicative difference of 1.02 in the odds ratio of cardiovascular death.

In the model including only Fibrinogen and the control variables, the estimate for the odds ratio with respect to fibrinogen was 1.003 (95% CI: 1.002, 1.005). This means it was estimated, on average, that a one mg/dl difference in blood fibrinogen was associated with a multiplicative difference of 1.004 in the odds ratio of cardiovascular death.

In the model including both inflammatory biomarkers, the estimate for the odds ratio with respect to CRP was 1.011 (95% CI: 0.997, 1.025). This means it was estimated, on average, that a one mg/l difference in C-reactive protein was associated with a multiplicative difference of 1.011 in the odds ratio of cardiovascular death, when adjusting for fibrinogen and the confounding variables.

In the same model including both inflammatory biomarkers, the estimate for the odds ratio with respect to fibrinogen was 1.003 (95% CI: 1.002, 1.005). This means it was estimated that, on average, a one mg/dl difference in blood fibrinogen was associated with a multiplicative difference of 1.003 in the odds ratio of cardiovascular death, when adjusting for CRP and the confounding variables.

The two estimated coefficients, and corresponding odds ratios, were very similar for both biomarkers between the partial and full models. A key difference, however, was the change in p-value for the coefficient of CRP between the partial and full models. In the model including only CRP an the control variables, the relationship between CRP and cardiovascular death was significant ($p < 0.0001$).

When adjusting for fibrinogen, the relationship did not remain significant ($p = 0.127$). Between the partial and full models, the relationship between fibrinogen and cardiovascular death did remain significant ($p < 0.0001$ for the partial model, and $p = 0.00012$ for the full model).

This change in significance indicates that CRP and fibrinogen are, within a scientific scheme of the variables, related to one another. This relationship could be direct, as one of the inflammatory biomarkers influencing the presence of the other (see Figure 3). This relationship could also be through a mediator, in which one inflammatory biomarker influences a third variable, which in turn influences the other biomarker. There could also be a confounding variable, which influences both inflammatory biomarkers, meaning when both are

included in the model, fibrinogen is more strongly associated with the relationship between the confounder and the outcome (see Figure 2). For these scenarios to be the case, the mediating or confounding variable would be unaccounted for in the group of six controlling variables included in all three of the experimental models.

Discussion

Using the significance of the corresponding coefficients among three logistic regression models, we conclude that there is an association between fibrinogen and cardiovascular death, among individuals aged 65 years and older. Additionally, C-reactive protein has an association with cardiovascular death, when fibrinogen is not adjusted for. This suggests an unaccounted-for relationship between C-reactive protein and fibrinogen. This relationship could take various forms, such as a direct relationship, a mediated relationship, or a common factor affecting both C-reactive protein and fibrinogen. With the design variables included in the model, we can conclude the existence of a relationship between fibrinogen and cardiovascular death, but we cannot make any reliable conclusion about the relationships between C-reactive protein and fibrinogen, or C-reactive protein and cardiovascular death.

For future work, a sensitivity analysis could help determine a more precise relationship between C-reactive protein and the other variables of interest. Furthermore, a different set of data could be collected, with a different set of variables, which could include variables not accounted for in the available data, that could also clarify the relationship between C-reactive protein, fibrinogen, and cardiovascular death.

Appendix

Below are full summaries for each model generated.

Table 4: The summary for the model with both only fibrinogen and the control variables

	Estimate	p-value	95% CI Low	95% CI Low
Intercept	-10.792	0.000	-12.307	-9.293
Fibrinogen	0.004	0.000	0.002	0.005
Age	0.091	0.000	0.075	0.107
Male	0.599	0.000	0.397	0.803
Race	-0.479	0.003	-0.800	-0.176
Estrogen	-0.330	0.268	-0.962	0.216
Type-2 Diabetes	0.801	0.000	0.566	1.031
BMI	0.009	0.458	-0.014	0.031

