SAMPLE DESIGN

2022 National Survey on Drug Use and Health Sample Design

The 2022 National Survey on Drug Use and Health (NSDUH) is a continuation of a multiyear sample design which was introduced by the Center for Behavioral Health Statistics and Quality (CBHSQ) in 2014. Similar to previous NSDUHs, the 2022 NSDUH sample design is a stratified, multi-stage area probability design. The sample design provides for estimates by state in all 50 states and the District of Columbia (DC). As shown in **Table 1**, the 2022 survey will have a sample designed to yield 4,560 completed interviews in California; 3,300 completed interviews each in Texas, New York, and Florida; 2,400 completed interviews each in Illinois, Pennsylvania, Ohio, and Michigan; 1,500 completed interviews each in Georgia, North Carolina, New Jersey, and Virginia; 967 completed interviews in Hawaii; and 960 completed interviews in each of the remaining 37 states and DC for a total national target sample size of 67,507 completed interviews. The sample will be selected from 6,000 area segments which vary in size according to state. For the 2022 NSDUH, each state's sample will be allocated to age groups as follows: 25 percent 12 to 17, 25 percent 18 to 25, 15 percent 26 to 34, 20 percent 35 to 49, and 15 percent 50 or older.

Table 1. Sample Sizes and Projected Respondents by State and Age Group, 2022 NSDUH

State	Age 12–17	Age 18–25	Age 26–34	Age 35–49	Age 50+	Total Age 12+	State Sampling Regions (SSRs)	Segment Size	No. of Segments
Total Population	16,877	16,877	10,126	13,501	10,126	67,507	750		6,000
CA	1,140	1,140	684	912	684	4,560	36	15.833	288
TX, NY, FL	825	825	495	660	495	3,300	30	13.75	240
IL, PA, OH, MI	600	600	360	480	360	2,400	24	12.5	192
GA, NC, NJ, VA	375	375	225	300	225	1,500	15	12.5	120
HI	242	242	145	193	145	967	12	10.07	96
Remaining 37 States and DC	240	240	144	192	144	960	12	10	96

Beginning in 2014 and continuing with the 2022 NSDUH, the sample will be designed to yield a minimum of 200 completed interviews in Kauai County, Hawaii over a three-year period. This will allow for Kauai County to be included as a separate entity in the production of substate estimates that are produced biennially and typically based on three years of data. To achieve this goal while maintaining precision at the state level, Kauai County will be treated separately from the remainder of Hawaii for sample allocation and sample size management purposes. The annual sample in Hawaii will consist of 67 completed interviews in Kauai and 900 completed interviews in the remainder of the state, for a total of 967 completed interviews each year.

Finally, the 2022 design includes the selection of Census block groups at the second stage of selection. This stage of selection was included beginning in 2014 in anticipation of moving to a hybrid address-based sample (ABS) design, which is planned for the 2022 NSDUH. ABS refers to the sampling of residential addresses from a list based on the U.S. Postal Service's Computerized Delivery Sequence file. ABS addresses are assigned to Census geographies using a process called geocoding. Compared to geocoding at the Census block level, geocoding accuracy improves significantly at the Census block group level in both rural and urban areas. Thus, in the hybrid ABS design, Census block groups will serve as geographic clusters in areas with sufficient mailing address coverage.

The selection of Census tracts at the first stage of selection and Census block groups at the second stage has the potential to reduce sampling variance by controlling the distribution of selected areas and reducing the chance of selecting neighboring and possibly similar areas within tracts and block groups. In addition, the merging of NSDUH data to external data sources for future analysis purposes is simplified when sampled areas are contained within tract and block group boundaries to the extent possible.

First, Second, and Third Stages of Selection: Census Tracts, Census Block Groups and Segments

A coordinated sample for the period 2014-2017 was selected down to the area segment level as a means of coordinating the overlap of sample areas from year to year. In anticipation of the next decennial census data being unavailable, a large reserve sample was also selected and is available for use in 2018 through 2022.

The "first level" of stratification within the coordinated design is states. The larger sample sizes obtained at the state level along with small area estimation (SAE) techniques or direct estimation techniques will enable the development of estimates for all states, for several demographic subgroups within each state (e.g., age group and gender), for some Core Based Statistical Areas (CBSAs), and for other small areas.

The "second level" of stratification within the coordinated design was defined as contiguous geographic areas (groups of Census tracts) within each state. Within each state, state sampling regions (SSRs) are of approximately equal size in terms of the population and allocated state sample, except in Hawaii where Kauai County is its own SSR and the remainder of the state is divided into equal-sized regions.

