

Introducing Educational Attainment to the Poststratification Adjustment in the National Survey on Drug Use and Health

Devon Cribb¹, Patrick Chen¹, Lanting Dai¹, Kathy Spagnola¹, Rong Cai², P. Mae Cooper², Jennifer M. Hoenig²

¹RTI International, 3040 East Cornwallis Road, Research Triangle Park, NC 27709

²Substance Abuse and Mental Health Services Administration, 5600 Fishers Lane, Rockville, MD 20857

Abstract

The National Survey on Drug Use and Health (NSDUH) provides national estimates of substance use and mental health among the civilian, noninstitutionalized population aged 12 or older in the United States. Since Quarter 4 of 2020, multimode (web and in-person) data collection has been employed in NSDUH. Adult web respondents had higher levels of educational attainment than adult in-person respondents, and educational attainment is often correlated with survey outcomes in NSDUH. To correct the imbalance of the educational attainment distributions across survey modes, educational attainment was added as a covariate to the poststratification adjustment in the 2020, 2021, and 2022 NSDUH weighting. Educational attainment proportions calculated from 1-year American Community Survey (ACS) data were used to derive control totals for the main effect and two-way interactions of educational attainment by demographic variables and by state. Two approaches for calculating educational attainment proportions, marginal distribution and cell distribution, were compared for accuracy across domains and summation of subdomains. The impact of excluding the institutionalized population and the active-duty military population from the 1-year ACS data was also investigated and is discussed.

Key Words: NSDUH, weighting, poststratification adjustment, educational attainment

1. Introduction

The National Survey on Drug Use and Health (NSDUH) is conducted annually and provides nationally representative estimates on tobacco, alcohol, and illicit drug use; substance use disorders; mental health issues; and receipt of mental health and substance use treatment among the civilian, noninstitutionalized population aged 12 or older in the United States. NSDUH uses a stratified five-stage cluster design (Center for Behavioral Health Statistics and Quality [CBHSQ], 2021) with a target of 67,500 respondents annually. In years prior to 2020, NSDUH data were collected from January to December by approximately 700 field interviewers completing 220,000 household screenings and 67,500 interviews. However, because of the coronavirus disease 2019 (COVID-19) pandemic, NSDUH in-person data collection was suspended on March 16, 2020. Ongoing COVID-19 infection rates in the United States made it nearly impossible to perform conventional in-person data collection at a response rate that could support an acceptable level of respondent sample size. Therefore, the Substance Abuse and Mental Health Services Administration approved multimode data collection (in person and web based) for the 2020 NSDUH beginning in Quarter 4 (October to December) of 2020. In-person data collection resumed on October 1, 2020 (in locations where COVID-19 infection metrics were sufficiently low), and web-based data collection began on October 30, 2020. Multimode data collection continued to be used in the 2021 2022, and 2023 NSDUHs and will be implemented in future NSDUHs.

The analysis weights for NSDUH are developed to obtain unbiased estimates. They are the product of 16 weight components. Each component either represents a design weight corresponding to one of the five selection stages, or adjusts for nonresponse, coverage, and extreme weight at the dwelling unit and person levels (CBHSQ, 2022). The analysis weights are benchmarked to the national target population counts for various demographic (age, gender, race/ethnicity) and geographic (state) domains in the last poststratification adjustment to reduce coverage bias and variance of survey estimates. The generalized exponential model (Folsom & Singh, 2000) is applied to perform nonresponse, poststratification, and extreme weight adjustments. Using multimode data collection in NSDUH impacted different aspects of the weighting process. This article discusses introducing educational attainment in the poststratification model to correct the imbalance of educational attainment for adult respondents in the NSDUH data when multimode data collection was introduced.

2. Educational Attainment Distributions in NSDUH

In Quarter 4, 2020, when multimode data collection was first introduced, about 93 percent of respondents completed the survey via the web. As shown in Table 1, in Quarter 4, 2020, educational attainment (less than high school, high school graduate, some college or associate's degree, and college graduate) showed a higher unweighted percentage of college graduates and a smaller proportion of adults with a high school education or less, when compared with results from prior years. Similar patterns of educational attainment distributions were observed in the 2021 NSDUH, where 54.6 percent of respondents completed the survey via the web. Because the 2021 NSDUH had more respondents completing the survey in person, the educational attainment distribution discrepancies were not as pronounced as they were in Quarter 4, 2020. Prior to the introduction of the web mode, the 2017–2019 and Quarter 1, 2020, NSDUHs maintained similar unweighted distributions of educational attainment.

