Final Project Soc 723 due April 2022

Brenda Onyango

4/1/2022

## Background

### Sociological Background on Breastfeeding

The individual decision to breastfeed an infant happens in the context of longstanding social forces including racial stereotypes and corporate advertising for formula with a global reach (Kirksey, 2021; Kaplan, 2008). In the United States, the scope of my analysis, breastfeeding rates are one of many maternal and child health disparities that fall along racial lines. In addition to the sociohistorical motivators/deterrents to breastfeeding, there are also concerns for economic well-being for the breastfeeding parent. Rippeyoung and Noonan document how breastfeeding rather than formula-feeding affects post-birth earnings (2012). Using data from the National Longitudinal Survey of Youth, they find that those who breastfeed for six months or longer experience longer wage loss than those who breastfeed for a shorter period or not at all (Rippeyoung and Noonan, 2012). Indeed, *both* the ability to breastfeed and the decision for breastfeeding duration has economic components, as well as biological. The dynamic between social forces and biological processes are also applicable to caries and other oral health outcomes across the lifecourse.

### Breastfeeding and Child Oral Health

A 2017 review summarized the extensive health benefits of breast milk that includes preventing teeth from being misaligned (Peres et al., 2017). The review also cites studies showing breastfeeding up to 12 months has been associated with fewer caries, but there is a positive association with caries when breastfeeding lasts longer (Peres et al., 2017). The authors of the review acknowledge methodological challenges in connecting breastfeeding behavior with oral health outcomes beyond infants aged 12-24 months. This present analysis attempts to understand the causal relationship between breastfeeding and caries. “Caries is a chronic disease process that… is an imbalance between destructive and oral forces…cavities are the holes in teeth that result from the underlying chronic caries process” (Centers for Medicare & Medicaid Services, 2015, p. 8-9).

### Current Question

The research question is does breastfeeding cause caries? Breastfeeding will be operationalized using definitions from the U.S. Centers for Disease Control and Prevention which collects several questions about breastfeeding via the National Immunization Survey landline sampling frame (U.S. CDC, n.d.). The indicators used in this study are “Infants breastfed at 12 months” and “Infants ever breastfed” (CDC2) from 2008; data from before 2009 in this set “represents children born over three years and each birth year can consist of respondent data from up to three years” (CDC2). These indicators are the independent variables. The dependent variable of child dental health is operationalized with data from the National Oral Health Surveillance System for third graders in 2016, who would be ~eight years old (CDC3). The indicator is “Caries Experience: Percentage of students with caries experience (treated or untreated tooth decay).”

The control variables in this analysis are fluoridation and child poverty rates from CDC1 and the National Women’s Law Center.

### Causal Diagram

Childhood caries formation and breastfeeding behavior have causes not included in the dataset used in this analysis but is captured in the diagram below. These determinants include:

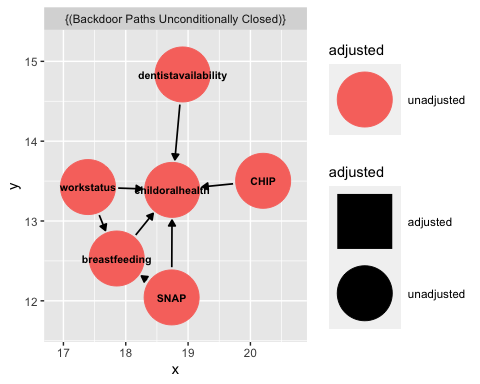
Supplemental Nutrition Assistance Program (SNAP): Colloquially known as food stamps, SNAP likely affects the physical ability to breastfeed and child oral health because people who are lactating require sufficient nutrition to produce milk and children’s diets is also a cause of caries.

Work status: The lactating parent’s work status (type of work, wages, and work hours) can be reasonably assumed to affect their ability to breastfeed and the ability to afford dental care for ones children.

Children’s Health Insurance Program (CHIP): CHIP provides more affordable health coverage for children in families that do not qualify for Medicaid. States vary in CHIP structure and eligibility, but all CHIP plans cover dental (HHS). Having dental coverage provides children the possibility of preventive dental services.