The design of the first stage of selection began with the construction of an area sample frame that contained one record for each census tract in the United States. If necessary, census tracts were aggregated until each tract had a minimum number of dwelling units. In California, Texas, New York, Florida, Illinois, Pennsylvania, Ohio, Michigan, Georgia, North Carolina, New Jersey, and Virginia, this minimum size requirement was 250 dwelling units in urban areas and 200 dwelling units in rural areas. The basis for the differing minimum dwelling unit requirement in urban and rural areas is that it is more difficult to meet the requirement in rural areas, and 200 dwelling units are sufficient to support one field test and two main study samples. In the remaining states and DC, the number of completed interviews per sampled area is smaller and, therefore, fewer dwelling units are needed. In these states, the minimum requirement was 150 dwelling units in

urban areas and 100 dwelling units in rural areas. After primary sampling units (PSUs; one or more census tracts) were formed, a sample was selected within each SSR with probabilities proportionate to a composite size measure and with minimum replacement. Additional implicit stratification was achieved by sorting the first-stage sampling units by a CBSA/SES (socioeconomic status) indicator[1] and by percent non-Hispanic white prior to selection.

For the second stage of selection, adjacent Census block groups were aggregated within selected PSUs as necessary to meet the minimum dwelling unit requirements (150 or 250 dwelling units in urban areas and 100 or 200 dwelling units in rural areas, according to state). After the resulting secondary sampling units (SSUs; one or more census block groups) were formed, one SSU was selected per sampled PSU with probability proportionate to a composite size measure.

For the third stage of sampling for the coordinated sample, each of the selected Census block groups was partitioned into clusters of dwelling units by aggregating adjacent Census blocks. Consistent with the terminology used in previous surveys, these geographic clusters of blocks are referred to as *segments*. Segments were formed so that they contain the same minimum number of dwelling units as the PSU (i.e., Census tracts) and SSU (i.e., Census block groups) to which they belong. That is, area segments contain at least 150 or 250 dwelling units in urban areas and 100 or 200 dwelling units in rural areas according to state. Segments were constructed using 2010 Decennial Census data supplemented with 2013 population projections obtained from outside sources. A sample *dwelling unit* in the survey refers to either a housing unit or a group quarters listing unit such as a dormitory room or a shelter bed.

For the coordinated sample, one segment was selected within each selected Census block group with probability proportionate to a composite size measure. As mentioned previously, segments were formed so that they would contain sufficient numbers of dwelling units to support one field test and two annual survey samples. This allows half of the segments used in any given year's main sample to be used again in the following year as a cost-saving measure and a means of improving the precision of measures of annual change. This also allows for any special supplemental sample or field test that SAMHSA may wish to conduct within the same segments.

A hybrid ABS design will be implemented for the 2022 NSDUH. First, ABS coverage estimates will be computed for all SSUs (one or more Census block groups) that were previously selected at the second stage of selection. In addition to having high expected ABS coverage (95 percent or more), SSUs will be required to meet separate criteria for group quarters and drop points.[2] SSUs that meet the ABS coverage criteria will then serve as sample segments and use the ABS frame. SSUs that do not meet the ABS coverage criteria will require field enumeration. For cost efficiency, the smaller geographic area that was previously selected at the third stage of selection will be field enumerated and considered the sample segment. The planned approach allows a hybrid ABS design to be deployed on a smaller scale, without significant risk, while realizing some of the cost and timeliness benefits.

---

[1] Four categories are defined as: (1) CBSA/low SES, (2) CBSA /high SES, (3) Non-CBSA /low SES, and (4) Non-CBSA /high SES.

[2] A drop point is a mail receptacle that is shared by multiple housing units (drop units). The ABS frame indicates the number of units at a drop point but does not include unit identifiers (e.g., apartment numbers). Drop points present additional challenges for sample implementation because of their one-to-many relationship to dwelling units.

A sample of 8 segments per SSR will be used for the 2022 survey year. These 8 segments will be randomly assigned to quarters and to two panels within each quarter. The panels used in the 2022 survey will be designated as panels 9 and 10. Panel 9 segments were field enumerated and used for the 2021 survey and will be used for the second time in the 2022 survey. Dwelling units that were not selected for the 2021 survey will be eligible for selection in the panel 9 segments in 2022. The panel 10 segments will be used for the 2022 survey only. Of the 3,000 panel 10 segments, 973 will use the ABS frame and 2,027 will be field enumerated.

Approximately one-fourth of the main sample of respondents will be collected from each calendar quarter. This design feature will help control the influence of seasonal variation on drug use prevalence estimates and other important survey outcome measures of interest.

Fourth Stage of Selection: Dwelling Units

The sampling frame for the fourth stage of sample selection will consist of dwelling units within selected segments. A dwelling unit is either a housing unit for a single household or one of the eligible noninstitutional group quarters listing units that are part of the defined target population. For ABS segments, the dwelling unit frame will consist of ABS addresses that geocode into the Census block group(s). For all other segments, specially trained field household listers will list all dwelling units and potential dwelling units within the segments. The listings will be based primarily on observations of the area segment and may include vacant dwelling units and units that appear to be dwelling units but may actually be used for nonresidential purposes. The objective of the listing is to obtain as complete a listing of eligible residential addresses as possible; any false positives for residences will be eliminated during the household screening process after the sample is selected.

