

# Is there an association between early childhood adversity and higher depression scores in US adults?

Vincent Yu

2024-6-18

```
#needed questionnaires
#install.packages("haven")
#install.packages("nhanesA")
#install.packages("tableone")
library(nhanesA)
library(haven)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(tableone)
library(dplyr)
library(tableone)
library(kableExtra)
```

```
##
##   'kableExtra'
##
## The following object is masked from 'package:dplyr':
##
##   group_rows
```

```
library(ggplot2)
library(broom)
#depress <- read_xpt("D:/_uoft_1/5 1/depress-and-childhood/data/DPQ_J.XPT")
demo <- nhanes("DEMO_J")
depress <- nhanes("DPQ_J")
#child <- read_xpt("D:/_uoft_1/5 1/depress-and-childhood/data/ECQ_J.XPT")
child <- nhanes("ECQ_J")
BMI <- nhanes("BMX_J")
```

The relationship between early childhood adversity and higher depression scores is complex and multifaceted. Women with a birth weight of less than 3 kg had an increased risk of depression at age 26 compared to those with a birth weight of more than 3.5 kg. Older maternal age is a predictor of adverse mental health symptoms, particularly in female offspring, suggesting a gender-specific effect.

Early exposure to significant stressors can lead to long-term changes in brain function and increase the risk of developing depressive disorders. Women appear to be more affected by these adversities, reporting higher rates of depression than men. Specific stressors such as parental divorce, frightening events, and physical abuse significantly increase the likelihood of depression. Parental influence, both genetic and environmental, plays a crucial role in shaping children's health outcomes, further emphasizing the need for supportive and stable family environments.

Based on these findings and considering the limited data available, I have formulated the following research questions for further investigation: How the participants was affected H1: Lower birth weight is associated with higher depression scores. ECD070A H2: Older maternal age at birth is associated with higher depression scores, particularly in female participants. ECD010 H3: Maternal smoking during pregnancy is associated with higher depression scores. ECQ020

And how these will affect the next generation H4: Higher BMI in participants is associated with higher depression scores. H5: Higher depression scores in participants are associated with adverse health outcomes in the next generation. MCQ080E I am not sure if this is the right direction and would greatly appreciate any guidance or suggestions you could provide.

```
# merge 2 datasets
depress <- nhanes("DPQ_J")%>% filter_all(all_vars(!is.na(.)))
# Replace value descriptions with code values
depress_replaced <- depress %>%
  mutate(across(starts_with("DPQ"), ~ recode(.,
                                            "Not at all" = 0,
                                            "Several days" = 1,
                                            "More than half the days" = 2,
                                            "Nearly every day" = 3)))
# Replace DPQ100 with the sum of DPQ010 to DPQ090
depress_replaced <- depress_replaced %>%
  mutate(DPQ100 = rowSums(select(., starts_with("DPQ0"))))

clean_data <- merge(demo, depress_replaced, by = "SEQN", all.x = T)
clean_data_T <- merge(clean_data, child, by = "SEQN", all.x = T)
clean_data_T <- merge(clean_data_T, BMI, by = "SEQN", all.x = T)
# filter missing and refused
```

#filter missing value still error: ModuleStopIfNoVarsLeft(vars): No valid variables. e.x how to handle missing value. none of the "SEQN" is matched

```
clean_data_F <- merge(depress, BMI, by = "SEQN", all.x = F, all.y = F)
clean_data_FF <- merge(depress, child, by = "SEQN", all.x = F, all.y = F)
```

```
demo_clean <- clean_data_T %>% select("SEQN", "DPQ100", "RIDAGEYR", "DMDHRGND", "RIDRETH1", "ECD070A", "ECD010", "MCQ080E")
colnames(demo_clean) <- c("SEQN", "Depress", "Age", "Gender", "Race", "birth_weight", "maternal_age", "Maternal_smoking", "Marital_status", "MCQ080E")
# Define variables, factorVars for categorical variable
vars <- c("Age", "Gender", "Race", "birth_weight", "maternal_age", "Maternal_smoking", "Marital_status", "MCQ080E")
factorVars <- c("Gender", "Race", "Maternal_smoking", "Marital_status")
table1 <- CreateTableOne(vars = vars, data = demo_clean, factorVars = factorVars)
```

```
# Convert to data frame for stargazer
table1_df <- as.data.frame(print(table1, printToggle = FALSE))
# Display the table in R
kable(table1_df, caption = "table.1 Demographics")
```

Table 1: table.1 Demographics

	Overall
n	9254
Age (mean (SD))	34.33 (25.50)
Gender = Female (%)	4625 (50.0)
Race (%)	
Mexican American	1367 (14.8)
Other Hispanic	820 ( 8.9)
Non-Hispanic White	3150 (34.0)
Non-Hispanic Black	2115 (22.9)
Other Race - Including Multi-Racial	1802 (19.5)
birth_weight (mean (SD))	281.41 (1634.07)
maternal_age (mean (SD))	173.52 (1191.73)
Maternal_smoking (%)	
Yes	398 (12.9)
No	2661 (86.1)
Don't know	33 ( 1.1)
Marital_status (%)	
Married	2737 (49.1)
Widowed	462 ( 8.3)
Divorced	641 (11.5)
Separated	202 ( 3.6)
Never married	1006 (18.1)
Living with partner	515 ( 9.2)
Refused	6 ( 0.1)
BMI (mean (SD))	26.58 (8.26)
adverse health outcomes in the next generation = No (%)	2170 (86.8)

