Final Project – CKD and CVD

Annika Balk

MA678

12/10/24

Contents

Background

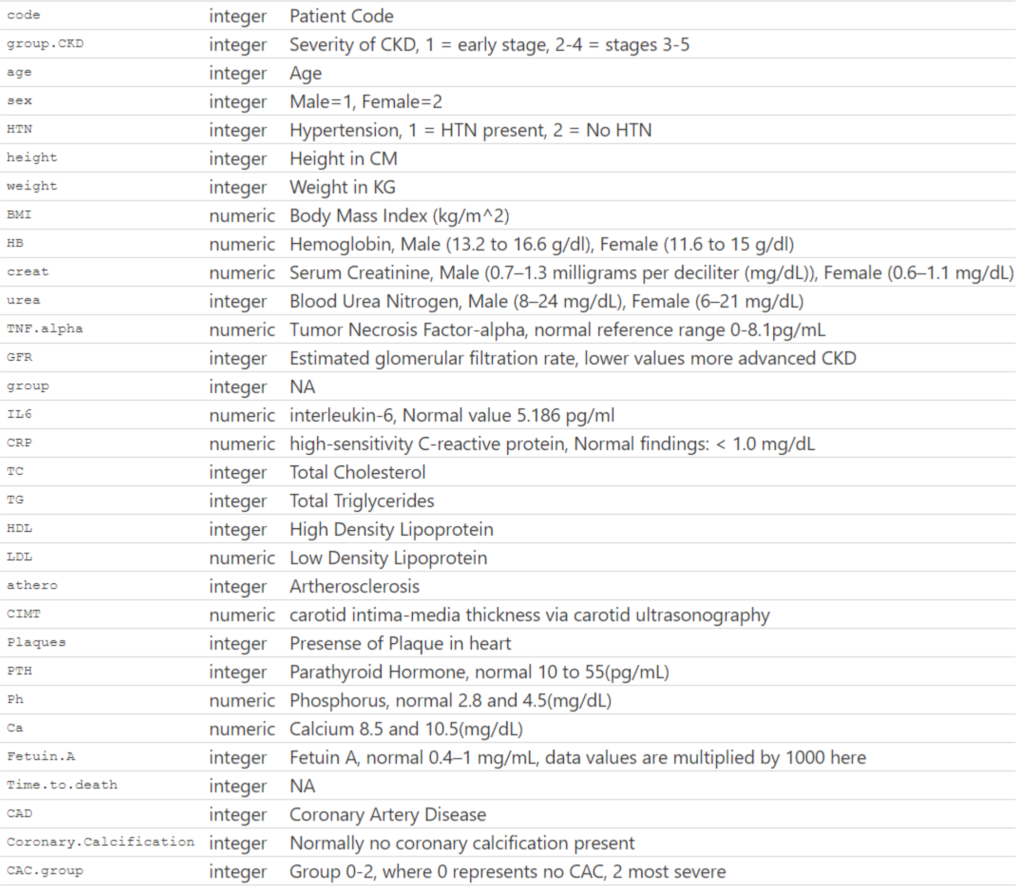
Chronic Kidney Disease (CKD) is a condition that affects multiple organ systems, with diabetes and hypertension being the primary drivers of its development. Diagnosis requires an eGFR below 60 mL/min/1.73 m^2 or signs of kidney damage that persist for at least three months [1]. When eGFR drops below 15 mL/min/1.73 m^2, the disease progresses to end-stage kidney disease (ESKD), where dialysis becomes essential to sustain life. As CKD worsens, the risk of life-threatening complications like cardiovascular disease (CVD) rises significantly, with CVD being the leading cause of early death in CKD patients [10]. Research has also shown that coronary artery calcification (CAC) becomes increasingly common as CKD advances. CAC, characterized by calcium buildup on arterial walls, hardens and thickens the arteries, making it harder for the heart to function and underscoring the strong link between CVD and CKD progression [11].

Objective

This project aims to investigate the relationship between coronary calcification (CAC) and the progression of Chronic Kidney Disease (CKD) to end-stage renal failure. I will use multinomial and hierarchical models to assess the predictive power of CAC across CKD stages, accounting for well-established confounders identified through referenced literature. The dataset for this analysis was sourced from the Harvard Dataverse.

Study Design, Measures, and Participants

The study contained 160 observations of 130 patients who ranged in CKD progression from mild to severe. Exclusions extend to current smokers and individuals with diabetes mellitus, inflammatory conditions, liver diseases, acute kidney failure, chronic hemodialysis, and cancer [2]. Measurements of kidney function and heart disease markers, including plaque buildup, were recorded. To ensure the independence of samples for my analysis, I removed 30 duplicate entries. While it's unclear why duplicate records were included in the initial dataset, exploratory analysis revealed that keeping them would violate model assumptions during the modeling process. A full list of variables is provided below.



