Association between eGFR and Different Methods of Access Ports

Bobby Shields, Samuel Jones Armoo, Augustine Ennin

April 22, 2025

Honor Statement

We certify here that the work on this exam is solely ours. We did not receive any assistance from others and we did not provide any assistance to others.

Bobby Shields 04/22/2025
Samuel Jones Armoo 04/22/2025
Augustine Ennin 04/22/2025

Extra Credit

Bobby Shields (add to final)

- + 3 from Homework 1 numerical problem.
- Course Evaluation

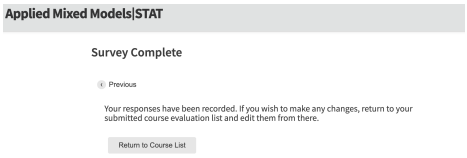

Figure 1: Proof of course evaluation for Bobby Shields.

Samuel Jones Armoo (add to final)

- +3 from Homework 1 numerical problem
- Course Evaluation:

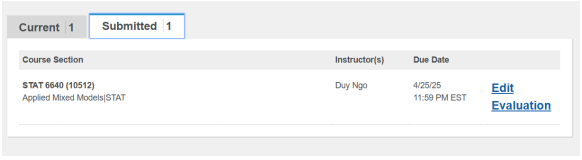

Figure 2: Course evaluation for Samuel Jones Armoo

Augustine Ennin (add to midterm)

- +3 from Homework 1 numerical problem
- Course Evaluation:

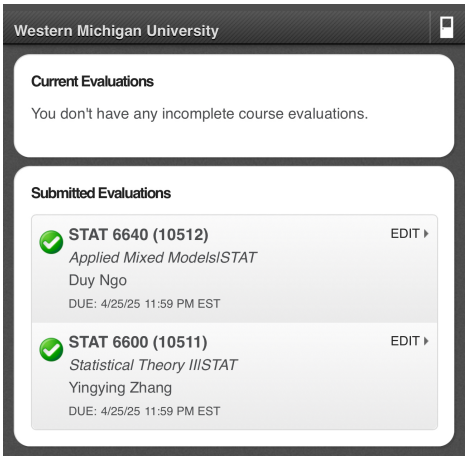

Figure 3: Course evaluation for Augustine Ennin

1 Abstract

The main goal of this study was to see if new methods of access ports for hemodialysis patients would improve their eGFR. Further, it was of interest to see if this association differed based on an individual's race. It was concluded that the venous fistula access port was significant in increasing an individual's eGFR when compared to the traditional graft port, and the standard fistula decreased eGFR when compared to the graft port. Further, it was concluded that race did not play a significant role in affecting the association between access port type and eGFR.

2 Introduction

The last phase of chronic kidney disease (CKD) known as end-stage renal disease (ESRD) defines the condition where the kidneys lose their ability to function sufficiently to require renal replacement therapy including dialysis or transplantation. Over 800,000 individuals with end-stage renal disease (ESRD) exist in the United States and more than 550,000 of these patients need dialysis treatment for survival (United States Renal Data System (USRDS), 2023) [1]. Hemodialysis stands as the primary dialysis method that needs dependable vascular access to perform blood exchange for waste extraction and fluid balance. The clinical outcomes of patients depend heavily on the selected vascular access type among AVF and AVG and CVC because these access types determine the chances of infection and thrombosis alongside hospitalization and mortality rates (National Kidney Foundation, 2020; Lee et al., 2021) [2] [3].

Current advances in vascular access technology and protocols still face the widespread clinical issue of access failure. The use of AVGs for urgent dialysis initiation results in the highest infection rates together with the highest rates of access failure (Allon & Robbin, 2002) [4]. Patients with vascular comorbidities face challenges when developing adequate AVFs because these access points require time to mature before use. The additional barriers that come from racial and demographic differences make access outcomes even more complicated. The racial disparities in vascular access utilization become evident through research which demonstrates African American patients receiving fewer AVFs while starting dialysis with AVGs compared to Caucasian patients (Patzer et al., 2014; Kazakova et al., 2022) [5] [6].

Clinical practice depends on the eGFR measurements for both disease progression evaluation and treatment effectiveness assessment of renal function. The widespread use of eGFR for kidney function assessment does not extend to understanding the effects of different vascular access types on both immediate and ongoing eGFR results among various patient demographics. The interpretation of kidney function becomes challenging due to access complications which require revisions because these revisions cause temporary disruptions in hemodynamic stability and eGFR accuracy (Moist et al., 2008) [7].

A total of 1,255 hemodialysis patients participated in this research which took place across multiple clinics throughout the United States. The research aims to address three essential questions regarding the relationship between access placement types and early eGFR results as well as eGFR changes over time by race and access type. The study explores methods for adjusting eGFR trends analysis to account for possible access complications that include revisions in the data. The research findings will help healthcare providers choose better vascular access methods and create individualized dialysis plans to enhance treatment results for ESRD patients.

3 Statistical Methods

First, it was determined that there were no missing data and that the data was balanced. Further, the variable `acctype` was modified to reflect the questions of interest. The baseline was changed to the graft port as this is the most widely used method. This then gives a comparison between the two fistula ports against the traditional graft port. This is more interesting since the two fistula ports are considered more new or complicated methods.

3.1 Quantifying for eGFR at 7 days

The first question of interest is to quantify the association between the eGFR after 7 days of placement and the access port type. Since only one point in time in the study is used, there are no repeated measurements that have to be considered. Thus, multiple linear regression will be the most useful model. Further, it was noted that eGFR at time 7 is skewed, so a log transformation was used to model the median and account for the skewness. An interaction between the access port and the race of the individual was also utilized to help see if the association between eGFR and the access port differed by race. To help validate the model, AIC and BIC were used to select the best overall model as lower values indicate better fitting models. Further, diagnostic plots were used to check model assumptions.