## what analysis model should I use

```
#model1 <- glm(DPQ100 ~ Maternal_smoking + Age + Gender + Race + birth_weight, data = demo_clean, famil
#model2 <- glm(DPQ100 ~ Maternal_smoking + Age + Gender + Race + maternal_age, data = demo_clean, famil
#model3 <- glm(DPQ100 ~ Maternal_smoking + Age + Gender + Race + Marital_status , data = demo_clean, fa
#model4 <- glm(DPQ100 ~ Maternal_smoking + Age + Gender + Race + BMI + adverse_health_outcomes_next_gen
# Summary of the model
#summary(model1)
```

```
# Create quantile groups for BMI
demo_clean_BMI <- demo_clean %>%
  mutate(BMI_quantiles = ntile(BMI, 4)) %>%
  mutate(BMI_quantiles = factor(BMI_quantiles, labels = c("Q1", "Q2", "Q3", "Q4")))
# Create Table 1 stratified by BMI quantiles
table2 <- CreateTableOne(vars = vars, strata = "BMI_quantiles", data = demo_clean_BMI, factorVars = fac
```

```
## Warning in ModuleReturnVarsExist(factorVars, data): The data frame does not
## have: Maternal_Smoking Marital_Status Dropped
```

```
# Convert to data frame for kableExtra
table2_df <- as.data.frame(print(table2, printToggle = FALSE))

# Use kableExtra for better formatting
kable(table2_df, caption = "Demographics")
```