Dentist Availability: The availability of dental experts in an area can affect parents’ education on how to take care of infant and young children’s oral health and the ability to access preventive services.

dentaldag <- dagitty::dagitty("dag {  
   
 childoralhealth <- breastfeeding   
 childoralhealth <- SNAP -> breastfeeding  
 childoralhealth <- workstatus -> breastfeeding  
 childoralhealth <- CHIP  
 childoralhealth <- dentistavailability  
   
   
 SNAP [exposure]  
 childoralhealth [outcome]   
}")  
  
tidy\_dagdrive <- tidy\_dagitty(dentaldag) #true exposure or predictor of interest is breastfeeding  
  
ggdag\_adjustment\_set(dentaldag, node\_size = 19, text\_size = 2.8, text\_col = 'black')



## Methods

The analysis is linear regression of “caries experience…” on “infants ever breastfed” and “caries experience…” on “infants breastfed at 12 months” controlled for fluoridation and child poverty rates. The most complete data for caries experience came from third graders in 2016 in eight states: Colorado, Connecticut, Florida, Georgia, Idaho, Louisiana, Oregon, and Vermont (CDC3). This data can be reasonably assumed to correspond to the 2008 breastfeeding data (when the third graders would have been born or around 1 year old).

### Data

To conduct this analysis I combined data on community water fluoridation (CDC1), breastfeeding (CDC2), child dental health (CDC3), and child poverty (National Women’s Law Center).

First I installed and loaded packages needed to import data.

Then I imported the datasets for the independent and dependent variables .

bfsixmonths <- read.csv("/Users/brendaonyango/Downloads//ExportCSV (6).csv") #breastfedthroughsixmonths  
  
bfever <- read.csv("/Users/brendaonyango/Downloads//ExportCSV (5).csv") #ever breastfed   
  
bfandformulasixmonths <- read.csv("/Users/brendaonyango/Downloads//ExportCSV (7).csv") #breastfed and formula before 6 months   
  
bftwelvemonths <- read.csv("/Users/brendaonyango/Downloads//ExportCSV (8).csv") #breastfed 12 months

childoralhealth <- read.csv("/Users/brendaonyango/Downloads//NOHSS\_Child\_Indicators (1).csv") #dependentvariables

#### Data Wrangling Indepedent and Dependent Variables

#### Dependent Variable

After importing the datasets, I need to clean and combine them.

I started with wrangling the childoralhealth set.

childoralhealth <- childoralhealth |> arrange(LocationAbbr) |> group\_by(LocationAbbr, SchoolYearStart, Grade)

childoralhealth2 <- childoralhealth |> filter(SchoolYearStart %in% c("2016")) |> group\_by(LocationAbbr, SchoolYearStart)

childoralhealthcleaner <- childoralhealth2 |> filter(Indicator == "Caries Experience: Percentage of students with caries experience (treated or untreated tooth decay)") |> filter(Grade == "Third Grade")

childoralhealthcleaner <- childoralhealthcleaner |> select(LocationDesc, SchoolYearStart, SchoolYearEnd, Grade, Indicator, Data\_Value, Low\_Confidence\_Interval, High\_Confidence\_Interval)

## Adding missing grouping variables: `LocationAbbr`

names(childoralhealthcleaner)[names(childoralhealthcleaner) == 'Data\_Value'] <- 'Percentevercav' #switching name of datavalue to percent

childoralhealthcleaner <- childoralhealthcleaner[-(6),] #removing IHS from data

I had the most extensive data for third graders who started in school in 2016; these students were approximately 8 years old meaning they were born around 2008. Therefore I used breastfeeding data from 2008. I’ll exclude the Indian Health Service and focus on 8 states.