3.2 Quantifying Association for all Repeated Measures

The next phase was to quantify the rate of change in eGFR for each access type using all repeated measurements. Further, the within subject rate of change needed to be modeled. Thus, a linear mixed effects model was the most useful approach in this scenario. Further, from exploratory data analysis, there seemed to be a difference in the rate of change after day 20 of the study. Thus, a piecewise component was added with a knot at time 20 to account for these differing rates of change. ANOVA was then used to help select which random effects should be included in the model.

4 Results

4.1 EDA

Before answering the researcher questions, it would be useful to explore the data to see what characteristics can be identified from this. The table below gives the mean, median, and range of some of the important covariates in the data:

Summary Statistics			
Variable	Mean	Median	Range [Min, Max]
`age_ssd`	62.43	66	[19,95]
`bmi`	25.95	24.42	[9.26,138.78]
`eGFR7`	1.9	1.3	[0.3,11.4]
`eGFR20`	9.854	9.6	[9,14.4]
`eGFR90`	15.89	15.40	[15,32.7]
`eGFR150`	20	18.2	[15,39.9]

Table 1: Summary statistics of some variables from data.

From this, it seems that some of the variables might be skewed such as: eGFR7, eGFR20, eGFR90, eGFR150, and age. This could cause issues that may need to be addressed especially when answering the research questions. Further, plots were created to see how the average eGFR behaves throughout the time of the study with respect to certain covariates. Some of the plots are given below:

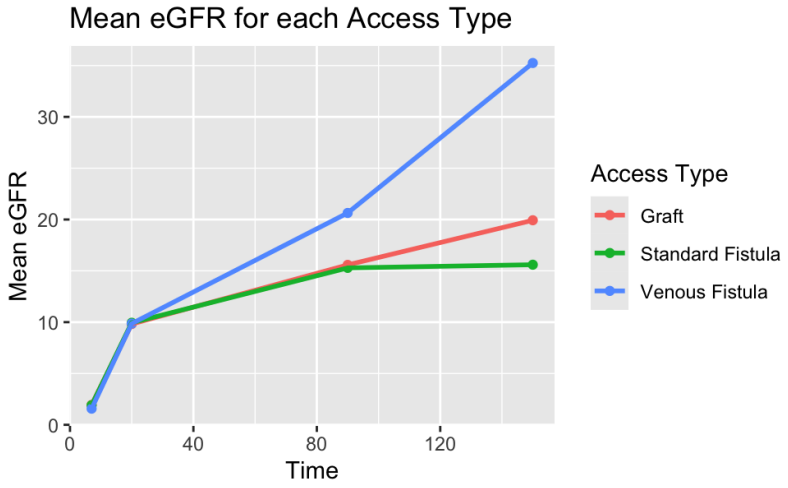

Figure 4: Plot of the average eGFR over the time with respect to the access port type.

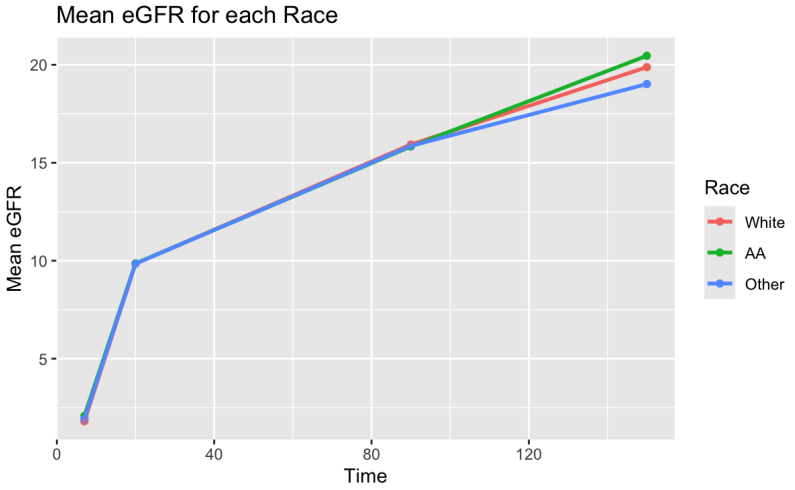

Figure 5: Plot of the average eGFR over the time with respect to the individual's race.

Figure 4 seems to indicate that the access port type may play a significant role in determining the average eGFR. More specifically, the venous fistula might increase the average eGFR while the standard might decrease the average eGFR when compared to the traditional graft port. Further, from Figure 5, the race of the individual will most likely not play a significant role in determining the average eGFR, though there is some more variation between races at the later end of the study. Thus, race should still most likely be considered. Other plots are included in appendix. These plots illustrate that other covariates did not seem to play an important role in characterizing the average eGFR.

4.2 Association of eGFR and Access Type at First Measurement

For answering the first research question, it is reasonable to use a multiple linear regression model to explain the relationship for eGFR at time 7 with the access type and other covariates. Again, the median will be modeled to account for the skewness of the eGFR at time 7. The next step was to consider what covariates to include. To answer the research question, access type and the interaction between access type and race will be included. Further, from preliminary research, age will also be

included. To determine if other covariates should be included, the BIC was used to determine the best model. Variables were removed from the beginning model based on plots of covariates over the time of the study (found in appendix). If there was little difference between categories over the time, these covariates were then removed first. The following table gives these values:

BIC Values for Different Linear Regression Models		
Model Number	Covariates	BIC
Fit 1	acctype , age_ssd, acctype:racegrp, smokegrp, bmi , dx_diab, female	2617.425
Fit 2	acctype , age_ssd, acctype:racegrp, smokegrp, bmi , female	2612.758
Fit 3	acctype , age_ssd, acctype:racegrp, bmi, female	2599.584
Fit 4	acctype , age_ssd, acctype:racegrp, bmi	2594.434
Fit 5	acctype , age_ssd, acctype:racegrp	2588.871 $\star$

Table 2: BIC values for different linear regression models with covariates included. Note: $\star$ indicates model with lowest BIC.

