Use of Single Years of Age in the National Survey on Drug Use and Health (NSDUH) Weighting to Improve Drug Prevalence Estimates

Patrick Chen,[1] Neeraja Sathe,[1] Michael Jones,[2] Lanting Dai,[1]
Jeff Laufenberg,[1] Ralph Folsom,[1] Harper Gordek[1]

[1]RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709
[2]Substance Abuse and Mental Health Services Administration, Office of Applied Studies,
1 Choke Cherry Road, Rockville, MD 20857

Abstract
In the weighting process for the National Survey on Drug Use and Health (NSDUH), several age groups (i.e., 12 to 17, 18 to 25, 26 to 34, 35 to 49, 50 to 64, and 65 or older) are currently used in the nonresponse and poststratification adjustments. Using 2004 to 2006 NSDUH data, we found that the response rate decreases as age increases in the 12 to 25 age range. Also, the prevalence rates for illicit drug, alcohol, and tobacco use increase almost linearly between ages 12 and 21, reaching a peak at age 21. Because the response rates and drug use prevalence rates change dramatically between the ages of 12 and 21, use of single years of age instead of age groups for both the nonresponse and poststratification adjustments should reduce nonresponse bias and the variance of estimates. This paper explores the use of single years of age between 12 and 25 in the 2006 NSDUH weighting process and discusses the prevalence rates and standard errors produced using the new set of weights.

Key Words: Age, weighting, NSDUH, drug use prevalence, nonresponse, post-stratification

1. Introduction to NSDUH

The National Survey on Drug Use and Health (NSDUH) provides national, State, and substate data on substance use and mental health in the civilian, noninstitutionalized U.S. population aged 12 or older. The survey is sponsored by the Substance Abuse and Mental Health Services Administration (SAMHSA), U.S. Department of Health and Human Services, and is planned and managed by SAMHSA's Office of Applied Studies (OAS). Data collection and analysis are currently conducted under contract with RTI International,[1] Research Triangle Park, North Carolina.

Since 1999, NSDUH has employed a 50-State design with an independent, multistage area probability sample for each of the 50 States and the District of Columbia. For the 50-State design, 8 States are designated as large sample States (California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas) with target sample sizes of 3,600 per year. For the remaining 42 States and the District of Columbia, the target sample size

[1] RTI International is a trade name of Research Triangle Institute.

is 900 per year. Approximately 140,000 household screenings and 67,500 interviews are completed annually. Most of the questions are administered with audio computer-assisted self-interviewing (ACASI). ACASI is designed to provide the respondent with a highly private and confidential mode for responding to questions in order to increase the level of honest reporting of illicit drug use and other sensitive behaviors. Less sensitive items are administered by interviewers using computer-assisted personal interviewing (CAPI).

1.1 NSDUH Sample Design for 2005 to 2009

A coordinated 5-year sample design was developed for the 2005 to 2009 NSDUHs. States are partitioned into State sampling (SS) regions: 12 SS regions for each of 43 small States and the District of Columbia, and 48 SS regions for the 8 large States listed above. A total of 900 SS regions are defined across the United States.

NSDUH has a four-stage design. The first stage of selection consists of selecting 48 census tracts per SS region. Segments are formed within a selected tract by aggregating adjacent census blocks. One segment is selected within each sampled tract (second-stage selection). After segments are selected and assigned, field enumeration is used to construct a list of all eligible dwelling units (DUs) within a segment. DU samples are selected from the list (third-stage selection) of eligible DUs. Using the roster information obtained from eligible members of the selected DU, 0, 1, or 2 persons are selected for the survey (fourth-stage selection). Sampling rates are predetermined for each State and age group.

1.2 NSDUH Weighting

The NSDUH sample weights are calculated for each respondent as a two-part process: (1) producing the design-based weights and (2) applying adjustment factors to obtain the final weights.

