Robert Wood Johnson Foundation

shadac
STATE HEALTH ACCESS DATA ASSISTANCE CENTER

MAY 2019

Four Methods for Calculating Income as a Percent of the Federal Poverty Guideline (FPG) in the Behavioral Risk Factor Surveillance System (BRFSS)

Author

Robert Hest, MPP

Research Fellow, State Health Access Data Assistance Center

Categorical vs. Continuous Income Measures

Categorical income measures ask respondents to report their income in several discrete categories, for example: "Is your annual household income from all sources…Less than $10,000, $10,000 to less than $15,000...$75,000 or more, etc."

Continuous income measures ask respondents to report their income as an exact dollar figure, for example: "What was this person's total income during the past 12 months? $___,____,____.00"

Stay updated by visiting us at **www.shadac.org** *or* **click here** *to join our mailing list.*

INTRODUCTION

Many surveys relevant to researchers evaluating the effects of health reform have income measures that lack needed specificity to precisely define the relevant eligible populations, such as those with incomes that make them eligible for Medicaid. Nevertheless, because these surveys are often the best (or only) source for a rich set of measures of health out-comes and health behaviors, researchers have employed various methods in order to best utilize the available income measures to examine potentially eligible populations.

The U.S. Centers for Disease Control and Prevention's (CDC) Behavioral Risk Factor Surveillance System (BRFSS) survey is one such resource; an annual household telephone survey of civilian non-institutionalized adults age 18 years or older that asks respondents about health behaviors, chronic health conditions, and the use of preventive health services. The BRFSS has a relatively large sample size, interviewing more than 400,000 respondents annually, and allows researchers to produce estimates for all 50 states and D.C.

The BRFSS' income measure asks respondents to report their total annual household income within eight possible categories. Because these categories do not align with the federal poverty guideline (FPG) thresholds used to determine eligibility for programs such as Medicaid expansion (up to 138% FPG) or the Affordable Care Act's (ACA) cost sharing reductions (up to 250% FPG) or premium tax credits (up to 400% FPG), this creates a problem for researchers who want to use the BRFSS to study health reform.*

To deal with this issue, researchers have typically chosen to assign a continuous income to the respondent based on the categorical income measure, choosing either the lower bound of each category,[1] the upper bound of each category,[2] or the midpoint of each category.[3] There is no clear consensus in the literature about which approach to assigning continuous income from categorical values is best, and we propose that the most appropriate method depends in part on the analytic issue at hand.

In this brief, we first outline how each method impacts the income distribution in the BRFSS overall and by state. We then use the Current Population Survey (CPS) to evaluate the impact of using different methods to assign continuous income from a categorical income variable. We chose the CPS because the survey is used broadly to report on income, and contains both a categorical and a continuous income variable. As a result, we can compare the impacts of different strategies of assigning continuous income from a categorical variable to actual, continuous income from the same data source. We then summarize our findings from the CPS and their implications for evaluating the impact of different health reform policies (e.g., Medicaid expansion) on health outcomes in the BRFSS.

* The BRFSS's household income measure is also problematic because eligibility for these programs is determined at the family level, not the household level, and the BRFSS does not have a measure of family income. Additionally, research shows that omnibus household income measures such as the one used in the BRFSS lead respondents to underreport their household income relative to aggregated household income measures. This leads to over-estimates of the poverty rate. See Davern et al., "The Effect of Income Question Design in Health Surveys on Family Income, Poverty and Eligibility Estimates."

Assigning Continuous Income in the BRFSS

Table 1 shows the BRFSS income categories and the relevant cut points for each method of assigning continuous income. As Table 1 demonstrates, if a respondent reported that their income was between $10,000 and $15,000, they would be assigned an income of $10,000 using the lower bound method, $15,000 using the upper bound method, or $12,500 using the midpoint method. The uniform distribution method randomly assign respondents an income value within their specified income category (e.g., $0-9,999) to create a uniform distribution of income within each category.

Income categories become wider as they increase. As a result, assignments in higher income categories are less precise. The top category, "$75,000 or more" has no upper bound, so we assigned an artificial upper bound of $100,000 to match the category size of the next lowest category (income $50,000 to less than $75,000).