Thus, from this table, model 5 will be selected which only includes the covariates access type, age, and the interaction between access type and race. The linear equation can be written out as the following:

$$E[\log(eGFR7_i)] = \beta_0 + \beta_1 acctype1_i + \beta_2 acctype3_i + \beta_3 acctype2 : racegrp2_i + \beta_4 acctype1 : racegrp2_i + \beta_5 acctype3 : racegrp2_i + \beta_6 acctype2 : racegrp3_i + \beta_7 acctype1 : racegrp3_i + \beta_8 acctype3 : racegrp3_i + \beta_9 age\_ssd_i$$

The following table then gives the estimates from the linear regression model:

Estimates from Least Squares Model			
Variable	**Estimate**	**Standard Error**	**95% Confidence Interval**
`(Intercept)`	2.4100	0.08930	(2.230 2.5800) $\star$
`acctype1`	-0.1550	0.05510	(-0.263, -0.0465) $\star$
`acctype3`	-0.2030	0.08440	(-0.368, -0.0372) $\star$
`age_ssd`	-0.0325	0.00127	(-0.0350, -0.0300) $\star$
`acctype2:racegrp2`	-0.0564	0.04920	(-0.153 , 0.0402)
`acctype1:racegrp2`	-0.1460	0.08790	(-0.319 , 0.0261)
`acctype3:racegrp2`	0.0813	0.16400	(-0.2410 , 0.4040)
`acctype2:racegrp3`	0.0665	0.09560	(-0.1210 , 0.2540)
`acctype1:racegrp3`	-0.0495	0.12900	(-0.3020 , 0.2030)
`acctype3:racegrp3`	0.3850	0.30600	(-0.2160 , 0.9850)

Table 3: Variable estimates from least squares result in R. Note: $\star$ indicates significant variables.

Thus, the significant variables are age and both access types. This corroborates with the preliminary research that age negatively effects eGFR. Further, it is expected for both access types to have significant effects on eGFR as indicated from Figure 4. However, it would be expected for access

type 3 to have a positive effect when compared to the graft port, but this is most likely due to only using one repeated measure and that all port types have eGFR values close at the beginning of the study. It is also seen from Table 3 that none of the interaction terms are significant. This could indicate that the association between the access port and eGFR does not differ by race.

4.3 Association for all Repeated Measures

For analyzing the relationship between all repeated measurements of eGFR and access type, a linear mixed effect model seemed to be most useful. The fixed effects will be the same as the covariates used in section 4.2, with the addition of a time variable and the variable timePlus for the piecewise component. The next step was to see what random effects might be used in the model. The following figure helps illustrate what these effects might be.

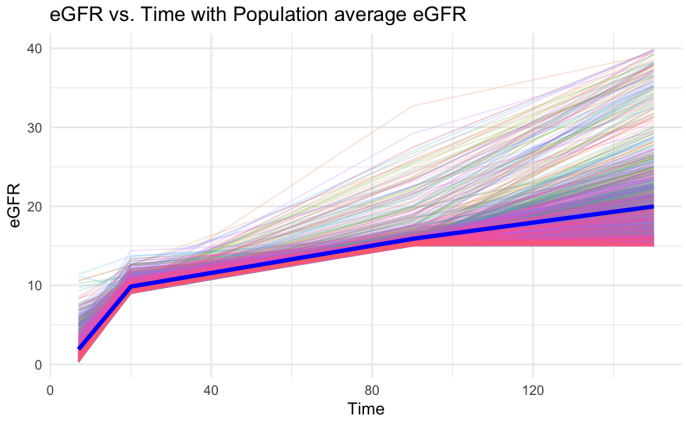

Figure 6: Plot of all individual eGFR over time. The average eGFR over the time is also shown (solid blue line).

This plot helps indicate that there should be an inclusion of a random intercept. Then ANOVA was used to see if a random effect on the variable time should be included. The procedure gave us the following results:

1. H_0 : Reduce Model (intercept only) vs. H_1 : Full model (intercept and time)
2. Likelihood Ratio: 1076.634
3. p-value: $< .0001$
4. We reject H_0 and conclude to use the full model with both intercept and time having random effects.

Thus, the full model now includes the following:

- Response: `eGFR`
- Covariates: `age_ssd, acctype, Time, timePlus, acctype:racegrp`
- Random Effects: `(Intercept), Time`

The following table gives the estimates of model.

Estimates from Mixed Effects Model			
Fixed Effects			
Variable	**Lower CI Bound**	**Estimate**	**Upper CI Bound**
`(Intercept)` ⋆	1.73165291	2.15198216	2.57231140
`acctype1` ⋆	-1.08733748	-0.85873163	-0.63012578
`acctype3` ⋆	3.01327201	3.36362535	3.71397870
`age_ssd` ⋆	-0.04589583	-0.04062187	-0.03534791
`Time` ⋆	0.42431082	0.43837005	0.45242929
`timePlus` ⋆	-0.31012032	-0.29751224	-0.28490416
`acctype2:racegrp2`	-0.18377533	0.02054837	0.22487207
`acctype1:racegrp2`	-0.60286113	-0.23786729	0.12712654
`acctype3:racegrp2`	-0.56432985	0.11826072	0.80085130
`acctype2:racegrp3`	-0.55391562	-0.15722852	0.23945858
`acctype1:racegrp3`	-0.62135125	-0.08616981	0.44901163
`acctype3:racegrp3`	-0.25609404	1.01356488	2.28322380
Random Effects			
Variable	**Lower CI Bound**	**Estimate**	**Upper CI Bound**
`sd((Intercept))`	1.25050324	1.38820931	1.54107964
`sd(Time)`	0.02932241	0.03111283	0.03301257
`cor((Intercept),Time)`	-0.98162515	-0.97189370	-0.957119935

Table 4: Estimate and confidence bounds of mixed effects model with random effect standard deviation estimates. Note: ⋆ indicates significant variables.