#### Independent Variable

Next it’s time to clean the independent variable. After some consideration I decided to focus on using breastfeed at twelve months because this data point better temporally connects breastfeeding to the cavity data. I will also use infantseverbreastfed to compare extremes between longest recorded length of breastfeeding from this dataset and never breastfeeding.

bftwelvemonthsclean <- bftwelvemonths |> filter(LocationAbbr == "FL"| LocationAbbr == "CO" | LocationAbbr == "CT" | LocationAbbr == "GA" | LocationAbbr == "ID" | LocationAbbr == "LA" | LocationAbbr == "OR" | LocationAbbr == "VT") #selecting same states from depedendent variable

names(bftwelvemonthsclean)[names(bftwelvemonthsclean) == 'Data\_Value'] <- 'PercentBFdozenmonths' #switching name of datavalue to percent bf at 12 months

bfeverclean <- bfeverclean |> select(YearStart, LocationAbbr, LocationDesc, Data\_Value, Low\_Confidence\_Limit, High\_Confidence\_Limit)

names(bfeverclean)[names(bfeverclean) == 'Data\_Value'] <- 'percenteverbreastfed' #switching name of datavalue to percent ever bf

Next I joined the breastfed at 12 months and ever breastfed sets.

bfdata <- right\_join(bftwelvemonthsclean, bfeverclean, by = "LocationAbbr")

Next I combined the datasets with independent and dependent variables.

combinedset <- left\_join(bfdata, childoralhealthcleaner, by = "LocationAbbr")

combinedset <- combinedset |> select(-Indicator)

#### Control Variables

Because it was readily available I decided to include state fluoridation and childhood poverty levels for 2016 to use as control variables to better isolate the effects of breastfeeding on cavity rates.

library(readxl)  
state\_fluoride <- read\_excel("state fluoride.xlsx",   
 col\_types = c("text", "numeric", "numeric"))

combinedset <- left\_join(combinedset, state\_fluoride, by = "LocationAbbr")

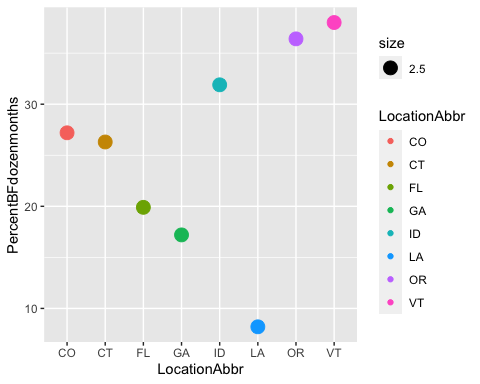
### Data Visualization

Before I can manipulate and visualize data I need to confirm that my combinedset object is a dataframe and not a list.

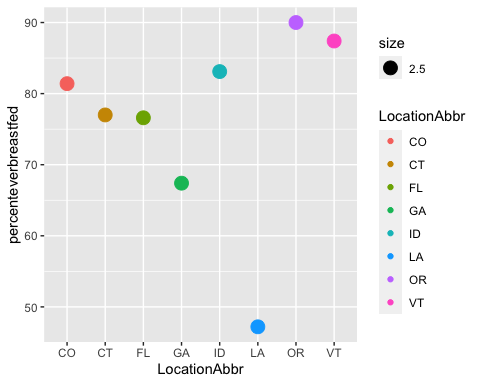
combinedsetfinal <- as.data.frame(combinedset) #ensures object I created in wrangling is a dataframe

#### Independent Variables

ggplot(combinedsetfinal) + geom\_point(aes(x= LocationAbbr, y=PercentBFdozenmonths, color = LocationAbbr, size = 2.5))



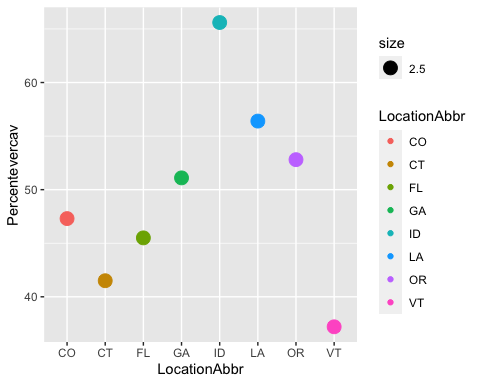
ggplot(combinedsetfinal) + geom\_point(aes(x= LocationAbbr, y=percenteverbreastfed, color =LocationAbbr, size = 2.5))



From the above plots we see variation in percent of children being breastfed at 12 months and ever being breastfed in 2008, respectively. Oregon and Vermont are among the highest a,ong both independent variables and Lousiiana is the least.

