Capstone Difference in Difference Report

Cindy Elder, Max Bi and Gabby Lopez

2025-02-24

# For ER Rates Population per Year

## All Patients in 2019 to 2022

Since the first Marshallese Community Health Worker (CHW) was hired in September 2019, we used the period of August 31, 2018 to August 31, 2019 as the pre-treatment populations of Marshallese and Non-Hispanic White patients at the Maple and Market clinics. We also used the same dates in 2021 and 2022 to find the post-treatment populations.

! Size of population

The population of CHAS patients has grown over time. Using data from all of the patients who visited the CHAS Maple and Market clinics during this time, we built a model that accounts for the correlation that results from patients who have been there in both 2019 and 2022, and includes patients who were not there in 2019 but joined by 2022.

We modeled the number of ER visits between June 1 and August 31, 2019, and the number for the same dates in 2022, using a saturated model with indicators for Marshallese, year, and the interaction of these 2 variables.

Using GEE GLM, we estimate that the expected value of ER visits per patients for our control population of Non-Hispanic White patients in 2019 is 0.0368721.

Similarly, we estimate the rate of ER visits per person visits for Marshallese in 2019 is 0.0490144. This is slightly higher than the Non-Hispanic White population, but we note that the p-value is above the alpha = 0.05 level (p-value = 0.063.)

! how to take out the p-value of this model?

Over the next three years, the rate of change for the ER visits Non-Hispanic White patients went up by 0.0032971 per year. This may not be not be surprising since this data was recorded during the COVID pandemic. However, we note that the rate of change for the Marshallese patients went up by a smaller amount during the same time, -0.0078504 per year. The p-values for these coefficients are both significant at less than 0.001.

Based on the Difference in Difference causal model, we interpret lower rate of ER visits for the Marshallese as the effect of the Marshallese Healthcare Workers.

! we can use causal language with DID. ! do we like our interpretations? Should I calculate the slope \*3 for the 3 year change?

## Existing Patients from 2019 (Balanced Panel ER Rate Per Year)

When we restrict the analysis to patients that were there the entire time period, from 2019 to 2022, we see similar trends.

! how do we pull out knitr tables with this model?

# ER Rates Population per Quarter

We also analyzed the model using only patients who visited the CHAS Maple and Market Clinics in the same quarter.

Pop size shrank by 25% for the Non-Hispanic White population, and about 33% for the Marshallese population.

315/476

## [1] 0.6617647

# 0.662  
 13804/18586

## [1] 0.7427096

# 0.743

We estimate the average ER rate for the Non-Hispanic White population in 2019

0.1761665. The Marshallese were higher at 0.2348178.

Per year, the rate for Non-Hispanic White patients decreased by -0.0114948 while the Marshallese patients ER rate decreased by -0.0522476.

## Existing Patients from 2019 (Balanced Panel ER Rate Per Quarter)

For those patients who were there in both 2019 and 2022, we estimate the ER rate for the Non-Hispanic White patients in 2019 was

0.1697034 and the Marshallese ER rate was 0.2307692

Per year, the Non-Hispanic White rates went down by -0.0069366 but the Marshallese rates went down faster, -0.0388796 per year. Both have statistically signficant p-values.

# For Primary Care Provider (PCP) Visits Rates per Year

## All Patients in 2019 to 2022

Again, using a GEE GLM model, we estimate that the average Primary Care Provider visit rate is 0.0840753. The average rate for the Marshallese is slightly lower but with a p-value =0.181 it was not statistically different.

Each year, the PCP rate went up slightly for the Non-Hispanic White group by 0.0159981, but it went down very slightly for the Marshallese, -0.001451. Both of these trends have significant p-values.

## Existing Patients from 2019 (Balanced Panel PCP Rate for the Years)

If we restrict the analysis to the primary care patients who visited the clinics in both 2019 and 2022, we find similar trends. The Marshallese had slightly higher rates of PCP visits but not statistically significant with a p-value = 0.27. During the three years of study, the current Non-Hispanic White patients had a statistically significant increase in their PCP rates, with an average increase of 0.0123829 per year.

However, again the Marshallese on average had a decreased PCP rate, with a change of -0.0113363 per year.

# PCP Rates Per Quarter

When we restrict our analysis to the population of patients who visited during the same quarter, the PCP rates show slightly different results. The Non-Hispanic White patients still have a higher PCP rate, but the Marshallese do have a statistically different rate (p-value of 0.019.) Both groups have fewer PCP over the next three years with statistically different rates. The Non-Hispanic White group PCP rate changed by 0.0128576 per year, while the Marshallese had a larger change of -0.0509458 per year.

