Investigating the Causal Relationship between Childhood Disabilities and Non-Parental Care

Anton Kireyev

2024-04-24

Table of contents

Objective:

The primary objective of this project is to explore and to attempt to establish causality between a child having a reported disability and their participation in non-parental care or program

arrangements. The project aims to contribute to understanding of the potential factors influencing childcare decisions and access to care for children with disabilities.

Background:

Childcare and program arrangements is a crucial aspect of childhood development and family support systems. Children with disabilities often require additional support and specialized services, yet there is a gap in understanding how disability status influences families participation in non-parental care settings. Establishing causality is critical to informing policy interventions and to developng support mechanisms for families of children with disabilities.

Methodology:

The project will utilize observational data from the **Early Childhood Program Participation (ECPP)** survey from 2019: found HERE. The dataset will contain information on children's disability status, participation in non-parental care or program arrangements, as well as relevant demographic and socio-economic information. Propensity score matching and entropy balancing will be used in this study to address potential confounding variables, minimize selection biases, and to increase internal validity of the results.

Establishing Treatment, Control, and Outcome Variables

"Treatment" Variable: DSBLTY

Control: TTLHHINC + AGE2018 + FHRSWK + HDHEALTH + ZIPLOCL + FFSTGN + ZIP18PO2 + ZIPBLHI2 + FEDUC

Outcome Variables: ANYCAREX, CAREHOURX

Goal: To attempt to establish causality between a child having a reported disability and the child participating in any non parental care or program arrangements.

> **ⓘ** Note
>
> The Variables mentioned above will be described in further detail in the "Data Management" Section below

Directed Acyclic Graph (DAG)

Below is a DAG designed in an attempt to describe the relevant causal relationships relevant to the effect of disabilities on the participation in non-parental care activities.
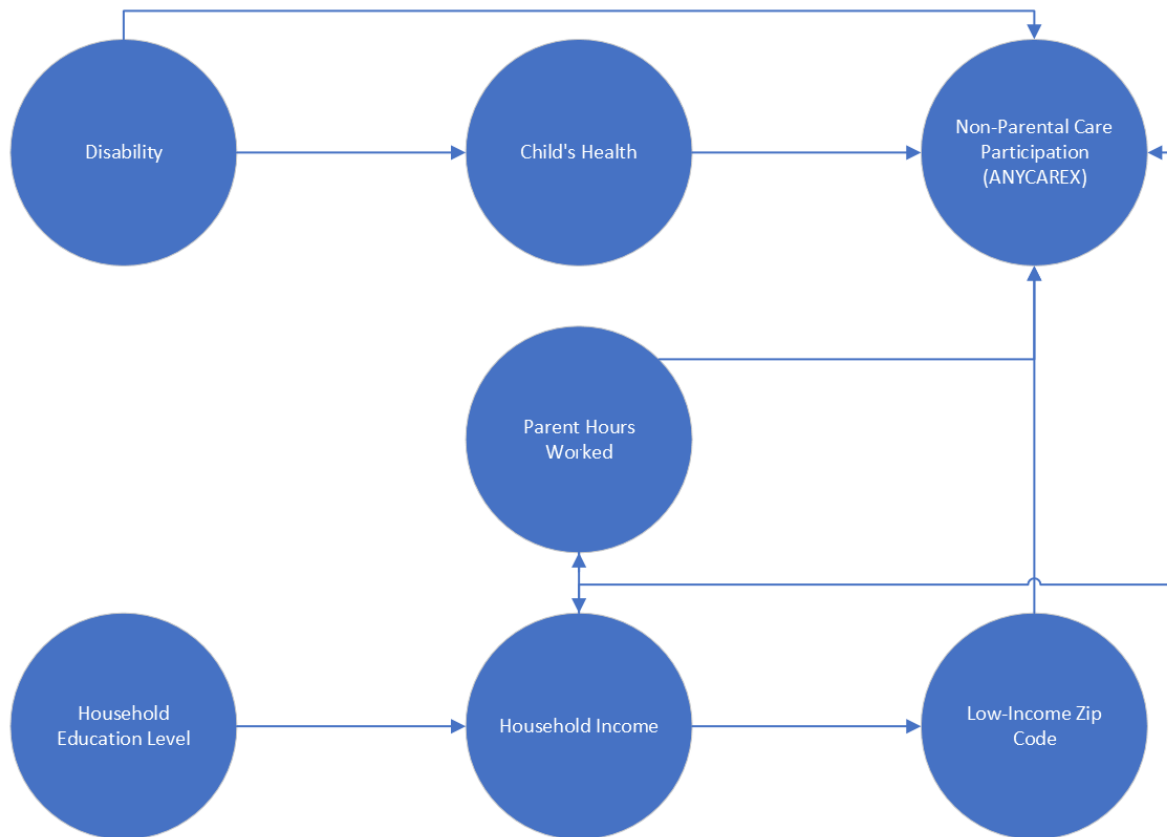

Figure 1: DAG

> **i** Note
>
> Note that Health is a mediator variable between DSBLTY and ANYCAREX, I believe for this to be true since a disability may cause a child's health to be reported differently compared to an instance if they did not have a disability. However, since health can also be independent to disability, I will treat it as a confounder in future analysis.

3

Data Management

```
# Importing Data
rm(list=ls())

NHES = read.csv('ecpp_pu_pert.csv')
```

Due to the large size of the data set and the number of variables present, I have limited the scope of the project to the following variables:

- **General Descriptive:**
  - AGE2018 : Child's Age as of Dec 31, 2018
  - ZIP18PO2 : percent of families in zip code with children under 18 below the poverty line (1 = < 5%, 2 = 5-9%, 3= 10-19%, 4= 20%+)
  - ZIPBLHI2 : Percent of persons in a zip code who were Black or Hispanic (1 = < 6%, 2= 6-15%, 3= 16-40%, 4= 41%+)
  - ZIPLOCL : Zip code classification by location
    * 11 = city large
    * 12 = city midsize
    * 13 = city small
    * 21 = suburb large
    * 22 = suburb midsize
    * 23 = suburb small
    * 31 = town fringe
    * 32= town distant
    * 33 = town remote
    * 41 = rural fringe
    * 42 = rural distant
    * 43 = rural remote
      · *This will be changed to: 1=City, 2=Suburb, 3= Fringe , 4=Distant, 5=Remote*

- **Parent/Guardian Information:**
  - P1EDUC : Highest Grade of school the primary parent completed.
    * 1 = completed up to 8th grade, 2 = HS, no diploma, 3 = HS diploma or GED, 4= vocational Diploma post HS, 5 = Some college, no degree, 6 = AA/AS, 7 = BA/BS, 8 = Graduate/Professional education, no degree, 9 = MA/MS, 10=PhD/EdD, 11 = MD,DDS,JD,LLB
    * *This will be changed to 0 - no diploma, 1 - parent has HS diploma, 2- parent has bachelors degree, 3 - parent has advanced degree.*
  - P2EDUC : Highest grade of school the secondary parent has completed.