The design-based weights reflect the four-stage sample design. There are four components to the design weights: Inverse of Probability of Selecting Census Tracts, Inverse of Probability of Selecting Segments, Inverse of Probability of Selecting DUs, and Inverse of Probability of Selecting Persons among DUs. These design-based weights then are adjusted for nonresponse (NR) and poststratification (PS) to reduce the bias and variance of estimates. Five separate adjustment factors are applied to the design-based weights: DU NR Adjustment, DU PS Adjustment, Selected Person PS Adjustment, Person NR Adjustment, and Person PS Adjustment. The final analysis weights are the product of all of the weight components.

The Generalized Exponential Model (GEM; Folsom & Singh, 2000) is used for all of the adjustment steps listed above. GEM is a flexible weighting routine that includes as special cases the well-known weighting-class-adjustment method (Little, 1986), the raking-ratio method (Deming & Stephen, 1940), bounded logit calibration (Deville & Särndal, 1992; Sautory, 1991), and logistic and exponential model approaches (Folsom, 1991; Folsom & Witt, 1994). It has a built-in feature that controls the adjustment factor to prevent the adjusted weights from being too extreme, and it provides a comprehensive weighting methodology for NR, PS, and extreme weight adjustments. The variables initially selected in the person-level NR model are either good predictors of response propensity or correlated with survey outcomes. The variables in the person-level PS model are determined by the analysis domains of interest and the availability of population estimates. For the NR model, several person characteristic variables, geographic characteristic variables, DU characteristics, and some segment-level variables are included. Also included are two-way and three-way interactions. For the PS model, several demographic variables, States, and quarters (as well as interactions of these) are included. Fitting one national model is computationally intensive and prohibitive, so nine model groups corresponding to the U.S. Census Bureau's divisions[2] are used. For details on the NSDUH weighting process and the different variables used in the NR and PR adjustments, see Chen et al. (2008).

2. Motivation to Use Single Years of Age in the Weight Adjustment

In the NSDUH weighting process, five age groups (12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older) are used in the person NR adjustment. In the person PS, the 50 or older age group is further split into the 50 to 64 and the 65 or older age groups. Age is a strong predictor of both response propensity and drug use prevalence. Sections 2.1 and 2.2 give the justification (using 2004 to 2006 NSDUH data) for using single years of age in the weighting calibration process.

2.1 Response Rates and Age

Figure 1 shows the relationship between age and response rates for the 2004 to 2006 NSDUHs. In general, as age increased, the response rate decreased. Also, a sharp decrease in response rates was observed in the age range from 12 to 25. Between 25 and 65, the response rates stayed relatively flat, then gradually decreased until age 75. After age 75, another sharp drop was observed. However, the decrease in the response rates between the ages of 12 to 25 is of most interest in the NSDUH sample. The response rate dropped linearly from almost 90 percent for ages 12 to 14 to just over 75 percent for ages 23 to 25.

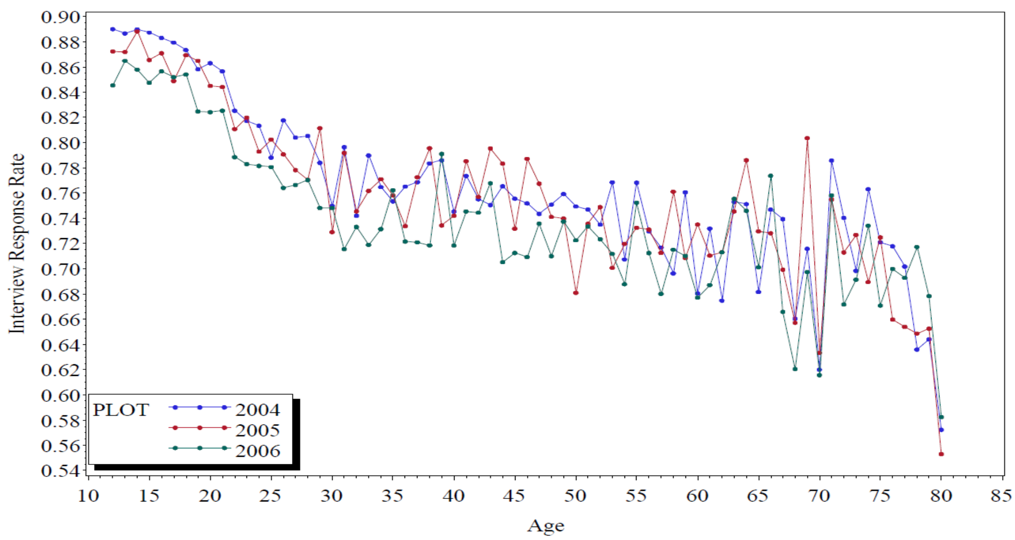

Figure 1: Interview response rates, by age: 2004, 2005, and 2006 NSDUHs (persons aged 80 or older are combined with persons aged 80)