Table 1. BRFSS Income Categories and Four Methods of Assigning Income

BRFSS Category	Category Size ($)	Lower Bound ($)	Midpoint ($)	Upper Bound ($)	Uniform Distribution ($)
Less than $10,000	10,000	0	5,000	10,000	0-9,999
$10,000 to less than $15,000	5,000	10,000	12,500	15,000	10,000-14,999
$15,000 to less than $20,000	5,000	15,000	17,500	20,000	15,000-19,999
$20,000 to less than $25,000	5,000	20,000	22,500	25,000	20,000-24,999
$25,000 to less than $35,000	10,000	25,000	30,000	35,000	25,000-34,999
$35,000 to less than $50,000	15,000	35,000	42,500	50,000	35,000-49,999
$50,000 to less than $75,000	25,000	50,000	62,500	75,000	50,000-74,999
$75,000 or more	∞	75,000	87,500	100,000	75,000-99,999

Source: SHADAC analysis of the 2017 Behavioral Risk Factor Surveillance System.

Comparing Results of Four Income Assignment Methods in the BRFSS

Table 2 shows the results of potentially implementing each of these four income assignment methods in the BRFSS. As shown, there is substantial variation in the distribution of income across methods, particularly in how the methods affect estimates at the lower- and upper-most income categories. Not surprisingly, the lower bound method results in the largest share of the population with incomes below 100% FPG, while the upper bound method skews the income distribution toward the 401%+ FPG category. The uniform distribution and midpoint method are similar in both their impact on the overall income distribution and difference from the lower and upper bound methods.

Table 2. Distribution of Income (%FPG) by Income Assignment Method, 2017

Income Category (% FPG)	Lower Bound (%)	Uniform Distribution (%)	Midpoint (%)	Upper Bound (%)
0-100	22.6	17.8	16.2	14.5
101-138	9.8	9.3	8.5	8.1
139-250	24.7	21.1	21.7	19.8
251-400	25.9	26.8	29.9	18.2
401+	17.0	25.0	23.7	39.3
Total	100.0	100.0	100.0	100.0

Source: SHADAC analysis of the 2017 Behavioral Risk Factor Surveillance System; Universe: Civilian Non-institutionalized Adults.

Table 3 shows the percent of adults with incomes at or below 100% FPG by state, comparing each income imputation method, again using data from the 2017 BRFSS. The table shows substantial variation in state-level poverty rates, and also substantial state variation in the difference in poverty rates produced by each income imputation method.

The lower bound method produces the highest rate of poverty across all states, followed in descending order by the uniform distribution, midpoint, and upper bound methods. In all states, the uniform distribution method produces a significantly lower poverty rate compared to the lower bound method; in nine states, the midpoint method produces a significantly lower poverty rate compared to the uniform distribution; and in 16 states, the upper bound method produces a significantly lower poverty rate compared to the midpoint method.

There is also substantial variation in the size of the difference between the lower-bound and upper-bound estimates. Massachusetts has the smallest difference at 4.5 percentage points, and West Virginia has the largest difference at 12.0 percentage points. The national average is 8.1 percentage points.

Table 3. Percent of Adults with Incomes at or Below 100% FPG by Income Assignment Method and State, 2017