- – P1HRSWK : Hours worked per week (Parent 1)
- – P2HRSWK : Hours worked per week (Parent 2)
- – PAR1FSTGN : 1st parent immigrant status (1 = first generation, 2 = 1.5 generation, 3 = neither)
- – PAR2FSTGN : 2nd parent immigrant status (1 = first generation, 2 = 1.5 generation, 3 = neither)

- **Household Information:**

  - – TTLHHINC : Total Household income (CATEGORICAL)
    - ∗ 1 = 0-10,000
    - ∗ 2 = 10,001-20,000
    - ∗ 3 = 20,001-30,000
    - ∗ 4 = 30,001-40,000
    - ∗ 5 = 40,001-50,000
    - ∗ 6 = 50,001-60,000
    - ∗ 7 = 60,001-75,000
    - ∗ 8 = 75,001-100,000
    - ∗ 9 = 100,001-150,000
    - ∗ 10 = 150,001-200,000
    - ∗ 11 = 200,001-250,000
    - ∗ 12 = 250,001+

- **Child Health / Disability:**

  - – HDHEALTH : Health of Child
    - ∗ (1 = Excellent, 2 = Very Good, 3 = Good, 4 = Fair, 5 = Poor)
  - – DSBLTY : Has disability
    - ∗ (1 = Yes, 2 = No)

- **Outcome Variables:**

  - – CAREHOURX : Total hours a week child is in non parental care
  - – ANYCAREX : child participates in any non parental care or program arrangements
    - ∗ (1 = yes, 2 = no) *(Adjusted to 0 = no, 1 = yes)*

By sub-setting I will focus solely on these variables.

```
# Sub-setting Data

variables = c('AGE2018', 'ZIP18PO2', 'ZIPLOCL', 'ZIPBLHI2', 'P1EDUC', 'P2EDUC', 'PAR1FSTGN

NHES_sub = NHES[ ,variables]
```

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max	
AGE2018	7	0	2.2	1.6	0.0	2.0	6.0	
ZIP18PO2	4	0	1.9	0.9	1.0	2.0	4.0	
ZIPLOCL	12	0	22.3	10.3	11.0	21.0	43.0	
ZIPBLHI2	4	0	2.7	1.0	1.0	3.0	4.0	
P1EDUC	11	0	6.3	2.3	1.0	7.0	11.0	
P2EDUC	12	0	4.8	3.5	-1.0	5.0	11.0	
PAR1FSTGN	3	0	2.7	0.7	1.0	3.0	3.0	
PAR2FSTGN	4	0	2.0	1.5	-1.0	3.0	3.0	
P1HRSWK	66	0	29.5	20.2	-1.0	40.0	80.0	
P2HRSWK	68	0	28.8	22.1	-1.0	40.0	80.0	
TTLHHINC	12	0	7.0	3.0	1.0	8.0	12.0	
HDHEALTH	5	0	1.4	0.6	1.0	1.0	5.0	
DSBLTY	2	0	1.9	0.3	1.0	2.0	2.0	
CAREHOURX	108	0	20.2	21.1	0.0	12.0	150.0	
ANYCAREX	2	0	1.3	0.5	1.0	1.0	2.0	

Observing Data

When viewing the data we can see that there is no missing data.

```
datasummary_skim(NHES_sub)
```

> **i** CAREHOURX has a maximum value of 150. This means that a child is under non-parental care for 150/168 hours of the week!

Data Management Steps

1. Adjust Response Variable to be (0,1)
2. Create Family Hours Worked (FHRSWK) Variable by combining p1hrsworked and p2hrsworked
3. Change ZIPLOCL : This will be changed to: 1-City, 2-Suburb, 3- Fringe , 4-Distant, 5-Remote
4. Change P1EDUC : Change into 3 separate dummy variables, 1 - parent has HS diploma, 2- parent has bachelors degree, 3 - parent has advanced degree. Combine with P2EDUC to create FEDUC (Family Education)

	Unique (#)	Missing (%)	Mean	SD	Min	Median	Max	
data	122	0	58.9	27.5	0.0	60.0	160.0	

5. Change PAR1FSTGN : Combine with PAR2FSTGN to create FFSTGN (Family Generational/Immigrant Status)

ANYCAREX

```
# Adjust Response Variable to be between 0-1. Change 2 (False, "No Care" into 0)
NHES_sub$ANYCAREX[NHES_sub$ANYCAREX == 2] = 0

# Adjust Disability Variable to be 0,1
NHES_sub$DSBLTY[NHES_sub$DSBLTY == 2] = 0
```

Family Hours Worked Per Week

```
# edit variables with -1 values (P1HRS,P2HRS per week, ANYCAREX), turn all -1 values to 0.
# Create new hours worked variable, FHRSWK, combining both parents worked hours per week.
NHES_sub$P1HRSWK[NHES_sub$P1HRSWK == -1] = 0
NHES_sub$P2HRSWK[NHES_sub$P2HRSWK == -1] = 0
NHES_sub$FHRSWK  = NHES_sub$P1HRSWK + NHES_sub$P2HRSWK

# Verify Results
datasummary_skim(NHES_sub$FHRSWK)


boxplot(NHES_sub$FHRSWK)
```
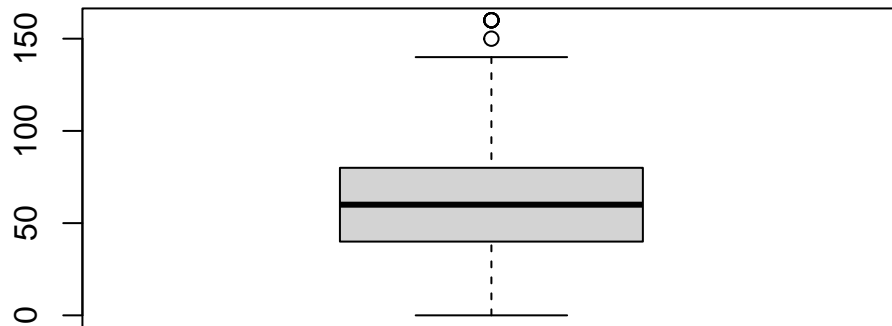
Location Descriptors

```
# Change ZIPLOCL : This will be changed to: 1-City, 2-Suburb, 3-Fringe , 4-Distant, 5-Remo
NHES_sub$ZIPLOCL[NHES_sub$ZIPLOCL == 11 | NHES_sub$ZIPLOCL == 12 | NHES_sub$ZIPLOCL == 13]
NHES_sub$ZIPLOCL[NHES_sub$ZIPLOCL == 21 | NHES_sub$ZIPLOCL == 22 | NHES_sub$ZIPLOCL == 23]
NHES_sub$ZIPLOCL[NHES_sub$ZIPLOCL == 31 | NHES_sub$ZIPLOCL == 41] = 3
NHES_sub$ZIPLOCL[NHES_sub$ZIPLOCL == 32 | NHES_sub$ZIPLOCL == 42] = 4
NHES_sub$ZIPLOCL[NHES_sub$ZIPLOCL == 33 | NHES_sub$ZIPLOCL == 43] = 5

# View Data
ggplot(NHES_sub, aes(x = ZIPLOCL)) +
  labs(title = "Household Location", x='Location Category', y='Number of Households') +
  geom_bar(fill='lightblue')
```