[2] The nine census divisions are New England, Middle Atlantic, South Atlantic, East South Central, West South Central, East North Central, West North Central, Mountain, and Pacific. For details, see http://www.census.gov/geo/www/us_regdiv.pdf.

2.2 Substance Use and Age

Past research has reported strong correlations between drug use prevalence and age in the younger age groups (Chen & Kandel, 1995; Kandel & Logan, 1984; Odom, Aldworth, & Wright, 2005; Perkonigg, Lieb, & Wittchen, 1998). Using 2004 to 2006 NSDUH data, we examined prevalence rates for many drug use outcome variables for the respondents in the 12 to 25 age range. Figure 2 shows the prevalence rates by age for three outcomes using 2006 NSDUH data. For past year alcohol use (ALCYR), the prevalence rate increased linearly from ages 12 to 21, reached a peak at age 21, then stayed approximately flat. The same pattern was observed for past year cigarette use (CIGYR). The prevalence rate for past year marijuana use (MRJYR) increased from ages 12 to 18 almost linearly, stayed relatively flat at ages 19 and 20, then decreased from ages 21 to 25 almost linearly. Similar patterns were observed for the 2004 and 2005 survey data.

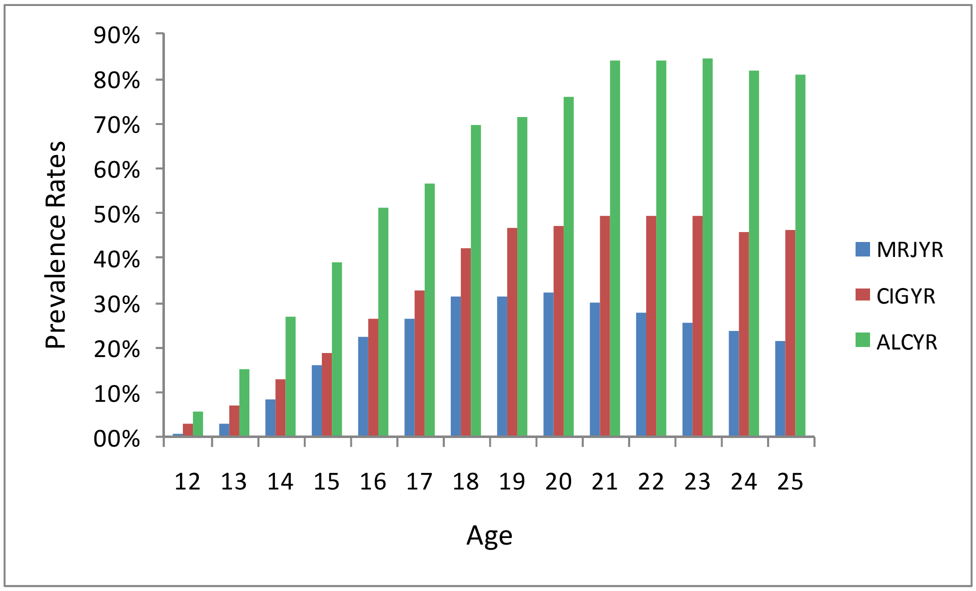

Figure 2: Prevalence rates for past year use of marijuana, cigarettes, and alcohol, by individual age (2006 NSDUH) (MRJYR = past year marijuana use, CIGYR = past year cigarette use, and ALCYR = past year alcohol use)

The strong correlation between single year ages, response rates, and drug use prevalence rates in the 12 to 25 age range suggests that use of single year ages in both the person-level NR and PS adjustments could further reduce NR and coverage bias, as well as the variance of estimates.