State	Lower Bound (%)	Uniform Distribution (%)	Midpoint (%)	Upper Bound (%)	Percentage Pt. Difference Upper vs. Lower
Alabama	27.6	19.9*	18.0	16.3	-11.3
Alaska	25.6	19.4*	19.5	18.5	-7.1
Arizona	24.3	18.8*	17.5	15.3*	-9.0
Arkansas	29.8	22.6*	20.6	18.6	-11.2
California	27.6	22.8*	21.2	19.5*	-8.1
Colorado	15.9	12.1*	10.7*	9.1*	-6.8
Connecticut	16.3	12.3*	11.1	9.8	-6.5
Delaware	13.2	9.3*	8.2	7.7	-5.5
District of Columbia	16.8	13.1*	12.1	11.0	-5.8
Florida	16.4	11.6*	10.0	8.6	-7.8
Georgia	20.3	15.3*	13.5	12.5	-7.8
Hawaii	24.8	21.4*	22.0	16.7*	-8.1
Idaho	21.5	15.5*	13.8	11.6*	-9.9
Illinois	21.2	17.2*	15.5	13.8	-7.4
Indiana	22.2	17.4*	16.1	14.0*	-8.2
Iowa	16.9	12.6*	11.3	9.7*	-7.2
Kansas	17.4	13.0*	11.4*	10.1*	-7.3
Kentucky	24.2	19.1*	17.8	16.3	-7.9
Louisiana	31.2	25.5*	23.4	21.4	-9.8
Maine	16.6	11.8*	9.9*	8.8	-7.8
Maryland	16.7	13.2*	12.4	11.1	-5.6
Massachusetts	12.8	9.8*	8.6	8.3	-4.5
Michigan	18.8	13.3*	12.2	10.4	-8.4
Minnesota	9.3	6.1*	5.2	4.7	-4.6
Mississippi	32.8	26.8*	23.6*	21.7	-11.1
Missouri	19.9	14.7*	13.1	11.9	-8.0
Montana	21.0	15.0*	13.2	11.9	-9.1
Nebraska	18.3	13.8*	12.2*	10.8*	-7.5
Nevada	25.9	20.9*	19.3	17.2	-8.7

State	Lower Bound (%)	Uniform Distribution (%)	Midpoint (%)	Upper Bound (%)	Percentage Pt. Difference Upper vs. Lower
New Hampshire	12.1	8.5*	7.4	6.5	-5.6
New Jersey	12.3	9.1*	8.2	7.2	-5.1
New Mexico	32.9	25.4*	23.4	21.2	-11.7
New York	24.5	19.8*	18.1*	16.5	-8.0
North Carolina	22.1	16.7*	14.6	12.9	-9.2
North Dakota	15.3	11.4*	10.1	8.8	-6.5
Ohio	16.9	12.3*	10.5	9.4	-7.5
Oklahoma	24.3	19.4*	17.8	15.8	-8.5
Oregon	21.1	16.0*	14.2	11.4*	-9.7
Pennsylvania	17.9	13.3*	11.6*	10.6	-7.3
Rhode Island	20.8	15.3*	13.7	12.4	-8.4
South Carolina	24.7	18.7*	16.9*	15.1*	-9.6
South Dakota	16.3	11.4*	9.9	8.8	-7.5
Tennessee	26.4	21.2*	18.9	16.4*	-10.0
Texas	29.9	24.3*	22.6	20.1*	-9.8
Utah	15.5	12.0*	10.8	9.0*	-6.5
Vermont	15.6	11.3*	9.9	8.6	-7.0
Virginia	14.4	11.0*	9.7	8.9	-5.5
Washington	17.6	14.0*	12.7	11.2*	-6.4
West Virginia	28.9	21.5*	19.2*	16.9*	-12.0
Wisconsin	15.8	11.5*	9.7	8.5	-7.3
Wyoming	11.6	7.7*	6.4	6.0	-5.6
United States	22.6	17.8*	16.2*	14.5*	-8.1

** Difference from column to the left statistically significant at 95% level.*
Source: SHADAC analysis of the 2017 Behavioral Risk Factor Surveillance System; Universe: Civilian Non-institutionalized Adults.

Results in the CPS of Four Methods of Assigning Income

Although it is helpful to understand the variation across these methods within the BRFSS, we also wanted to compare the outcomes of each method to the outcome produced by a measure of continuous income. The CPS provides a helpful data source to make this comparison because it contains both categorical and continuous income information. In order to evaluate the impact of each method, we first recoded the CPS's categorical income category to match the BRFSS categories. We then assigned a continuous income value based on each of the lower, midpoint, upper bound, and uniform distribution methods as described previously. Finally, we assigned FPG based on family size, calculated income as a percent of FPG, and compared the results to income as a percent of FPG calculated based on the actual continuous income measure observed in the CPS data set, which for the purposes of this analysis will be considered the "true" or reference income distribution.

Table 4 shows the results of our analysis by FPG categories typically used to determine eligibility for Medicaid or ACA coverage subsidies. As the table demonstrates, similar to our findings in the BRFSS, there is substantial variation across methods on the overall income distribution. However, there is also variation in how each method measures up against the "true" distribution based on continuous income.