Household Location
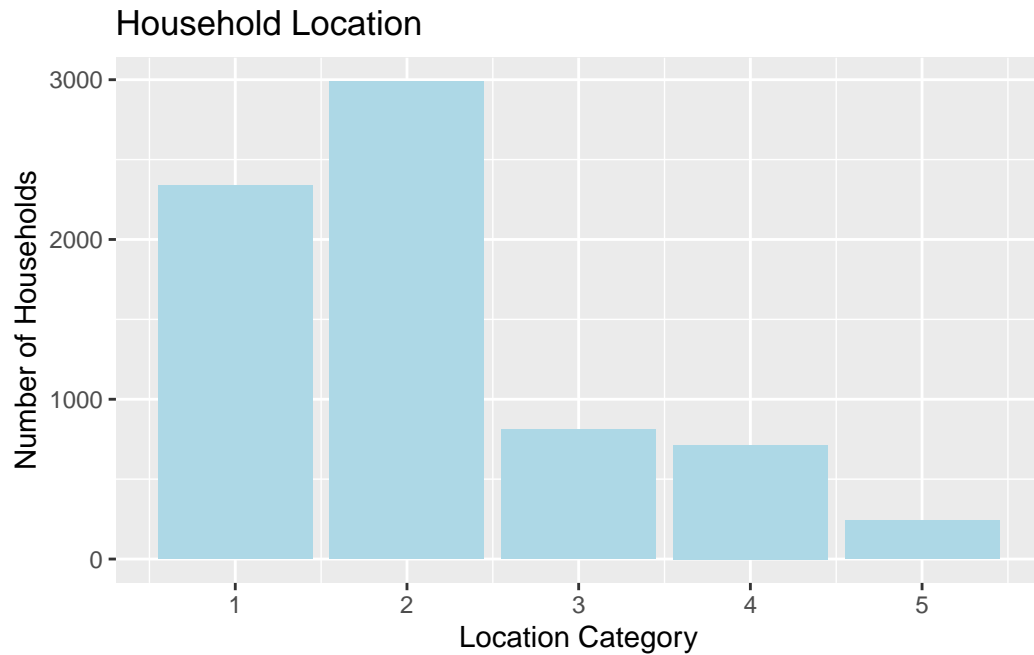
```
# After viewing the data it makes sense to combine categories 3,4,5 into one category: Urb
NHES_sub$ZIPLOCL[NHES_sub$ZIPLOCL == 4 | NHES_sub$ZIPLOCL == 5] = 3

ggplot(NHES_sub, aes(x = ZIPLOCL)) +
  labs(title = "Household Location", x='Location Category', y='Number of Households') +
  geom_bar(fill='lightblue')
```

Household Location
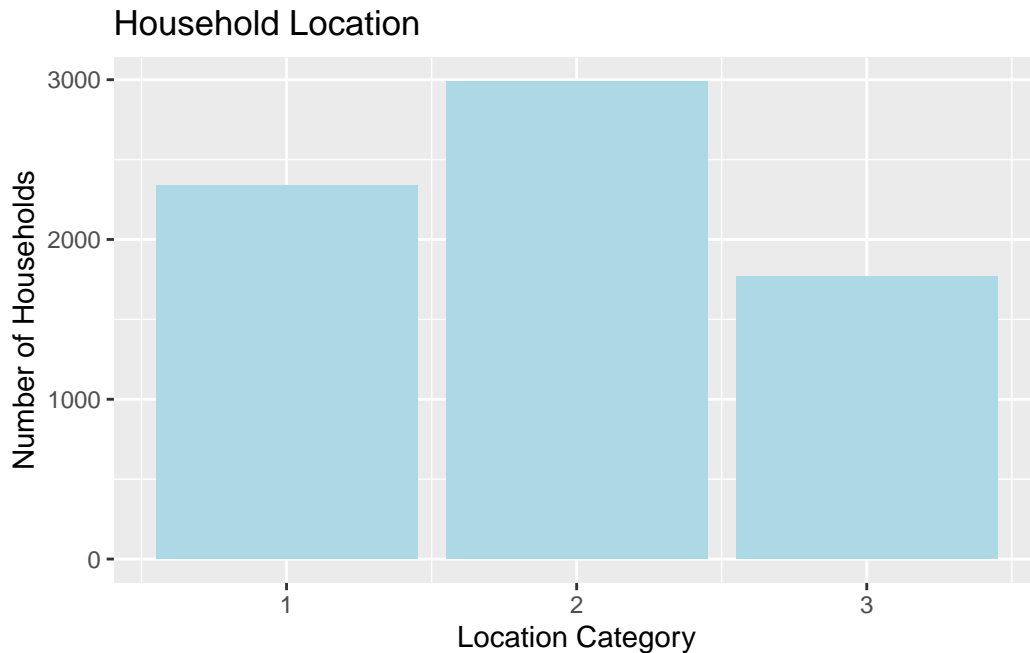
Parent Education Degree

```
# P1EDUC : Change into 4 separate dummy variables, 0 - parent has no degree/diploma, 1 - p

# What the code does: Uses an ifelse() to determine education status, assigns to one of 4

# NOTE: DO NOT RUN THE ifelse() MULTIPLE TIMES, IT WILL OVERWITE ITSELF. Yes, I learned th

NHES_sub$P1EDUC = ifelse(NHES_sub$P1EDUC <= 2, 0, NHES_sub$P1EDUC)
NHES_sub$P1EDUC = ifelse(NHES_sub$P1EDUC <= 6 & NHES_sub$P1EDUC > 2, 1, NHES_sub$P1EDUC)
NHES_sub$P1EDUC = ifelse(NHES_sub$P1EDUC == 7, 2, NHES_sub$P1EDUC)
NHES_sub$P1EDUC = ifelse(NHES_sub$P1EDUC > 7, 3, NHES_sub$P1EDUC)

# P2EDUC : Do the same as P1EDUC

NHES_sub$P2EDUC = ifelse(NHES_sub$P2EDUC <= 2, 0, NHES_sub$P2EDUC)
NHES_sub$P2EDUC = ifelse(NHES_sub$P2EDUC <= 6 & NHES_sub$P2EDUC > 2, 1, NHES_sub$P2EDUC)
NHES_sub$P2EDUC = ifelse(NHES_sub$P2EDUC == 7, 2, NHES_sub$P2EDUC)
NHES_sub$P2EDUC = ifelse(NHES_sub$P2EDUC > 7, 3, NHES_sub$P2EDUC)

# Create Graphs of Data
```
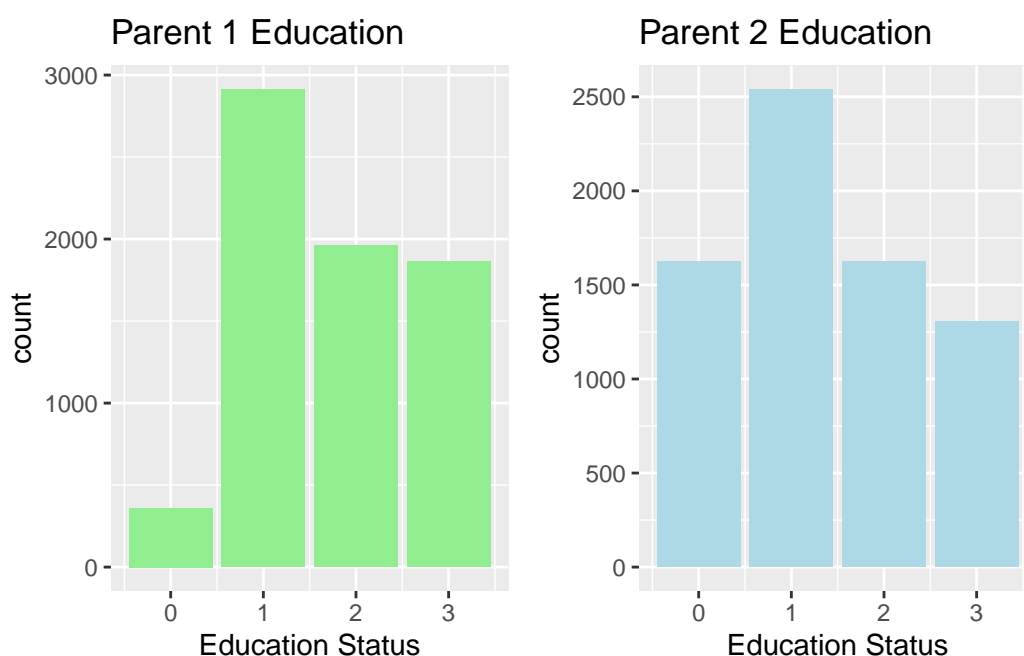
```
p1 = ggplot(NHES_sub, aes(x = P1EDUC)) +
  labs(title = "Parent 1 Education", x='Education Status', y='count') +
  geom_bar(fill="lightgreen") +
  scale_x_continuous(breaks = c(0:3))
p2 = ggplot(NHES_sub, aes(x = P2EDUC)) +
  labs(title = "Parent 2 Education", x='Education Status', y='count') +
  geom_bar(fill="lightblue") +
  scale_x_continuous(breaks = c(0:3))

# View Data
grid.arrange(p1,p2, nrow=1)
```


