BRFSS 2019: Adverse Childhood Experiences and Adult Health Indicators

A reproducible exploratory analysis using a weighted subsample

Aryan Bhojani

```
library(tidyverse)
library(ggplot2)
library(scales)
```

Project overview

This project explores whether adverse childhood experiences (ACEs) are associated with adult outcomes captured in the 2019 Behavioral Risk Factor Surveillance System (BRFSS), including self-rated general health, smoking status, and self-reported diagnosis of a depressive disorder. Because BRFSS is extremely large, I use a reproducible weighted subsample and focus on a small set of variables to build a clean analysis dataset and compute descriptive comparisons across subgroups.

Data import

```
selected_vars <- c(
"_SEX", "_AGEG5YR",
"GENHLTH", "ACEPRISN",
"ACEDRUGS", "ACEDRINK",
"ACEDEPRS", "ADDEPEV3",
"_SMOKER3", "_LLCPWT"
)

brfss <- readr::read_csv(
  params$brfss_path,
```

```
  col_select = tidyselect::all_of(selected_vars)
)

head(brfss)
```

```
# A tibble: 6 x 10
  `_SEX` `_AGEG5YR` GENHLTH ACEPRISN ACEDRUGS ACEDRINK ACEDEPRS ADDEPEV3
   <dbl>      <dbl>   <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
1      2         13       3        2        2        2        2        2
2      2         11       4        2        2        1        2        2
3      2         10       3        2        2        2        2        2
4      2         13       4       NA       NA       NA       NA        2
5      2         13       2        2        2        2        2        2
6      2         14       2       NA       NA       NA       NA        2
# i 2 more variables: `_SMOKER3` <dbl>, `_LLCPWT` <dbl>
```

Variable preparation

Some BRFSS variables begin with underscores and are less convenient to work with directly.
I rename them into simpler, more descriptive names.

```
brfss <- brfss |>
rename(
sampling_weights = any_of("_LLCPWT"),
SEX              = any_of("_SEX"),
AGEG5YR          = any_of("_AGEG5YR"),
SMOKER3          = any_of("_SMOKER3")
)
```

Weighted subsample

To keep computation lightweight while preserving representativeness, I draw a reproducible
weighted sample.

```
set.seed(params$seed)

samp <- brfss |>
slice_sample(n = params$sample_n, weight_by = sampling_weights)
```

```
colMeans(is.na(samp))
```

```
            SEX          AGEG5YR          GENHLTH         ACEPRISN
         0.0000           0.0000           0.0000           0.7495
       ACEDRUGS         ACEDRINK         ACEDEPRS         ADDEPEV3
         0.7491           0.7490           0.7488           0.0000
        SMOKER3 sampling_weights
         0.0000           0.0000
```

Tidying and recoding

BRFSS encodes many categorical responses numerically. I convert them to interpretable factor
variables using codebook-consistent category labels.

Age group

```
age_codes <- c(
"1" = "18-24", "2" = "25-29", "3" = "30-34",
"4" = "35-39", "5" = "40-44", "6" = "45-49",
"7" = "50-54", "8" = "55-59", "9" = "60-64",
"10" = "65-69", "11" = "70-74", "12" = "75-79",
"13" = "80+", "14" = "Unsure/refused/missing"
)

samp_mod1 <- samp |>
mutate(
age_group = factor(
as.character(AGEG5YR),
levels = names(age_codes),
labels = age_codes,
ordered = TRUE
)
)

head(samp_mod1)
```

```
# A tibble: 6 x 11
```

```
      SEX AGEG5YR GENHLTH ACEPRISN ACEDRUGS ACEDRINK ACEDEPRS ADDEPEV3 SMOKER3
    <dbl>   <dbl>   <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>   <dbl>
1     1       2       2        2        2        2        1        2       4
2     2       4       3        2        2        2        2        2       2
3     2       8       3       NA       NA       NA       NA        2       9
4     1       9       2        1        2        2        2        2       4
5     2       7       3       NA       NA       NA       NA        2       4
6     2       3       1       NA       NA       NA       NA        2       4
# i 2 more variables: sampling_weights <dbl>, age_group <ord>
```

