BRFSS 2024 Variable Documentation

Behavioral Risk Factor Surveillance System

Centers for Disease Control and Prevention

Generated on: October 03, 2025

This document provides comprehensive documentation for all variables extracted from the CDC BRFSS 2024 dataset, based on the '2024_CDC_DATA_CODEBOOK.HTML' file and mapped to the generated CSV columns in 'BRFSS_2024_Readable_Columns.csv'.

Data Source Information

- **Dataset**: Behavioral Risk Factor Surveillance System (BRFSS) 2024
- **Source**: Centers for Disease Control and Prevention (CDC)
- **Original File**: LLCP2024.XPT (SAS Transport Format)
- **Codebook**: 2024_CDC_DATA_CODEBOOK.HTML
- **Processed CSV**: BRFSS_2024_Readable_Columns.csv

Variable Mapping and Documentation

Core Variables Extracted

The following 29 variables were extracted from the original BRFSS 2024 dataset and given readable column names:

---

1. Patient Identification

CSV Column: `Patient_ID`

Original SAS Variable: `SEQNO`
Data Type: Character (Column 36-45)
Description: Annual Sequence Number - Unique identifier for each survey respondent
Section: Record Identification
Question: Annual Sequence Number
Use: Primary key for linking records and statistical analysis

---

2. Demographics

CSV Column: `Age_Group_5yr`

Original SAS Variable: `_AGEG5YR`
Data Type: Numeric (Calculated Variable)
Description: Age groups categorized in 5-year intervals
Section: Calculated Demographics
Question: Derived from reported age
Categories:

- 18-24 years
- 25-29 years
- 30-34 years
- (continues in 5-year increments)
- 80+ years

CSV Column: `Sex`

Original SAS Variable: `SEXVAR`
Data Type: Numeric (Column 88)
Description: Sex of Respondent
Section: Respondent Sex
Question: Sex of Respondent (determined from landline/cell phone introductions)
Values:

- 1 = Male
- 2 = Female

CSV Column: `Household_Income_Category`

Original SAS Variable: `INCOME3`
Data Type: Numeric (Column 204-205)
Description: Annual household income from all sources
Section: Demographics (Core Section 7)
Question: "Is your annual household income from all sources:"
Categories:

- Less than $15,000
- $15,000 to less than $25,000
- $25,000 to less than $35,000

- $35,000 to less than $50,000
- $50,000 to less than $100,000
- $100,000 to less than $200,000
- $200,000 or more
- Refused/Don't know

CSV Column: `Weight_Pounds`

Original SAS Variable: `WEIGHT2`
Data Type: Numeric (Column 207-210)
Description: Reported weight in pounds without shoes
Section: Demographics (Core Section 7)
Question: "About how much do you weigh without shoes?"
Note: If respondent answers in metrics, 9 is placed in first column

CSV Column: `Height_Feet_Inches`

Original SAS Variable: `HEIGHT3`
Data Type: Numeric (Column 211-214)
Description: Reported height in feet and inches without shoes
Section: Demographics (Core Section 7)
Question: "About how tall are you without shoes?"
Note: If respondent answers in metrics, 9 is placed in first column

---

_

3. Health Status

CSV Column: `General_Health_Status`

Original SAS Variable: `GENHLTH`
Data Type: Numeric (Column 101)
Description: Self-reported general health status
Section: Health Status (Core Section 1)
Question: "Would you say that in general your health is:"
Values:

- 1 = Excellent
- 2 = Very good
- 3 = Good

- 4 = Fair
- 5 = Poor

CSV Column: `Mental_Health_Days`

Original SAS Variable: `MENTHLTH`

Data Type: Numeric (Column 104-105)

Description: Number of days mental health was not good in past 30 days

Section: Healthy Days (Core Section 2)

Question: "Now thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?"

Range: 1-30 days, 88 = None, 77 = Don't know/Not sure, 99 = Refused

CSV Column: `Poor_Physical_Health_Days`

Original SAS Variable: `POORHLTH`

Data Type: Numeric (Column 106-107)

Description: Days poor physical or mental health kept from usual activities

Section: Healthy Days (Core Section 2)

Question: "During the past 30 days, for about how many days did poor physical or mental health keep you from doing your usual activities, such as self-care, work, or recreation?"

Range: 1-30 days, 88 = None, 77 = Don't know/Not sure, 99 = Refused

---

_

4. Chronic Health Conditions

CSV Column: `Diabetes_Status`

Original SAS Variable: `DIABETE4`

Data Type: Numeric (Column 149)

Description: Ever told you had diabetes

Section: Chronic Health Conditions (Core Section 6)

Question: "(Ever told) (you had) diabetes?"