Now, in order to make work with this variable easier, I will now combine the two variables to create a household education "score".

The score will range from 0-6, with a 0 meaning that neither parent has a high school diploma and a 6 meaning both parents have post-graduate degrees.

```
# Create new variable, FEDUC (Family Education), which combines P1EDUC and P2EDUC
NHES_sub$FEDUC = NHES_sub$P1EDUC + NHES_sub$P2EDUC
```

```
# Create Graph of Data
ggplot(NHES_sub, aes(x = FEDUC)) +
  labs(title = "Family Education", x='Education Status', y='count') +
  geom_bar(fill="lightyellow") +
  scale_x_continuous(breaks = c(0:6))
```


Family Education
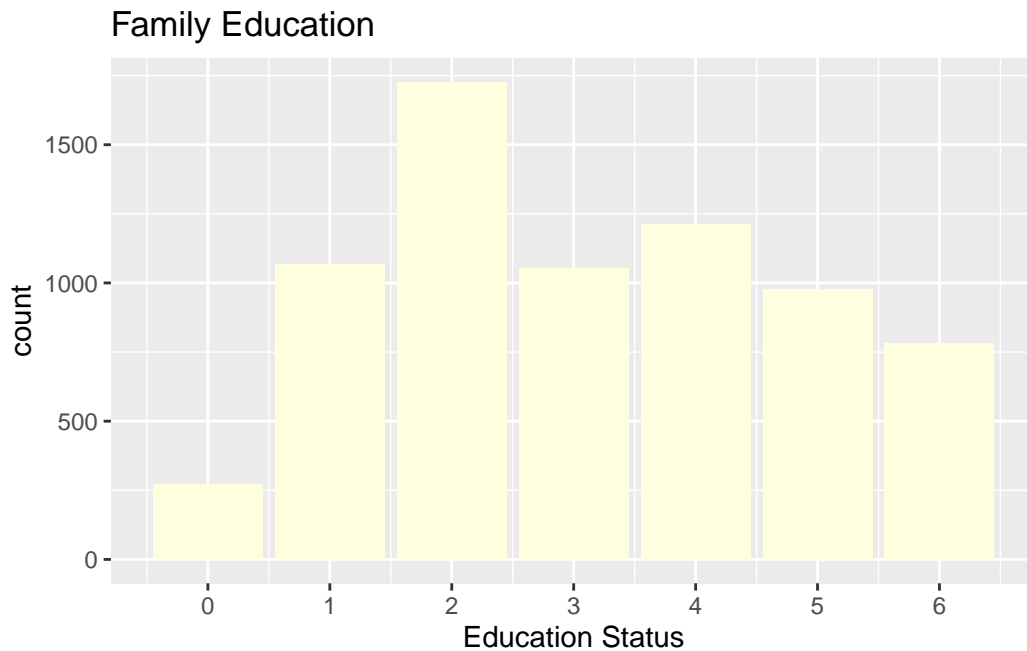
Household Income

Viewing the average household income shows that a majority of the data is around category 8
- $75,001-100,000 income.