Sex, general health, and smoking status

```
sex_codes <- c(
"1" = "M",
"2" = "F",
"7" = "Unsure",
"9" = "Refused"
)

samp_mod2 <- samp_mod1 |>
mutate(
SEX = factor(
as.character(SEX),
levels = names(sex_codes),
labels = sex_codes
)
)

health_codes <- c(
"1" = "Excellent",
"2" = "Very good",
"3" = "Good",
"4" = "Fair",
"5" = "Poor",
"7" = "Unsure",
"9" = "Refused"
)

samp_mod3 <- samp_mod2 |>
mutate(
GENHLTH = factor(
```

```
as.character(GENHLTH),
levels = names(health_codes),
labels = health_codes,
ordered = TRUE
)
)

smoke_codes <- c(
"1" = "Daily",
"2" = "Some days",
"3" = "Former",
"4" = "Never",
"9" = "Unsure/refused/missing"
)

samp_mod4 <- samp_mod3 |>
mutate(
SMOKER3 = factor(
as.character(SMOKER3),
levels = names(smoke_codes),
labels = smoke_codes,
ordered = TRUE
)
)

head(samp_mod4)
```

```
# A tibble: 6 x 11
  SEX   AGEG5YR GENHLTH   ACEPRISN ACEDRUGS ACEDRINK ACEDEPRS ADDEPEV3 SMOKER3
  <fct>   <dbl> <ord>        <dbl>    <dbl>    <dbl>    <dbl>    <dbl> <ord>
1 M           2 Very good        2        2        2        1        2 Never
2 F           4 Good             2        2        2        2        2 Some days
3 F           8 Good            NA       NA       NA       NA        2 Unsure/r~
4 M           9 Very good        1        2        2        2        2 Never
5 F           7 Good            NA       NA       NA       NA        2 Never
6 F           3 Excellent       NA       NA       NA       NA        2 Never
# i 2 more variables: sampling_weights <dbl>, age_group <ord>
```

Recode ACE items consistently

ACE items share a common response structure, so I recode them with a single reusable function and apply it across all variables that start with "ACE".

```r
recode_ace <- function(x) {
case_match(
as.character(x),
"1" ~ "Yes",
"2" ~ "No",
"7" ~ "Unsure",
"9" ~ "Refused",
.default = NA_character_
) |>
factor(levels = c("Yes", "No", "Unsure", "Refused"))
}

samp_mod5 <- samp_mod4 |>
mutate(across(starts_with("ACE"), recode_ace))

head(samp_mod5)
```

```
# A tibble: 6 x 11
  SEX   AGEG5YR GENHLTH   ACEPRISN ACEDRUGS ACEDRINK ACEDEPRS ADDEPEV3 SMOKER3
  <fct>   <dbl> <ord>     <fct>    <fct>    <fct>    <fct>       <dbl> <ord>
1 M           2 Very good No       No       No       Yes             2 Never
2 F           4 Good      No       No       No       No              2 Some days
3 F           8 Good      <NA>     <NA>     <NA>     <NA>            2 Unsure/r~
4 M           9 Very good Yes      No       No       No              2 Never
5 F           7 Good      <NA>     <NA>     <NA>     <NA>            2 Never
6 F           3 Excellent <NA>     <NA>     <NA>     <NA>            2 Never
# i 2 more variables: sampling_weights <dbl>, age_group <ord>
```

Derived indicators

Any ACE reported

I define a single indicator that flags whether a respondent answered "Yes" to any ACE item.

```
samp_mod7 <- samp_mod5 |>
mutate(
adverse_conditions = if_any(
starts_with("ACE"),
~ coalesce(. == "Yes", FALSE)
)
)

head(samp_mod7)
```

```
# A tibble: 6 x 12
  SEX   AGEG5YR GENHLTH   ACEPRISN ACEDRUGS ACEDRINK ACEDEPRS ADDEPEV3 SMOKER3
  <fct>   <dbl> <ord>     <fct>    <fct>    <fct>    <fct>       <dbl> <ord>
1 M           2 Very good No       No       No       Yes             2 Never
2 F           4 Good      No       No       No       No              2 Some days
3 F           8 Good      <NA>     <NA>     <NA>     <NA>            2 Unsure/r~
4 M           9 Very good Yes      No       No       No              2 Never
5 F           7 Good      <NA>     <NA>     <NA>     <NA>            2 Never
6 F           3 Excellent <NA>     <NA>     <NA>     <NA>            2 Never
# i 3 more variables: sampling_weights <dbl>, age_group <ord>,
#   adverse_conditions <lgl>
```

ACE missingness and module restriction

Missing ACE values typically indicate the module was not administered to that respondent. I track this explicitly.