Values:

- 1 = Yes
- 2 = Yes, but female told only during pregnancy
- 3 = No
- 4 = No, pre-diabetes or borderline diabetes
- 7 = Don't know/Not sure

- 9 = Refused

CSV Column: `CVD_Stroke_History`

Original SAS Variable: `CVDSTRK3`
Data Type: Numeric (Column 140)
Description: Ever diagnosed with a stroke
Section: Chronic Health Conditions (Core Section 6)
Question: "(Ever told) (you had) a stroke."
Values:

- 1 = Yes
- 2 = No
- 7 = Don't know/Not sure
- 9 = Refused

CSV Column: `Heart_Attack_History`

Original SAS Variable: `CVDINFR4`
Data Type: Numeric (Column 138)
Description: Ever diagnosed with heart attack
Section: Chronic Health Conditions (Core Section 6)
Question: "(Ever told) you had a heart attack, also called a myocardial infarction?"
Values:

- 1 = Yes
- 2 = No
- 7 = Don't know/Not sure
- 9 = Refused

CSV Column: `Coronary_Heart_Disease`

Original SAS Variable: `CVDCRHD4`
Data Type: Numeric (Column 139)
Description: Ever diagnosed with angina or coronary heart disease
Section: Chronic Health Conditions (Core Section 6)
Question: "(Ever told) (you had) angina or coronary heart disease?"
Values:

- 1 = Yes
- 2 = No
- 7 = Don't know/Not sure
- 9 = Refused

---

5. Tobacco Use

CSV Column: `Smoked_100_Cigarettes_Lifetime`

Original SAS Variable: `SMOKE100`

Data Type: Numeric (Column 241)

Description: Smoked at least 100 cigarettes in entire life

Section: Tobacco Use (Core Section 11)

Question: "Have you smoked at least 100 cigarettes in your entire life? [Note: 5 packs = 100 cigarettes]"

Values:

- 1 = Yes
- 2 = No
- 7 = Don't know/Not sure
- 9 = Refused

CSV Column: `Smoking_Frequency`

Original SAS Variable: `SMOKDAY2`

Data Type: Numeric (Column 242)

Description: Frequency of days now smoking

Section: Tobacco Use (Core Section 11)

Question: "Do you now smoke cigarettes every day, some days, or not at all?"

Values:

- 1 = Every day
- 2 = Some days
- 3 = Not at all
- 7 = Don't know/Not sure
- 9 = Refused

CSV Column: `Smoking_Status`

Original SAS Variable: `_RFSMOK3`

Data Type: Numeric (Calculated Variable)

Description: Current smoking status (calculated variable)

Section: Calculated Risk Factors

Values:

- 1 = Current smoker
- 2 = Former smoker
- 3 = Never smoked

- 9 = Don't know/Refused/Missing

---

6. Alcohol Consumption

CSV Column: `Alcohol_Days_Per_Month`

Original SAS Variable: `ALCDAY4`
Data Type: Numeric (Column varies)
Description: Days per month when alcohol was consumed
Section: Alcohol Consumption
Question: Number of days in past 30 days when had alcoholic beverages
Range: 101-199 = Days per week, 201-299 = Days per month, 888 = No drinks, 777 = Don't know, 999 = Refused

CSV Column: `Binge_Drinking_Episodes`

Original SAS Variable: `DRNK3GE5`
Data Type: Numeric (Column varies)
Description: Binge drinking frequency (5+ drinks for men, 4+ for women)
Section: Alcohol Consumption
Question: How many times in past 30 days had 5 or more drinks (men) or 4 or more drinks (women)
Range: 1-76 times, 88 = None, 77 = Don't know, 99 = Refused

CSV Column: `Max_Drinks_Single_Occasion`

Original SAS Variable: `MAXDRNKS`
Data Type: Numeric (Column varies)
Description: Maximum number of drinks on any occasion in past 30 days
Section: Alcohol Consumption
Question: Maximum number of alcoholic drinks consumed on any single occasion
Range: 1-76 drinks, 77 = Don't know, 99 = Refused

---

7. Physical Activity and Exercise

CSV Column: `Exercise_Past_30_Days`

Original SAS Variable: `EXERANY2`

Data Type: Numeric (Column 113)

Description: Exercise in past 30 days

Section: Exercise (Physical Activity) (Core Section 4)

Question: "During the past month, other than your regular job, did you participate in any physical activities or exercises such as running, calisthenics, golf, gardening, or walking for exercise?"

Values:

- 1 = Yes
- 2 = No
- 7 = Don't know/Not sure
- 9 = Refused

CSV Column: `Exercise_Guidelines_Met`

Original SAS Variable: `_TOTINDA`

Data Type: Numeric (Calculated Variable)

Description: Meets aerobic physical activity guidelines

Section: Calculated Physical Activity

Values:

- 1 = Meets guidelines
- 2 = Does not meet guidelines
- 9 = Don't know/Refused/Missing

CSV Column: `Difficulty_Walking_Stairs`

Original SAS Variable: `DIFFWALK`

Data Type: Numeric (Column 218)

Description: Difficulty walking or climbing stairs

Section: Disability (Core Section 8)

Question: "Do you have serious difficulty walking or climbing stairs?"