## Existing Patients from 2019 (Balanced Panel PCP Rate Per Quarter)

! This model output confuses me. The average number of PCP visits for those who visited that quarter between the Non-Hispanic White was 0.4060131. In this analysis, the Marshallese patients were not statistically different, with a p-value of 0.96. Also the rate of change for the Non-Hispanic Whites was not statistically different than zero. There was no significant change. However, the rate of change for the Marshallese over the three years was significant, at -0.069398 per year, with a p-value < 0.001.

summary(gee\_mod\_DID\_ER\_yr)

##   
## Call:  
## geeglm(formula = ER ~ marsh \* year\_center, family = gaussian,   
## data = did\_visit\_types\_year, id = UniqueIdentifier, corstr = "exchangeable",   
## scale.fix = T)  
##   
## Coefficients:  
## Estimate Std.err Wald Pr(>|W|)   
## (Intercept) 0.0368721 0.0006125 3623.862 < 2e-16 \*\*\*  
## marsh 0.0121422 0.0065208 3.467 0.0626 .   
## year\_center 0.0032971 0.0002948 125.106 < 2e-16 \*\*\*  
## marsh:year\_center -0.0111476 0.0024908 20.030 7.62e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation structure = exchangeable   
## Scale is fixed.  
##   
## Link = identity   
##   
## Estimated Correlation Parameters:  
## Estimate Std.err  
## alpha 0.04002 0.001027  
## Number of clusters: 186070 Maximum cluster size: 2

summary(gee\_mod\_DID\_ER\_yr\_bal)

##   
## Call:  
## geeglm(formula = ER ~ marsh \* year\_center, family = gaussian,   
## data = did\_visit\_types\_year\_balanced, id = UniqueIdentifier,   
## corstr = "exchangeable", scale.fix = T)  
##   
## Coefficients:  
## Estimate Std.err Wald Pr(>|W|)   
## (Intercept) 0.037391 0.000671 3105.06 < 2e-16 \*\*\*  
## marsh 0.025770 0.008986 8.22 0.00413 \*\*   
## year\_center 0.003522 0.000344 104.63 < 2e-16 \*\*\*  
## marsh:year\_center -0.012391 0.003628 11.67 0.00064 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation structure = exchangeable   
## Scale is fixed.  
##   
## Link = identity   
##   
## Estimated Correlation Parameters:  
## Estimate Std.err  
## alpha 0.0425 0.00115  
## Number of clusters: 132115 Maximum cluster size: 2

summary(gee\_mod\_DID\_ER\_qtr)

##   
## Call:  
## geeglm(formula = ER ~ marsh \* year\_center, family = gaussian,   
## data = did\_visit\_types\_qtr, id = UniqueIdentifier, corstr = "exchangeable",   
## scale.fix = T)  
##   
## Coefficients:  
## Estimate Std.err Wald Pr(>|W|)   
## (Intercept) 0.17617 0.00281 3936.18 < 2e-16 \*\*\*  
## marsh 0.05865 0.02712 4.68 0.031 \*   
## year\_center -0.01149 0.00115 100.73 < 2e-16 \*\*\*  
## marsh:year\_center -0.04075 0.00952 18.31 1.9e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation structure = exchangeable   
## Scale is fixed.  
##   
## Link = identity   
##   
## Estimated Correlation Parameters:  
## Estimate Std.err  
## alpha -0.153 0.00694  
## Number of clusters: 50601 Maximum cluster size: 2

summary(gee\_mod\_DID\_ER\_qtr\_bal)

##   
## Call:  
## geeglm(formula = ER ~ marsh \* year\_center, family = gaussian,   
## data = did\_visit\_types\_qtr\_balanced, id = UniqueIdentifier,   
## corstr = "exchangeable", scale.fix = T)  
##   
## Coefficients:  
## Estimate Std.err Wald Pr(>|W|)   
## (Intercept) 0.16970 0.00336 2551.94 < 2e-16 \*\*\*  
## marsh 0.06107 0.03390 3.24 0.072 .   
## year\_center -0.00694 0.00150 21.32 3.9e-06 \*\*\*  
## marsh:year\_center -0.03194 0.01377 5.38 0.020 \*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation structure = exchangeable   
## Scale is fixed.  
##   
## Link = identity   
##   
## Estimated Correlation Parameters:  
## Estimate Std.err  
## alpha -0.198 0.00596  
## Number of clusters: 26851 Maximum cluster size: 2

summary(gee\_mod\_DID\_PCP\_yr)

