Trends in Functional Impairment

Taylor Keating

01/06/2021

Introduction

Minority groups in the United States experience greater challenges on average due to lack of access to healthcare, fewer employment opportunities, and many other reasons. The Americans' Changing Lives (ACL) survey series is a longitudinal study that began following 3,617 Black and White Americans aged 25 years and up in 1986. We will be using the data from this survey series to assess the trends in odds of functional impairment over time and to determine if this trend is different among racial groups.

Methods

Descriptive Analysis: To investigate the trend in functional impairment over time for both racial groups, we plot the percentage functionally impaired at each data collection wave for both White Americans and African Americans in **Figure 1**. We will then provide plots showing the patterns of missing outcome measurements over the years, as well as plots showing the percentage of missing outcomes, stratified by race, for each wave year in **Figure 2**. Lastly, to investigate further the missingness mechanism we will present present the percentage functionally impaired in each year, stratified by race and whether or not individuals had any missing observations in **Table 1**.

Inferential Analysis: To investigate whether past functional impairment status is an important predictor of current functional impairment status, we will first fit a marginal logistic regression model using GEE with a working independence covariance model using all available data. We will use binary functional impairment as the outcome, an interaction between race and year, and adjust for baseline age, sex, socioeconomic status, and past functional impairment status. The estimates on the log-odds scale from this model will be presented in **Table 2**. Using this model, we will test whether changes in odds of functional impairment over time vary by racial group (adjusting for all other covariates including past functional impairment status) and will test whether past functional impairment status is an important predictor of current functional impairment status. To investigate whether the trends in odds of functional impairment over time varies across racial groups, we will fit a 2nd marginal logistic regression model using GEE with a working independence covariance matrix. We will still use binary functional impairment as the outcome, an interaction between race and year, but will only adjust for baseline age, sex, and socioeconomic status. The primary analysis will be performed using all available data. A sensitivity analysis will be performed using Inverse-Probability-Weighting (IPW) to account for the missing data. The IPW-GEE method will use a missingness model for functional impairment with covariates of race, year, sex, socioeconomic status, baseline age, and past functional impairment status. The estimates from these two models on the log-odds scale will be presented in **Table 3**. We will test whether there is a difference in functional impairment at baseline between African Americans and White Americans and whether changes in odds of functional impairment over time vary by racial group.

Results

Descriptive Analysis: From **Figure 1** we see that the percentage functionally impaired increases over time from 1986 to 2011 for both White Americans and African Americans, but with this rate of increase being greater for African Americans. In **Figure 2** we see that there is no missing outcome measurements in 1986, over 25% missing in 2002, and between 15-20% missing in all other years. We also see that about 60% of participants completed all five survey years, 7% missed only the last survey in 2011, but that the pattern of missingness is non-monotone. We finally notice that the percentage of missing outcomes is higher for Africans Americans than for White Americans in every year with missing data, leading us to believe that missingness could depend on race. In **Table 1** we see that the percentage functionally impaired at baseline, in 1989, and in 1994 is much higher for African Americans who have complete data compared to those with missing data at some point in the study, but this trend is not seen in White Americans. This pattern could mean that missingness depends on past or present functional impairment status, which is important to remember while considering the validity of our inferences later.

Inferential Analysis: The results from the GEE model adjusting for past functional impairment status are presented in **Table 2** on the log-odds scale. We estimate that among African Americans of the same sex, baseline age, socio-economic status, and past functional impairment status, the odds of functional impairment is 8.4% higher per year (95% CI: 6.6% higher to 10.3% higher). We estimate that among White Americans of the same sex, baseline age, socio-economic status, and past functional impairment status, the odds of functional impairment is 7.9% higher per year (95% CI: 6.5% higher to 9.2% higher). We do not have evidence that the odds ratio of functional impairment over time is different between the two races when adjusting for sex, baseline age, socioeconomic status, and past functional impairment status ($p=0.62$). We estimate that when comparing groups of the same sex, baseline age, socio-economic status, race, and in the same year, the odds of functional impairment is 14.9 times higher for a group that was functionally impaired at the last measurement compared to a group that was not (95% CI: 11.5 to 19.3 times higher). We conclude that this odds ratio is not equal to 1 ($p<0.001$) and therefore past functional impairment status is an important predictor of current functional impairment status.