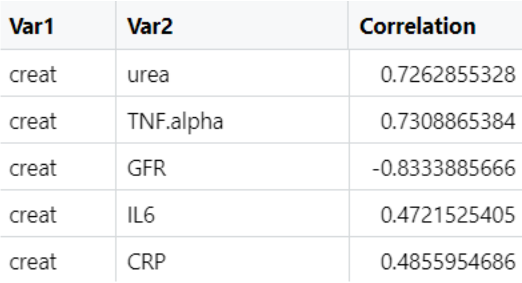
code	integer	Patient Code
group.CKD	integer	Severity of CKD, 1 = early stage, 2-4 = stages 3-5
age	integer	Age
sex	integer	Male=1, Female=2
HTN	integer	Hypertension, 1 = HTN present, 2 = No HTN
height	integer	Height in CM
weight	integer	Weight in KG
BMI	numeric	Body Mass Index (kg/m^2)
HB	numeric	Hemoglobin, Male (13.2 to 16.6 g/dl), Female (11.6 to 15 g/dl)
creat	numeric	Serum Creatinine, Male (0.7–1.3 milligrams per deciliter (mg/dL)), Female (0.6–1.1 mg/dL)
urea	integer	Blood Urea Nitrogen, Male (8–24 mg/dL), Female (6–21 mg/dL)
TNF.alpha	numeric	Tumor Necrosis Factor-alpha, normal reference range 0-8.1pg/mL
GFR	integer	Estimated glomerular filtration rate, lower values more advanced CKD
group	integer	NA
IL6	numeric	interleukin-6, Normal value 5.186 pg/ml
CRP	numeric	high-sensitivity C-reactive protein, Normal findings: < 1.0 mg/dL
TC	integer	Total Cholesterol
TG	integer	Total Triglycerides
HDL	integer	High Density Lipoprotein
LDL	numeric	Low Density Lipoprotein
athero	integer	Artherosclerosis
CIMT	numeric	carotid intima-media thickness via carotid ultrasonography
Plaques	integer	Presense of Plaque in heart
PTH	integer	Parathyroid Hormone, normal 10 to 55(pg/mL)
Ph	numeric	Phosphorus, normal 2.8 and 4.5(mg/dL)
Ca	numeric	Calcium 8.5 and 10.5(mg/dL)
Fetuin.A	integer	Fetuin A, normal 0.4–1 mg/mL, data values are multiplied by 1000 here
Time.to.death	integer	NA
CAD	integer	Coronary Artery Disease
Coronary.Calcification	integer	Normally no coronary calcification present
CAC.group	integer	Group 0-2, where 0 represents no CAC, 2 most severe

Figure 1 List Of Variables In Dataset

Methods

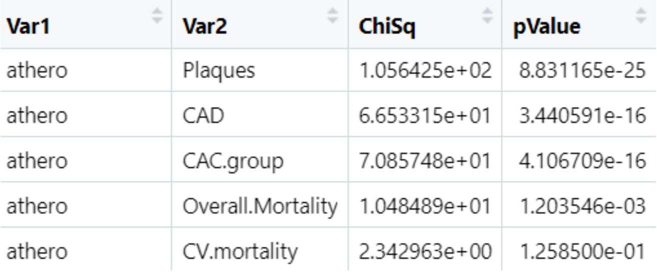
Exploratory Data Analysis

It was imperative to understand relationships between variables within the same organ systems. I had suspicion that variables related to the same organ disease for would be highly collinear and thus making estimates unstable should I include both in future models. To help find these relationships I found all possible combinations of categorical associations and all continuous correlations. Shown below are small chunks of output from both data frames generated in R.



Var1	Var2	Correlation
creat	urea	0.7262855328
creat	TNF.alpha	0.7308865384
creat	GFR	-0.8333885666
creat	IL6	0.4721525405
creat	CRP	0.4855954686

Figure 2 Correlations of Continuous Variables



Var1	Var2	ChiSq	pValue
athero	Plaques	1.056425e+02	8.831165e-25
athero	CAD	6.653315e+01	3.440591e-16
athero	CAC.group	7.085748e+01	4.106709e-16
athero	Overall.Mortality	1.048489e+01	1.203546e-03
athero	CV.mortality	2.342963e+00	1.258500e-01

Figure 3 Chi-Sq Test for Associations in Categorical Variables

Seen in figures 2 and 3, kidney measures such as creatinine and urea are strongly related to each other, likewise categorical variables such as presence of atherosclerosis and the presence of plaques are indeed strongly associated. This then prompted the analysis of graphical representations that show the relation of CKD progression to cardiovascular diseases in the data set. As seen in figures 5 and 6 below there is a natural progression of having a higher CAC_group (higher group number indicates more calcfication) as disease progresses similarly to the presence of coronary artery disease.