##   
## Call:  
## geeglm(formula = PCP ~ marsh \* year\_center, family = gaussian,   
## data = did\_visit\_types\_year, id = UniqueIdentifier, corstr = "exchangeable",   
## scale.fix = T)  
##   
## Coefficients:  
## Estimate Std.err Wald Pr(>|W|)   
## (Intercept) 0.084075 0.000896 8813.17 < 2e-16 \*\*\*  
## marsh -0.010641 0.007963 1.79 0.18   
## year\_center 0.015998 0.000453 1247.24 < 2e-16 \*\*\*  
## marsh:year\_center -0.017449 0.003240 29.01 7.2e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation structure = exchangeable   
## Scale is fixed.  
##   
## Link = identity   
##   
## Estimated Correlation Parameters:  
## Estimate Std.err  
## alpha -0.0172 0.00149  
## Number of clusters: 186070 Maximum cluster size: 2

summary(gee\_mod\_DID\_PCP\_yr\_bal)

##   
## Call:  
## geeglm(formula = PCP ~ marsh \* year\_center, family = gaussian,   
## data = did\_visit\_types\_year\_balanced, id = UniqueIdentifier,   
## corstr = "exchangeable", scale.fix = T)  
##   
## Coefficients:  
## Estimate Std.err Wald Pr(>|W|)   
## (Intercept) 0.092133 0.001019 8179.7 < 2e-16 \*\*\*  
## marsh 0.012383 0.011320 1.2 0.27   
## year\_center 0.012858 0.000535 577.2 < 2e-16 \*\*\*  
## marsh:year\_center -0.024194 0.004713 26.4 2.8e-07 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation structure = exchangeable   
## Scale is fixed.  
##   
## Link = identity   
##   
## Estimated Correlation Parameters:  
## Estimate Std.err  
## alpha 0.00487 0.00153  
## Number of clusters: 132115 Maximum cluster size: 2

summary(gee\_mod\_DID\_PCP\_qtr)

##   
## Call:  
## geeglm(formula = PCP ~ marsh \* year\_center, family = gaussian,   
## data = did\_visit\_types\_qtr, id = UniqueIdentifier, corstr = "exchangeable",   
## scale.fix = T)  
##   
## Coefficients:  
## Estimate Std.err Wald Pr(>|W|)   
## (Intercept) 0.43243 0.00367 13855.33 < 2e-16 \*\*\*  
## marsh -0.07211 0.03077 5.49 0.01910 \*   
## year\_center -0.00590 0.00154 14.68 0.00013 \*\*\*  
## marsh:year\_center -0.04505 0.01121 16.16 5.8e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation structure = exchangeable   
## Scale is fixed.  
##   
## Link = identity   
##   
## Estimated Correlation Parameters:  
## Estimate Std.err  
## alpha -1.02 0.00147  
## Number of clusters: 50601 Maximum cluster size: 2

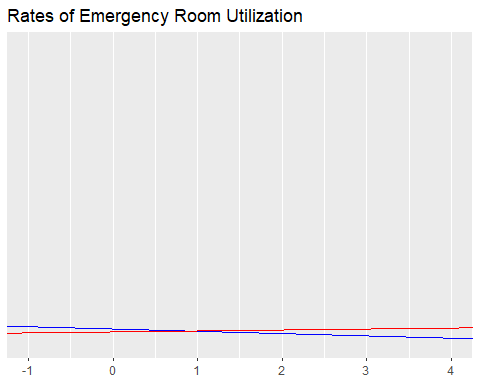
summary(gee\_mod\_DID\_PCP\_qtr\_bal)

##   
## Call:  
## geeglm(formula = PCP ~ marsh \* year\_center, family = gaussian,   
## data = did\_visit\_types\_qtr\_balanced, id = UniqueIdentifier,   
## corstr = "exchangeable", scale.fix = T)  
##   
## Coefficients:  
## Estimate Std.err Wald Pr(>|W|)   
## (Intercept) 0.40601 0.00439 8542.11 < 2e-16 \*\*\*  
## marsh -0.00217 0.03953 0.00 0.96   
## year\_center 0.00244 0.00202 1.47 0.23   
## marsh:year\_center -0.07184 0.01645 19.08 1.3e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Correlation structure = exchangeable   
## Scale is fixed.  
##   
## Link = identity   
##   
## Estimated Correlation Parameters:  
## Estimate Std.err  
## alpha -0.719 0.00347  
## Number of clusters: 26851 Maximum cluster size: 2