Table 1: Unweighted Distributions of Educational Attainment: Among Adults Aged 18 or Older; Percentages, 2017–2021

<i>Educational Attainment</i>	<i>2017</i>	<i>2018</i>	<i>2019</i>	<i>2020</i> <i>Quarter 1</i>	<i>2020</i> <i>Quarter 4</i>	<i>2021</i>
Less Than High School	12.2	12.2	11.7	12.4	6.3	8.9
High School Graduate	25.9	26.1	26.2	24.7	18.6	23.1
Some College or Associate's Degree	33.8	33.8	33.6	33.4	30.1	29.9
College Graduate	28.1	27.9	28.5	29.5	45.0	38.1

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2017–2019; Quarters 1 and 4, 2020; and 2021.

Table 2 shows weighted distributions of educational attainment compared with distributions from prior NSDUH years and distributions from the corresponding 1-year American Community Survey (ACS) data (U.S. Census Bureau, 2023a, 2023b, 2023c). Because of the impact of the COVID-19 pandemic on the 2020 ACS data collection, the U.S. Census Bureau did not officially release the standard set of 1-year data products for the 2020 ACS. Instead, the U.S. Census Bureau released a set of 1-year estimates using experimental data because of concerns about poor data quality and unreliable estimates (U.S. Census Bureau, 2021). Thus, the official 2020 ACS 1-year data are not available, and data from the 1-year 2019 ACS restricted-use file (RUF), are used for comparisons in Table 2 and in poststratification adjustments for the 2020 and 2021 NSDUHs (described in Section 3).

Table 2: NSDUH and ACS Weighted Distributions of Educational Attainment: Among Adults Aged 18 or Older; Percentages, 2017–2021

	2017		2018		2019		2020		2021
<i>Educational Attainment</i>	<i>NSDUH¹</i>	<i>ACS</i>	<i>NSDUH¹</i>	<i>ACS</i>	<i>NSDUH¹</i>	<i>ACS</i>	<i>Quarter 1 NSDUH¹</i>	<i>Quarter 4 NSDUH¹</i>	<i>NSDUH¹</i>
Less Than High School	12.5	12.1	12.4	11.8	12.0	11.5	11.9	8.4	11.4
High School Graduate	24.2	27.7	24.7	27.5	24.3	27.6	23.4	19.5	24.2
Some College or Associate's Degree	31.1	30.8	31.0	30.7	30.8	30.3	30.7	28.8	28.8
College Graduate	32.2	29.4	31.9	30.1	32.9	30.6	34.0	43.3	35.6

ACS = American Community Survey; NSDUH = National Survey on Drug Use and Health.

¹ Weighted distributions for the 2017–2019 NSDUHs are based on final analysis weights after poststratification adjustment. Weighted distributions for Quarter 1, 2020; Quarter 4, 2020; and 2021 are based on nonresponse-adjusted weights (no poststratification adjustment is included).

Source: SAMHSA, Center for Behavioral Health Statistics and Quality, National Survey on Drug Use and Health, 2017–2019; Quarters 1 and 4, 2020; and 2021; U.S. Census Bureau, American Community Survey (RUF), 2017–2019.

Although educational attainment was not included in the poststratification adjustment models in the 2017–2019 NSDUHs, the weighted distributions of educational attainment are comparable and consistent and align well with ACS distributions. The discrepancies in the unweighted distributions for Quarter 4, 2020, and 2021 (as shown in Table 1) were reduced after person-level nonresponse adjustment. For example, the proportion of adults with a high school education or less increased from 6.3 percent to 8.4 percent, and the proportion of college graduates decreased from 45.0 percent to 43.3 percent in Quarter 4, 2020 (Tables 1 and 2). However, the differences in the distributions are still noticeable (as shown in Table 2 for the 2020 and 2021 NSDUHs). These differences needed to be further corrected by adding educational attainment variables to the poststratification adjustment models.

3. Methods

Six domains of educational attainment were used in the poststratification adjustment models: (1) educational attainment main effect, (2) interactions with state, (3) interactions with age group (18–25, 26–34, 35–49, 50 or older), (4) interactions with race (White, Black or African American, Other), (5) interactions with gender (male, female), and (6) interactions with Hispanic ethnicity (Hispanic or Latino, not Hispanic or Latino). Two approaches were used to estimate educational attainment control totals used in the poststratification adjustment models: the marginal distribution method and the cell distribution method.