#### Dependent Variable

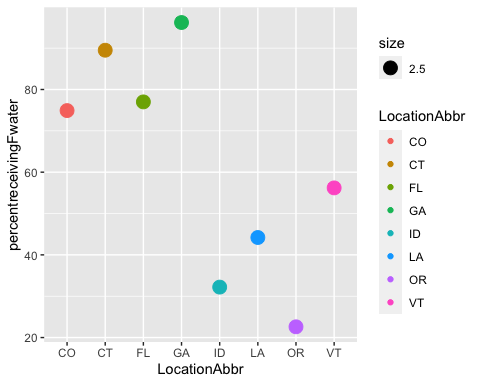
ggplot(combinedsetfinal) + geom\_point(aes(x= LocationAbbr, y=Percentevercav, color =LocationAbbr, size = 2.5))



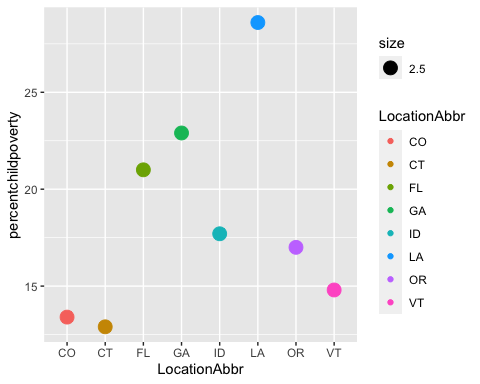
From this visualization we see variation in percent ever having caries (treated or untreated) among third graders in 2016. Idaho has the highest and Vermont the lowest.

### Control Variables

ggplot(combinedsetfinal) + geom\_point(aes(x= LocationAbbr, y=percentreceivingFwater, color =LocationAbbr, size = 2.5))



ggplot(combinedsetfinal) + geom\_point(aes(x= LocationAbbr, y=percentchildpoverty, color =LocationAbbr, size = 2.5))



Louisiana had the highest child (individuals under age 18) poverty in 2016 (28.6%). Oregon had the lowest level of the percent of the population being served by a community water system receiving fluoridated water.

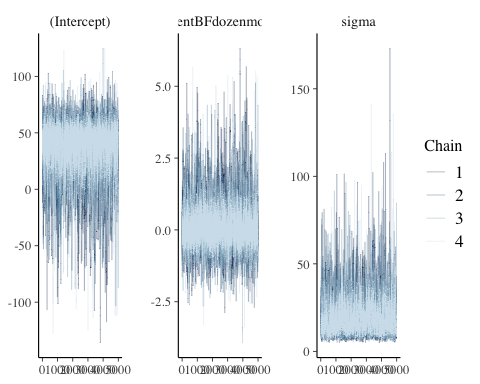
These data visualizations show variation across the independent, dependent, and control variables and a sense of values to use in a simulation.

### Model Simulation

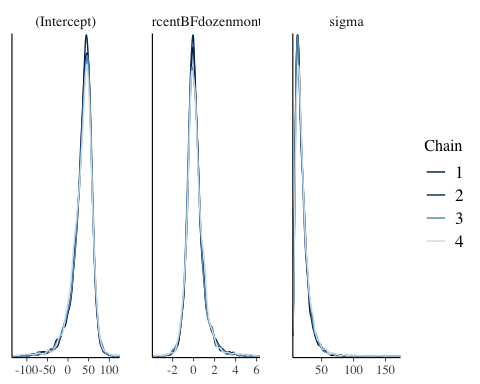
To inform prior assumptions for the model, I found the summary statistics for the complete set of breastfeeding data.

#simulating posterior   
  
cav\_model <- stan\_glm(Percentevercav ~ PercentBFdozenmonths, data = combinedsetfinal,  
 family = gaussian,  
 prior\_intercept = normal(25, 7), #from summary statistics I see consistent bf at 12 months is around 20-25 and used range rule for sd estimate   
 prior = normal(5, 3), #assuming that every increase in breastfeeding results in breastfeeding increasing by 5 percentage points with a standard deviation of 3  
 prior\_aux = exponential(0.0008),  
 chains = 4, iter = 5000\*2, seed = 100)

mcmc\_trace(cav\_model, size = 0.1) #trace plots of parallel MCMC chains



mcmc\_dens\_overlay(cav\_model) #densityplot



#diagnostics for first simulation  
  
rhat(cav\_model)

## (Intercept) PercentBFdozenmonths sigma   
## 1.001166 1.000818 1.002203

neff\_ratio(cav\_model)

## (Intercept) PercentBFdozenmonths sigma   
## 0.23520 0.26735 0.19515

The chains in this simulation were fast-mixing and look stable. The R-hat is approximately 1 which supports chain stability.