Overall accuracy: In terms of coming closest to matching the reference distribution across the majority of FPG categories, the upper-bound method performs best. The upper-bound method outperforms the other methods in estimating the share of the population in the 251-400% FPG and 401%+ FPG categories and comes nearly as close to matching the reference distribution's estimates of the 101-138% and 139-250% categories, as do the other methods. However, it performs poorly relative to all but the lower-bound method in matching the lowest FPG categories.

Accuracy for Medicaid & Cost Sharing-eligible Categories: The uniform distribution and midpoint methods perform best in estimating the 0-100% and 100-138% FPG populations. They are also similar in their accuracy for matching the reference distribution in the share of the population potentially eligible for cost sharing subsidies (139-250% FPG).

Table 4. Distribution of Income (% FPG) by Income Assignment Method

Income Assignment Method	Percent	Difference from Reference (Percentage Point)
Lower Bound		
0-100% FPG (Medicaid)	14.6	3.7
101-138% FPG (Medicaid)	7.4	1.6
139-250% FPG (Cost Sharing)	26.6	7.9
251-400% FPG (Premium Subsidies)	29.9	9.7
401%+ FPG	21.4	-22.8
Total	100.0	
Uniform Distribution		
0-100% FPG	11.4	0.5
101-138% FPG	6.7	0.8
139-250% FPG	20.5	1.7
251-400% FPG	30.8	10.6
401%+ FPG	30.7	-13.6
Total	100.0	
Midpoint		
0-100% FPG	10.1	-0.9
101-138% FPG	6.2	0.4
139-250% FPG	20.8	2.0
251-400% FPG	33.6	13.4
401%+ FPG	29.3	-15.0
Total	100.0	
Upper Bound		
0-100% FPG	9.3	-1.7
101-138% FPG	5.3	-0.5
139-250% FPG	17.2	-1.6
251-400% FPG	20.8	0.6
401%+ FPG	47.5	3.2
Total	100.0	
Reference Income		
0-100% FPG	10.9	
101-138% FPG	5.8	
139-250% FPG	18.8	
251-400% FPG	20.2	
401%+ FPG	44.3	
Total	100.0	

Definition: 2017 family income as a percent of the 2017 federal poverty guidelines (FPG) for the civilian, non-institutionalized population of adults age 18 or older.
Source: SHADAC analysis of the 2018 Current Population Survey's Annual Social and Economic Supplements (CPS-ASEC) public use microdata file.

CONCLUSION

As these findings demonstrate, the best method for assigning continuous income from categorical income depends, in part, on which populations are most relevant for the analysis. If the goal is to assign income to most closely match the overall distribution of income derived from actual continuous income, the upper-bound method is best. However, for analyses focused on lower-income populations, either the uniform distribution or midpoint methods perform better. In other words, if researchers are focusing on questions about the impact of Medicaid expansion or cost-sharing subsidies on health outcomes in the BRFSS, they should probably consider implementing the uniform distribution or midpoint methods to derive continuous income. Our results suggest that the lower-bound method performs poorly across the income distribution.

References

Benitez, Joseph A., and Eric E. Seiber. (2018). US Health Care Reform and Rural America: Results From the ACA's Medicaid Expansions. *Journal of Rural Health*, *34*(2), 213–22. doi:10.1111/jrh.12284.

Centers for Disease Control and Prevention (CDC). (2015). BRFSS Statistical Brief on the Health Care Access Module, 2013 and 2014 [PDF file]. Retrieved from https://www.cdc.gov/brfss/data_documentation/pdf/2013-2014_hcs.pdf.

Davern, Michael, Holly Rodin, Timothy J. Beebe, and Kathleen Thiede Call. (2005). The Effect of Income Question Design in Health Surveys on Family Income, Poverty and Eligibility Estimates. *Health Services Research*, *40*(5), 1534-1552. doi:10.1111/j.1475-6773.2005.00416.x.

Hawaii Health Data Warehouse. (n.d.) HHDW BRFSS Poverty Level Methodology [PDF file] Retrieved from http://www.hhdw.org/wp-content/uploads/BRFSS-Poverty-Level-Methodology.pdf.