> **i** The Annual Household Income is Categorical! Which makes it a non-exact category,
> and only gives us the ability to examine ranges of income.

```
ggplot(NHES_sub, aes(x=TTLHHINC)) +
  geom_bar(fill='lightblue') +
  labs(title = "Total Annual Household Income", x='Total Income Category', y='Number of Ho
  scale_x_continuous(breaks = c(1:12))
```

Total Annual Household Income
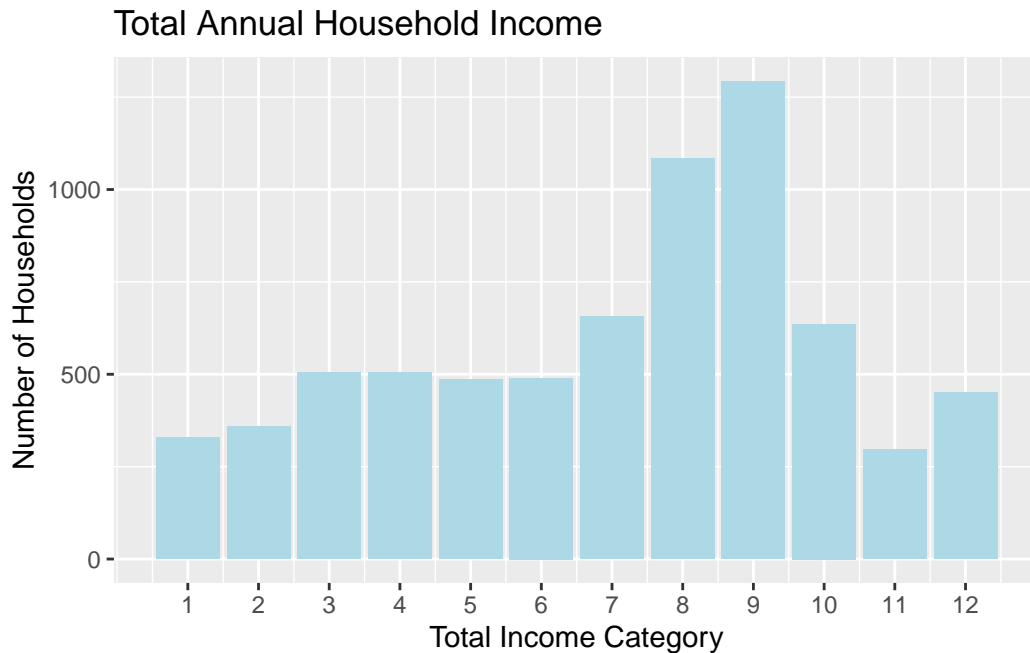
1st Generation Parents

The PAR1FSTGN is a categorical variable with the following categories: 1- first generation, 2- 1.5 generation, 3-neither For ease of use and interpretation purposes I will be turning this variable into a dummy variable to represent 1st generation status. Groups 1 and 2 will be merged to be TRUE, and "neither" will be FALSE.

Since this variable only represents the first parent, I will also use the variable for the second parent and combine the two to create a variable with 3 levels. (0 - neither first generation, 1 - one parent first gen, 2 - both parents first gen)

```
NHES_sub$PAR1FSTGN = ifelse(NHES_sub$PAR1FSTGN < 3, 1, 0)
NHES_sub$PAR2FSTGN = ifelse(NHES_sub$PAR2FSTGN < 3, 1, 0)

NHES_sub$FFSTGN = NHES_sub$PAR1FSTGN + NHES_sub$PAR2FSTGN
table(NHES_sub$PAR1FSTGN)
```

```
   0    1
5730 1362
```

```
table(NHES_sub$PAR2FSTGN)
```

```
    0    1
4612 2480
```

```r
table(NHES_sub$FFSTGN)
```

```
   0    1    2
4342 1658 1092
```

```r
ggplot(NHES_sub, aes(x=FFSTGN)) +
  geom_bar(fill='lightblue') +
  labs(title = "Family Generation", x='Family Generation Status', y='count') +
  scale_x_continuous(breaks = c(0:2))
```
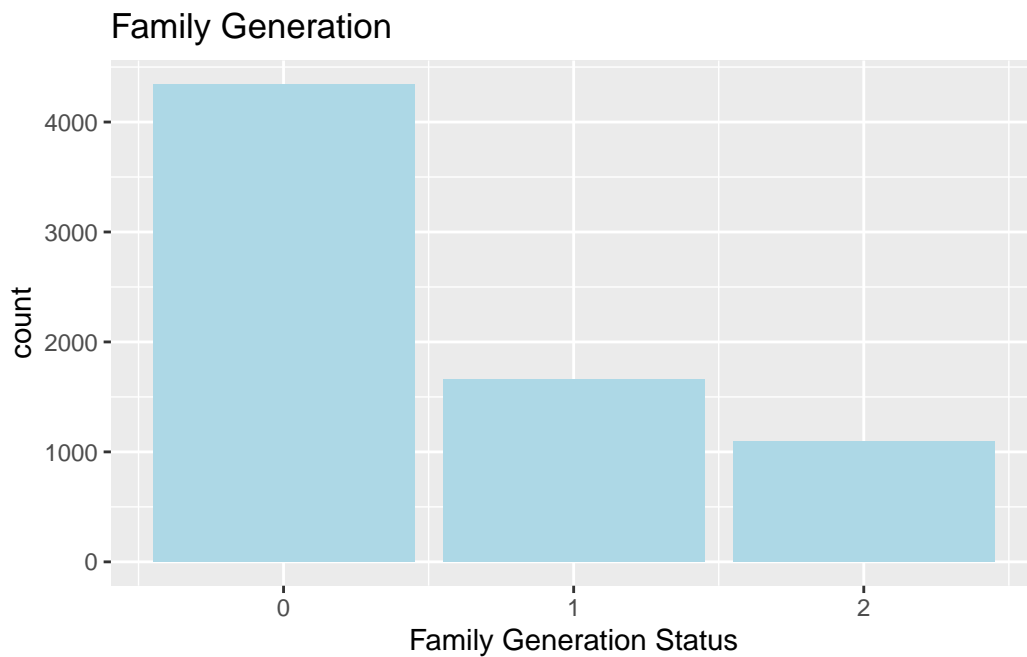

Building Models

```r
# Simple Linear Regressions
# Predicting ANYCAREX on various predictors
simplemodels = list()
```

```r
simplemodels[['Income']] = lm(ANYCAREX ~ TTLHHINC, data = NHES_sub)
simplemodels[['Age']] = lm(ANYCAREX ~ AGE2018, data = NHES_sub)
simplemodels[['Disability']] = lm(ANYCAREX ~ DSBLTY, data = NHES_sub)
simplemodels[['Parent Hours Worked']] = lm(ANYCAREX ~ FHRSWK, data = NHES_sub)
simplemodels[['Child Health']] = lm(ANYCAREX ~ HDHEALTH, data = NHES_sub)
simplemodels[['Location Type']] = lm(ANYCAREX ~ ZIPLOCL, data = NHES_sub)
simplemodels[['1st Generation']] = lm(ANYCAREX ~ FFSTGN, data = NHES_sub)
simplemodels[['% Low Income']] = lm(ANYCAREX ~ ZIP18PO2, data = NHES_sub)
simplemodels[['% Black/Hispanic']] = lm(ANYCAREX ~ ZIPBLHI2, data = NHES_sub)

# Looking at education status, the FEDUC variables somewhat averages the two parent's educ

simplemodels[['Family Education']] = lm(ANYCAREX ~ FEDUC, data = NHES_sub)
simplemodels[['Parent1 Education']] = lm(ANYCAREX ~ P1EDUC, data = NHES_sub)
simplemodels[['Parent2 Education']] = lm(ANYCAREX ~ P2EDUC, data = NHES_sub)

modelsummary(simplemodels, stars=TRUE, title="Simple Linear Models, Predicting ANYCAREX on
```