The estimates now show that the access types are significant including the time, the knot at time 20, and the age. Further, these results align with the plots from the exploratory data analysis and preliminary research. Further, all of the interaction terms are not significant. A plot of the predicted values can be given to see how well the random effects are fitting the within subject rates of change. This plot is given below:

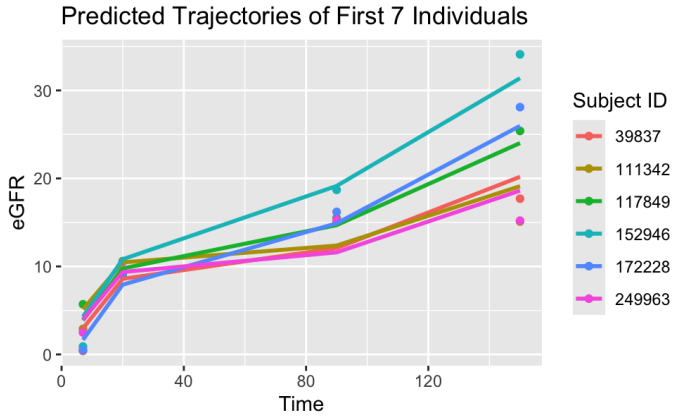

Figure 7: Plot of predicted values (lines) compared to observed values (points) for the first seven subjects.

The plot indicates that the model seems to do fairly well at predicting towards the beginning of the study, but starts to vary from the observed values at the later end of it. However, overall trends seem to be performing fairly well.

5 Discussion

5.1 eGFR and Access Placement at Time 7

The main question at hand was to quantify the association between access type and eGFR at time 7. From Table 3, we have the following conclusions of the significant terms:

- For individuals of the same age and race, having a standard fistula port decreases median eGFR by 0.155 when compared to the graft port.
- For individuals of the same age and race, receiving a venous fistula port decreases median eGFR by 0.203 when compared to the graft port.
- The difference in median eGFR between two individuals who receive the same access port and are of the same race is -0.0325.
- All interaction terms between access port and race are not significant. Thus, the association does not differ by race.

The last point corroborates with more recent research as it was once believed that race played an important factor on eGFR, but is no longer viewed to do so [8]. Model diagnostic found plots in the appendix also indicate that the model satisfies linear regression assumptions, verifying the significance of the findings.

5.2 eGFR and Access Placement for all Repeated Measurements

The main question of interest here was to quantify the association between eGFR and access type using all longitudinal data. Further, modeling the within subject rate of change and seeing if the association differed by race was also of interest. From analysis, there are several conclusions that can be made:

- Receiving the standard fistula port seems to decrease average eGFR when compared to the graft port. While the venous fistula port increased average eGFR when compared to the graft port.
- Age again seems to decrease average eGFR. This again aligns with preliminary research.
- Time is significant in increasing an individual's eGFR. This makes sense that as an individual starts treatment, their kidney function should improve.
- Further, there is a drop in the rate of change after 20 days, which is shown by the significance of the `timePlus` variable.
- Again, all interactions between access type and race are not significant. It can be concluded that race does not differ the association between eGFR and access port type.

Further, from the prediction plot, the estimates of the random effects seem to be doing well at capturing the within subject rates of change, especially at the early stages of the study.

5.3 Limitations

The study contains multiple important limitations which should be considered during interpretation of the results.

First, the researchers monitored kidney function at four times during the five months. However, dialysis patients often have sessions up to three times per week and are usually on dialysis for life or until they receive a kidney transplant. Thus, more time measurements may be needed to fully model the relationship between eGFR and access port type since the wide intervals and lack thereof may not fully capture the association.

Finally, one major flaw of the study is the lack of representation of the venous fistula port. Due to the difficulty of the transplant, especially on those that are smokers, have diabetes, or are overweight, this category has a lower count. This lower count is especially clear when comparing to the other access port types. This under representation may give biased estimates in favor of the venous fistula type. Further, due to the set of conditions that affect the installation of the venous fistula port, healthier individuals are more likely to receive this port. Thus, those with this access type will tend to have higher eGFR rates due to their health. Further, those who are healthier may have more resources or are more likely to be active participants in their treatment, which would also increase eGFR. Thus, more representation of the venous fistula port may be needed to fully capture the true relationship.

5.4 Future Work

One question of interest that could not be answered due to lack of information from the data was the impact that port revisions or replacements on eGFR. While this data is not given, there a few approaches that could be implemented to handle this issue.

The first method would be to create an indicator variable that measure the status of the port at every time of measurement. This could then indicate if the port was revised, replaced, or functioning at the time of the measurements. The upsides of this approach is that this variable could be included in the mixed effects model and its coefficient estimate could be analyzed to see if it has a significant impact on eGFR. The drawbacks to this approach is that it requires more accurate tracking and another variable to be recorded. This can increase the cost of the study or lead to more issues such as missing data.

The second approach that could be taken is to record the time of a revision or replacement, and then take the difference of the time of measurement and this recording. This will then scale the time variable such that negative values indicate before revision/replacement, 0 indicates replacement/revision, and positive values indicate after revision/replacement. While this would make the repeated measurements unbalanced, the mixed effects model can handle this. Further, a piecewise component could then be added at time 0 to see if the rate of change of eGFR was affected after revision. The upsides to this method is that it has an easy implementation to the current methods used, and it can give useful visuals and interpretations. The biggest drawback to this approach is that it is only effective if all individuals only experience one replacement or revision. If this occurs more than once, then there is no way for this new variable to account for that, causing a loss in information.

References

(1) United States Renal Data System (USRDS). (2023). 2023 USRDS Annual Data Report: Epidemiology of Kidney Disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases.

(2) National Kidney Foundation. (2020). KDOQI Clinical Practice Guidelines for Vascular Access: 2019 Update. American Journal of Kidney Diseases, 75(4 Suppl 2), S1–S164.

(3) Lee, T., Mokrzycki, M., Moist, L., Maya, I. D., Vazquez, M., & Lok, C. E. (2021). Standardized definitions for hemodialysis vascular access. Seminars in Dialysis, 34(1), 47–53.