To keep ACE comparisons meaningful, I filter to respondents with complete ACE module responses.

```
samp_mod8 <- samp_mod7 |>
mutate(
adverse_missing = if_any(starts_with("ACE"), ~ is.na(.))
)

samp_mod9 <- samp_mod8 |>
filter(!adverse_missing)
```

Depression diagnosis indicator

I convert the depressive disorder item into a logical indicator.

```
samp_mod10 <- samp_mod9 |>
mutate(
depression = case_match(
as.character(ADDEPEV3),
"1" ~ TRUE,
"2" ~ FALSE,
"7" ~ NA,
"9" ~ NA,
.default = NA
)
)

head(samp_mod10)
```

```
# A tibble: 6 x 14
  SEX   AGEG5YR GENHLTH   ACEPRISN ACEDRUGS ACEDRINK ACEDEPRS ADDEPEV3 SMOKER3
  <fct>   <dbl> <ord>     <fct>    <fct>    <fct>    <fct>       <dbl> <ord>
1 M           2 Very good No       No       No       Yes             2 Never
2 F           4 Good      No       No       No       No              2 Some days
3 M           9 Very good Yes      No       No       No              2 Never
4 F           1 Very good No       No       No       No              2 Never
5 M          12 Good      No       No       No       No              2 Never
6 F           5 Good      No       No       No       No              2 Never
# i 5 more variables: sampling_weights <dbl>, age_group <ord>,
#   adverse_conditions <lgl>, adverse_missing <lgl>, depression <lgl>
```

Final analysis dataset

I keep only the cleaned variables and the derived indicators needed for descriptive analysis.

```
final_brfss_data <- samp_mod10 |>
select(
SEX, age_group, GENHLTH, SMOKER3,
adverse_conditions, depression
)

head(final_brfss_data)
```

```
# A tibble: 6 x 6
  SEX   age_group GENHLTH   SMOKER3   adverse_conditions depression
```

```
    <fct> <ord>    <ord>     <ord>     <lgl>          <lgl>
1 M     25-29    Very good Never     TRUE           FALSE
2 F     35-39    Good      Some days FALSE          FALSE
3 M     60-64    Very good Never     TRUE           FALSE
4 F     18-24    Very good Never     FALSE          FALSE
5 M     75-79    Good      Never     FALSE          FALSE
6 F     40-44    Good      Never     FALSE          FALSE
```

Descriptive results

Overall ACE prevalence (among ACE-module respondents)

```
mean_ace <- mean(final_brfss_data$adverse_conditions, na.rm = TRUE)
mean_ace
```

```
[1] 0.3662141
```

ACE prevalence by self-rated general health

```
ace_health <- final_brfss_data |>
group_by(GENHLTH) |>
summarise(
adverse_conditions = mean(adverse_conditions, na.rm = TRUE),
missing_GENHLTH    = mean(is.na(GENHLTH)),
missing_adverse    = mean(is.na(adverse_conditions)),
.groups = "drop"
)

print(ace_health, n = 50)
```

```
# A tibble: 7 x 4
  GENHLTH   adverse_conditions missing_GENHLTH missing_adverse
  <ord>                  <dbl>           <dbl>           <dbl>
1 Excellent              0.299               0               0
2 Very good              0.342               0               0
3 Good                   0.360               0               0
4 Fair                   0.447               0               0
```

```
5 Poor                     0.562              0              0
6 Unsure                   0.333              0              0
7 Refused                  0.5                0              0
```

```r
ace_smoking <- final_brfss_data |>
group_by(SMOKER3) |>
summarise(
adverse_conditions = mean(adverse_conditions, na.rm = TRUE),
.groups = "drop"
) |>
arrange(SMOKER3)

ace_smoking
```

```
# A tibble: 5 x 2
  SMOKER3                   adverse_conditions
  <ord>                                  <dbl>
1 Daily                                  0.543
2 Some days                              0.466
3 Former                                 0.369
4 Never                                  0.321
5 Unsure/refused/missing                 0.5
```

```r
ace_depression <- final_brfss_data |>
group_by(depression) |>
summarise(
adverse_conditions = mean(adverse_conditions, na.rm = TRUE),
.groups = "drop"
)

print(ace_depression, n = 50)
```

```
# A tibble: 3 x 2
  depression adverse_conditions
  <lgl>                   <dbl>
1 FALSE                   0.306
2 TRUE                    0.618
3 NA                      0.294
```

```
# Stratified check: health vs ACE, accounting for sex

ace_health_sex <- final_brfss_data |>
group_by(GENHLTH, SEX) |>
summarise(
adverse_conditions = mean(adverse_conditions, na.rm = TRUE),
.groups = "drop"
)

ace_health_sex_wide <- ace_health_sex |>
pivot_wider(
names_from = SEX,
values_from = adverse_conditions
)

print(ace_health_sex_wide, n = 50)
```

```
# A tibble: 7 x 3
  GENHLTH       M     F
  <ord>     <dbl> <dbl>
1 Excellent 0.249 0.354
2 Very good 0.300 0.383
3 Good      0.343 0.378
4 Fair      0.387 0.5
5 Poor      0.463 0.642
6 Unsure    0.5   0
7 Refused   1     0
```