Propensity Score Matching

```r
# Estimating the propensity score
logistic_model = glm(formula = DSBLTY ~ TTLHHINC + AGE2018 + FHRSWK + HDHEALTH + ZIPLOCL +

# make predictions and store in 'predictions'
predictions = predict(logistic_model, NHES_sub)

# column bind with predictions
NHES_sub$predictions = predictions

# Use matchit to match treatment/control on defined observable co-variates
library(MatchIt)
```

Warning: package 'MatchIt' was built under R version 4.3.3

```r
match_model <- matchit(DSBLTY ~ TTLHHINC + AGE2018 + FHRSWK + HDHEALTH + ZIPLOCL + FFSTGN

# Create a data frame with only matched observations
matched <- match.data(match_model)
```

Table 1: Simple Linear Models, Predicting ANYCAREX on various predictors

	Income	Age	Disability	Parent Hours Worked	Child Health	Location Type
(Intercept)	0.437***	0.503***	0.665***	0.363***	0.613***	0.709***
	(0.014)	(0.009)	(0.006)	(0.012)	(0.013)	(0.015)
TTLHHINC	0.034***					
	(0.002)					
AGE2018		0.079***				
		(0.003)				
DSBLTY			0.104***			
			(0.018)			
FHRSWK				0.005***		
				(0.000)		
HDHEALTH					0.044***	
					(0.009)	
ZIPLOCL						−0.018*
						(0.007)
FFSTGN						
ZIP18PO2						
ZIPBLHI2						
FEDUC						
P1EDUC						
P2EDUC						
Num.Obs.	7092	7092	7092	7092	7092	7092
R2	0.047	0.070	0.005	0.097	0.004	0.001
R2 Adj.	0.047	0.069	0.004	0.097	0.004	0.001
AIC	9024.6	8856.6	9334.8	8642.2	9340.9	9361.9
BIC	9045.2	8877.2	9355.4	8662.8	9361.5	9382.5
Log.Lik.	−4509.290	−4425.276	−4664.404	−4318.124	−4667.469	−4677.952
RMSE	0.46	0.45	0.47	0.44	0.47	0.47

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Common Support

> **i**   Conclusion: The common support is great!

```
# See Common Support
ggplot(matched,aes(x=predictions)) +
  geom_histogram(data=subset(matched,DSBLTY == 1),fill = "black", alpha = 0.7) +
  geom_histogram(data=subset(matched,DSBLTY == 0),fill = "blue", alpha = 0.7) +
  xlab("Probability of Treatment") +
  ggtitle("Common Support") +
  theme_bw()
```

```
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
`stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```


Common Support
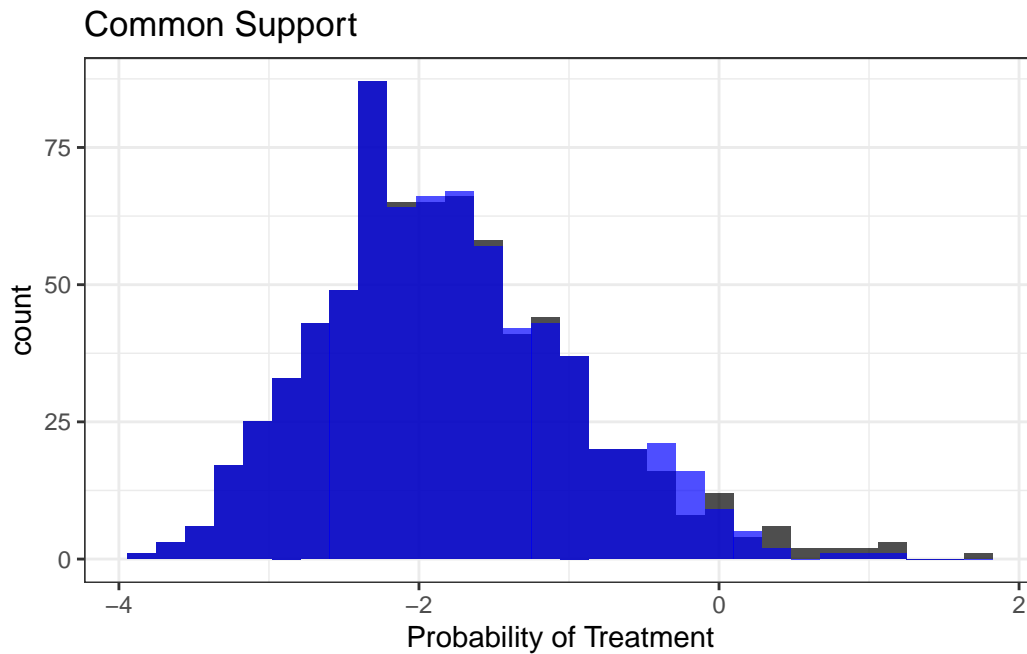
Analyzing Models

Conclusion: By reviewing the differences in treatment and control groups of the PSM models it is visible that the PSM modeling method worked and created treatment and control groups which are very similar.

Table 2: Raw Differences between Treatment and Control Groups (Disabled/Not Disabled)

	Income	Age	Parent Hours Worked	Child Health	Location Type	1st Genera
(Intercept)	7.118***	2.101***	59.467***	1.356***	1.916***	0.541**
	(0.038)	(0.019)	(0.344)	(0.008)	(0.009)	(0.009)
DSBLTY	−0.810***	0.747***	−5.626***	0.447***	0.033	0.005
	(0.116)	(0.060)	(1.069)	(0.025)	(0.029)	(0.029)
Num.Obs.	7092	7092	7092	7092	7092	7092
R2	0.007	0.021	0.004	0.045	0.000	0.000
R2 Adj.	0.007	0.021	0.004	0.044	0.000	0.000
AIC	35 671.3	26 304.5	67 119.1	13 596.8	16 172.5	15 972.0
BIC	35 691.9	26 325.1	67 139.7	13 617.4	16 193.1	15 992.0
Log.Lik.	−17 832.669	−13 149.251	−33 556.546	−6795.396	−8083.268	−7982.98
RMSE	2.99	1.55	27.46	0.63	0.76	0.75

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