(4) Allon, M., & Robbin, M. L. (2002). Increasing arteriovenous fistulas in hemodialysis patients: problems and solutions. Kidney International, 62(4), 1109–1124.

(5) Patzer, R. E., Amaral, S., Wasse, H., Volkova, N., Kleinbaum, D., McClellan, W., & Pastan, S. (2014). Neighborhood poverty and racial disparities in kidney transplant wait listing. American Journal of Transplantation, 14(2), 349–358.

(6) Kazakova, S., Carrero, J. J., & Caskey, F. J. (2022). Disparities in vascular access use and outcomes among hemodialysis patients. Clinical Journal of the American Society of Nephrology, 17(3), 385–393.

(7) Moist, L. M., Lok, C. E., Vachharajani, T. J., et al. (2008). Vascular access function and disability: what's the link? American Journal of Kidney Diseases, 51(1), 125–131.

(8) Uppal P, Golden BL, Panicker A, Khan OA, Burday MJ. The Case Against Race-Based GFR. Dela J Public Health. 2022 Aug 31;8(3):86-89. doi: 10.32481/djph.2022.08.014. PMID: 36177174; PMCID: PMC9495470.

Appendix

R Code

```
# Libraries needed for analysis
library(tidyverse)
library(nlme)

### Import Data
setwd("~/Documents/WMU/SPRING 2025/STAT 6640/Final")
esrd = read.csv("esrd.csv")
head(esrd)

### Change categorical variables to factor type
esrd = esrd %>%
mutate_at(vars(acctype, female, racegrp, smokegrp, dx_diab),
as.factor) %>%
mutate(acctype = fct_relevel(acctype,"2"))
summary(esrd)

### Create long format of data
esrd.long = esrd %>%
pivot_longer(cols = 13:16,
names_to = "Time",
values_to = "eGFR") %>%
mutate(Time = case_when(
Time == "eGFR7" ~ 7,
Time == "eGFR20" ~ 20,
Time == "eGFR90" ~ 90,
Time == "eGFR150" ~ 150,
))
head(esrd.long)

### EDA

## Plot for access type
acctype.result = esrd.long %>%
group_by(acctype, Time) %>%
summarise(mean(eGFR))
acctype.result

ggplot(acctype.result, aes(x = Time, y =`mean(eGFR)`, color = acctype)) +
geom_point() +
geom_line(linewidth = 1) +
scale_color_discrete(name = "Access Type",
labels = c("Graft", "Standard Fistula",
"Venous Fistula")) +
labs(title = 'Mean eGFR for each Access Type', x = 'Time',y = 'Mean eGFR')

## Plot for race
race.result = esrd.long %>%
group_by(racegrp, Time) %>%
summarise(mean(eGFR))
race.result
```

```
ggplot(race.result, aes(x = Time, y =‘mean(eGFR)‘, color = racegrp)) +
geom_point() +
geom_line(linewidth = 1) +
scale_color_discrete(name = "Race", labels = c("White", "AA", "Other")) +
labs(title = 'Mean eGFR for each Race', x = 'Time',y = 'Mean eGFR')

## Plot for diabetes type
diab.result = esrd.long %>%
group_by(dx_diab, Time) %>%
summarise(mean(eGFR))
diab.result

ggplot(diab.result, aes(x = Time, y =‘mean(eGFR)‘, color = dx_diab)) +
geom_point() +
geom_line(linewidth = 1) +
scale_color_discrete(name = "Diabetes Status",
labels = c("Type I", "Type II")) +
labs(title = 'Mean eGFR for each Diabetes Type', x = 'Time',
y = 'Mean eGFR')

## Plot for gender type
sex.result = esrd.long %>%
group_by(female, Time) %>%
summarise(mean(eGFR))
sex.result

ggplot(sex.result, aes(x = Time, y =‘mean(eGFR)‘, color = female)) +
geom_point() +
geom_line(linewidth = 1) +
scale_color_discrete(name = "Gender", labels = c("Male", "Female")) +
labs(title = 'Mean eGFR for each Gender', x = 'Time',y = 'Mean eGFR')
## Plot for smoking status
smoke.result = esrd.long %>%
group_by(smokegrp, Time) %>%
summarise(mean(eGFR))
smoke.result

ggplot(smoke.result, aes(x = Time, y =‘mean(eGFR)‘, color = smokegrp)) +
geom_point() +
geom_line(linewidth = 1) +
scale_color_discrete(name = "Smoking Status",
labels = c("Never Smoked", "Former Smoker",
"Current Smoker")) +
labs(title = 'Mean eGFR for each Smoking Status', x = 'Time',
y = 'Mean eGFR')

## Histogram for eGFR7 and log(eGFR7)
hist(esrd$eGFR7)
hist(log(esrd$eGFR7))

### Linear Model Code

## BIC table
```

```
fit.lm1 = lm(log(eGFR7) ~  acctype + age_ssd +
acctype:racegrp +smokegrp + bmi + dx_diab +
female,
data = esrd)
fit.lm2 = lm(log(eGFR7) ~  acctype + age_ssd +
acctype:racegrp +smokegrp + bmi + female,
data = esrd)
fit.lm3 = lm(log(eGFR7) ~  acctype + age_ssd +
acctype:racegrp + female + bmi,
data = esrd)
fit.lm4 = lm(log(eGFR7) ~  acctype + age_ssd +
acctype:racegrp + bmi,
data = esrd)
fit.lm5 = lm(log(eGFR7) ~  acctype + age_ssd +
acctype:racegrp,
data = esrd)

BIC.lm.tab = data.frame(model = c("Fit 1", "Fit 2","Fit 3",
"Fit 4", "Fit 5"),
BIC  = (c(BIC(fit.lm1),BIC(fit.lm2),BIC(fit.lm3),
BIC(fit.lm4),BIC(fit.lm5))))
BIC.lm.tab

## Summary of results and diagnostic plots
signif(coef(summary(fit.lm5)),3)
signif(confint(fit.lm5, level = 0.95),3)
plot(fit.lm5)

### All longitudinal data

## Lattice plot for all subjects
time.result = esrd.long %>%
group_by(Time) %>%
summarise(mean(eGFR))

ggplot(esrd.long, aes(x = Time, y = eGFR, group = usrds_id,
color = factor(usrds_id))) +
geom_line(alpha = 0.2, size = 0.4) +
geom_line(data = time.result, aes(x = Time, y = `mean(eGFR)`),
color = "blue", size = 1.3, inherit.aes = FALSE) +
labs(title = 'eGFR vs. Time with Population average eGFR', x = 'Time',
y = 'eGFR') +
theme_minimal() +
theme(legend.position = "none")

## Creating piecewise variable
library(nlme)
esrd.long$timePlus = esrd.long$Time * as.numeric(esrd.long$Time > 20)
head(esrd.long)

## Using ML on reduced and full to determine best model
fit.lme1 = lme(fixed = eGFR ~ acctype + age_ssd + Time  + acctype:racegrp,
random = ~ Time | usrds_id,
data = esrd.long,
```

```
control = lmeControl(
maxIter = 100,
msMaxIter = 100,
opt = "optim"
),
method = "ML")
fit.lme2 = lme(fixed = eGFR ~ acctype + age_ssd + Time + timePlus +
acctype:racegrp ,
random = ~ 1 | usrds_id,
data = esrd.long,
control = lmeControl(
maxIter = 100,
msMaxIter = 100,
opt = "optim"
),
method = "ML")
anova(fit.lme1,fit.lme2)

## Use REML for fit.lme1 to get less biased estimates
fit.lme3 = lme(fixed = eGFR ~ acctype + age_ssd + Time + timePlus +
acctype:racegrp ,
random = ~ Time | usrds_id,
data = esrd.long,
control = lmeControl(
maxIter = 100,
msMaxIter = 100,
opt = "optim"
),
method = "REML")
signif(coef(summary(fit.lme3)),3)
intervals(fit.lme3)

## Diagnostic plots for lme model
qqnorm(fit.lme3)
hist(fit.lme3$residuals)
plot(fit.lme3)

## Plot for within subject prediction
esrd.long$predict = predict(fit.lme3)
esrd.subset = subset(esrd.long, usrds_id<259677)

ggplot(esrd.subset, aes(x=Time,y=eGFR,col=factor(usrds_id))) +
geom_point(alpha=2) + labs(
title = "Predicted Trajectories of First 7 Individuals",
x ="Time", y ='eGFR') +
geom_line(data=esrd.subset, aes(x=Time, y=predict), lwd=1) +
scale_color_discrete(name = "Subject ID")
```