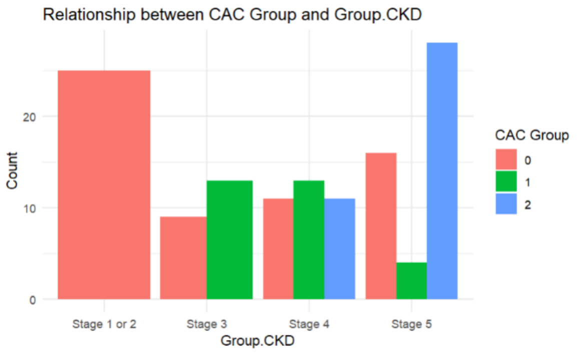


Figure 5 CAC Group by CKD Stage

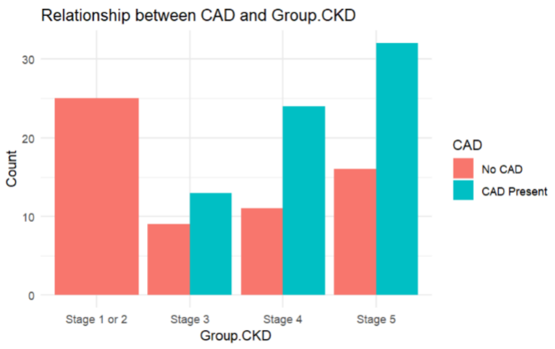


Figure 6 Presence of CAD by CKD Stage

There were several more graphical representations of the data created, reference appendix. Due to the nature of the research question I will be mainly focused on the relationships of CKD and CAC with extra exploration into other cardiovascular diseases and measurements.

Note on Imputation

In the dataset, the first 25 measurements of eGFR were missing, all of which corresponded to patients categorized as having Stage 1 or Stage 2 kidney disease. While I initially explored multiple imputation methods, the imputed values tended to reflect more severe illness due to the influence of lower values in the dataset. To address this logically, I identified the normal eGFR ranges for individuals with Stage 1 and Stage 2 CKD and assigned values by drawing from a random uniform distribution within these ranges. This approach ensured that the imputed values aligned with the expected health status of these patients. These values will later be used to perform hierarchical multilevel modeling.

Modeling

In the early modeling phase, while focused primarily on CAC group as a predictor I was also exploring whether other cardiovascular complications might have stronger predictive power for CKD progression. Using CKD stage as the response variable, I fit five multinomial regression models. The first model was the null model, where the predictive power is based solely on the group proportions for each CKD stage. The remaining models included coronary artery disease (CAD), carotid intima-media thickness (CIMT), and CAC group as predictors. A fifth model was developed specifically to assess whether an interaction term between age and hypertension was justified. Models were evaluated using AIC and reduction is residual deviance.

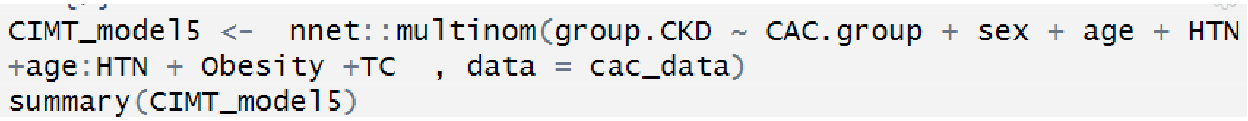
In all models, I controlled for age, sex, hypertension, and obesity and total cholesterol. Women generally exhibit higher rates of CKD but experience slower progression than men before menopause [3]. Hypertension and diabetes are well-established contributors to CKD progression [1]. Although diabetes data was unavailable, obesity was included as a proxy confounder. Additionally, CKD risk increases with age, and there is a known interaction between age and hypertension, which I sought to evaluate in this context. Lastly, Chen et.al considers total cholesterol (TC) an, "atherosclerotic cardiovascular disease risk factor [10]", the correlation between CAC group and TC sits at .47.

