## Overview

This document provides step-by-step instructions to reproduce the study examining the association between marital status and serious psychological distress (SPD) using the 2023 California Health Interview Survey (CHIS) dataset. The study employs logistic regression models to assess the impact of marital status while adjusting for key sociodemographic and behavioral factors.

## Data Source

* The dataset used in this study is from the California Health Interview Survey (CHIS) 2023.
* The data includes information from approximately 20,000 households in California, collected via web and telephone interviews.
* The survey is conducted in multiple languages and covers children, adolescents, and adults.
* One adult per household is randomly selected to participate.
* Data can be accessed from the CHIS website: https://healthpolicy.ucla.edu/chis

## Study Sample

* Total unweighted sample: 21,655 non-institutionalized adults after excluding individuals who were frail/ill and 65 or over.
* The study population was categorized into married and unmarried groups, and a sensitivity analysis using a tertiary classification of marital status was conducted.

## Variables and Definitions

### Outcome Variable

* Serious Psychological Distress (SPD)
  + Measured using the K6 scale, a validated screening tool for nonspecific psychological distress.
  + SPD was assessed over two timeframes:
    - SPD in the past month
    - SPD in the past year

### Exposure Variable

* Marital Status
  + Binary Classification:
    - Married
    - Not married (never married, living with a partner, widowed, separated, or divorced)
  + Tertiary Classification (for sensitivity analysis):
    - Married
    - Never married
    - Other (living with partner, widowed, separated, or divorced)

### Covariates

The following covariates were included in the logistic regression models:

* Age (categorized as 18–25 [reference], 26–34, 35–44, 45–54, 55–64, 65+)
* Gender (male, female [reference])
* Race/Ethnicity (Hispanic, Non-Hispanic White, Non-Hispanic African American, Non-Hispanic American Indian/Alaskan Native, Non-Hispanic Asian, Other/two or more races)
* Education (less than high school, high school graduate, some college, college or more)
* Poverty Level (categorized as <100% FPL, 100%-199% FPL, 200%-399% FPL, ≥400% FPL)
* Substance Use and Mental Health (whether the respondent needed help for emotional/mental or alcohol/drug problems in the past year)

## Statistical Analysis

* Software Used: RStudio 4.0.2
* Main Analysis: Logistic regression models estimating the association between marital status and SPD.
  + Crude models: Unadjusted odds ratios (PORs) for SPD by marital status.
  + Adjusted models: PORs controlling for age, gender, race/ethnicity, education, poverty, and substance use.
* Sensitivity Analysis: Logistic regression models using the tertiary classification of marital status.

## Steps to Reproduce the Study

1. Obtain the Data: Download the 2023 CHIS dataset from the CHIS website.
2. Data Cleaning & Preprocessing:
   * Exclude respondents 65+ and those classified as frail/ill.
   * Define marital status as binary and tertiary variables.
   * Recode categorical covariates according to study definitions.
3. Descriptive Statistics:
   * Calculate frequencies and percentages for demographic variables (Table 1).
   * Summarize SPD prevalence across marital status groups.
4. Regression Analysis:
   * Run logistic regression models for SPD in the past month and past year.
   * Adjust for covariates.
5. Sensitivity Analysis:
   * Perform logistic regression using the tertiary marital status classification.

## Contact for Questions

For issues related to data access, refer to the CHIS website. For questions regarding analysis steps, ensure proper consultation with a statistician or epidemiologist familiar with survey data.

## R Code for Reproducibility

To ensure full reproducibility, the following R code was used:

# Load required packages

library(haven)

library(tidyverse)

library(dplyr)

library(labelled)

library(tableone)

# Load CHIS 2023 dataset

adult\_chis\_2023 <- read\_sas("C:/Users/sonal/Desktop/2023 Adult SAS/adult.sas7bdat")

# Select variables from original data

clean\_data <- adult\_chis\_2023 %>%

select(SRAGE\_P1, SRSEX, OMBSRR\_P1, BINGE30, DISTRESS, DSTRS30, DSTRS12,

AHEDC\_P1, WRKST\_P1, POVLL2\_P1V2, MARIT, MARIT2, RAKEDW0)

# Create marital status binary variable

clean\_data <- clean\_data %>%

mutate(marital\_status = ifelse(MARIT == 1, "Married", "Not Married"))

# Logistic regression for SPD in past month

logit\_spd\_month <- glm(DSTRS30 ~ marital\_status + SRAGE\_P1 + SRSEX + OMBSRR\_P1 +

AHEDC\_P1 + POVLL2\_P1V2 + BINGE30,

data = clean\_data, family = binomial)

summary(logit\_spd\_month)

# Logistic regression for SPD in past year

logit\_spd\_year <- glm(DSTRS12 ~ marital\_status + SRAGE\_P1 + SRSEX + OMBSRR\_P1 +

AHEDC\_P1 + POVLL2\_P1V2 + BINGE30,

data = clean\_data, family = binomial)

summary(logit\_spd\_year)