After preparing the ABS and field enumerated frames, a sample of dwelling units will be selected within each segment. After accounting for eligibility, nonresponse, and the fifth-stage sample selection procedures, it was determined that roughly 300,671 dwelling units will need to be selected in order to obtain a sample of 67,507 responding persons distributed by state and age-group as shown in **Table 1**.

Fifth Stage of Selection: Persons

After dwelling units are selected within each segment, each selected dwelling unit will be mailed an invitation to participate in the survey by web. If the selected dwelling unit does not complete the web screener or has one or more pending interviews, a field interviewer will visit the dwelling unit a few weeks later to complete the screener and/or interview(s) in-person. During the web or in-person screener, a roster of all persons residing in the dwelling unit will be obtained and used to select 0, 1, or 2 persons for the survey. Sampling rates will be pre-set by age group and State. Roster information will be entered directly into the electronic screening instrument which will automatically implement this fifth stage of selection based on the State and age group sampling parameters.

One exciting consequence of using an electronic screening instrument in the survey is the ability to efficiently sample from all possible pairs of respondents within a dwelling unit while preserving the target sampling rates for individuals within 5 age groups (12 to 17, 18 to 25, 26 to

34, 35 to 49, and 50 or older). Using an adaptation of Brewer's method for samples of size 2, a sample of 0, 1, or 2 persons will be selected from each dwelling unit. As a consequence, *any* two survey-eligible people within a dwelling unit will have a known chance of being selected, that is, all survey eligible pairs of people will have some nonzero chance of being selected. This feature of the design is of interest to survey researchers because for example, it will allow analysts to examine how the drug use propensity of one individual in a family will relate to the drug use propensity of another family member residing in the same dwelling unit (e.g., the relationship of drug use between a parent and child).

As illustrated in **Table 2**, at the fifth stage of selection, roughly 114,687 people will be selected from within 168,673 screened and eligible dwelling units. Assuming a 66% screening completion rate and a 59% interview completion rate across both survey modes, these sample sizes are sufficient to obtain the desired 67,507 person respondents.

Table 2. Summary of 2022 Main Study Design

Statistic	Web	In-Person	2022 NSDUH Total	2022 Rate
State Sampling (SS) Regions			750	
Segments			6,000	
Selected Lines	33,074	267,597	300,671	
Eligible Dwelling Units	33,073	223,066	256,139	0.85
Completed Screening Interviews	33,073	135,600	168,673	0.66
Selected Persons	22,488	92,199	114,687	0.68
Completed Interviews	12,818	54,689	67,507	0.59

Expected Precision of Survey Estimates

The multistage, stratified 2022 survey design has been designed to achieve acceptable precision for various person subpopulations of interest. The allocation of persons per state and age group (12-17, 18-25, 26-34, 35-49, and 50 or older) was also taken as a requirement to support direct estimation in some large sample states and SAE in the remaining states. **Table 3** shows the projected relative standard errors for selected prevalence measures.

Table 3. Estimated Prevalences and Relative Standard Errors for Key Measures by Demographic Domain

Measure	Domain	Prevalence	Projected RSE
Past Month Alcohol Use	12+	0.5077	0.0072
Past Month Alcohol Use	12-20	0.1851	0.0276
Past Month Alcohol Use	50+	0.4919	0.0127
Past Month Alcohol Use	API,12+	0.3781	0.0415
Past Month Alcohol Use	AIAN, 12+	0.3210	0.0847
Past Month Alcohol Use	Pregnant, 12-44	0.0956	0.2105
Past Month Binge Alcohol Use	18-25	0.3427	0.0145
Past Month Binge Alcohol Use	12+	0.2392	0.0116
Past Month Marijuana Use	12+	0.1148	0.0172
Past Month Marijuana Use	12-17	0.0738	0.0372
Past Month Marijuana Use	18-25	0.2297	0.0193
Past Month Marijuana Use	50+	0.0647	0.0468
Past Month Marijuana Use	API,12+	0.0482	0.0994
Past Month Marijuana Use	AIAN, 12+	0.1202	0.1367
Past Month Marijuana Use	Pregnant, 12-44	0.0542	0.2943
Past Month Cigarette Use	12-17	0.0230	0.0702
Past Month Cigarette Use	12+	0.1667	0.0163
Past Month Pain Reliever Misuse	18-25	0.0124	0.0897
Past Month Pain Reliever Misuse	12+	0.0102	0.0580
Past Year Alcohol Disorder	12+	0.0527	0.0250
Past Year Illicit Drug Disorder	12+	0.0302	0.0293
Past Year Substance Use Disorder	50+	0.0411	0.0588
Past Year Specialty Substance Use Treatment	12+	0.0096	0.0628
Past Year SMI	18+	0.0524	0.0258
Past Year MDE	18+	0.0783	0.0209

Note: Prevalence estimates are based on data from the 2019 NSDUH. Projected RSEs were determined using 2014 through 2022 state and age sample allocations in a variance component model. All model components were updated using 2019 NSDUH data.