Marginal Distribution Method: Nine model groups corresponding to the nine census divisions are used for NSDUH person-level weighting. Main effects and two-way interactions with educational attainment were added to each model group. The marginal educational attainment distributions for six domains were obtained from the ACS. Available marginal educational attainment distributions were based on the 1-year 2019 ACS-RUF, including main effects, state by educational attainment, and gender by educational attainment. Marginal educational attainment distributions for age group by educational attainment, race by educational attainment, and Hispanic ethnicity by educational attainment were based on the 1-year 2019 ACS public use file (PUF). Control totals for educational attainment were then estimated by multiplying the marginal educational attainment distributions by the civilian, noninstitutionalized population counts in each domain. Table 3 lists the six domains and data sources for the marginal educational attainment distributions. This method was applied to calculate the control totals for educational attainment in the 2020 NSDUH weighting process.

Table 3: Educational Attainment by Domains in Each Model Group

<i>Domain</i>	<i>Description</i>	<i>Data Source</i>
Main Effect	Educational attainment (four levels)	ACS-RUF
State by Educational Attainment	Varies among model groups	ACS-RUF
Age Group by Educational Attainment	18–25, 26–34, 35–49, 50+	ACS-PUF
Race by Educational Attainment	White, Black or African American, Other	ACS-PUF
Gender by Educational Attainment	Male, female	ACS-RUF
Hispanic Ethnicity by Educational Attainment	Hispanic or Latino, not Hispanic or Latino	ACS-PUF

ACS = American Community Survey; PUF = public use file; RUF = restricted-use file.

Table 4 demonstrates how to calculate the control totals for educational attainment in each domain by using the marginal educational attainment distribution method. This example is for state by educational attainment, but the same method was applied to other domains.

Table 4: Example of Calculating Control Totals for Educational Attainment Using the Marginal Distribution Method for Connecticut

<i>State</i>	<i>Educational Attainment</i>	<i>2020 Civilian, Noninstitutionalized Population Counts for Adults Aged 18 or Older (from U.S. Census Bureau)</i>	<i>2019 ACS Marginal Distribution (from ACS-RUF)</i>	<i>Educational Attainment Control Totals in Poststratification Adjustment</i>
Connecticut	Less than high school	2,794,746	9.49%	265,221 = $(9.49\% \times 2,794,746)$
	High school graduate		27.10%	757,376
	Some college or associate's degree		26.99%	754,302
	College graduate		36.42%	1,017,847

ACS = American Community Survey; RUF = restricted-use file.

Source: U.S. Census Bureau, Census, 2020; U.S. Census Bureau, American Community Survey (RUF), 2019.

Because the domain-specific marginal educational attainment distributions and population counts were used to estimate the control totals for educational attainment, the sum of control totals for each educational attainment category calculated from the six domains did not match perfectly for the New England census division (Table 5). For example, the control totals of “less than high school” from the main effect were 1,024,040, whereas the sum of “less than high school” control totals from interactions of race by educational attainment were 976,299. Generalized exponential calibration models had the intercept term, so some variables were excluded from the model as reference covariates. When variables explicitly included were all retained in the models, the reference covariates also should have been controlled after adjustment. However, this was not the case when control totals for educational attainment were inconsistent across domains. Thus, weighted sums of reference covariates of two-way educational attainment interactions did not match the control totals after adjustment, even when all variables were retained in the models. This result was undesirable.

Table 5: Inconsistency in Control Totals of Educational Attainment Using the Marginal Distribution Method for the New England Census Division
Population Totals from the Marginal Distribution Method

<i>Educational Attainment and Domain</i>	<i>Main Effect</i>	<i>Race Interaction</i>	<i>Gender Interaction</i>	<i>Hispanic Ethnicity Interaction</i>	<i>Age Group Interaction</i>	<i>State Interaction</i>
Less Than High School	1,024,040	976,299	1,023,843	1,015,503	1,012,272	1,023,769
High School Graduate	3,138,903	3,151,655	3,138,385	3,137,258	3,135,368	3,138,646
Some College or Associate's Degree	3,148,600	3,180,196	3,148,909	3,153,574	3,149,684	3,148,551
College Graduate	4,494,345	4,497,737	4,494,750	4,499,551	4,508,563	4,494,921
Total	11,805,887	11,805,887	11,805,887	11,805,887	11,805,887	11,805,887

Source: U.S. Census Bureau, Census, 2020; U.S. Census Bureau, American Community Survey (RUF), 2019; U.S. Census Bureau, American Community Survey (PUF), 2019.