Plots from Code

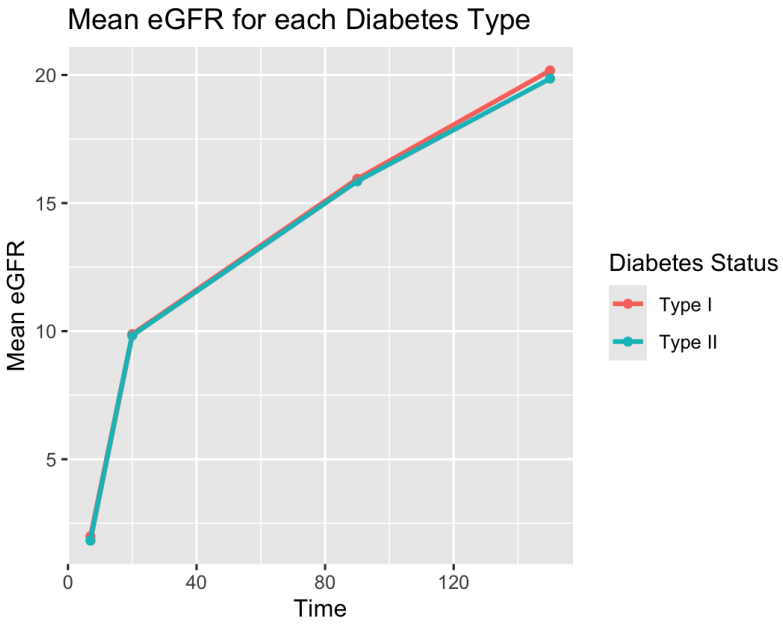

Figure 8: Plot of the average eGFR over time separated by diabetes type.

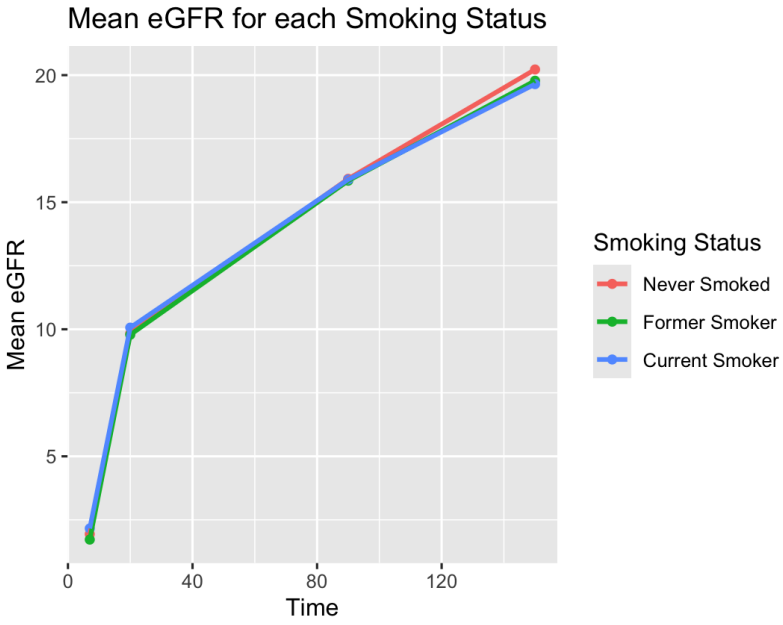

Figure 9: Plot of the average eGFR over time separated by smoking status.