Table 5: The summary for the model with only CRP and the control variables

	Estimate	p-value	95% CI Low	95% CI Low
Intercept	-9.880	0.000	-11.352	-8.420
C-Reactive Protein	0.026	0.000	0.014	0.038
Age	0.094	0.000	0.077	0.110
Male	0.565	0.000	0.363	0.768
Race	-0.421	0.008	-0.740	-0.121
Estrogen	-0.449	0.131	-1.080	0.096
Type-2 Diabetes	0.781	0.000	0.546	1.013
BMI	0.010	0.371	-0.012	0.033

Table 6: The summary for the model with both inflammatory biomarkers

	Estimate	p-value	95% CI Low	95% CI Low
Intercept	-10.613	0.000	-12.143	-9.097
C-Reactive Protein	0.011	0.127	-0.003	0.025
Fibrinogen	0.003	0.000	0.002	0.005
Age	0.091	0.000	0.075	0.108
Male	0.590	0.000	0.387	0.794
Race	-0.475	0.003	-0.796	-0.172
Estrogen	-0.360	0.228	-0.993	0.188
Type-2 Diabetes	0.787	0.000	0.551	1.019
BMI	0.008	0.502	-0.015	0.030

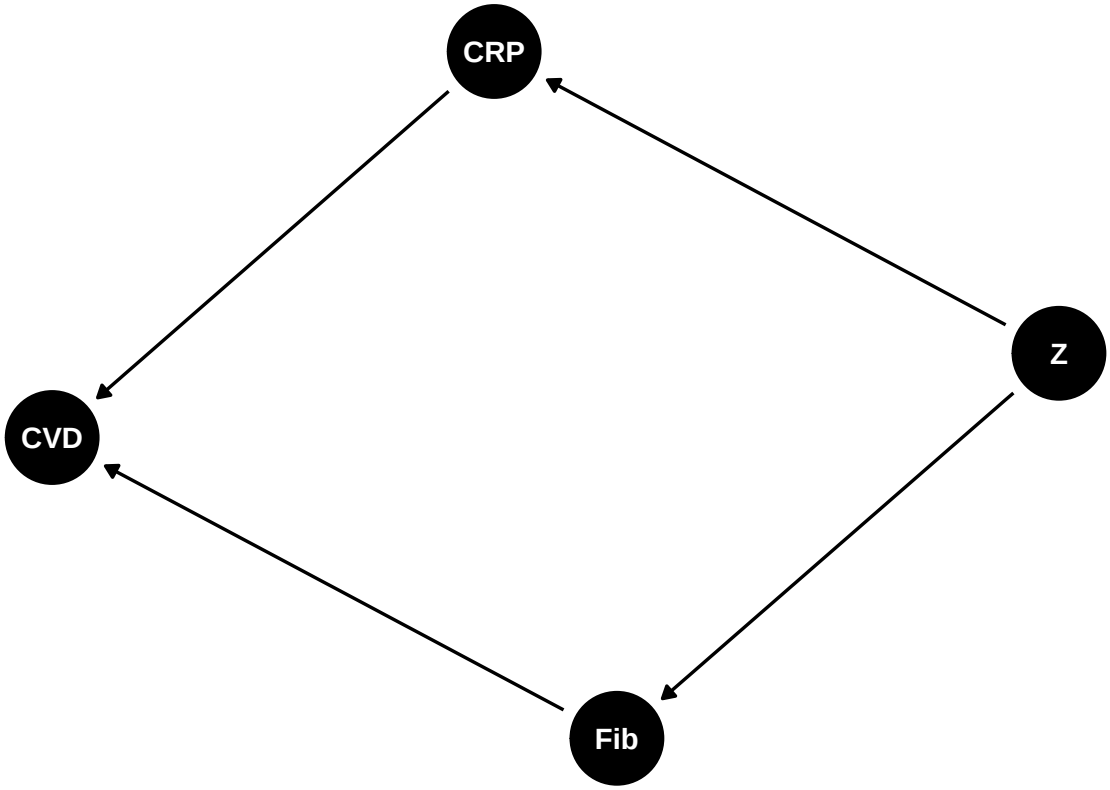

Figure 2: DAG illustrating a potential configuration of variables, with Z as a common influence