We next will discuss the results and inference from our 2nd GEE model without adjusting for past functional impairment status. The results from both the primary analysis using available data and the sensitivity analysis using IPW GEE are presented in **Table 3** on the log-odds scale. From the available data analysis, we estimate that when comparing groups of the same sex, socioeconomic status, and baseline age, the odds of functional impairment in 1986 was 14% lower for White Americans than for African Americans (95% CI: 38% lower to 18% higher). Using IPW GEE, this estimate was 17% lower (95% CI: 40% lower to 15% higher). We do not have evidence that there is a difference in functional impairment at baseline in 1986 between White Americans and African Americans of the same sex, socioeconomic status, and baseline age in both the primary and sensitivity analyses (p=0.34 and p=0.26). We next estimate that among African Americans of the same sex, baseline age, and socioeconomic status, the odds of functional impairment is 11.2% higher per year (95% CI: 10.0% higher to 12.5% higher). Using IPW GEE, this estimate was 10.6% higher per year (95% CI: 9.1% higher to 12.1% higher). We estimate that among White Americans of the same sex, baseline age, and socioeconomic status, the odds of functional impairment is 8.9% higher per year (95% CI: 7.9% higher to 10.0% higher). Using IPW GEE, this estimate was 8.4% higher per year (95% CI: 7.3% higher to 9.5% higher). We reject the null hypothesis that the odds ratio of functional impairment over time comparing groups of the same sex, baseline age, and socioeconomic status is the same between the two races in both the primary and sensitivity analyses (p=0.006 and p=0.02).

Discussion

One of the first key results is that there was no significant difference in functional impairment at baseline comparing African Americans and White Americans of the same sex, socioeconomic status, and baseline age. We also found that previous functional impairment status is an important predictor of current functional impairment status. We next found that there was a difference in the trend of functional impairment over time between African Americans and White Americans of the same sex, socioeconomic status, and baseline age. The inference from the primary analysis using GEE with available data was the same as that from the sensitivity analysis using IPW GEE. However, the IPW GEE method produced slightly different estimates with higher standard errors. As well, the GEE with available data analysis only gives us valid inference if the data are missing completely at random (MCAR). On the other hand, the IPW GEE analysis gives us valid inference if the data are missing at random (MAR) and the missingness mechanism was correctly specified.

Tables and Figures

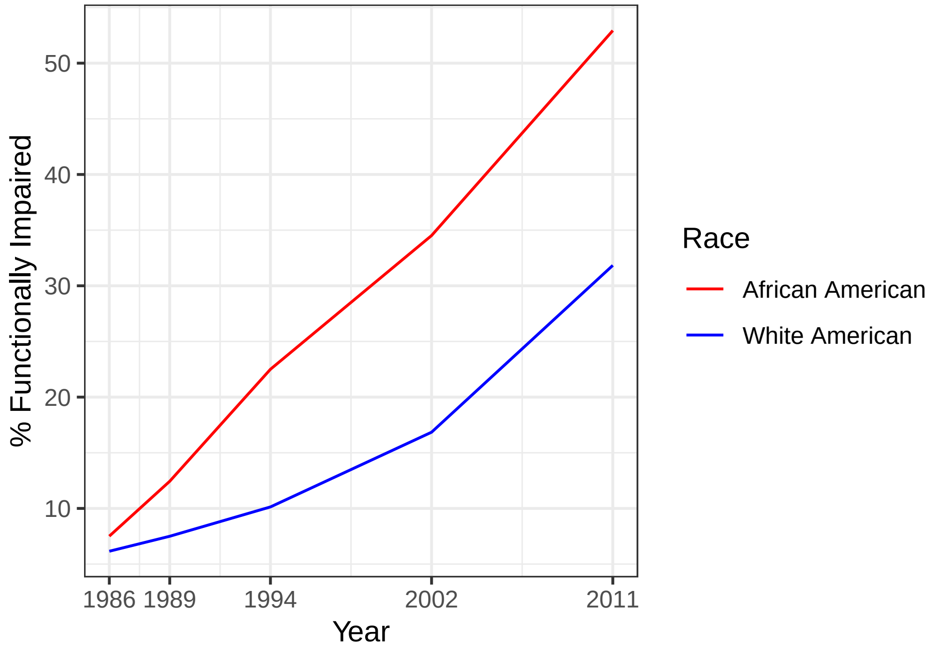
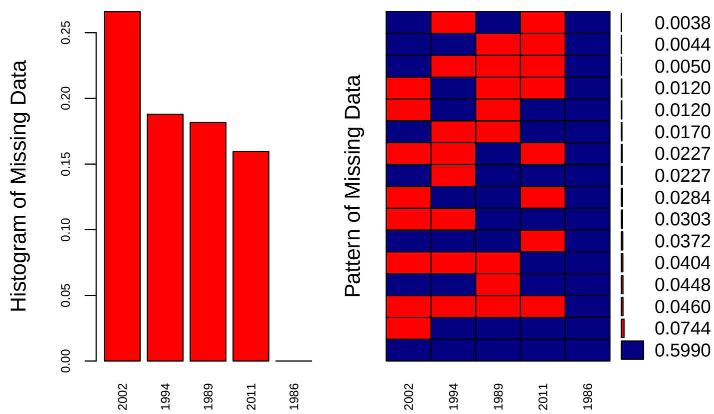