Table 2: Demographics

	Q1	Q2	Q3	Q4	p	test
n	2002	2001	2001	2001		
Age (mean (SD))	13.19 (16.01)	38.28 (23.52)	47.92 (20.86)	47.33 (18.79)	<0.001	
Gender = Female (%)	1081 (54.0)	945 (47.2)	915 (45.7)	1023 (51.1)	<0.001	
Race (%)					<0.001	
Mexican American	266 (13.3)	248 (12.4)	334 (16.7)	327 (16.3)		
Other Hispanic	145 ( 7.2)	163 ( 8.1)	226 (11.3)	170 ( 8.5)		
Non-Hispanic White	677 (33.8)	636 (31.8)	657 (32.8)	694 (34.7)		
Non-Hispanic Black	470 (23.5)	407 (20.3)	410 (20.5)	581 (29.0)		
Other Race - Including Multi-Racial	444 (22.2)	547 (27.3)	374 (18.7)	229 (11.4)		
birth_weight (mean (SD))	326.39 (1759.07)	332.11 (1775.46)	297.49 (1683.58)	130.23 (1110.24)	0.793	
maternal_age (mean (SD))	249.76 (1465.67)	91.18 (774.28)	258.30 (1507.43)	26.83 (6.87)	0.089	
Maternal_smoking (%)					0.003	
Yes	180 (11.3)	65 (15.1)	34 (19.8)	17 (21.0)		
No	1391 (87.3)	362 (84.2)	136 (79.1)	64 (79.0)		
Don't know	23 ( 1.4)	3 ( 0.7)	2 ( 1.2)	0 ( 0.0)		
Marital_status (%)					<0.001	
Married	89 (30.4)	700 (52.2)	923 (53.5)	876 (48.2)		
Widowed	30 (10.2)	96 ( 7.2)	134 ( 7.8)	146 ( 8.0)		
Divorced	38 (13.0)	130 ( 9.7)	202 (11.7)	213 (11.7)		
Separated	8 ( 2.7)	50 ( 3.7)	54 ( 3.1)	78 ( 4.3)		
Never married	98 (33.4)	257 (19.2)	246 (14.3)	328 (18.1)		
Living with partner	30 (10.2)	106 ( 7.9)	165 ( 9.6)	173 ( 9.5)		
Refused	0 ( 0.0)	2 ( 0.1)	0 ( 0.0)	3 ( 0.2)		
BMI (mean (SD))	17.06 (1.86)	23.27 (1.56)	28.39 (1.56)	37.59 (6.21)	<0.001	
adverse health outcomes in the next generation = No (%)	1535 (96.3)	324 (75.3)	85 (49.4)	25 (30.9)	<0.001	

```
model11 <- glm(Depress ~ BMI, data = demo_clean)
summary(model11)
```

```
##
## Call:
## glm(formula = Depress ~ BMI, data = demo_clean)
##
## Coefficients:
```

```
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.356059    0.304015  11.039 < 2e-16 ***
## BMI         0.051064    0.009762   5.231 1.79e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 18.82757)
##
##      Null deviance: 62703  on 3304  degrees of freedom
## Residual deviance: 62187  on 3303  degrees of freedom
##    ( 5949    )
## AIC: 19084
##
## Number of Fisher Scoring iterations: 2
```

```
data <- demo_clean %>% filter(!is.na(Depress))%>%
  mutate(Depress = rowSums(select(., starts_with("Depress"))))

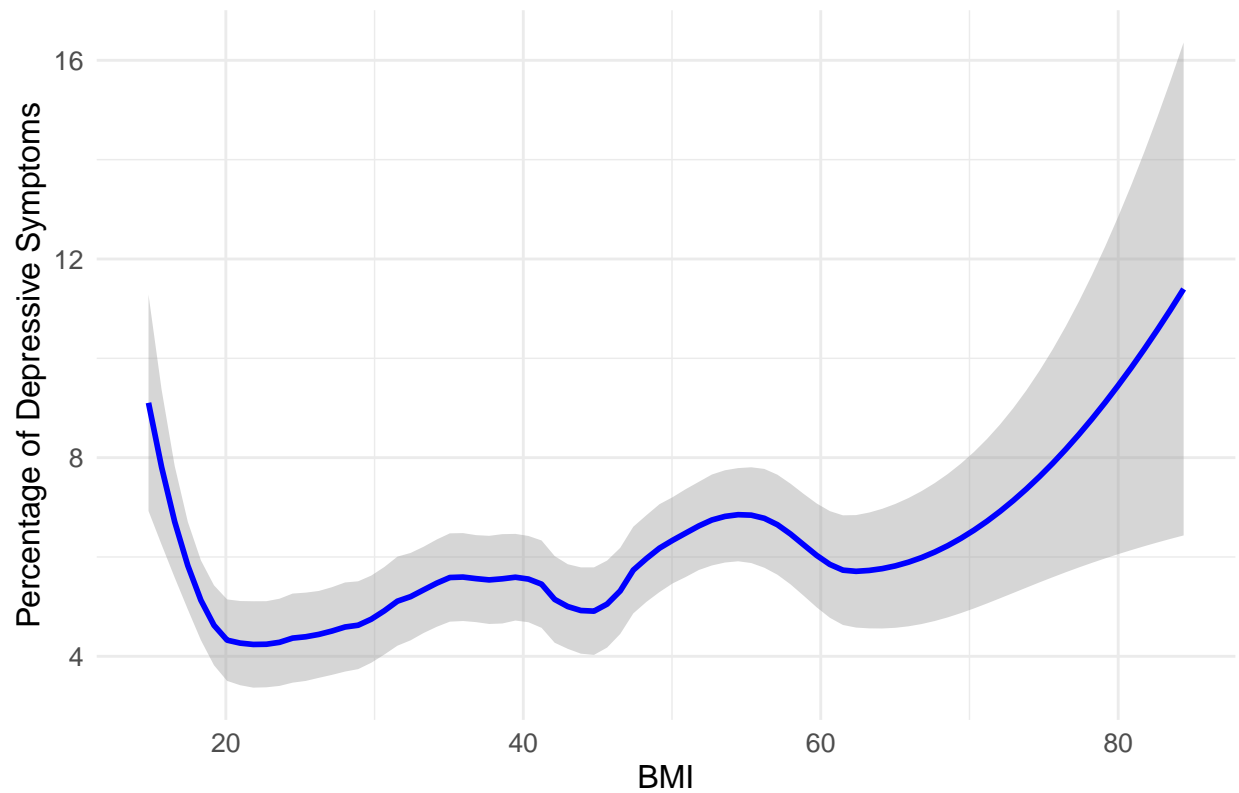
# Calculate percentage of depressive symptoms for each BMI integer category
percentage_data <- data %>%
  group_by(BMI) %>%
  summarise(depress_count = sum(Depress),
            total_count = n(),
            percentage_depress = (depress_count / total_count))

# Create the plot
ggplot(percent_data, aes(x = BMI, y = percentage_depress)) +
  geom_smooth(color = "blue", method = "loess", span = 0.3) +
  labs(title = "Percentage of Depressive Symptoms by BMI",
       x = "BMI",
       y = "Percentage of Depressive Symptoms") +
  theme_minimal() +
  theme(axis.title = element_text(size = 12),
        axis.text = element_text(size = 10),
        plot.title = element_text(size = 14, face = "bold", hjust = 0.5))
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

```
## Warning: Removed 1 row containing non-finite outside the scale range
## (`stat_smooth()`).
```

## Percentage of Depressive Symptoms by BMI



```
# Create the density plot for BMI
ggplot(demo_clean, aes(x = BMI)) +
  geom_density(fill = "blue", alpha = 0.3) +
  labs(title = "Density Plot of BMI", x = "BMI", y = "Density") +
  theme_minimal() +
  theme(axis.title = element_text(size = 12),
        axis.text = element_text(size = 10),
        plot.title = element_text(size = 14, face = "bold", hjust = 0.5))
```

```
## Warning: Removed 1249 rows containing non-finite outside the scale range
## (`stat_density()`).
```