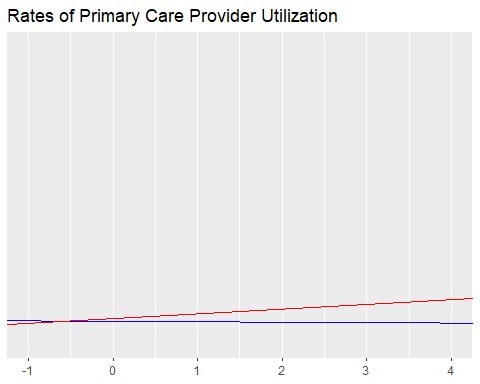
#knitr::kable(twobytwo, caption = "Contingency table of whether patient achieved complete remission by treatment group.")

# Some Visuals that need work

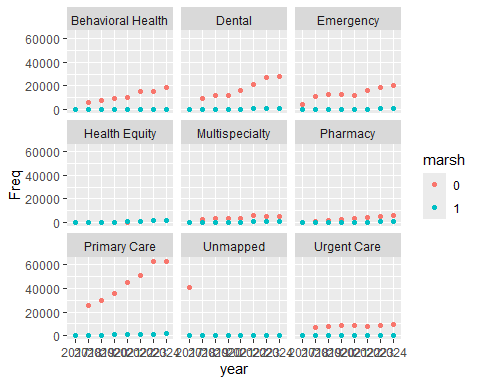
# Visuals   
  
#ER   
#colnames(did\_visit\_types\_year)  
  
# no! takes forever to load!   
#   
# did\_visit\_types\_year %>%   
# ggplot(aes(x = Date, y = ER, col = marsh))+  
# geom\_jitter()+  
  
  
ggplot()+  
 geom\_abline( intercept =(gee\_mod\_DID\_ER\_yr$coefficients[1] + gee\_mod\_DID\_ER\_yr$coefficients[2]),   
 slope= (gee\_mod\_DID\_ER\_yr$coefficients[3] + gee\_mod\_DID\_ER\_yr$coefficients[4]), col = "blue")+ # line for Marshallese  
 geom\_abline( intercept = (gee\_mod\_DID\_ER\_yr$coefficients[1]),   
 slope= gee\_mod\_DID\_ER\_yr$coefficients[3], col = "red")+ # line for Non-Hispanic White  
 xlim(c(-1,4))+  
 labs(title= "Rates of Emergency Room Utilization")



ggplot()+  
 geom\_abline( intercept =(gee\_mod\_DID\_PCP\_yr$coefficients[1] + gee\_mod\_DID\_PCP\_yr$coefficients[2]),   
 slope= (gee\_mod\_DID\_PCP\_yr$coefficients[3] + gee\_mod\_DID\_PCP\_yr$coefficients[4]), col = "blue")+ # line for Marshallese  
 geom\_abline( intercept = (gee\_mod\_DID\_PCP\_yr$coefficients[1]),   
 slope= gee\_mod\_DID\_PCP\_yr$coefficients[3], col = "red")+ # line for Non-Hispanic White  
 xlim(c(-1,4))+  
 labs(title= "Rates of Primary Care Provider Utilization")



# graph number of all visits?   
#table\_all\_visit\_types\_wider <- read.csv("~/BIOST CLASSES/597 Capstone with Lloyd Mancl/597 Capstone/CHAS-capstone/Analysis Data/table\_all\_visit\_types\_wider.csv")  
  
  
table\_all\_visit\_types <- as.data.frame( with(all\_visit\_types, table(ServiceLine,year, marsh)) ) #, useNA = "always"  
  
  
# this is a good visual idea but we need rates instead of total counts to show any Marshallese on the scale!   
table\_all\_visit\_types %>% filter( year != 2025, year != 2016) %>%  
 ggplot(aes(x = year, y = Freq, col = marsh))+  
 geom\_point()+  
 facet\_wrap(~ServiceLine)



# add a column with total population size for each year, then mutate to get the rate?   
# can adjust code from other DIDVisuals.R file