The hierarchical models were a natural progression from the multinomial regression models. After confirming that CAC group was a significant predictor, I opted to treat it as a hierarchical level rather than a standard predictor. This decision was inspired by Yun et al., who implemented a similar hierarchical structure for CAC group in their analysis [4]. Using this framework, I fit three models: no pooling, partial pooling, and complete pooling. These models allowed for varying levels of information sharing across CAC groups. Model evaluation included comparing mean squared error (MSE) alongside standard model diagnostics and performance metrics for each approach. I would like to note that I had difficulty fitting a multinomial hierarchical model in R so I chose to fit the model based on eGFR which is the metric directly responsible for CKD classification, in this case the continuous variable was easier to handle.

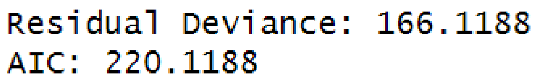
Results

Multinomial Regression

Below is the output for the final model chosen and its predictive power. The null model showed an AIC of 354 and null deviance of 348.



```
CIMT_model5 <-  nnet::multinom(group.CKD ~ CAC.group + sex + age + HTN
+age:HTN + Obesity +TC  , data = cac_data)
summary(CIMT_model5)
```



```
Residual Deviance: 166.1188
AIC: 220.1188
```

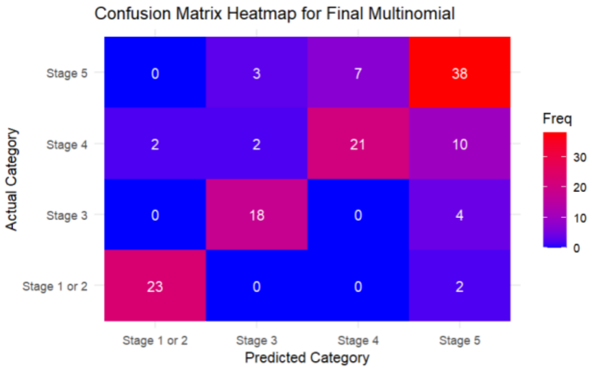
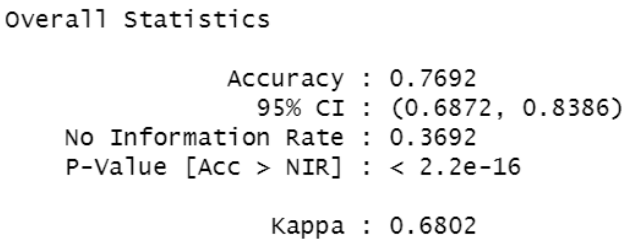


Figure 7 Confusion Matrix for Final Model



```
Overall Statistics

               Accuracy : 0.7692
                 95% CI : (0.6872, 0.8386)
    No Information Rate : 0.3692
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.6802
```

Figure 8 Accuracy Statistics for Final Model

$$Number\ Needed\ for\ 20\%\ Dec\ In\ Deviance = 348 * .8 = 278$$

$$Actual\ Dec\ In\ Deviance = 1 - (166.12/348) \approx 52\%$$

With an accuracy of 76.92% it demonstrates strong relationships between CAC and CKD stage, particularly given the complexity of predicting a four-category response. Additionally, the 52% reduction in residual deviance, combined with the lowest AIC value among the tested models, further shows that CAC group contributes strong predictive value to the progression of CKD when controlling for confounders. This model includes the interaction and had a slightly lower AIC then the model without the interaction so to motivate keeping it I generated a visual representation that would show the distribution with and without the interaction term in the model. If they look the same there seems to be no need to have the interaction. However, in figure 8 it is clearly seen they are different motivating me to keep the interaction in the model.

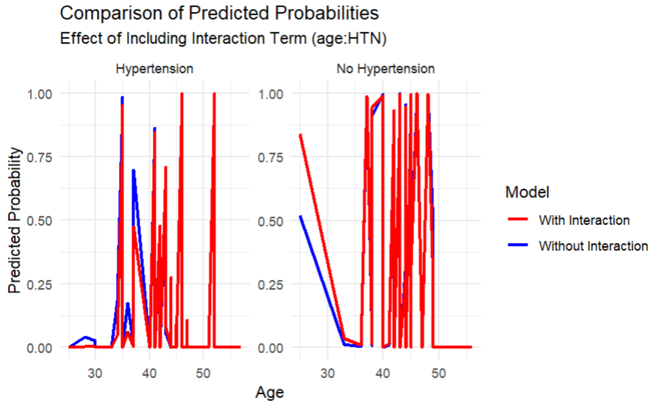
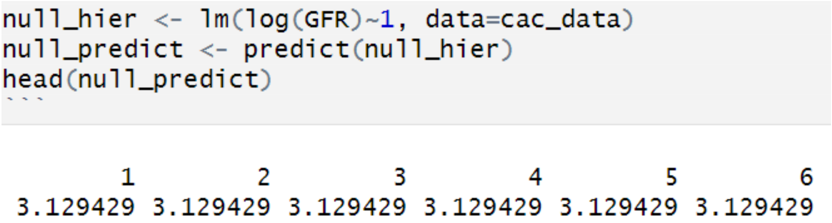