Cell Distribution Method: To overcome the inconsistency in the educational attainment control totals caused by the marginal distribution method in the 2020 NSDUH weighting, an alternative method was explored to calculate control totals for educational attainment. The cell distribution method calculates educational attainment distributions for each cell formed by the intersection of every level of state, age group, race, gender, and Hispanic ethnicity (total of 2,488 cells, 51 by 4 by 3 by 2 by 2) using 1-year 2019 ACS-PUF data. This level of detail is not available using the ACS-RUF. For example, for the cell intersection of “Connecticut, 18–25, White, Hispanic or Latino, Male,” the cell educational attainment distributions are 17.43 percent for less than high school, 35.54 percent for high school graduate, 39.21 percent for some college or associate’s degree, and 7.82 percent for college graduate (Table 6). These cell educational attainment distributions were then multiplied by the cell population count (2020 census target population, 31,731) to estimate cell control totals for educational attainment (Table 6). For cells with zero counts in the 1-year 2019 ACS-PUF, average educational attainment distributions for the entire state were used instead. Cell educational attainment control totals were then aggregated to the six domains. Applying the cell distribution method allows the sums of control totals for each educational attainment category calculated from all six domains to match precisely. The cell distribution method was employed in the poststratification adjustment for 2021 NSDUH weighting.

Table 6: Example of Calculating Control Totals for Educational Attainment Using the Cell Distribution Method for Connecticut

<i>State</i>	<i>Age Group</i>	<i>Race</i>	<i>Hispanic Ethnicity</i>	<i>Gender</i>	<i>Educational Attainment</i>	<i>2020 Cell Target Population Count (from U.S. Census Bureau)</i>	<i>Cell Educational Attainment Distribution (from ACS-PUF)</i>	<i>Cell Control Totals</i>
Connecticut	18–25	White	Hispanic or Latino	Male	Less than high school	31,731	17.43%	5,531
					High school graduate		35.54%	11,278
					Some college or associate's degree		39.21%	12,441
					College graduate		7.82%	2,481

ACS = American Community Survey; PUF = public use file.

Source: U.S. Census Bureau, Census, 2020; U.S. Census Bureau, American Community Survey (PUF), 2019.

4. Marginal and Cell Distribution Method Comparison Results

For comparing the control totals of the marginal and cell distribution methods, 2020 civilian, noninstitutionalized population estimates and educational attainment proportions based on the 1-year 2019 ACS data were used. The control totals for main effects and two-way interactions from all nine model groups calculated from the marginal distribution method (M_{total}) and cell distribution method (C_{total}) were compared, and the absolute relative differences ($100 \times \text{abs}(C_{total} - M_{total})/M_{total}$) were calculated. Table 7 shows the distributions of the absolute relative differences, by domain and overall. In general, control totals calculated from both methods are comparable. The average absolute relative differences for main effect, age group by educational attainment, and gender by educational attainment were less than 1 percent. They were 1.35 percent for state by educational attainment and 2.45 percent for Hispanic ethnicity by educational attainment. Some large absolute relative differences were observed in race by educational attainment, particularly in the less than high school and Other race categories. The largest absolute relative difference (32.37 percent) occurred in “Other race, less than high school, Model Group 9 (Pacific Census Division).” Large differences can be caused by unstable educational attainment distributions in cells with small sample sizes in the 1-year 2019 ACS-PUF data. The average absolute relative difference for race by educational attainment was 6.19 percent.

Table 7: Absolute Relative Differences (in percentages) of Control Totals from the Marginal and Cell Distribution Methods

<i>Domain</i>	<i>Number of Covariates¹</i>	<i>Min</i>	<i>P25</i>	<i>Median</i>	<i>P75</i>	<i>Max</i>	<i>Mean</i>
Main Effect	36	0.02%	0.18%	0.64%	1.06%	4.19%	0.93%
State by Educational Attainment	204	0.01%	0.38%	0.79%	1.77%	8.62%	1.35%
Age Group by Educational Attainment	144	0.02%	0.23%	0.50%	0.87%	8.48%	0.90%
Race by Educational Attainment	108	0.02%	0.36%	1.37%	8.48%	32.37%	6.19%
Gender by Educational Attainment	72	0.02%	0.25%	0.62%	1.09%	4.34%	0.97%
Hispanic Ethnicity by Educational Attainment	72	0.01%	0.16%	0.48%	3.48%	12.96%	2.45%
Overall	636	0.01%	0.28%	0.67%	1.78%	32.37%	2.13%

Max = maximum; Min = minimum; P25 = 25th percentile; P75 = 75th percentile.