3. Recalibrating with Single Years of Age

We adopted the following approach for using single years of age as predictors in the person-level NR and PS adjustments for each of the nine census division-level model groups:

- added single years of age in the main effect (12,…,25) for both person-level NR and PS in place of the 12 to 17 and the 18 to 25 age groups;
- kept all of the other variables in the models;
- used age groups 12 to 17, 18 to 25, 26 to 34, 35 to 49, and 50 or older in the interactions for NR; and

- used age groups 12 to 17, 18 to 25, 26 to 34, 35 to 49, 50 to 64, and 65 or older in the interactions for PS.

After recalibrating the weights using single years of age as the main effects in the NR and PS adjustments, we checked the distribution of the recalibrated weights against the current analysis weights that used age groups as the main effects in the GEM models. Table 1 shows the weight distributions of the two sets of weights.

Table 1: Weight Distribution of Original Weight and Recalibrated Weight (2006 NSDUH)

	Overall (n = 67,802)		*12 to 17 (n = 22,871)*		*18 to 25 (n = 21,948)*	
Statistics	*Current weight*	*Recalibrated weight*	*Current weight*	*Recalibrated weight*	*Current weight*	*Recalibrated weight*
Minimum	1	1	1	1	2	2
25%	758	755	477	480	619	614
Median	1,531	1,532	881	879	1,193	1,195
75%	3,945	3,953	1,483	1,482	1,976	1,975
Maximum	94,918	89,435	16,157	10,677	19,719	15,289
Mean	3,629	3,629	1,110	1,110	1,492	1,492
Max/Mean	26.16	24.64	14.56	9.62	13.22	10.25
UWE[1]	3.32	3.30	1.68	1.68	1.70	1.69

[1] UWE = unequal weighting effect.

The distribution of the weights did not change much overall or for the two age groups (12 to 17 and 18 to 25). The measures of location exhibited minimal change. However, the most extreme weights were affected by the recalibration, as seen by the reduced maximum weight for each group. That and the slightly smaller UWE offer evidence that the recalibration reduced the variation in the weights.

4. Results

We examined the estimated numbers of past year users and the prevalence rates for a selected set of outcomes using the recalibrated weights (which used single year of age variables in the NR and PS adjustments) and compared them with the estimates based on the current weights (without the single year of age variables). Then we produced tables for several demographic domains, age groups, races/ethnicities, and by gender. For the 12 to 17 and the 18 to 25 age groups, we also looked at single year of age.

Appendices 1 to 3 show the prevalence rates and associated standard errors for past year alcohol, cigarette, and marijuana use based on the current analysis weights and the recalibrated weights. The differences between the estimated numbers of users and prevalence rates for past year alcohol use based on the two sets of weights were statistically significant for the 12 to 17 and 18 to 25 age groups, for the individual ages 12 and 15, and for every individual age in the 18 to 25 age group (see Appendix 1). Very few differences were observed in the percentages for the single years of age. For other domains, only the estimated numbers of users for non-Hispanic whites were different among the two sets of estimates. Results for past year cigarette use and past year marijuana use were similar to those for past year alcohol use (see Appendices 2 and 3);

many significant differences were observed for single years of age in the estimated numbers, but only a few in the percentages.

The difference between the estimated numbers of past year users by weighting type was more often statistically significant at single years of age compared with the estimated prevalence rates. We also computed "pseudo-mean square errors (MSEs)" (Biemer, 2010) of the estimated numbers of users for the three past year drug use outcomes. We use the term pseudo-MSE to emphasize that no true numbers of users are available, which precludes the possibility of computing true MSEs for the estimated numbers of users. The pseudo-MSE approach assumes that the estimated numbers of users based on the recalibrated weights were essentially unbiased. In that case, the MSE of the estimate is given by its sampling variance. Further, the bias of the estimates based on the current weights is estimated by the difference $\hat{t}_1 - \hat{t}_2$ and its MSE (or pseudo-MSE) by:

$$MSE_{\hat{t}_1} = (\hat{t}_1 - \hat{t}_2)^2 - \mathrm{var}(\hat{t}_2) + 2[\mathrm{var}(\hat{t}_1)\,\mathrm{var}(\hat{t}_2)]^{1/2} \qquad (1)$$

where $\hat{t}_1$ and $\hat{t}_2$ are the estimated numbers of users based on the current weights and the recalibrated weights, respectively, and $\mathrm{var}(\hat{t}_1)$ and $\mathrm{var}(\hat{t}_2)$ are their associated sampling variance estimates. This MSE estimate assumes further that the correlation between $\hat{t}_1$ and $\hat{t}_2$ is essentially one.