Figure 1: Plot of % Functionally Impaired Over Time by Race

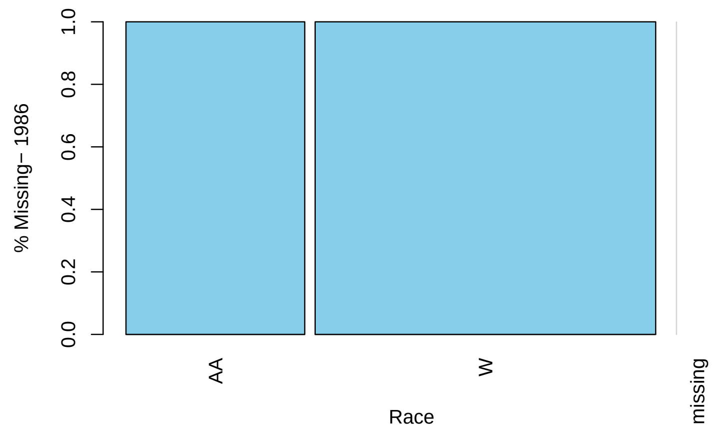
```
##
##  Variables sorted by number of missings:
##  Variable     Count
##       2002 0.2660782
##       1994 0.1878941
##       1989 0.1815889
##       2011 0.1595208
##       1986 0.0000000
```

Table 1: % Functionally Impaired for Individuals with Complete vs Missing Data

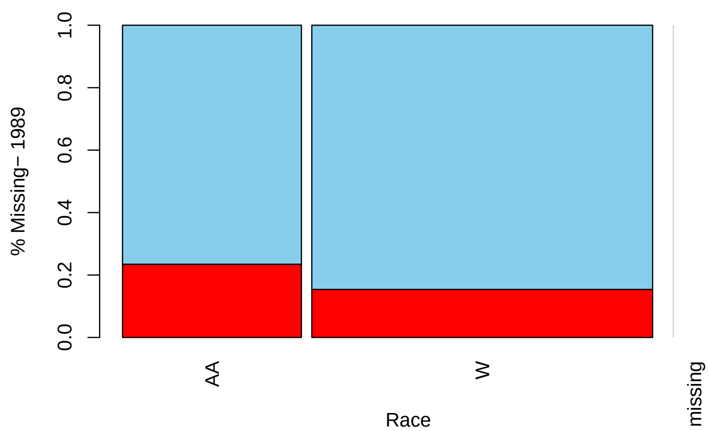
Race	Missing Observations	Number	% Impaired 1986	% Impaired 1989	% Impaired 1994	% Impaired 2002	% Impaired 2011
AA	FALSE	256	0.105	0.152	0.266	0.348	0.520
AA	TRUE	290	0.048	0.080	0.153	0.337	0.543
W	FALSE	694	0.063	0.065	0.104	0.166	0.291
W	TRUE	346	0.058	0.113	0.093	0.183	0.415