```
# Differences in Treatment/Control variables before matching
models_cov = list()

models_cov[['Income']] = lm(TTLHHINC ~ DSBLTY, data = NHES_sub)
models_cov[['Age']] = lm(AGE2018 ~ DSBLTY, data = NHES_sub)
models_cov[['Parent Hours Worked']] = lm(FHRSWK ~ DSBLTY, data = NHES_sub)
models_cov[['Child Health']] = lm(HDHEALTH ~ DSBLTY, data = NHES_sub)
models_cov[['Location Type']] = lm(ZIPLOCL ~ DSBLTY, data = NHES_sub)
models_cov[['1st Generation']] = lm(FFSTGN ~ DSBLTY, data = NHES_sub)
models_cov[['% Low Income']] = lm(ZIP18PO2 ~ DSBLTY, data = NHES_sub)
models_cov[['% Black/Hispanic']] = lm(ZIPBLHI2 ~ DSBLTY, data = NHES_sub)
models_cov[['Family Education']] = lm(FEDUC ~ DSBLTY, data = NHES_sub)

modelsummary(models_cov, stars=TRUE, title="Raw Differences between Treatment and Control

# Examine differences between treatment and control after matching
# Put the models into model summary
library(modelsummary)

models_matched <- list()
models_matched[['Income']] = lm(TTLHHINC ~ DSBLTY, data = matched)
models_matched[['Age']] = lm(AGE2018 ~ DSBLTY, data = matched)
models_matched[['Parent Hours Worked']] = lm(FHRSWK ~ DSBLTY, data = matched)
```

Table 3: Differences between treatment and control group, Matched Sample

	Income	Age	Parent Hours Worked	Child Health	Location Type	1st Generatio
(Intercept)	6.370***	2.906***	53.944***	1.769***	1.923***	0.552***
	(0.114)	(0.053)	(1.064)	(0.032)	(0.028)	(0.027)
DSBLTY	−0.061	−0.058	−0.103	0.034	0.027	−0.005
	(0.161)	(0.076)	(1.505)	(0.045)	(0.040)	(0.038)
Num.Obs.	1472	1472	1472	1472	1472	1472
R2	0.000	0.000	0.000	0.000	0.000	0.000
R2 Adj.	−0.001	0.000	−0.001	0.000	0.000	−0.001
AIC	7496.7	5274.7	14 082.4	3753.4	3381.5	3218.9
BIC	7512.6	5290.6	14 098.3	3769.3	3397.3	3234.8
Log.Lik.	−3745.337	−2634.336	−7038.189	−1873.715	−1687.729	−1606.452
RMSE	3.08	1.45	28.86	0.86	0.76	0.72

$+ \text{ p} < 0.1, \text{ * p} < 0.05, \text{ ** p} < 0.01, \text{ *** p} < 0.001$

```
models_matched[['Child Health']] = lm(HDHEALTH ~ DSBLTY, data = matched)
models_matched[['Location Type']] = lm(ZIPLOCL ~ DSBLTY, data = matched)
models_matched[['1st Generation']] = lm(FFSTGN ~ DSBLTY, data = matched)
models_matched[['% Low Income']] = lm(ZIP18PO2 ~ DSBLTY, data = matched)
models_matched[['% Black/Hispanic']] = lm(ZIPBLHI2 ~ DSBLTY, data = matched)
models_matched[['Family Education']] = lm(FEDUC ~ DSBLTY, data = matched)

modelsummary(models_matched, stars=TRUE, title="Differences between treatment and control
```

Estimating Effects

```
models_effects <- list()
models_effects[['Effects (Aftercare)']] <- lm(ANYCAREX ~ DSBLTY, matched)
models_effects[['Effects (Aftercare) + Controls']] <- lm(ANYCAREX ~ DSBLTY + TTLHHINC + AG
models_effects[['Effects (Aftercare Weekly Hours)']] <- lm(CAREHOURX ~ DSBLTY, matched)
models_effects[['Effects (Aftercare Weekly Hours) + Controls']] <- lm(CAREHOURX ~ DSBLTY +
modelsummary(models_effects, stars=TRUE, title="Effects w/PSM")
```

In the PSM model we can see a 6.9 percentage point increase, on average, in attendance for children with disabilities in non-parental care programs. This number increases to 7.4 percentage points once we control for other variables such as income, age, and others.

Interestingly, there is no observable difference in treatment and control units when examining

Table 4: Effects w/PSM

	Effects (Aftercare)	Effects (Aftercare) + Controls	Effects (Aftercare Weekly Hours)	Effe
(Intercept)	0.700***	0.193*	21.216***	
	(0.016)	(0.076)	(0.789)	
DSBLTY	0.069**	0.074***	0.931	
	(0.023)	(0.021)	(1.116)	
TTLHHINC		0.017***		
		(0.005)		
AGE2018		0.071***		
		(0.007)		
FHRSWK		0.004***		
		(0.000)		
HDHEALTH		0.040**		
		(0.012)		
ZIPLOCL		−0.024		
		(0.016)		
FFSTGN		0.044**		
		(0.016)		
ZIP18PO2		0.002		
		(0.014)		
ZIPBLHI2		−0.020		
		(0.013)		
FEDUC		−0.005		
		(0.009)		
Num.Obs.	1472	1472	1472	
R2	0.006	0.156	0.000	
R2 Adj.	0.005	0.150	0.000	
AIC	1768.4	1546.2	13 200.2	
BIC	1784.3	1609.8	13 216.0	
Log.Lik.	−881.212	−761.110	−6597.077	
RMSE	0.44	0.41	21.39	

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

the effect of a child having a disability on the numbers of hours they spend in a program.

Entropy Balancing

```
# INSERT VARIABLES YOU WANT TO BALANCE IN THE MODEL HERE
variables <- c('AGE2018', 'ZIP18PO2', 'ZIPLOCL', 'ZIPBLHI2', 'FHRSWK', 'FFSTGN', 'TTLHHINC

weights <- ebalance(Treatment = NHES_sub$DSBLTY,X = NHES_sub[,variables])
```

Converged within tolerance

```
  # Treatment option is the treatment dummy
  # X is variables from dataset I want to keep in model

# The ebalance function only keeps the weights of all the control group data. The treatmen

# Subset data by treatment and control
NHES_sub_control <- subset(NHES_sub, DSBLTY==0)
entropy_weights <- weights$w
NHES_sub_control <- cbind(NHES_sub_control,entropy_weights)

# treatment group only with entropy weights
NHES_sub_treatment <- subset(NHES_sub,DSBLTY==1)
NHES_sub_treatment$entropy_weights <- 1

# bind treatment and control
NHES_sub_EB<-rbind(NHES_sub_control,NHES_sub_treatment)
```

Analyzing Models

Here are the "regular" un-weighted Models, taken from the PSM section from earlier.

```
# Show Un-weighted Models
modelsummary(models_cov, stars=TRUE, title="Covariate Balance WITHOUT entropy weights")
```

Next, I create the weighted models in order to compare to the weighted ones.