For past year alcohol use, the MSE of the estimates using the recalibrated weights was smaller by about 20 percent for the 12 to 17 age group and 24 percent for the 18 to 25 age group (see Table 2). The MSE was reduced at each single year of age, and the MSE dropped by 86 percent for age 25.

Table 2: Mean Square Errors (MSEs) of the Estimated Numbers of Past Year Alcohol Users (2006 NSDUH)

	Current weight		*Recalibrated weight*		
Domain (age)	*Estimated number ($\hat{t}_1$)*	*MSE($\hat{t}_1$)*[1]	*Estimated number ($\hat{t}_2$)*	*MSE($\hat{t}_2$)*[2]	*Relative diff of MSE*
12 to 17	8,361	130	8,286	104	-19.90%
12	224	19	231	18	-6.80%
13	621	34	622	32	-5.76%
14	1,157	43	1,146	37	-14.21%
15	1,776	108	1,684	44	-59.36%
16	2,209	61	2,215	44	-27.98%
17	2,374	64	2,387	44	-30.78%
18 to 25	25,814	181	25,928	138	-23.56%
18	3,187	329	2,870	48	-85.41%
19	2,987	120	2,905	47	-60.81%
20	3,062	96	3,093	46	-51.91%
21	3,601	214	3,405	40	-81.32%
22	3,325	100	3,384	38	-62.18%
23	3,331	148	3,456	38	-74.25%
24	3,215	200	3,399	44	-78.03%
25	3,105	320	3,415	44	-86.25%

[1] MSE($\hat{t}_1$) is estimated as in Formula (1). [2] MSE($\hat{t}_2$) is estimated as the sampling error [$\mathrm{var}(\hat{t}_2)$].

For past year cigarette use, the recalibration using single years of age reduced the MSEs of the estimates by about 14 percent for the 12 to 17 age group and by 4 percent for the 18 to 25 age group (see Table 3). The largest reduction in MSE was seen among 18 year olds (up to 80 percent). For past year marijuana use, a similar pattern of reduction in MSE was observed (see Table 4). The only exception was for age 12, where the MSE increased 8 percent after recalibration; this may have been caused by the low prevalence at age 12, so the age 12 estimates are unstable.

Table 3: Mean Square Errors (MSEs) of the Estimated Numbers of Past Year Cigarette Users (2006 NSDUH)

	Current weight		*Recalibrated weight*		
Domain (age)	*Estimated number* ($\hat{t}_1$)	*MSE*($\hat{t}_1$)[1]	*Estimated number* ($\hat{t}_2$)	*MSE*($\hat{t}_2$)[2]	*Relative diff of MSE*
12 to 17	4,315	99	4,270	85	-13.96%
12	119	15	125	15	-1.31%
13	288	23	283	21	-6.83%
14	547	29	545	28	-3.62%
15	845	58	803	34	-41.06%
16	1,143	50	1,132	41	-18.65%
17	1,374	50	1,382	41	-17.72%
18 to 25	15,390	163	15,430	156	-4.28%
18	1,933	229	1,715	46	-79.87%
19	1,960	91	1,904	51	-43.98%
20	1,904	77	1,932	55	-28.39%
21	2,112	141	1,992	49	-65.13%
22	1,950	80	1,994	54	-32.88%
23	1,945	100	2,015	53	-46.83%
24	1,804	121	1,906	52	-56.91%
25	1,783	199	1,971	51	-74.36%

[1] MSE($\hat{t}_1$) is estimated as in Formula (1). [2] MSE($\hat{t}_2$) is estimated as the sampling error [$\mathrm{var}(\hat{t}_2)$].

