



Montana BRFSS Data User Guide

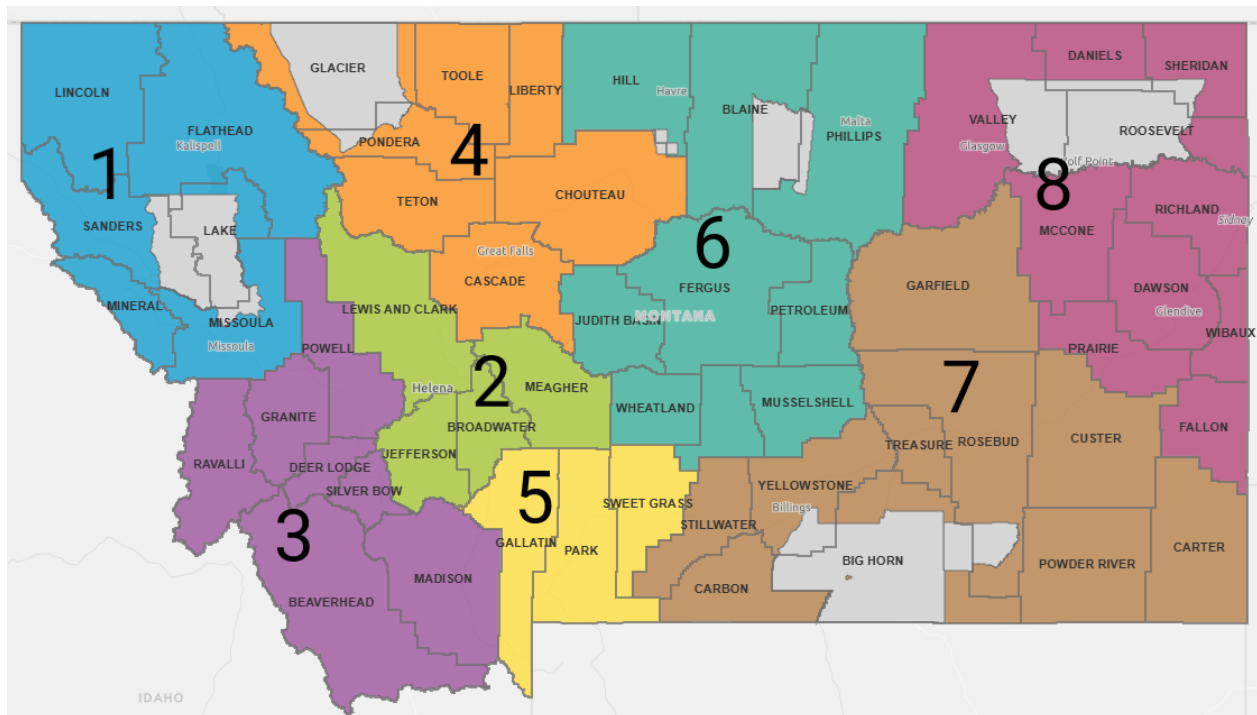
The Behavioral Risk Factor Surveillance System (BRFSS) is a state-based telephone survey that has gathered information from non-institutionalized Montana adults since 1984. The primary focus of these surveys is on behaviors that are linked with the leading causes of death. BRFSS is the only source for estimating the prevalence of behaviors and chronic conditions at the state level. The MT BRFSS is supported by cooperative agreement with the [Centers for Disease Control and Prevention \(CDC\)](#) and is part of a nation-wide system to collect comparable data across all states as well as allow each state to customize the survey for their unique needs.

Sampling and Weighting Methodology

BRFSS is conducted continuously using Random Digit Dialing (RDD) techniques for both landline telephones and cell phones. The sampling design consists of two frames, one for landline and one for cellular. The landline sample is taken from listed and unlisted numbers at a ratio of 1:4. The cellular frame is entirely unlisted. In MT we currently strive for 70% of completed surveys to come from the cellular frame.

Since landline telephones are often shared, household sampling is used in the landline telephone sample. Household sampling requires interviewers to collect information on the number of adults living in a residence and then select randomly from all eligible adults (see [questionnaire](#)). Cell phone respondents are treated as single adult households and therefore do not require household sampling. The samples are fully overlapping, so that any eligible person in the landline frame may also be eligible in the cell phone frame. Note that the BRFSS is a sample with replacement. It is possible, therefore, for a single household/respondent to be eligible and appear in a sample more than once within a year.