Figure 8 Justification for Interaction Term

Hierarchical Model

Initially I fit the null model which yield simply the group mean for log(GFR) for each observation. The following models in addition to CAC group include confounders sex, age, hypertension, and obesity.



```
null_hier <- lm(log(GFR)~1, data=cac_data)
null_predict <- predict(null_hier)
head(null_predict)
```

```
       1        2        3        4        5        6 
3.129429 3.129429 3.129429 3.129429 3.129429 3.129429 
```

When partial pooling was initially performed on GFR and model assumptions were checked I found strong evidence of non-normality and heteroscedasticity, upon further investigation GFR was right skewed and after performing a log transform the model was refit and assumptions seemed to be well corrected with a slight possibility of heteroscedasticity still. See model violations in appendix. Below you will find the graphical representation of all three pooling methods:

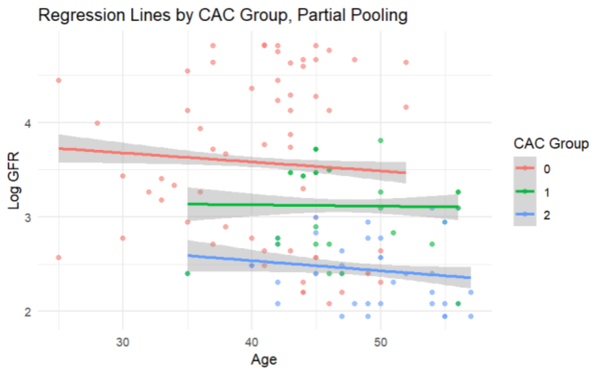


Figure 9 Partial Pooling

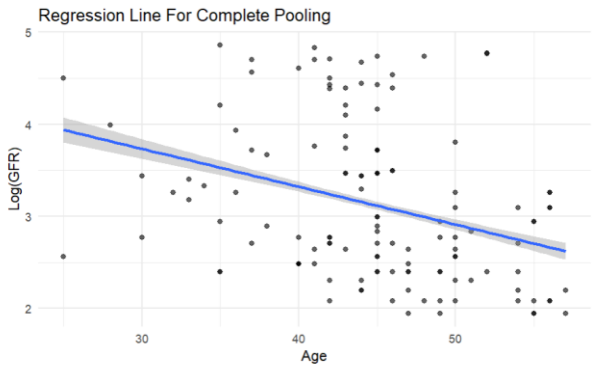


Figure 10 Complete Pooling

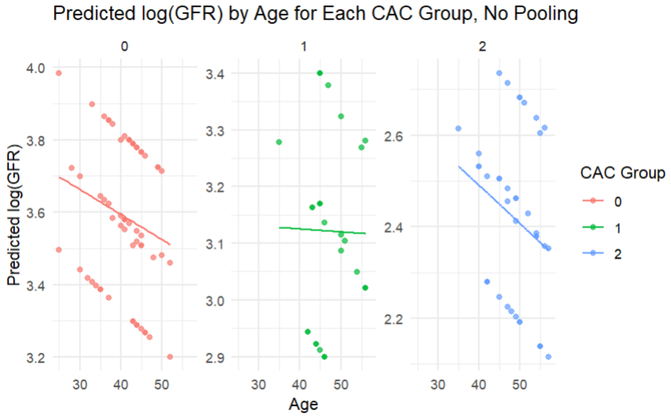
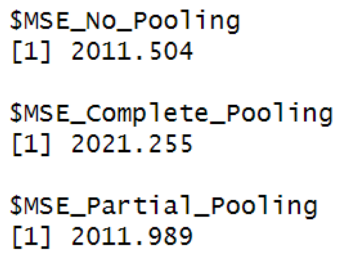


Figure 11 No Pooling

There MSE’s are reported as the following:



```
$MSE_No_Pooling
[1] 2011.504

$MSE_Complete_Pooling
[1] 2021.255

$MSE_Partial_Pooling
[1] 2011.989
```

Although no pooling yields the smallest MSE in theory, the practical decision leans towards choosing the no pooling model. However, I believe treating CAC group as a level in the hierarchical model does not add significant value to this dataset. Its lack of additional insight suggests that the multinomial model is a more effective approach for demonstrating the relationship between CKD progression and coronary calcification.