```
# Create Weighted Models
models_weighted = list()
```

Table 5: Covariate Balance WITHOUT entropy weights

	Income	Age	Parent Hours Worked	Child Health	Location Type	1st Genera
(Intercept)	7.118***	2.101***	59.467***	1.356***	1.916***	0.541**
	(0.038)	(0.019)	(0.344)	(0.008)	(0.009)	(0.009)
DSBLTY	−0.810***	0.747***	−5.626***	0.447***	0.033	0.005
	(0.116)	(0.060)	(1.069)	(0.025)	(0.029)	(0.029)
Num.Obs.	7092	7092	7092	7092	7092	7092
R2	0.007	0.021	0.004	0.045	0.000	0.000
R2 Adj.	0.007	0.021	0.004	0.044	0.000	0.000
AIC	35 671.3	26 304.5	67 119.1	13 596.8	16 172.5	15 972.0
BIC	35 691.9	26 325.1	67 139.7	13 617.4	16 193.1	15 992.0
Log.Lik.	−17 832.669	−13 149.251	−33 556.546	−6795.396	−8083.268	−7982.98
RMSE	2.99	1.55	27.46	0.63	0.76	0.75

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
models_weighted[['Income']] = lm(TTLHHINC ~ DSBLTY, data = NHES_sub_EB, weights=entropy_we
models_weighted[['Age']] = lm(AGE2018 ~ DSBLTY, data = NHES_sub_EB, weights=entropy_weight
models_weighted[['Parent Hours Worked']] = lm(FHRSWK ~ DSBLTY, data = NHES_sub_EB, weights
models_weighted[['Child Health']] = lm(HDHEALTH ~ DSBLTY, data = NHES_sub_EB, weights=entr
models_weighted[['Location Type']] = lm(ZIPLOCL ~ DSBLTY, data = NHES_sub_EB, weights=entr
models_weighted[['1st Generation']] = lm(FFSTGN ~ DSBLTY, data = NHES_sub_EB, weights=entr
models_weighted[['% Low Income']] = lm(ZIP18PO2 ~ DSBLTY, data = NHES_sub_EB, weights=entr
models_weighted[['% Black/Hispanic']] = lm(ZIPBLHI2 ~ DSBLTY, data = NHES_sub_EB, weights=
models_weighted[['Family Education']] = lm(FEDUC ~ DSBLTY, data = NHES_sub_EB, weights=ent

modelsummary(models_weighted, stars=TRUE, title="Covariate Balance WITH entropy weights")
```

We can see that Entropy Balancing completely removed differences in treatment and control
groups through weights.

Outcomes

```
# EB Effects
models_outcomes = list()

models_outcomes[['Aftercare Attendance - No Weight']] = lm(ANYCAREX ~ DSBLTY, data = NHES_
models_outcomes[['Aftercare Weekly Hours - No Weight']] = lm(CAREHOURX ~ DSBLTY, data = NH
```

Table 6: Covariate Balance WITH entropy weights

	Income	Age	Parent Hours Worked	Child Health	Location Type	1st Genera
(Intercept)	6.308***	2.848***	53.841***	1.803***	1.950***	0.546**
	(0.052)	(0.025)	(0.482)	(0.015)	(0.013)	(0.012)
DSBLTY	0.000	0.000	0.000	0.000	0.000	0.000
	(0.073)	(0.035)	(0.682)	(0.021)	(0.018)	(0.017)
Num.Obs.	7092	7092	7092	7092	7092	7092
R2	0.000	0.000	0.000	0.000	0.000	0.000
R2 Adj.	0.000	0.000	0.000	0.000	0.000	0.000
AIC	40 600.4	29 989.3	72 216.0	22 841.6	20 838.5	20 007.8
BIC	40 621.0	30 009.9	72 236.6	22 862.2	20 859.1	20 028.4
Log.Lik.	−20 297.198	−14 991.672	−36 104.976	−11 417.781	−10 416.238	−10 000.9
RMSE	3.09	1.70	27.97	0.76	0.76	0.75

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

```
models_outcomes[['Aftercare Attendance - Weighted']] = lm(ANYCAREX ~ DSBLTY, data = NHES_s
models_outcomes[['Aftercare Weekly Hours - Weighted']] = lm(CAREHOURX ~ DSBLTY, data = NHE

modelsummary(models_outcomes, stars=TRUE, title="Effect Estimates")
```

In the weighted model for a disabilities impact on Aftercare attendance it is shown that there is a 7.5 percentage point increase, on average, in attendance if a child has a disability. Children who have a disability also have increased attendance time, by roughly 1.5 hours per week, on average.

By creating a linear regression model and linear probability model to predict the effect of a child having a disability on their participation in an non-parental care program it is evident that there is a strong positive correlation between a child having a disability and them spending time in non-parental care programs.

Adding the entropy balancing weights to the model slightly lessens the effect of the "treatment" variable (DSBLTY) but still results in a model with a high positive correlation between children with disabilities and participation in after school programs.

Table 7: Effect Estimates

	Aftercare Attendance - No Weight	Aftercare Weekly Hours - No Weight	Aftercare Attendan
(Intercept)	0.665***	19.929***	0.694**
	(0.006)	(0.265)	(0.007)
DSBLTY	0.104***	2.218**	0.075**
	(0.018)	(0.823)	(0.010)
Num.Obs.	7092	7092	7092
R2	0.005	0.001	0.007
R2 Adj.	0.004	0.001	0.007
AIC	9334.8	63 404.9	13 013.6
BIC	9355.4	63 425.5	13 034.2
Log.Lik.	−4664.404	−31 699.432	−6503.80
RMSE	0.47	21.13	0.47

+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001

Conclusions

So, what does this all mean?

By utilizing a few different causal inference techniques the goal is to increase internal validity in
the models. The goal of PSM and Entropy Balancing is to create treatment and control groups
which are similar, the only difference being the treatment variable. Once this is established it
is then possible to create models that connect the treatment variable to the outcome variable
with greater certainty.

By examining the models it is evident that there is a strong positive correlation between
a child having a disability and their participation in a non-parental care activity. While
there is a strong correlation here we cannot say immediately that just having a disability
guarantees greater rates of participation in these activities. There may be multiple reasons for
participation, such as after school programs having additional or more specialized resources
for children with disabilities.

Limitations

Some limitations of this study, particularly in the data, is the survey nature of data. Since the
data was collected through a survey there may be a certain element of non-response bias and
the limited sample size of the survey. This survey attempts to cover the entirety of the untied
states, with four different census regions (Northeast, South, Midwest, West). Since there are

only 7000 rows of data in this data set (7000 households), it may be difficult to extrapolate or apply this data on a national scale.

If future studies were to be completed on this data set, there could be a greater focus made on examining data for specific regions, for example, sub-setting into 4 data sets, one for each region.

Additionally, this study used the DSBLTY variable as a "treatment" variable - in this data set there are also variables available for many different types of disability (Speech/Language impairments, Physical disabilities, etc). It may be worthwhile to examine the effects of different types of disabilities on participation in programs.