The target sample size can change each year, depending on budget, but we strive to have at least 6,000 completed interviews each year. MT stratified our sample by the 5 [health planning regions](#) with a sixth stratum to oversample counties with American Indian reservations through 2024 data collection. Starting in 2025, we will stratify by a different 8 region scheme that will allow more granular sub-state analysis and still include the American Indian oversample. The map below shows which counties are included in each region with American Indian reservation areas in light gray.



CDC weights each state’s data to account for the probably of selection (the design weight) and then rakes the design weight to be representative of the state’s non-institutionalized adult population (the final weight). The MT BRFSS raking methodology includes 12 margins:

- sex by age
- race/ethnicity
- education
- marital status
- tenure (rent/own home)
- sex by race/ethnicity
- age group by race/ethnicity
- phone ownership
- region
- region by age group
- region by sex
- and region by race/ethnicity.

This weighting methodology and the inclusion of cellular numbers into the sample were adopted in 2011. These two large methodological changes caused all previous years to not be directly comparable to 2011 forward. All trend lines should be broken between 2010 and 2011. For a full explanation of the changes implemented in 2011 and the effects on estimates see this [MMWR article from June 2012](#).

For full details of the weighting formula refer to each year’s documentation on the [CDC Annual Survey Data](#) page.

BRFSS Questionnaire Composition

The BRFSS questionnaire has three main sections: Core, Optional modules, and State-added questions.

The Core is set by CDC each year and must be asked exactly as written in every state. Some questions are included every year and are called the Fixed Core and other questions are rotated every other year and are called the Rotating Core. When a question is included in the Core you will be able to compare estimates for MT to other states and to the national median. Estimates from core questions are also available on the [CDC Prevalence and Trends Website](#).

Optional modules are developed and supported by CDC programs, but they are not required to be asked in every state. If the entire, unchanged optional module is used than it will be comparable across the states collecting data and the data will be processed by CDC. If an optional module is changed in any way, the questions will be considered state-added questions. To see which states have included a specific optional module in a specific year refer to the [CDC Annual Survey Data](#) page and consult the documentation for the year of interest. Below is part of the 2022 BRFSS Modules Used by Category report. Modules Used by State is also available for each year.

CDC – BRFSS – 2022 BRFSS Modules Used by Category

[Print](#)

Modules by Category				
Module	Description	Data Set	Data Weight	State(s)
Adverse Childhood Experiences	Combined Land Line and Cell Phone data	LLCP2022	_LLCPWT	Arkansas, Florida, Iowa, Nevada, North Dakota, Oregon, South Dakota, Virginia
	Combined Land Line and Cell Phone data,	LLCP22V1	_LCPWTV1	Arizona, Ohio

State-added questions are unique to each state and will not be comparable to any other states. State-added questions are also not included in the code book provided by CDC. The data for state-added questions must be parsed into individual variables by the BRFSS epidemiologist each year and variable names are assigned at that time. MT BRFSS staff try to use variable names consistently over time for the same state-added questions but the system for naming these variables has changed through various staff changes over the years. For

information about the variable names and coding of state-added questions contact MT BRFSS staff at mtbrfss@mt.gov.

States may choose to split their questionnaire. With a split questionnaire, all respondents are asked the core questions but the optional modules and state added questions are split between two (or three) different tracks. This allows more optional modules and state added questions to be included without making the questionnaire too long for any individual respondent. However, it also means you are reducing the sample size by half (or a third) for all questions that are split. It also requires the use of different final weights when analyzing the questions that were split. Montana has only split our questionnaire one time, in 2013. If you need to use 2013 data refer to the [2013 CDC documentation](#) for how to use the data from multiple version questionnaires.

Data Dictionary and Code Books

All needed documentation for each year of data is available on the [CDC Annual Survey Data](#) website. The public use data files available here include all core and optional module data for all participating states and territories each year. To filter the data to only Montana responses, select data where `_STATE=30`, the FIPS code for Montana. If you are interested in a state added question that is not available in the public use data, you may request the data by completing our [Access Request Form](#). Requesters must log into the form using the the "Citizen login". If you do not already have a State of Montana Okta account, you may create an account [here](#).

Analysis Methods

BRFSS is a complex sample survey so analyzing the data requires specific procedures to account for the sample design and to incorporate the final weight. As such, only certain statistical procedures can be used. Below is list of the complex sample survey procedures available in some of the most common statistical software packages.

- **SAS** (version 9.0 or higher) – *proc surveyfreq, surveymeans, surveyreg, surveylogistic*
- **SPSS** (with Complex Samples Add-in) - *CSPlan then - CSDescriptives, CSTables, CSTabulate, CSGLM, CSLogistic*
- **Sudaan** – *proc crosstab, descript, ratio, regress, logistic*
- **Stata** – *svyset, then svy: mean, proportion, ratio, total, regress, logit, etc.*
- **R** – survey package

The majority of analysts at DPHHS use SAS software so the remaining discussion will focus on SAS procedures, but the concepts can be applied to any software.