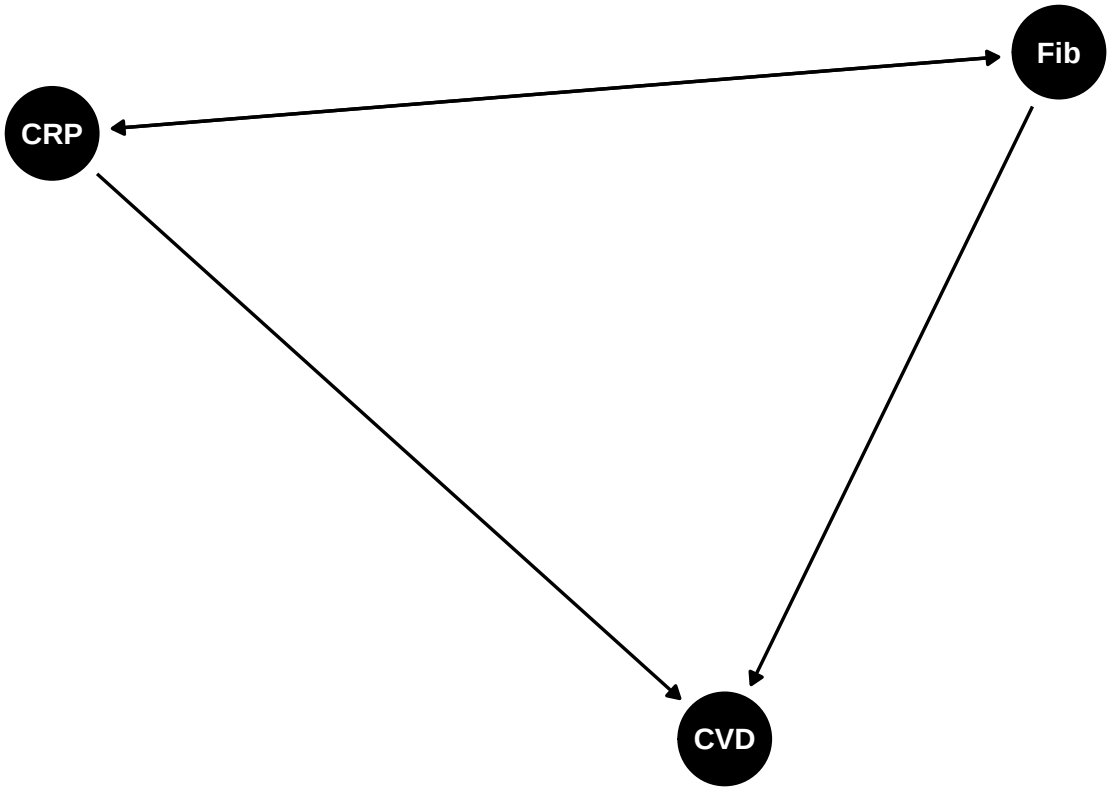

Figure 3: DAG illustrating a potential relationship between CRP and Fibrinogen

Code

The full code used to generate this analysis is shown below:

```
knitr::opts_chunk$set(echo = TRUE)
library(ggplot2)
library(dplyr)
library(tidyverse)
library(magrittr)
library(knitr)
library(ggdag)
library(kableExtra)
library(broom)
library(MASS)
inflamm <- read.delim("inflamm.txt")
inflamm <- na.omit(inflamm)
#Partitioning data for death and no death, to make table
inflamm_death = inflamm[which(inflamm$cvddth == 1),]
inflamm_no_death = inflamm[which(inflamm$cvddth == 0),]
#Table with non-binary variables
non_binary_mean_death <- inflamm_death[c(-1,-2,-10)] %>%
  select_if(function(col) length(unique(col)) > 2) %>%
  summarize_all(mean)
non_binary_sd_death <- inflamm_death[c(-1,-2,-10)] %>%
  select_if(function(col) length(unique(col)) > 2) %>%
  summarize_all(sd)


non_binary_mean_no_death <- inflamm_no_death[c(-1,-2,-10)] %>%
  select_if(function(col) length(unique(col)) > 2) %>%
  summarize_all(mean)
non_binary_sd_no_death <- inflamm_no_death[c(-1,-2,-10)] %>%
  select_if(function(col) length(unique(col)) > 2) %>%
  summarize_all(sd)

table_no_header <- round(cbind(t(rbind(non_binary_mean_death, non_binary_sd_death)),t(rbind(non_binary_m
row.names(table_no_header) <- c("Age", "BMI", "Systolic Blood Pressure (mmHg)", "Anke-Arm Ratio", "Chol
table_no_header %>% kable( caption = "Table of continuous variables included in the study", col.names =
  #pack_rows(start_row = 1, end_row = length(kable),
            #group_label = "Contunious Variables") %>%
  add_header_above(c(" " = 1, "Died of CD" = 2, "Did not die of CD" = 2)) %>%
  kable_styling(position = "center",latex_options = "hold_position")
#Table with binary variables
rem_vec <- c(-17,-18)
#Separate for cvddeath
binary_mean_death <- 100*inflamm_death[rem_vec] %>%
  select_if(function(col) length(unique(col)) <= 2) %>%
  summarize_all(mean)
binary_count_death <- inflamm_death[rem_vec] %>%
  select_if(function(col) length(unique(col)) <= 2) %>%
  summarize_all(sum)

binary_mean_no_death <- 100*inflamm_no_death[rem_vec] %>%
  select_if(function(col) length(unique(col)) <= 2) %>%
```

```
  summarize_all(mean)
binary_count_no_death <- inflamm_no_death[rem_vec] %>%
  select_if(function(col) length(unique(col)) <= 2) %>%
  summarize_all(sum)