5. Conclusions

It was relatively easy to fit the single years of age (12 to 25) main effects in GEM at both the person-level NR and PS steps. Adding single years of age (12 to 25) in the person-level NR and PS adjustments did not change the weight distribution dramatically (see Table 1). Differences between the estimated numbers and percentages of users for the younger age groups and single years of age (12 to 25) using the two sets of weights were observed, while the impact for other demographic domains was minimal (see Appendices 1 to 3). The effect on the estimated numbers of users (counts) was greater than on the estimated percentages for the 12 to 17 and the 18 to 25 age groups and for the single years of age. There was also evidence that the recalibration with single years of age (12 to 25) reduced errors (lower MSE using recalibrated weights) for persons aged 12 to 25 in the NSDUH sample (see Tables 2 to 4).

Table 4: Mean Square Errors (MSEs) of the Estimated Number of Past Year Marijuana Users (2006 NSDUH)

Domain (age)	*Current weight*		*Recalibrated weight*		*Relative diff of MSE*
	Estimated number ($\hat{t}_1$)	*MSE*($\hat{t}_1$)[1]	*Estimated number* ($\hat{t}_2$)	*MSE*($\hat{t}_2$)[2]	
12 to 17	3,348	83	3,319	78	-6.27%
12	40	10	42	11	8.39%
13	136	16	131	15	-5.13%
14	363	24	366	24	-0.77%
15	723	56	681	31	-44.30%
16	974	43	975	38	-11.05%
17	1,112	46	1,124	40	-13.61%
18 to 25	9,169	169	9,090	148	-12.24%
18	1,433	166	1,279	45	-72.97%
19	1,309	72	1,269	45	-37.86%
20	1,295	66	1,303	51	-23.11%
21	1,283	95	1,211	47	-50.43%
22	1,096	58	1,112	47	-18.32%
23	1,001	61	1,033	44	-27.31%
24	933	73	986	46	-36.77%
25	820	91	897	46	-49.29%

[1] MSE($\hat{t}_1$) is estimated as in Formula (1). [2] MSE($\hat{t}_2$) is estimated as the sampling error [$\mathrm{var}(\hat{t}_2)$].

Acknowledgments

The authors thank SAMHSA's Art Hughes and Jonaki Bose for their support.

References

Biemer, P. P. (2010). Overview of design issues: Total survey error. In P. Marsden & J. Wright (Eds.), *Handbook of survey research* (2nd ed., pp. 27-57). Bingley, United Kingdom: Emerald Group Publishing, LTD.

Chen, K., & Kandel, D. B. (1995). The natural history of drug use from adolescence to the mid-thirties in a general population sample. *American Journal of Public Health, 85*, 41-47.

Chen, P., Dai, L., Gordek, H., Laufenberg, J., Liu, B., Sathe, N. S., & Westlake, M. (2008, January). Person-level sampling weight calibration. In *2006 National Survey on Drug Use and Health: Methodological resource book* (Section 12, prepared for the Substance Abuse and Mental Health Services Administration, Office of Applied Studies, under Contract No. 283-2004-00022, Phase I, Deliverable No. 39, RTI 0209009.274.002). Research Triangle Park, NC: RTI International.

Deming, W. E., & Stephan, F. F. (1940). On a least square adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, *11*, 427-444.

Deville, J. C., & Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association, 87*(418), 376-382.

Folsom, R. E. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction. In *Proceedings of the 1991 American Statistical Association, Survey Research Methods Section, Atlanta, GA* (pp. 197-201). Alexandria, VA: American Statistical Association.

Folsom R. E., & Singh, A. C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. In *Proceedings of the 2000 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Indianapolis, IN* (pp. 598–603). Alexandria, VA: American Statistical Association.

Folsom, R. E., & Witt, M. B. (1994). Testing a new attrition nonresponse adjustment method for SIPP. In Proceedings of the 1994 Joint Statistical Meetings, American Statistical Association, Social Statistics Section, Toronto, Ontario, Canada (pp. 428-433). Alexandria, VA: American Statistical Association.