¹ Number of covariates from all nine model groups; for example, main effects: $4 \times 9 = 36$; age group by educational attainment: $4 \times 4 \times 9 = 144$.

Source: U.S. Census Bureau, Census, 2020; U.S. Census Bureau, American Community Survey (RUF), 2019; U.S. Census Bureau, American Community Survey (PUF), 2019.

5. Discussion

Hundreds of NSDUH tables are published every year. In 2019 and previous years, several key measures were provided by educational attainment. Some key measures provided in the annual NSDUH detailed tables are significantly different between levels of educational attainment (testing results not published). Because associations between NSDUH measures and educational attainment exist, it is imperative to ensure that weighted distributions of educational attainment in NSDUH are comparable and consistent across years. Thus, adding educational attainment to the poststratification adjustment models to correct the imbalance of educational attainment distributions was necessary in the 2020–2022 NSDUHs and will continue to be done in future NSDUHs.

The cell distribution method uses the cell distributions of educational attainment by crossing state, age group, race, Hispanic ethnicity, and gender based on 1-year ACS data, then by multiplying the target population estimates by the cell educational attainment proportions to obtain the educational attainment control totals for each cell. Those cell control totals can then be aggregated to control totals in all six domains, such as main effect and two-way interactions with state, age group, race, Hispanic ethnicity, and gender. The cell distribution method provides consistent control totals of educational attainment; thus, it is preferable to the marginal distribution method. The marginal distribution method was used in 2020 NSDUH weighting, and the cell distribution method was used in the 2021 and 2022 NSDUH weighting and will be used in weighting for future NSDUHs.

For the cell distribution method, when calculating cell educational attainment distributions using 1-year ACS-PUF data, the total population or the non-institutionalized population and the non-active-duty military population can be used. An investigation was conducted to assess the impact of removing the institutionalized population and the active-duty military population from 1-year ACS-PUF population counts in estimating educational attainment control totals on weights and survey estimates. Using educational attainment control totals that derived from educational attainment proportions from 1-year ACS-PUF data after excluding the institutionalized population and the active-duty military population has negligible impact on analysis weights and survey estimates. However, excluding the institutionalized population and the active-duty military population from 1-

year ACS-PUF data is consistent with what is done for the NSDUH target population. Thus, in 2022, educational attainment control totals were calculated from the educational attainment proportions after excluding the institutionalized population and the active-duty military population from the 1-year ACS-PUF data. This exclusion will also be applied to calculate educational attainment proportions in future NSDUHs.

Acknowledgments

This work was funded by the Substance Abuse and Mental Health Services Administration (SAMHSA), U.S. Department of Health and Human Services (HHS), under Contract No. HHSS283201700002C with RTI International. Marlon Daniel served as the government project officer and as the contracting officer representative.

Disclaimer

The views expressed in this document do not necessarily reflect the official policies of SAMHSA or HHS, nor does mention of trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

References

- Center for Behavioral Health Statistics and Quality. (2021). *2020 National Survey on Drug Use and Health (NSDUH) methodological resource book, Section 2: Sample design report*. <https://www.samhsa.gov/data/report/nsduh-2020-sample-design-report>
- Center for Behavioral Health Statistics and Quality. (2022). *2020 National Survey on Drug Use and Health (NSDUH) methodological resource book, Section 11: Person-level sampling weight calibration*. <https://www.samhsa.gov/data/report/nsduh-2020-person-level-sampling-weight-calibration-report>
- Folsom, R. E., & Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section* (pp. 598–603). http://www.asasrms.org/Proceedings/papers/2000_099.pdf
- U. S. Census Bureau. (2021). *An assessment of the COVID-19 pandemic's impact on the 2020 ACS 1 year data*. https://www.census.gov/content/dam/Census/library/working-papers/2021/acs/2021_CensusBureau_01.pdf
- U.S. Census Bureau. (2023a). *2017 American Community Survey 1-year public use microdata samples* [SAS data file]. <https://data.census.gov/mdat/#/search?ds=ACSPUMS1Y2017>
- U.S. Census Bureau. (2023b). *2018 American Community Survey 1-year public use microdata samples* [SAS data file]. <https://data.census.gov/mdat/#/search?ds=ACSPUMS1Y2018>
- U.S. Census Bureau. (2023c). *2019 American Community Survey 1-year public use microdata samples* [SAS data file]. <https://data.census.gov/mdat/#/search?ds=ACSPUMS1Y2019>