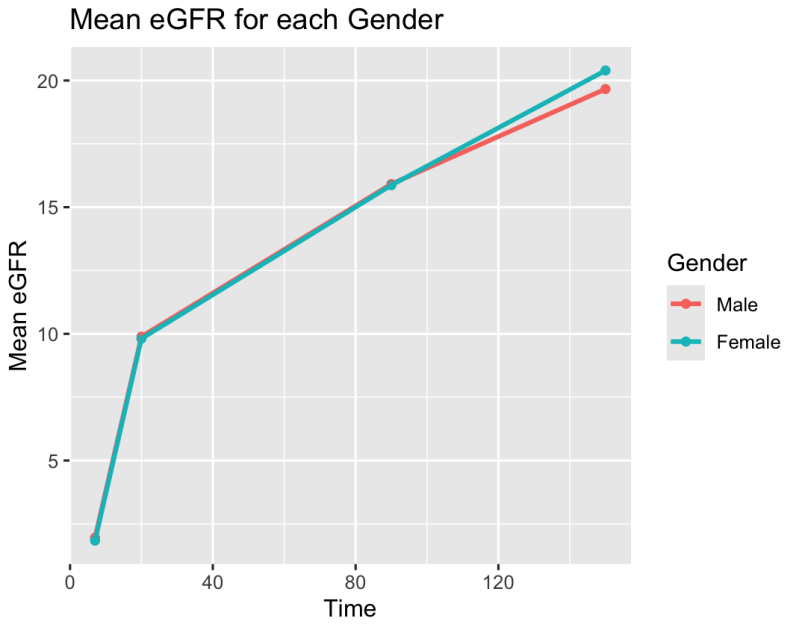

Figure 10: Plot of the average eGFR over time separated by gender.

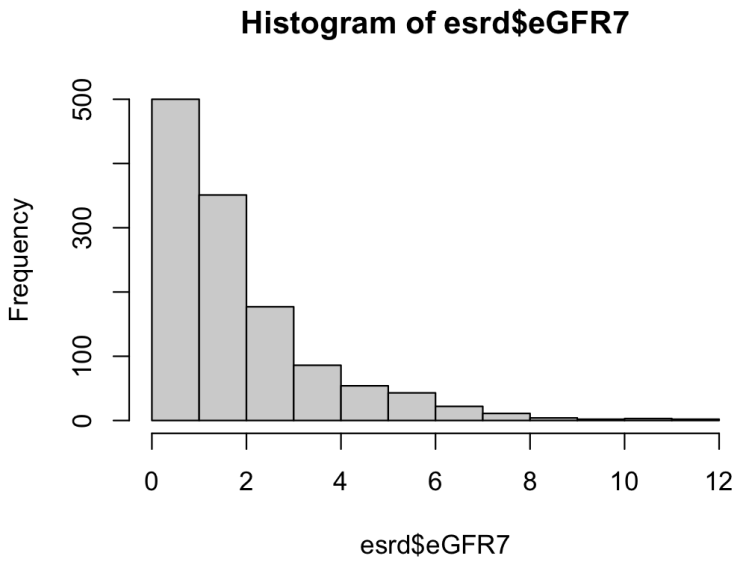

Figure 11: Histogram of eGFR7.

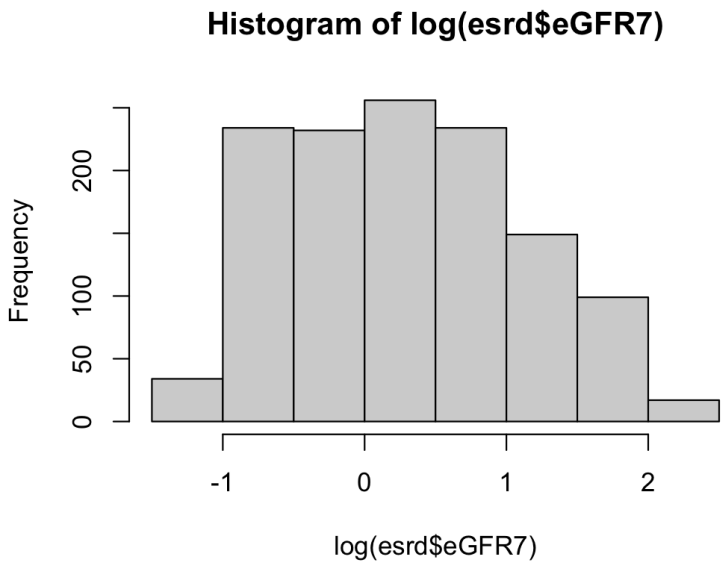

Figure 12: Histogram of log of eGFR7.

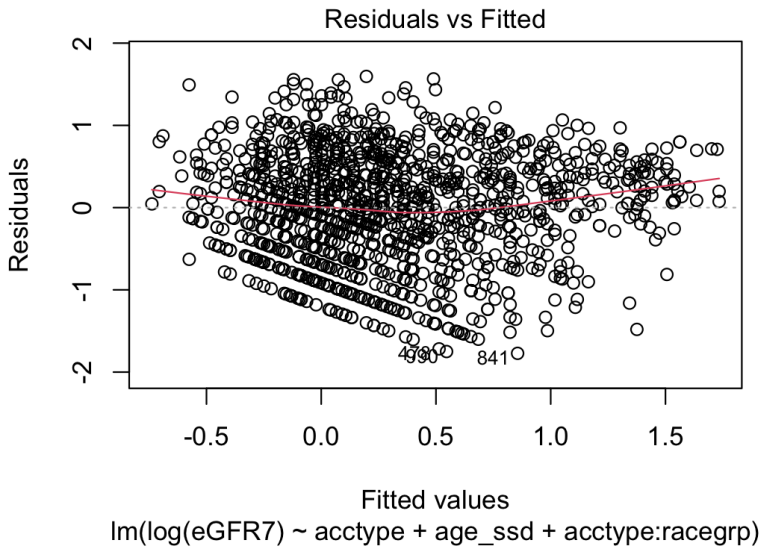

Figure 13: Residual plot for linear regression plot.