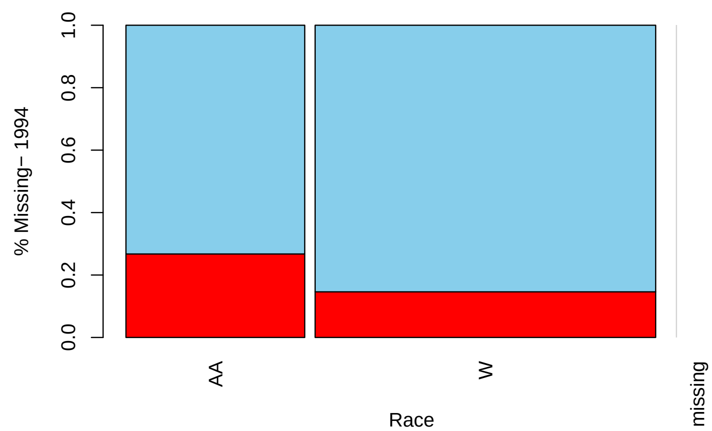
(a) Missingness pattern of outcomes.



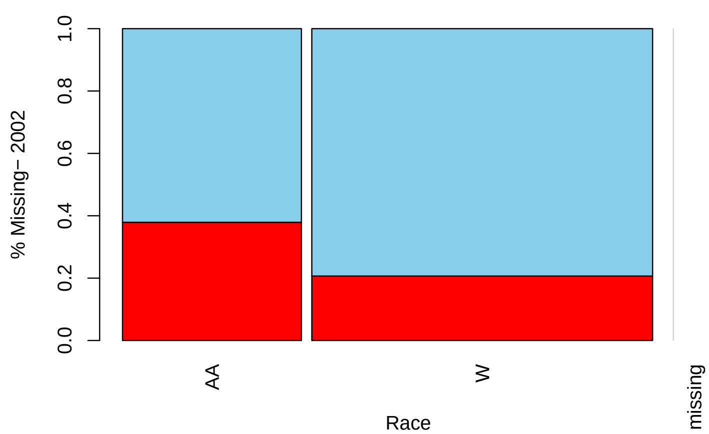
(b) Missingness pattern of 1986 outcomes, stratified by race.



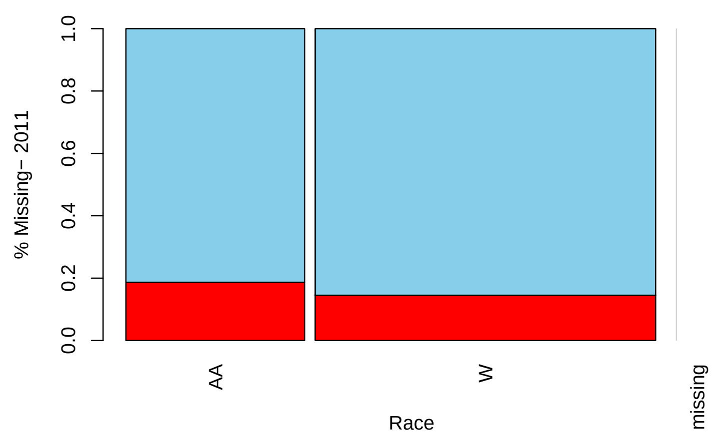
(c) Same as (b), but for 1989.



(d) Same as (b), but for 1994.



(e) Same as (b), but for 2002.



(f) Same as (b), but for 2011.