Discussion of results

Using the filtered ACE-module respondents in this subsample, the overall prevalence of report-ing at least one ACE is 36.6%. This value should be interpreted as a descriptive summary for this analysis subset (not a causal statement).

Relationship with self-rated general health

There is a clear gradient between self-rated health and ACE reporting. Respondents in excel-lent health report >=1 ACE at about 29.9%, while respondents in poor health report >=1 ACE at about 56.2%. In other words, ACE reporting is substantially more common among

those reporting worse general health. This pattern is consistent with the idea that early-life adversity is associated with later-life health risks, though this analysis does not establish direction or causality.

Relationship with smoking status

ACE reporting also varies by smoking category. In this subsample, the prevalence is highest among current smokers: daily smokers are at about 54.3% and some-days smokers are at about 46.6%. Former smokers are lower (36.9%) and never smokers are lowest (32.1%). This suggests that ACE exposure is more common among groups with higher smoking prevalence, which could reflect long-run behavioral or socioeconomic pathways that are not measured here.

Relationship with depression diagnosis

The largest contrast appears for depression diagnosis. Respondents who report a depressive disorder diagnosis have ACE prevalence around 61.8%, compared to about 30.6% among those without a diagnosis. This is a strong descriptive association in the survey responses and motivates deeper modeling that adjusts for confounders.

Stratified comparison by sex

When stratifying the health gradient by sex, the overall pattern remains: ACE prevalence tends to be higher in worse health categories for both groups. For example, in excellent health the prevalence is about M: 24.9% vs F: 35.4%, and in poor health it is about M: 46.3% vs F: 64.2%. Because these are subgroup summaries from a subsample, differences should be treated as exploratory rather than definitive. The 'Unsure'/'Refused' categories likely have very small sample sizes, so I do not interpret those subgroup percentages.

Overall, the descriptive results show consistent associations between ACE reporting and multiple adult health indicators in BRFSS 2019. The next section visualizes the health gradient by sex to make this pattern easier to see.

```
# Visualization

ggplot(ace_health_sex, aes(x = GENHLTH, y = adverse_conditions, fill = SEX)) +
geom_col(position = "dodge") +
labs(
x = "Self-rated general health",
```

```
y = "Proportion reporting \u22651 ACE",
fill = "Sex",
title = "ACEs by self-rated health, stratified by sex (BRFSS 2019 subsample)"
) +
scale_y_continuous(limits = c(0, 1), labels = scales::percent_format(accuracy = 1)) +
theme_minimal()
```

ACEs by self−rated health, stratified by sex (BRFSS 2019 sul

Interpretation

Across respondents who were administered the ACE module, the descriptive summaries suggest that reporting at least one ACE is associated with:

- poorer self-rated general health

- higher smoking prevalence among current smokers relative to former/never smokers

- higher prevalence among respondents reporting a depressive disorder diagnosis

These patterns are descriptive (non-causal) and reflect associations in the observed survey responses.

Limitations and next steps

- **Module availability**: ACE questions are not asked everywhere, so this analysis reflects the subset of respondents who received the module.

- **Self-report and nonresponse**: Health and life experiences are self-reported and may include reporting bias.

- **Confounding**: Many plausible confounders (e.g., socioeconomic status, education, race/ethnicity, region) are not included here.

- **Inference scope**: Results here are intended as exploratory summaries rather than population estimates.

Potential next steps include incorporating additional covariates, using survey-weighted estimation end-to-end (rather than only for sampling), and fitting models to adjust for confounding while keeping interpretation appropriately non-causal.