Levy, Helen, and David Meltzer. (2007). The Impact of Health Insurance on Health. *Annual Review of Public Health*, *29*(1), 399-409. doi:10.1146/annurev.publhealth.28.021406.144042.

Lopez, Russ. (2004). Income Inequality and Self-Rated Health in US Metropolitan Areas: A Multi-Level Analysis. *Social Science and Medicine*, *59*(12), 2409-2419. doi:10.1016/j.socscimed.2004.03.033.

Nasseh, Kamyar, and Marko Vujicic. (2014). The Effect of Growing Income Disparities on U.S. Adults' Dental Care Utilization. *Journal of the American Dental Association*, *145*(5), 435-442. doi:10.14219/jada.2014.1.

Sabik, Lindsay M., Wafa W. Tarazi, and Cathy J. Bradley. (2015). State Medicaid Expansion Decisions and Disparities in Women's Cancer Screening. *American Journal of Preventive Medicine*, *48*(1), 98-103. doi:10.1016/j.amepre.2014.08.015.

Simon, Kosali, Aparna Soni, and John Cawley. (2017). The Impact of Health Insurance on Preventive Care and Health Behaviors: Evidence from the First Two Years of the ACA Medicaid Expansions. *Journal of Policy Analysis and Management*, *36*(2), 390-417. doi:10.1002/pam.21972.

State Health Access Data Assistance Center. (2019). FPG vs. FPL: What's the Difference? *SHADAC Blog*, Retrieved from https://www.shadac.org/news/fpg-vs-fpl-whats-difference.

Yue, Dahai, Petra W. Rasmussen, and Ninez A. Ponce. (2018). Racial/Ethnic Differential Effects of Medicaid Expansion on Health Care Access. *Health Services Research*, *53*(5), 3640-3656. doi:10.1111/1475-6773.12834.

APPENDIX

Implementing Income Imputation Methods in the BRFSS

Each of these strategies were implemented using microdata from the publically available 2017 BRFSS data, using data from respondents residing in the contiguous United States, Alaska, and Hawaii. All analysis was conducted using STATA version 15.1.

Assigning Continuous Income

The first step in the analysis is to assign each respondent a continuous income value based on their response to the BRFSS categorical income measure. Respondents with missing income information (16.7% of sample) were dropped from the analysis. The code below shows how continuous incomes were assigned using the midpoint method. The same code was used for the lower-bound and upper-bound strategies, but using the income values shown in Table 1.

```
generate income_mid = .
replace income_mid = 5000 if INCOME2 == 1
replace income_mid = 12500 if INCOME2 == 2
replace income_mid = 17500 if INCOME2 == 3
replace income_mid = 22500 if INCOME2 == 4
replace income_mid = 30000 if INCOME2 == 5
replace income_mid = 42500 if INCOME2 == 6
replace income_mid = 62500 if INCOME2 == 7
replace income_mid = 87500 if INCOME2 == 8
```

The code below shows how the uniform distribution strategy was implemented. The STATA random number function "runiform" was used to randomly assign respondents an income value from the uniform distribution based on their relevant income category. Using the "set seed" function allows the same results to be later reproduced. In this case the seed was set to a random, four-digit number generated by www.random.org.

```
set seed 3849
gen income_unif = .
replace income_unif = runiform(0,9999.99) if INCOME2 == 1
replace income_unif = runiform(10000,14999.99) if INCOME2 == 2
replace income_unif = runiform(15000,19999.99) if INCOME2 == 3
replace income_unif = runiform(20000,24999.99) if INCOME2 == 4
replace income_unif = runiform(25000,34999.99) if INCOME2 == 5
replace income_unif = runiform(35000,49999.99) if INCOME2 == 6
replace income_unif = runiform(50000,74999.99) if INCOME2 == 7
replace income_unif = runiform(75000,100000) if INCOME2 == 8
```

Calculating Household Size

An individual's FPG is based on their family size. Because the BRFSS lacks a measure of family size, this analysis uses household size to determine the FPG.