Figure 2: Missingness Patterns

Table 2: Log Odds Estimates for GEE Model adjusting for Past Functional Impairment

	Estimate	Std Error	p-value
Intercept	-164.459	17.256	0.000
White American	10.195	21.480	0.635
Year	0.081	0.009	0.000
Male	-0.170	0.100	0.090
Middle SES Class	-0.610	0.136	0.000
Upper SES Class	-1.161	0.134	0.000
Baseline Age	0.044	0.006	0.000
Past Functional Impairment Status	2.701	0.132	0.000
White American:Year	-0.005	0.011	0.619

Table 3: Log Odds Estimates for GEE Available and IPW GEE Models

	GEE Available Estimate	GEE Available Std Err	GEE Available p-value	IPW GEE Estimate	IPW GEE Std Err	IPW GEE p-value
Intercept	-3.732	0.316	0.000	-3.538	0.350	0.000
White American	-0.154	0.163	0.342	-0.189	0.167	0.258
Years after 1986	0.106	0.006	0.000	0.101	0.007	0.000
Male	-0.350	0.110	0.001	-0.341	0.127	0.007
Middle SES Class	-0.664	0.145	0.000	-0.845	0.163	0.000
Upper SES Class	-1.454	0.143	0.000	-1.556	0.161	0.000
Baseline Age	0.059	0.007	0.000	0.058	0.008	0.000
White American : Years after 1986	-0.021	0.008	0.006	-0.020	0.008	0.020