Recoding and Calculating Variables

Nearly every variable included in BRFSS data has code options for "don't know/unsure" and "refused", usually 7 and 9 respectively. It is BRFSS convention to eliminate don't know and refused responses from the denominator for analysis and treat these responses the same as missing or if the question was not asked. Please note that the percent and weighted percent listed in the code book DO include don't know and refused response options in the denominator so they will be slightly lower than results that are published by MT BRFSS or CDC. To ensure these responses are properly left out of denominators, it may be useful to recode analysis variables to set don't know and refused responses to missing in the data step. There are several options for accomplishing this task, but I find one of the most efficient methods is by use of arrays. In the following example all don't know and refused responses are set to missing for the needed variables using separate arrays for each code structure (77 and 99, 7 and 9, or just 9). Also note that one variable, `_AGEG5YR`, has a non-standard missing code so a single 'if' statement is used to recode.

```
libname brfs22 '\\state.mt.ads\HHS\Shared\PHSD\DIV-SHARE\BRFSSData\2022 MT BRFSS\2022  
Data Files';
```

```
data work.mt2022;  
set brfs22.mt2022_brfss_v110723;
```

```

if _AGEG5YR=14 then _AGEG5YR=.;

array twodigit (3) PRIMINSR INCOME3 MARIJAN1;
do i=1 to 3;
  if twodigit(i) in (77, 99) then twodigit(i)=.;
end;

array rfvar (14) _PHYS14D _MENT14D _HLTHPLN _RFHLTH _MICHD _CASTHM1 _RFDRHV8 _RFBING6
_CURECI2 _RFSMOK3 VETERAN3 _EDUCAG MARITAL EMPLOY1;
do i=1 to 14;
  if rfvar(i)=9 then rfvar(i)=.;
end;

array hlthydy (10) PERSDOC3 MEDCOST1 CHECKUP1 DECIDE DIFFWALK DIFFDRES DIFFALON BLIND
DEAF DRNKANY6;
do i=1 to 10;
  if hlthydy(i) in (7, 9) then hlthydy(i)=.;
end;

```

While it is advised to remove “don’t know/unsure” or “refused” from the analysis, analysts may choose to include them. In the event an analyst decides that including don’t know or refused responses is relevant to their analysis, they should be sure to clearly state this decision and the reasoning behind it when reporting the results.

Calculated variables use the results of multiple questions to determine a measure of interest. CDC provides many calculated variables using the Core and Optional Module questions. Analysts should always look through the provided calculated variables before attempting to calculate the needed measure independently. This will ensure consistent calculation methods and comparability between states and the national median. If CDC or MT BRFSS does not provide a calculated variable to meet your needs here are a few considerations.

- **Be very mindful of your target denominator.** Is the measure intended to measure prevalence among all adults or among a specific demographic group? Then determine if the skip patterns used for your questions of interest match your target demographic. Is anyone left out of the skip pattern? Is anyone included that shouldn’t be?
- **Ensure the statistical procedure you are using handles missing data in the way you would prefer.** Do you want observations with a missing value for any of the variables of interest to be left missing in the calculated variable or can they be included if at least one of the variables has a valid response? As an example, consider how you might calculate the number of chronic conditions reported. The chronic health conditions (CHC) section of the core asks about the lifetime prevalence of 11 chronic conditions. After recoding each condition to be 0=condition not present and 1=condition is present, you have two options in SAS:
 - **the sum function:** `numb_chc = sum(DIABETE4, CHCKDNY2, HAVARTH5, CHCCOPD3, CHCSCNCR, CHCOCNCR, CVDSTRK3, CVDCRHD4, CVDINFR4, ADDEPEV3, _casthm1);`
 - **using the + operator to add all the CHC variables:** `numb_chc = DIABETE4 + CHCKDNY2 + HAVARTH5 + CHCCOPD3 + CHCSCNCR + CHCOCNCR + CVDSTRK3 + CVDCRHD4 + CVDINFR4 + ADDEPEV3 + _casthm1;`

These two options seem equivalent, but the sum function will return the sum of non-missing variables even if some of the variables are missing and the + operator will return a missing value if any of the variables are missing. Either may be appropriate for your analysis but it is essential that you are aware of the distinction and intentionally choose what is correct for you.