The BRFSS has three variables that are combined to determine household size: HHADULT (the number of adults in the household for cellphone respondents), NUMADULT (the number of adults in the household for landline respondents), and CHILDREN (the number of children under 18 years of age in the household for all respondents). Observations with missing information for any of these variables (12.2% of the sample) or with more than 14 individuals in the household [†] were dropped from the analysis.

[†] This follows the methodology applied by the CDC when determining household size for the purposes of calculating income as a percent of poverty. See the CDC brief, "BRFSS Statistical Brief on the Health Care Access Module, 2013 and 2014."

The code below first cleans the HHADULT and CHILDREN variables, setting non-responses to "missing" and recoding the "no children" response to "0" for the CHILDREN variable. The code then sums each respondent's NUMADULT, cleaned HHADULT, and cleaned CHILDREN variables, setting the resulting variable to "missing" if any of the three input variables are missing. This variable is then set to "missing" if it is equal to 0 or if is greater than 14 or if either NUMADULT or HHADULT_CLEAN are missing.

```
gen hhadult_clean = .
replace hhadult_clean = hhadult if hhadult != 77 & hhadult != 99

gen children_clean = .
replace children_clean = children if children != 99
replace children_clean = 0 if children == 88

egen hh_size = rowtotal(numadult hhadult_clean children_clean), missing
replace hh_size = . if hh_size == 0 | hh_size > 14
replace hh_size = . if hhadult_clean == . & numadult == .
```

Calculating Income as a Percent of the Federal Poverty Guideline (FPG)

In addition to family size, an individual's FPG is also determined by their state of residence. The FPG is higher in Alaska and Hawaii than it is in the 48 contiguous states and D.C.

The code below creates the STATEGROUP and YEAR variable to merge a base and increment FPG variable for each respondent in the dataset. These FPG base and increment variables are contained in a separate dataset based on the Federal Poverty Guidelines (FPG) released by the U.S. Department of Health and Human Services.

```
gen stategroup = .
replace stategroup = 1 if _STATE != 2 & _STATE != 15 & !mi(_STATE)
replace stategroup = 2 if _STATE == 2 // Alaska
replace stategroup = 3 if _STATE == 15 // Hawaii

gen year = 2017

merge m:1 year stategroup using "C:\...\HHS Poverty Guidelines.dta"
drop if _merge == 2
drop _merge year stategroup
```

The section of code below uses the merged base and increment variables to create a FPG for each respondent based on their state and household size. The FPG is created by adding the base FPG and an additional increment FPG for each additional household member.

```
gen fpg_guideline = .
replace fpg_guideline = base + (increment * (hh_size - 1)) if !mi(hh_size)
```

The code below produces income as a percent of FPG for each of the four methods of imputing continuous income in the BRFSS by dividing income by the FPG and multiplying the result by 100.

```
gen fpg_low = 100 * (income_low / fpg_guideline)
gen fpg_mid = 100 * (income_mid / fpg_guideline)
gen fpg_up = 100 * (income_up / fpg_guideline)
gen fpg_unif = 100 * (income_unif / fpg_guideline)
```

These FPG variables can then be used to create policy-relevant categorical variables; for example, for individuals at or below 100% FPG.

[1] Nasseh and Vujicic, "The Effect of Growing Income Disparities on U.S. Adults' Dental Care Utilization."

[2] Levy and Meltzer, "The Impact of Health Insurance on Health"; Simon, Soni, and Cawley, "The Impact of Health Insurance on Preventive Care and Health Behaviors: Evidence from the First Two Years of the ACA Medicaid Expansions"; Yue, Rasmussen, and Ponce, "Racial/Ethnic Differential Effects of Medicaid Expansion on Health Care Access."

[3] Benitez and Seiber, "US Health Care Reform and Rural America: Results From the ACA's Medicaid Expansions"; Centers for Disease Control and Prevention (CDC), "BRFSS Statistical Brief on the Health Care Access Module, 2013 and 2014"; Hawaii Health Data Warehouse, "HHDW BRFSS Poverty Level Methodology"; Lopez, "Income Inequality and Self-Rated Health in US Metropolitan Areas: A Multi-Level Analysis"; Sabik, Tarazi, and Bradley, "State Medicaid Expansion Decisions and Disparities in Women's Cancer Screening."