raw_binary_table <- cbind(t(rbind(binary_count_no_death,binary_mean_no_death)),
      t(rbind(binary_count_death,binary_mean_death)))
raw_binary_table <- round(raw_binary_table,3)
row.names(raw_binary_table) <- c("Sex (Male = 1)", "African-American", "Smoking status (Y = 1)", "Estrog
raw_binary_table %>% kable(caption = "Table of binary variables included in the study", col.names = c("
    add_header_above(c(" " = 1, "Died of CD" = 2, "Did not die of CD" = 2)) %>%
  kable_styling(position = "center", "striped",latex_options = "hold_position")
full_dag <- dagify("CVD" ~ "Chol.",
                "CVD" ~ "CRP" + "Fib." + "AAI" +
                  "Prev_Dis." + "Age" + "Male" +
                  "SBP",
                "AAI" ~ "Prev_Dis." + "Smoker",
                "BMI" ~ "Male",
                "CRP" ~ "BMI" + "Race" + "Male",
                "Diab." ~ "Male" + "Estrogen" + "Race",
                "Prev_Dis." ~ "BMI" + "Smoker",
                "Chol." ~ "BMI" + "Male" + "Estrogen" +
                  "Diab.",
                "Fib." ~ "Race" + "Diab.",
                "SBP" ~ "BMI" + "Male" +  "Race" +
                  "Estrogen" + "Smoker",
                "Smoker" ~ "Male",
                "Estrogen" ~ "Male" + "Age",
                outcome = "CVD",
                exposure = c("Fib.", "CRP"))
ggdag(full_dag, edge_type = "link_arc") +
  theme_dag_blank()
crp_glm <- glm(cvddth ~ crp + age + male + bkrace + estrogen +
                 diab2 + bmi, data = inflamm, family = binomial)

fib_glm <- glm(cvddth ~ fib + age + male + bkrace + estrogen +
                 diab2 + bmi, data = inflamm, family = binomial)

full_biomarker_glm <- glm(cvddth ~ crp + fib + age + male + bkrace + estrogen +
                            diab2 + bmi, data = inflamm, family = binomial)
#round(exp(confint(crp_glm, "crp")), 3)
#round(exp(confint(full_biomarker_glm, "crp")), 3)
crp_col <- c("1.025 (1.014, 1.038)", "--", "1.012 (0.997, 1.025)")
#round(exp(confint(fib_glm, "fib")), 3)
#round(exp(confint(full_biomarker_glm, "fib")), 3)
fib_col <- c("--", "1.004 (1.002, 1.005)", "1.003 (1.002  1.005)")

raw_conf_table <- cbind(crp_col, fib_col)
row.names(raw_conf_table) <- c("CRP-only model", "Fibrinogen-only model", "Both biomarker model")
raw_conf_table %>% kable(col.names = c("CRP", "Fibrinogen"),
                         align = "c", booktabs = T, caption = "Confidence interval of the expoentiated