Values:

- 1 = Yes
- 2 = No
- 7 = Don't know/Not sure
- 9 = Refused

---

8. Body Mass Index (BMI)

CSV Column: `BMI_Category`

Original SAS Variable: `_BMI5`
Data Type: Numeric (Calculated Variable)
Description: Body Mass Index categories
Section: Calculated BMI
Categories:

- Underweight (BMI < 18.5)
- Normal weight (BMI 18.5-24.9)
- Overweight (BMI 25.0-29.9)
- Obese (BMI ≥ 30.0)

CSV Column: `BMI_Category_Alt`

Original SAS Variable: `_BMI5CAT`
Data Type: Numeric (Calculated Variable)
Description: Alternative BMI categorization
Section: Calculated BMI
Note: Alternative categorization system for BMI ranges

CSV Column: `BMI_Risk_Factor`

Original SAS Variable: `_RFBMI5`
Data Type: Numeric (Calculated Variable)
Description: BMI risk factor classification
Section: Calculated Risk Factors
Values:

- 1 = Normal weight
- 2 = Overweight or obese
- 9 = Don't know/Refused/Missing

---

9. Preventive Care

CSV Column: `Flu_Vaccination`

Original SAS Variable: `FLUSHOT7`
Data Type: Numeric (Column varies)
Description: Received flu vaccination in past 12 months
Section: Immunization
Question: "During the past 12 months, have you had either a flu shot or a flu vaccine that was sprayed in your nose?"
Values:

- 1 = Yes
- 2 = No
- 7 = Don't know/Not sure
- 9 = Refused

---

10. Geographic and Survey Administration

CSV Column: `State_Code`

Original SAS Variable: `_STATE`
Data Type: Numeric (Column 1-2)
Description: State FIPS Code
Section: Record Identification
Question: State FIPS Code
Use: Identifies the state where the survey was conducted
Range: Standard FIPS state codes (01-56)

CSV Column: `Primary_Sampling_Unit`

Original SAS Variable: `_PSU`
Data Type: Numeric (Column 36-45)
Description: Primary Sampling Unit (Equal to Annual Sequence Number)
Section: Record Identification
Question: Primary Sampling Unit for statistical sampling
Use: Used for complex survey design analysis and weighting

CSV Column: `Regional_Health_Prevalence`

Original SAS Variable: `_RFHLTH`
Data Type: Numeric (Calculated Variable)
Description: Regional health prevalence indicator
Section: Calculated Risk Factors
Use: Derived measure for regional health comparisons

Data Quality and Missing Values

Standard Missing Value Codes Across Variables:

- **7 or 77**: Don't know/Not sure
- **9 or 99**: Refused to answer
- **88**: None/Never (for count variables)
- **Blank**: Not applicable/System missing

Survey Design Considerations:

- **Sample Design**: Complex multistage probability sampling
- **Weighting**: Data should be weighted for population estimates
- **Geographic Coverage**: All 50 states, DC, and territories
- **Survey Mode**: Landline and cellular telephone interviews

---

Usage Notes

For Analysis:

1. **Always use appropriate survey weights** when making population estimates
2. **Account for complex survey design** in statistical analyses
3. **Consider missing data patterns** when interpreting results
4. **State-level estimates** require adequate sample sizes

Variable Relationships:

- `Patient_ID` and `Primary_Sampling_Unit` are identical (both use SEQNO)
- BMI variables (`BMI_Category`, `BMI_Category_Alt`, `BMI_Risk_Factor`) are calculated from height and weight
- `Smoking_Status` is derived from `Smoked_100_Cigarettes_Lifetime` and `Smoking_Frequency`
- Age groupings are calculated from individual age responses

Data Linkage:

- Use `Patient_ID` (SEQNO) for linking with other BRFSS modules
- `State_Code` for linking with state-level geographic or policy data
- Survey year is 2024 for temporal analyses

References

- **CDC BRFSS Website**: https://www.cdc.gov/brfss/
- **2024 BRFSS Overview**: https://www.cdc.gov/brfss/annual_data/annual_2024.html
- **BRFSS Methodology**: https://www.cdc.gov/brfss/data_documentation/index.htm
- **Original Codebook**: 2024_CDC_DATA_CODEBOOK.HTML

This documentation was generated based on the CDC BRFSS 2024 Codebook and the variable extraction process documented in `xpt_to_csv_data_sample.py`. For the most current information, always refer to the official CDC BRFSS documentation.