Conclusion

Discussion

The multinomial model yielded quite compelling results, demonstrating that the presence of coronary artery calcification (CAC) can be linked to the progression of later stages of kidney disease. Although CAC was the primary predictor of interest, the inclusion of other covariates was necessary to adjust for confounding factors. Even with this, the model still highlighted a meaningful association between CAC and CKD progression. On the other hand, while hierarchical models revealed some distinctions across CAC groups, they did not provide substantial insights into the disease progression in CKD patients. The lack of significant additional information from treating CAC group as a hierarchical level suggests that a more straightforward multinomial model is likely more effective in capturing the relationship between CKD and coronary calcification. I also came across model validity issues in the complete pooling case, even after transformation of the response variable. It seems overwhelmingly the multinomial model performed better in this project.

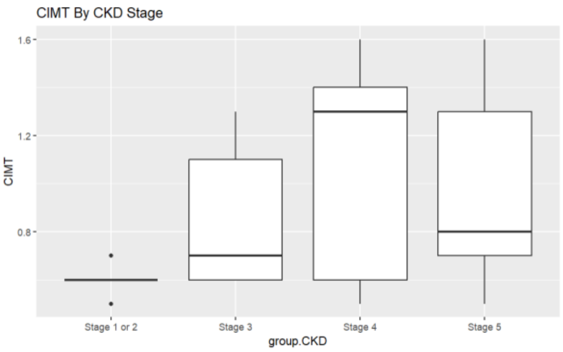
Model Validation

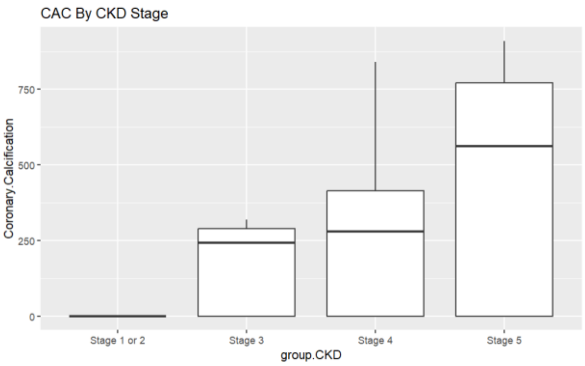
There are several papers that motivated my work on this project. There is research about CAC links to CKD but there seems to be a consensus that more needs to be done to understand the, “temporal association between declining renal function, the appearance and growth of calcification across various cardiovascular beds, the presence of other confounding risk factors, and cardiovascular morbidity [11].” This same paper verifies that a mere 7% of patients on dialysis (ESKD) do not exhibit coronary calcification, this aligns with my findings those who present with CAC have will often times be farther in progression of renal decline [11]. In another study conducted by the journal for cardiology they found, “Of 451 participants who had a CAC score of greater than 100, 109 (24.2%) had an eGFR of less than 30 mL/min/1.73 m2, 158 (35.0%) had an eGFR of 30 to 44 mL/min/1.73 m2, 131 (29.0%) had an eGFR of 45 to 59 mL/min/1.73 m2, and 53 (11.8%) had an eGFR of 60 mL/min/1.73 m2 [10]”. Further justifying that those who present with CAC will have low eGFR and thus later stage kidney disease. It seems that my work coincides with the literature that is currently out there on the associations between CKD and CAC.

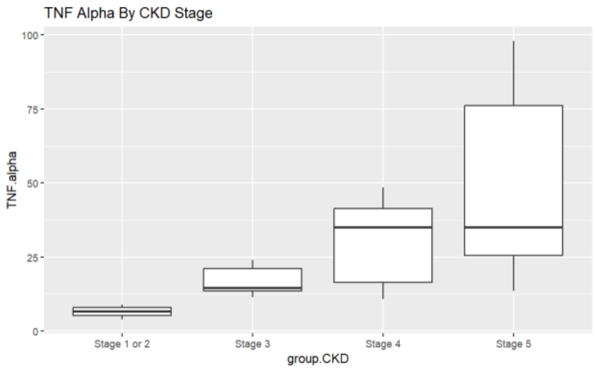
Thank you for taking the time to read my project, I will see you next semester. The appendix has many more figures and model assumptions that I could not fit in the main report.