  kable_styling(position = "center",latex_options = "hold_position")
exp(summary(crp_glm)$coefficients[2])
```

```r
exp(confint(crp_glm, "crp"))

summary(crp_glm)$coefficients[2,]
exp(summary(fib_glm)$coefficients[2])
summary(fib_glm)
exp(confint(fib_glm, "fib"))
summary(full_biomarker_glm)$coefficients[c(2,3),]
exp(summary(full_biomarker_glm)$coefficients[2])
exp(summary(full_biomarker_glm)$coefficients[3])
fib_glm_table <- round(tidy(fib_glm, conf.int = TRUE)[,c(2,5,6,7)],3)
row.names(fib_glm_table) <- c("Intercept","Fibrinogen", "Age", "Male", "Race", "Estrogen", "Type-2 Diabe
fib_glm_table %>% kable(caption = "The summary for the model with both only fibrinogen and the control v
                        col.names = c("Estimate", "p-value", "95% CI Low", "95% CI Low")) %>%
  kable_styling(position = "center", latex_options = "hold_position")
crp_glm_table <- round(tidy(crp_glm, conf.int = TRUE)[,c(2,5,6,7)],3)
row.names(crp_glm_table) <- c("Intercept", "C-Reactive Protein", "Age", "Male", "Race", "Estrogen", "Typ
crp_glm_table %>% kable(caption = "The summary for the model with only CRP and the control variables", b
                        col.names = c("Estimate", "p-value", "95% CI Low", "95% CI Low")) %>%
  kable_styling(position = "center", "striped",latex_options = "hold_position")
full_glm_table <- round(tidy(full_biomarker_glm, conf.int = TRUE)[,c(2,5,6,7)],3)
row.names(full_glm_table) <- c("Intercept", "C-Reactive Protein","Fibrinogen", "Age", "Male", "Race", "E
full_glm_table %>% kable(caption = "The summary for the model with both inflammatory biomarkers", bookta
                        col.names = c("Estimate", "p-value", "95% CI Low", "95% CI Low")) %>%
  kable_styling(position = "center", "striped", latex_options = "hold_position")
common_factor_dag <- dagify("CVD" ~ "CRP" + "Fib",
                            "CRP" ~ "Z",
                            "Fib" ~ "Z",
                            outcome = "CVD",
                            exposure = "Z")
ggdag(common_factor_dag, edge_type = "link_arc") +
  theme_dag_blank()
mutual_influence_dag <- dagify("CVD" ~ "CRP" + "Fib",
                            "CRP" ~ "Fib",
                            "Fib" ~ "CRP",
                            outcome = "CVD"
                           )
ggdag(mutual_influence_dag, edge_type = "link_arc") +
  theme_dag_blank()
```