- **Check that calculations functioned as expected** by performing simple frequency and crosstab analysis on the input and output variables used.

SAS Survey Procedures

As stated previously, BRFSS data requires the use of complex sample survey procedures to correctly produce weighted estimates and variance tables. To do this SAS survey procedures include additional statements that specify the sample and weighting variables. These statements are the same across all SAS survey procedures: [surveyfreq](#), [surveymeans](#), [surveyreg](#), [surveylogistic](#). I will show an example using *surveyfreq* as it is the most common of these. This code uses *surveyfreq* to find the prevalence of ever being diagnosed with arthritis among all adults and includes confidence limits (CL) in the output.

```
PROC SURVEYFREQ DATA=work.mt2022 VARMETHOD=TAYLOR;  
STRATA _STSTR;  
CLUSTER _PSU;  
WEIGHT _LLCPWT;  
TABLE HAVARTH5 / CL;  
RUN;
```

The VARMETHOD option determines which variance estimation method should be used. For BRFSS you should specify the Taylor Series. The STRATA and CLUSTER statements are used to specify sampling design. _STSTR should be used in STRATA and _PSU should be used in CLUSTER. The WEIGHT statement is where you specify the weighting variable. In most cases you will use _LLCPWT but there are a few exceptions. For any questions using the random child selection module _CLLCPWT should be used. When analyzing multiple years of data weights will need to be adjusted (see following section). The table statement functions exactly like a regular FREQ procedure. As such, you can modify the table statement to perform cross tabulations just like you would in any FREQ procedure. See the SAS documentation linked in each procedure names above for full details and many helpful options.

Combining Multiple Years

It may be necessary to combine multiple years of data to increase sample size for a particular analysis. Especially if the population of interest is a relatively small proportion of adults (i.e. adults who have been diagnosed with diabetes) or if you want to analyze by multiple demographic subgroups at once (i.e. females aged 18 to 24). **The unweighted denominator must be at least 50 for all analyses.**

When combining multiple years of data, the first consideration is to ensure the data was collected in a comparable way across all years. Refer to the data dictionary and questionnaire for each year of data to be included on the [CDC Annual Survey Data](#) page. If the variable name has remained the same across years you can be assured that the question has not changed. Even if the variable name changes it may be appropriate to still combine the data if the analysts determines that the change(s) to the question are not large enough to create a meaningful difference in the data. However, this should only be done after careful consideration and examining the results of all effected years of data. Please note that variable names sometime change even with no change to question wording because of a change to question order, skip pattern, or a change to a previous question that effected the skip pattern. Be sure you have consulted the questionnaires for the years of interest to understand all these possible changes before determining if data from the two years can be combined. You can also consult the "Comparability of Data CDC" report posted for each survey year on the [CDC Annual Survey Data](#) page.

Once the data is combined the final weight must be adjusted before analysis. Because weights are calculated to be representative of the total adult population for each year combining the unadjusted weights would effectively add the total population for each year together and inflate the weighted frequency (the estimated number of adults effected). Consider the following example. In 2022, the weighted frequency of adults who reported any marijuana use in the past 30 days was about 160,000 (20% of all adults). If this prevalence remained about the same for 3 years and we combined those years of data together the unadjusted weights for all the years combined would produce a weighted frequency of about 480,000 adults although the percent of adults would still be about 20%. To adjust the weights, each year's weight should be multiplied by the fraction of the total

sample size contributed by that year. Let's say we were combining data from 2020, 2021, and 2022. The sample size was 6,315 in 2020, 6,243 in 2021, and 7,048 in 2022 for a total 3-year sample size of 19,606. In this example the final weights for each year would be adjusted as follows.

$$_LLCPWT(2020) * (6,315 / 19,606) = _LLCPWT(2020) * 0.32$$

$$_LLCPWT(2021) * (6,243 / 19,606) = _LLCPWT(2021) * 0.32$$

$$_LLCPWT(2022) * (7,048 / 19,606) = _LLCPWT(2022) * 0.36$$

If the sample size for each year is basically equivalent, you can simply divide the final weight for all years by the number of years included.

Geographic Analysis

BRFSS is primarily designed to produce statewide prevalence estimates and that is how results are nearly always reported. However, since MT also includes the [health planning regions](#) in our sampling strata it is also possible to get estimates for these regions for 2024 data and all previous years. Starting with 2025 data we will be able to analyze the new regions shown in the [Sampling and Weighting Methodology](#) section. The CDC weighting formula also allows analysis by urban rural classification, variable `_URBNRRL`. This variable uses the [2013 NCHS urban-rural classification scheme](#). In Montana we only have 3 levels, Small Metro, Micropolitan, and Non-core (also called Rural). The MT counties included in each level are listed in the table below.