Code Appendix

```
# Packages Setup
library(knitr)
knitr::opts_chunk$set(echo = FALSE)
library(tidyverse)
library(reshape2)
library(GGally)
library(VIM)
library(geepack)
library(multcomp)
library(wgeesel)



#-------------------------
# Data
acl_wide<- read_csv("Data/acl_subset.csv")[,-1]
acl_wide$race<- as.factor(acl_wide$race)

acl_long<- melt(acl_wide, id=c("id","sex", "race", "ses","age"))
acl_long<- arrange(acl_long, id, variable)
acl_long <- acl_long %>% mutate(year= case_when(variable=="w1" ~ 1986,
                                                variable=="w2" ~ 1989,
                                                variable=="w3" ~ 1994,
                                                variable=="w4" ~ 2002,
                                                variable=="w5" ~ 2011))
acl_long$year_from_baseline<- acl_long$year - 1986
acl_long$lag1value<- ylag(acl_long$id, acl_long$value, 1)
acl_long$missing<- ifelse(is.na(acl_long$value),1,0)

acl_wide$miss<- apply(is.na(acl_wide), 1, any)
#--------------------------
# Summary Table
```

```r
summary_tab<- acl_long %>% group_by(race,year) %>%
  summarise(nobs=n()-sum(missing),
            nmissing=sum(missing),
            n_impaired=sum(value, na.rm=TRUE),
            percent_impaired=mean(value, na.rm=TRUE))

# plot of % functionally impaired over time by race
ggplot(data=summary_tab, mapping=aes(x=year, y=100*percent_impaired,
                                     group=race, colour=race)) +
  geom_line() +
  labs(x="Year", y="% Functionally Impaired", colour="Race") +
  scale_colour_manual(values=c("red", "blue"),
                      labels=c("African American", "White American")) +
  scale_x_continuous(breaks=c(1986,1989,1994,2002,2011)) +
  theme_bw()
#----------------------------
# Missing Data Patterns
aggr_plot<- aggr(acl_wide[,6:10],
      col=c("navyblue","red"),
      numbers=TRUE,
      sortVars=TRUE,
      labels=c("1986","1989","1994","2002","2011"),
      cex.axis=0.7,
      gap=3,
      ylab=c("Histogram of Missing Data","Pattern of Missing Data"))
spineMiss(as.data.frame(acl_wide[, c("race", "w1")]),
          xlab="Race", ylab="% Missing- 1986")
spineMiss(as.data.frame(acl_wide[,c("race", "w2")]),
          xlab="Race", ylab="% Missing- 1989")
spineMiss(as.data.frame(acl_wide[,c("race", "w3")]),
          xlab="Race", ylab="% Missing- 1994")
spineMiss(as.data.frame(acl_wide[,c("race", "w4")]),
```

```
          xlab="Race", ylab="% Missing- 2002")
spineMiss(as.data.frame(acl_wide[,c("race", "w5")]),
          xlab="Race", ylab="% Missing- 2011")
# Complete vs Missing observations comparison

missing_comparison_tab<- acl_wide %>% group_by(race, miss) %>%
  summarise(N=n(),
            percent_1986=mean(w1),
            percent_1989=mean(w2, na.rm=TRUE),
            percent_1994=mean(w3, na.rm=TRUE),
            percent_2002=mean(w4, na.rm=TRUE),
            percent_2011=mean(w5, na.rm=TRUE))
missing_comparison_tab[,4:8]<- round(missing_comparison_tab[,4:8], digits=3)
colnames(missing_comparison_tab)<- c("Race", "Missing Observations", "Number",
                                     "% Impaired 1986",
                                     "% Impaired 1989",
                                     "% Impaired 1994",
                                     "% Impaired 2002",
                                     "% Impaired 2011")
kable(missing_comparison_tab,
      caption= "% Functionally Impaired for Individuals with Complete vs Missing Data")
#--------------------------
# Models

#------------------------
#Q2 Modeling

#gee model with available data and adjusting for past functional impairment status
acl_long_subset<- acl_long[complete.cases(acl_long),]
gee_with_past<- geeglm(value ~ race*year + sex + ses + age + lag1value,
                            id=id,
                            data=acl_long_subset,
```

```
                            family=binomial(link="logit"),
                            corstr="independence")

  #Estimates and CI's for OR per year for each race
gee_with_past_raceAA_peryear<- glht(gee_with_past, "year = 0")
ci_gee_with_past_raceAA_peryear<- (confint(gee_with_past_raceAA_peryear)$confint)[1,]
exp_ci_gee_with_past_raceAA_peryear<- exp(ci_gee_with_past_raceAA_peryear)

gee_with_past_raceW_peryear<- glht(gee_with_past, "year + raceW:year = 0")
ci_gee_with_past_raceW_peryear<- (confint(gee_with_past_raceW_peryear)$confint)[1,]
exp_ci_gee_with_past_raceW_peryear<- exp(ci_gee_with_past_raceW_peryear)

  #Estimates and CI's for OR by lag1value
gee_with_past_lag1value<- glht(gee_with_past, "lag1value = 0")
ci_gee_with_past_lag1value<- (confint(gee_with_past_lag1value)$confint)[1,]
exp_ci_gee_with_past_lag1value<- exp(ci_gee_with_past_lag1value)

#Table of Estimates for Q2 (log-odds scale)
summary_gee_with_past<- (summary(gee_with_past))$coefficients
summary_gee_with_past<- summary_gee_with_past[,-3]
summary_gee_with_past<- as.data.frame(round(summary_gee_with_past, digits=3))
row.names(summary_gee_with_past)<- c("Intercept", "White American","Year",