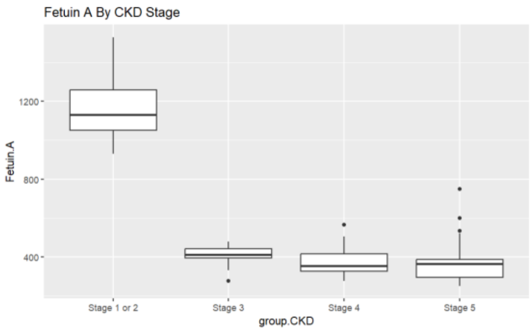
Appendix

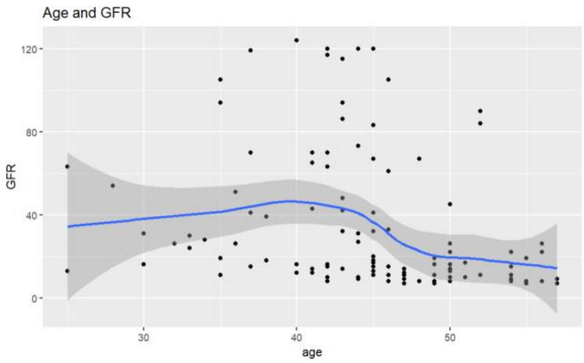
Exploratory Data Analysis Additional Plots:

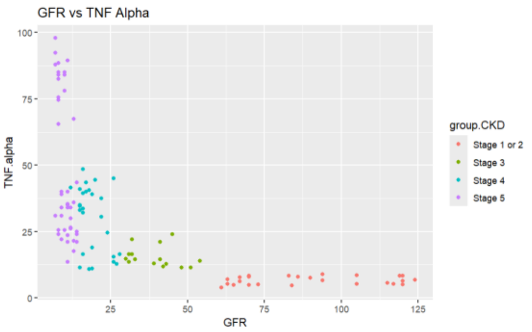




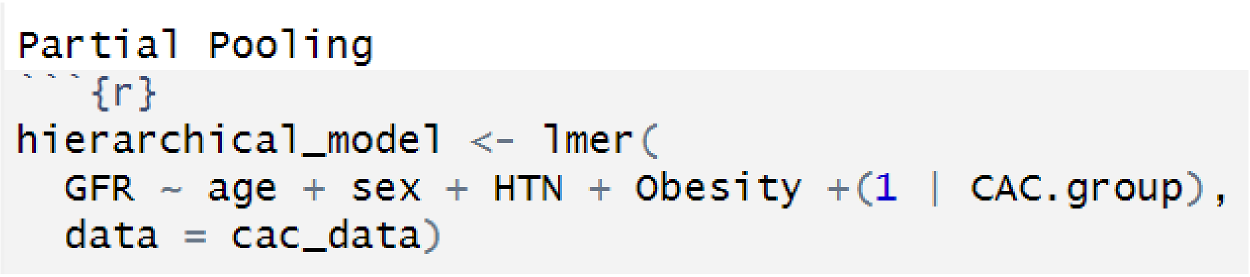




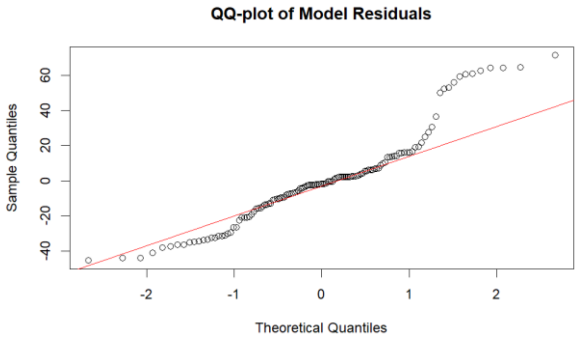


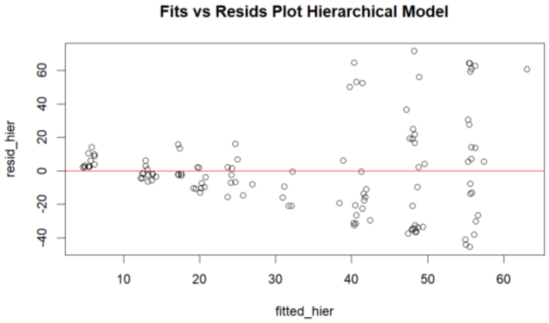


Model Violations for Initial Hierarchical Model:

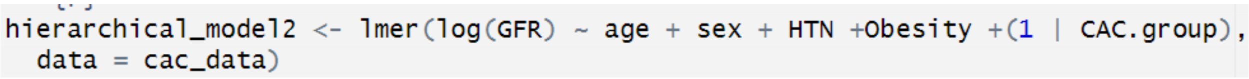


~~~~
Partial Pooling
```{r}
hierarchical_model <- lmer(
  GFR ~ age + sex + HTN + Obesity +(1 | CAC.group),
  data = cac_data)
~~~~

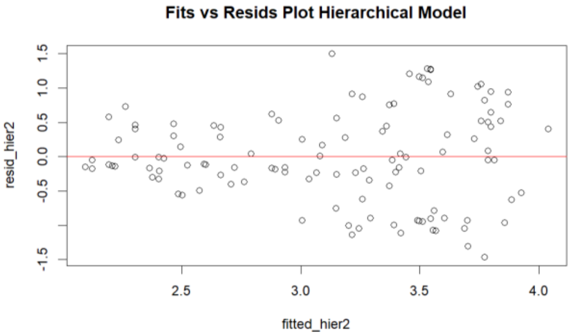


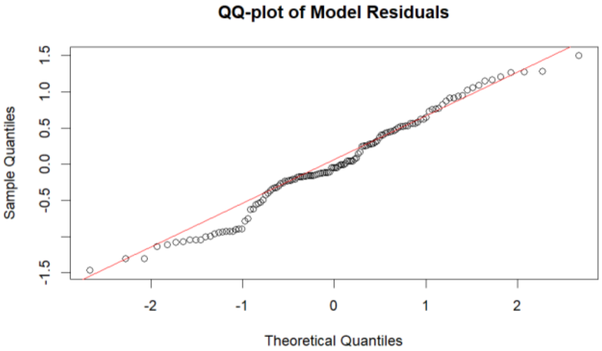


Model Assumption Check after Log(GFR) transform:



```
hierarchical_model2 <- lmer(log(GFR) ~ age + sex + HTN +Obesity +(1 | CAC.group),
  data = cac_data)
```





References

1. Webster AC, Nagler EV, Morton RL, Masson P. Chronic kidney disease. *Lancet*. 2017;389(10075):1238–1252. doi: 10.1016/S0140-6736(16)32064-5.

2. Mohamed ON, Mohamed MRM, Hassan IG, et al. The relationship of fetuin-A with coronary calcification, carotid atherosclerosis, and mortality risk in non-dialysis chronic kidney disease. *J Lipid Atheroscler*. 2024;13(2):194–211. doi: 10.12997/jla.2024.13.2.194.

3. Ahmed SB, Dumanski SM. Do sex and gender matter in kidney and cardiovascular disease? *American Journal of Kidney Diseases*. 2021;78(2):177–179. https://doi.org/10.1053/j.ajkd.2021.05.002. doi: 10.1053/j.ajkd.2021.05.002.

4. Yun H, Joo YS, Kim HW, et al. Coronary artery calcification score and the progression of chronic kidney disease. *J Am Soc Nephrol*. 2022;33(8):1590–1601. doi: 10.1681/ASN.2022010080.

5. Nehring SM GA, Patel BC. C reactive protein. https://www.ncbi.nlm.nih.gov/books/NBK441843/. Updated 2023. Accessed 11/12, 2024.

6. Said EA, Al-Reesi I, Al-Shizawi N, et al. Defining IL-6 levels in healthy individuals: A meta-analysis. *J Med Virol*. 2021;93(6):3915–3924. doi: 10.1002/jmv.26654.

7. Li G, Wu W, Zhang X, et al. Serum levels of tumor necrosis factor alpha in patients with IgA nephropathy are closely associated with disease severity. *BMC Nephrology*. 2018;19(1):326. https://doi.org/10.1186/s12882-018-1069-0. doi: 10.1186/s12882-018-1069-0.

8. Lousa I, Reis F, Santos-Silva A, Belo L. The signaling pathway of TNF receptors: Linking animal models of renal disease to human CKD. *Int J Mol Sci*. 2022;23(6):3284. doi: 10.3390/ijms23063284. doi: 10.3390/ijms23063284.

9. Moe SM, Chen NX. Pathophysiology of vascular calcification in chronic kidney disease. *Circ Res*. 2004;95(6):560–567. https://doi.org/10.1161/01.RES.0000141775.67189.98. doi: 10.1161/01.RES.0000141775.67189.98.

10. Chen J, Budoff MJ, Reilly MP, et al. Coronary artery calcification and risk of cardiovascular disease and death among patients with chronic kidney disease. *JAMA Cardiol*. 2017;2(6):635–643. https://doi.org/10.1001/jamacardio.2017.0363. Accessed 11/8/2024. doi: 10.1001/jamacardio.2017.0363.

11. Hutcheson JD, Goettsch C. Cardiovascular calcification heterogeneity in chronic kidney disease. *Circ Res*. 2023;132(8):993–1012. https://doi.org/10.1161/CIRCRESAHA.123.321760. doi: 10.1161/CIRCRESAHA.123.321760.