#posterior summary statistics   
  
tidy(cav\_model, effects = c("fixed", "aux"),  
 conf.int = TRUE, conf.level = 0.80)

## # A tibble: 4 × 5  
## term estimate std.error conf.low conf.high  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 40.1 18.3 5.97 59.4   
## 2 PercentBFdozenmonths 0.0269 0.609 -0.709 1.09  
## 3 sigma 16.3 10.6 9.33 31.9   
## 4 mean\_PPD 40.6 10.2 26.1 49.8

Based on this model, the posterior median relationship is 40.100 + 0.026X which can be interpreted as that for every one percentage point increase in breastfeeding at 12 months, cavities increase by 0.026 percentage points. Note the confidence interval for the coefficient includes 0.

cav\_model\_2 <- stan\_glm(Percentevercav ~ percenteverbreastfed, data = combinedsetfinal, #doing second simulation for second independent variable  
 family = gaussian,  
 prior\_intercept = normal(25, 7), #from summary statistics I see consistent bf at 12 months is around 20-25 and used range rule for sd estimate   
 prior = normal(5, 3), #assuming that every increase in breastfeeding results in breastfeeding increasing by 5 percentage points with a standard deviation of 3  
 prior\_aux = exponential(0.0008),  
 chains = 4, iter = 5000\*2, seed = 100)

#diagnostics for second simulation   
rhat(cav\_model\_2)

## (Intercept) percenteverbreastfed sigma   
## 1.000265 1.000246 1.000478

neff\_ratio(cav\_model\_2)

## (Intercept) percenteverbreastfed sigma   
## 0.24230 0.25905 0.19685

#posterior summary statistics for second model   
tidy(cav\_model\_2, effects = c("fixed", "aux"),  
 conf.int = TRUE, conf.level = 0.80)

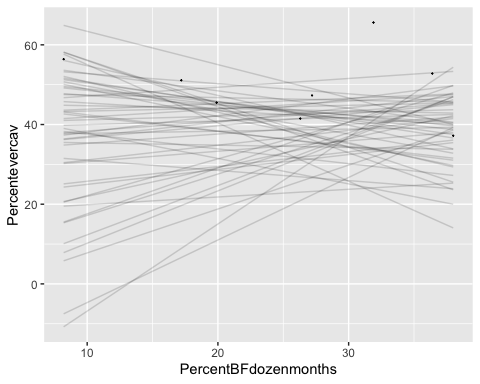
## # A tibble: 4 × 5  
## term estimate std.error conf.low conf.high  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 (Intercept) 44.9 34.4 -15.6 86.4   
## 2 percenteverbreastfed -0.0491 0.431 -0.606 0.670  
## 3 sigma 15.9 9.92 9.30 30.7   
## 4 mean\_PPD 40.9 10.0 26.6 50.0

Based on this model, the posterior median relationship is 44.91 - 0.0491X which can be interpreted as that for every one percentage point increase in ever being breastfed, cavities decrease by 0.0491 percentage points. Note the confidence intervals include 0 for the intercept and coefficient.

#### Simulated model lines

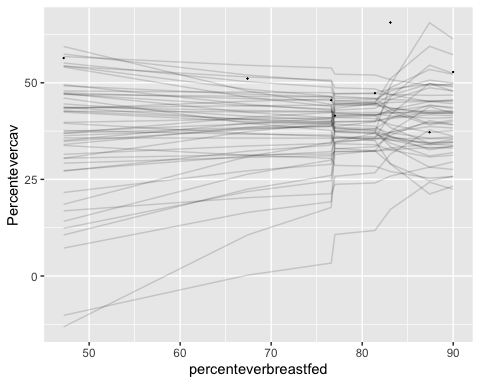
combinedsetfinal |>   
 add\_fitted\_draws(cav\_model, n = 50) %>%  
 ggplot(aes(x = PercentBFdozenmonths, y = Percentevercav)) +  
 geom\_line(aes(y = .value, group = .draw), alpha = 0.15) +   
 geom\_point(data = combinedsetfinal, size = 0.05)