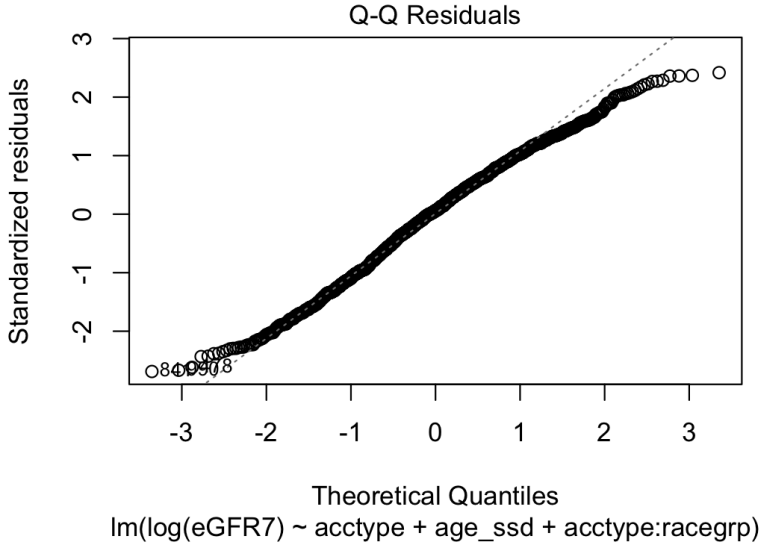

Figure 14: QQ plot for linear regression model.

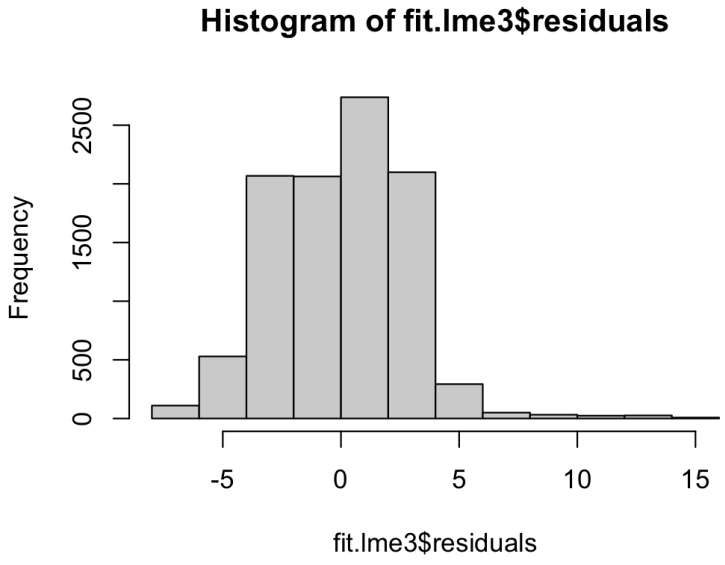

Figure 15: Histogram of residuals for mixed effects model.

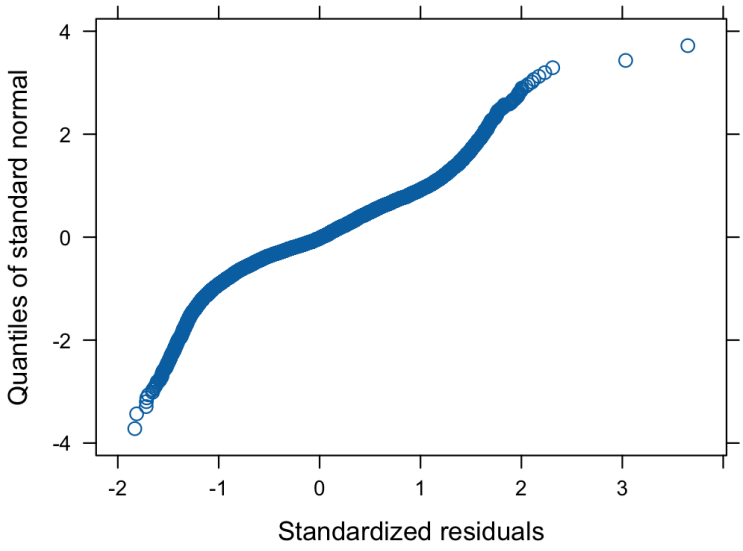

Figure 16: QQ plot for mixed effects model.

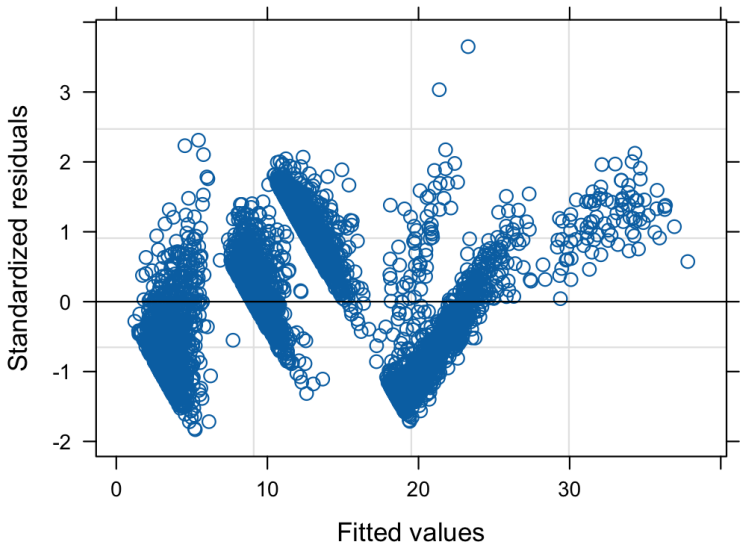

Figure 17: Residual plot for mixed effects model.