                                     "Male","Middle SES Class","Upper SES Class",
                                     "Baseline Age","Past Functional Impairment Status",
                                     "White American:Year")
colnames(summary_gee_with_past)<- c("Estimate", "Std Error", "p-value")
kable(summary_gee_with_past,
      caption= "Log Odds Estimates for GEE Model adjusting for Past Functional Impairment")
#-----------------------
#Q3 Modeling

#available data model
```

```r
acl_long_available<- acl_long[complete.cases(acl_long$value),]
gee_available<- geeglm(value ~ race*year_from_baseline + sex + ses + age,
                       id=id,
                       data=acl_long_available,
                       family=binomial(link="logit"),
                       corstr="independence")

  #Estimates and CI's for OR between races at baseline (year=1986)
gee_available_raceW_est<- glht(gee_available, "raceW = 0")
ci_gee_available_raceW_est<- (confint(gee_available_raceW_est)$confint)[1,]
exp_ci_gee_available_raceW_est<- exp(ci_gee_available_raceW_est)

  #Estimates and CI's for OR per year for each race
gee_available_raceAA_peryear<- glht(gee_available, "year_from_baseline = 0")
ci_gee_available_raceAA_peryear<- (confint(gee_available_raceAA_peryear)$confint)[1,]
exp_ci_gee_available_raceAA_peryear<- exp(ci_gee_available_raceAA_peryear)

gee_available_raceW_peryear<- glht(gee_available,
                                   "year_from_baseline + raceW:year_from_baseline = 0")
ci_gee_available_raceW_peryear<- (confint(gee_available_raceW_peryear)$confint)[1,]
exp_ci_gee_available_raceW_peryear<- exp(ci_gee_available_raceW_peryear)



#IPW model
  #make monotone dropout in wide format
acl_wide_ipw<- acl_wide %>% dplyr::select(!miss)
acl_wide_ipw[is.na(acl_wide_ipw$w1), 7:10]<- NA
acl_wide_ipw[is.na(acl_wide_ipw$w2), 8:10]<- NA
acl_wide_ipw[is.na(acl_wide_ipw$w3), 9:10]<- NA
acl_wide_ipw[is.na(acl_wide_ipw$w4), 10]<- NA
  #reshape into long format and add variables needed
acl_long_ipw<- melt(acl_wide_ipw, id=c("id","sex", "race", "ses","age"))
```

```
acl_long_ipw<- arrange(acl_long_ipw, id, variable)
acl_long_ipw <- acl_long_ipw %>% mutate(year= case_when(variable=="w1" ~ 1986,
                                                        variable=="w2" ~ 1989,
                                                        variable=="w3" ~ 1994,
                                                        variable=="w4" ~ 2002,
                                                        variable=="w5" ~ 2011))
acl_long_ipw$year_from_baseline<- acl_long_ipw$year - 1986
acl_long_ipw$lag1value<- ylag(acl_long_ipw$id, acl_long_ipw$value, 1)
acl_long_ipw$R<- ifelse(is.na(acl_long_ipw$value),0,1)
  #ipw gee model
gee_ipw<- wgee(model= value ~ race*year_from_baseline + sex + ses + age,
               id=acl_long_ipw$id,
               data=acl_long_ipw,
               family=binomial(link="logit"),
               corstr="independence",
               scale=NULL,
               mismodel= R ~ race + year_from_baseline + sex + ses + age + lag1value)

  #Estimates and CI's for OR between races at baseline (year=1986)
lambda<- c(0,1,0,0,0,0,0,0)
exp_gee_ipw_raceW_est<- exp(lambda %*% gee_ipw$beta)
exp_ci_gee_ipw_raceW_est<- exp(lambda %*% gee_ipw$beta + qnorm(c(0.025,0.975)) *
                                 c(sqrt(lambda %*% gee_ipw$var %*% lambda)))

  #Estimates and CI's for OR per year for each race
lambda1<- c(0,0,1,0,0,0,0,0)
lambda2<- c(0,0,1,0,0,0,0,1)
exp_gee_ipw_raceAA_peryear<- exp(lambda1 %*% gee_ipw$beta)
exp_ci_gee_ipw_raceAA_peryear<- exp(lambda1 %*% gee_ipw$beta + qnorm(c(0.025,0.975)) *
                                      c(sqrt(lambda1 %*% gee_ipw$var %*% lambda1)))
exp_gee_ipw_raceW_peryear<- exp(lambda2 %*% gee_ipw$beta)
exp_ci_gee_ipw_raceW_peryear<- exp(lambda2 %*% gee_ipw$beta + qnorm(c(0.025,0.975)) *
```

```r
                                        c(sqrt(lambda2 %*% gee_ipw$var %*% lambda2)))

#Comparison of primary and sensitivity table (log-odds scale) for Q3
summary_gee_available<- (summary(gee_available))$coefficients
summary_gee_available<- summary_gee_available[,-3]
summary_gee_ipw<- summary(gee_ipw)
summary_gee_ipw<- cbind(summary_gee_ipw$beta, summary_gee_ipw$se.robust, summary_gee_ipw$p)

comparison_tab<- cbind(summary_gee_available, summary_gee_ipw)
comparison_tab<- as.data.frame(round(comparison_tab, digit=3))
row.names(comparison_tab)<- c("Intercept","White American", "Years after 1986",
                              "Male","Middle SES Class", "Upper SES Class",
                              "Baseline Age", "White American : Years after 1986")
colnames(comparison_tab)<- c("GEE Available Estimate", "GEE Available Std Err",
                             "GEE Available p-value","IPW GEE Estimate",
                             "IPW GEE Std Err", "IPW GEE p-value")
kable(comparison_tab, caption= "Log Odds Estimates for GEE Available and IPW GEE Models")
```