## Warning: `fitted\_draws` and `add\_fitted\_draws` are deprecated as their names were confusing.  
## Use [add\_]epred\_draws() to get the expectation of the posterior predictive.  
## Use [add\_]linpred\_draws() to get the distribution of the linear predictor.  
## For example, you used [add\_]fitted\_draws(..., scale = "response"), which  
## means you most likely want [add\_]epred\_draws(...).



combinedsetfinal |>   
 add\_fitted\_draws(cav\_model, n = 50) %>%  
 ggplot(aes(x = percenteverbreastfed, y = Percentevercav)) +  
 geom\_line(aes(y = .value, group = .draw), alpha = 0.15) +   
 geom\_point(data = combinedsetfinal, size = 0.05)

## Warning: `fitted\_draws` and `add\_fitted\_draws` are deprecated as their names were confusing.  
## Use [add\_]epred\_draws() to get the expectation of the posterior predictive.  
## Use [add\_]linpred\_draws() to get the distribution of the linear predictor.  
## For example, you used [add\_]fitted\_draws(..., scale = "response"), which  
## means you most likely want [add\_]epred\_draws(...).



The simulation indicates that data from 8 states as percentage points may not provide enough variation as is ideal for model simulation.

## Results

## Discussion

### Limitations

% white population; dental workforce availability; children diet; whether propensity to breastfeed correlates to other health behaviors like diet or if there’s a biological pathway (e.g. sugar and acid content of milk v formula v early baby food )

## References

Centers for Disease Control and Prevention1. Community Water Fluoridation: 2016 Flouridation Statistics [online]. [accessed April 07, 2022]. URL: <https://www.cdc.gov/fluoridation/statistics/2016stats.htm>

Centers for Disease Control and Prevention2. National Center for Chronic Disease Prevention and Health Promotion, Division of Nutrition, Physical Activity, and Obesity. Data, Trend and Maps Category: Breastfeeding [online]. [accessed Apr 03, 2022]. URL: <https://www.cdc.gov/nccdphp/dnpao/data-trends-maps/index.html>.

Centers for Diseease Control and Prevention National Oral Health Surveillance System (NOHSS)3. NOHSS Child Indicators Oral Health. [access Apr 03, 2022]. URL: <https://chronicdata.cdc.gov/Oral-Health/NOHSS-Child-Indicators/qcai-zfj9>

Centers for Medicare & Medicaid Services.(2015). Reducing Early Childhood Tooth Decay: Approaches in Medicaid. [online PowerPoint]. [accessed April 08, 2022]. URL: <https://www.medicaid.gov/sites/default/files/2019-12/learninglabslides12.pdf>

Kaplan, D. L., & Graff, K. M. (2008). Marketing breastfeeding—reversing corporate influence on infant feeding practices. Journal of Urban Health, 85(4), 486-504.

Kirksey, K. (2021). A social history of racial disparities in breastfeeding in the United States. Social Science & Medicine, 289, 114365.

National Women’s Law Center. Poverty Rates State by State, 2016. [accessed April 08, 2022]. URL: <https://nwlc.org/wp-content/uploads/2017/09/Poverty-Rates-State-by-State-2016.pdf>

Rippeyoung, P. L., & Noonan, M. C. (2012). Is breastfeeding truly cost free? Income consequences of breastfeeding for women. American Sociological Review, 77(2), 244-267.

Peres, K. G., Chaffee, B. W., Feldens, C. A., Flores-Mir, C., Moynihan, P., & Rugg-Gunn, A. (2018). Breastfeeding and Oral Health: Evidence and Methodological Challenges. Journal of dental research, 97(3), 251–258. <https://doi.org/10.1177/0022034517738925>

U.S. Department of Health and Human Services (HHS). Medicaid & CHIP: The Children’s Health Insurance Program (CHIP). [accessed April 08, 2022]. URL: <https://www.healthcare.gov/medicaid-chip/childrens-health-insurance-program/>