Small Metro	Micropolitan	Non-core
Billings Metropolitan Area <ul style="list-style-type: none"> Yellowstone (111) Carbon (009) Golden Valley (037) 	Kalispell Micropolitan Area <ul style="list-style-type: none"> Flathead (029) 	All remaining counties
Great Falls Metropolitan Area <ul style="list-style-type: none"> Cascade (013) 	Bozeman Micropolitan Area <ul style="list-style-type: none"> Gallatin (031) 	
Missoula Metropolitan Area <ul style="list-style-type: none"> Missoula (063) 	Helena Micropolitan Area <ul style="list-style-type: none"> Jefferson (043) Lewis and Clark (049) 	
	Butte Micropolitan Area <ul style="list-style-type: none"> Silver Bow (093) 	

County-level analysis is only possible if the county has a total sample size of at least 500 in a single year. With this minimum sample size, CDC is able to add margins to adjust the final weight at the county level as well as at the region level, and the final weight should still be representative. County-level estimates are rarely possible in MT with a single year of data. Larger population counties: Flathead, Missoula, Lewis & Clark, Cascade, Silver Bow, Gallatin, and Yellowstone, do sometimes reach this sample size in a single year but less populated counties will not. Additionally, county of residence is not included in the public use data available online. If you need county for a specific project please contact the BRFSS Epidemiologist, mtbrfss@mt.gov, to determine if BRFSS data is the right resource for your project and, if so, for assistance requesting the county data.

CDC also produces small area estimates for all US counties and several types of sub-county areas using a combination of BRFSS data and US Census data. These data are available through the [Places](#) website. Small area estimates are calculated using multilevel statistical modeling and are not a direct estimate of prevalence like BRFSS state-level data is. For more detail about the methods used to calculate Places data please refer to

their [methodology](#). Small area estimates are very useful for needs assessments and program planning. However, they are not appropriate for program evaluation.

Age-adjusting

BRFSS data is most often reported as unadjusted or crude prevalence estimates to reflect the true prevalence in the population of interest. However, when comparing prevalence between groups it is sometimes appropriate to age-adjust. Analysts should consider age-adjusting when 1) the risk factor of interest varies by age AND 2) the two groups being compared have different age distributions. On the 2016 BRFSS data documentation site CDC posted instruction for [direct age adjustment](#) using SUDDAN. This resource refers to 2016 variables, but the methods can be applied to other years of data as well. This [article](#) has detailed instructions and example code for age-adjusting survey data using SAS. Additionally, the [CDC Prevalence and Trends](#) query site offers the option to get age-adjusted rates on all Core questions for the overall adult population across time (where multiple years of data are available).

Suppression Rules

BRFSS suppression rules are all set to ensure the statistical stability of the estimates. There are three ways to check the stability of an estimate. Estimate must meet all three requirements to be reported.

- 1) The unweighted count of the denominator is more than 50
- 2) The half-width of the 95% confidence interval is less than 10
- 3) The relative standard error (RSE) is less than 30%. RSEs are calculated by dividing the standard error by the estimate.

Citations

Montana BRFSS estimates you calculated or pulled from DPHHS dashboards or websites should be cited as follows.

In-text citation: (Montana BRFSS, [year(s)])

Full citation in the reference section: Behavioral Risk Factor Surveillance System, Montana Department of Public Health and Human Services. [year(s)]

Data taken from published reports should cite the original report.

National BRFSS estimates should be cited using the Centers for Disease Control and Prevention (CDC) citation.

Centers for Disease Control and Prevention (CDC). Behavioral Risk Factor Surveillance System Survey Data. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, [appropriate year].

BRFSS Resources

- [CDC Statistical Briefs](#): guidance on conducting analyses for particular subjects.
- [CDC BRFSS Data User Guide](#): information on national BRFSS protocols and general analysis principles.
- [CDC Annual Survey Data](#): National data and documentation for each year from 1988 forward.
- [CDC Prevalence and Trends Data](#): Data query system that provides state-level estimates for core questions by gender, age-group, race/ethnicity, household income, educational attainment, and year. The national median is also available for comparison to state-level estimates.

- [Chronic Disease Indicators](#): provides comparisons between two or more geographic areas (states, metropolitan/micropolitan areas, or regions within states) on chronic conditions and risk behaviors.
- [Places](#): provides health data for small areas including counties, census designated places, census tracts, and ZIP code tabulation areas (ZCTAs) using multilevel statistical models.