Kandel, D. B., & Logan, J. A. (1984). Patterns of drug use from adolescence to young adulthood: I. Periods of risk for initiation, continued use, and discontinuation. *American Journal of Public Health, 74*, 660-666.

Little R. J. (1986). Survey nonresponse adjustment. *International Statistics Review, 50*, 139-157.

Odom, D., Aldworth, J., & Wright, D. (2005, May). *2002 National Survey on Drug Use and Health: Summary of UEAT analysis of field interventions* (RTI 7190, prepared for the Office of Applied Studies, Substance Abuse and Mental Health Services Administration, under Contract No. 283-98-9008). Research Triangle Park, NC: RTI International.

Perkonigg, A., Lieb, R., & Wittchen H. (1998). Prevalence of use, abuse and dependence of illicit drugs among adolescents and young adults in a community sample. *European Addiction Research, 4*(1-2), 58-66.

Sautory, O. (1991). Redressements d'échantillons d'enquêtes auprès des ménages par calage sur marges. *Proceedings, Journée de Méthodologie Statistique*, 299-326. [Available as a PDF at http://jms.insee.fr/files/documents/2008/616_1-JMS1991_S4-3_SAUTORY.PDF]

Appendix 1: Estimated Numbers and Percentages of Past Year Alcohol Users (2006 NSDUH)

	Using current weight				*Using recalibrated weight*			
Domain	*Esti-mated number*	*SE*	*Esti-mated percent-age*	*SE*	*Esti-mated number*	*SE*	*Esti-mated percent-age*	*SE*
Age/Age Group								
12 to 17	8,361	106	32.9	0.42	8,286[b]	104	32.6[b]	0.41
12	224	18	5.7	0.45	231[b]	18	5.6[a]	0.45
13	621	34	15.0	0.77	622	32	14.9	0.77
14	1,157	42	27.0	0.87	1,146	37	27.0	0.86
15	1,776	59	39.1	1.02	1,684[b]	44	39.1	1.03
16	2,209	64	51.0	1.04	2,215	44	50.9	1.01
17	2,374	66	56.7	1.05	2,387	44	56.7	1.05
18 to 25	25,814	140	78.8	0.43	25,928[b]	138	79.2[b]	0.42
18	3,187	104	69.5	1.16	2,870[b]	48	69.6	1.16
19	2,987	105	71.5	1.17	2,905[b]	47	71.9	1.16
20	3,062	112	76.0	1.18	3,093[b]	46	76.3	1.13
21	3,601	113	83.9	0.97	3,405[b]	40	83.7	0.98
22	3,325	106	84.2	0.95	3,384[b]	38	84.2	0.95
23	3,331	100	84.4	0.93	3,456[b]	38	84.5	0.93
24	3,215	93	82.0	1.13	3,399[b]	44	82.1	1.07
25	3,105	93	80.9	1.09	3,415[b]	44	81.3	1.06
26 to 34	27,221	242	77.5	0.69	27,185	243	77.4	0.69
35 to 49	48,913	395	74.9	0.60	48,995	392	75.0	0.60
50 to 64	34,043	506	65.7	0.98	33,988	502	65.6	0.97
65 or Older	18,063	474	50.7	1.33	18,016	476	50.5	1.33
Race/Ethnicity								
Hispanic	19,365	318	57.9	0.95	19,263	314	57.6	0.94
Non-Hispanic White	119,433	703	70.4	0.41	119,559[b]	700	70.4	0.41
Non-Hispanic Black	16,262	313	55.1	1.06	16,339	310	55.2	1.05
Non-Hispanic Others	7,355	259	55.2	1.94	7,234	249	55.4	1.91
Gender								
Male	83,275	569	69.8	0.48	83,262	567	69.8	0.47
Female	79,140	630	62.5	0.50	79,134	624	62.5	0.49

[a] Difference in the estimated number/percentage produced from the two sets of weights is significant at $\alpha = 0.05$.

[b] Difference in the estimated number/percentage produced from the two sets of weights is significant at $\alpha = 0.01$.

Appendix 2: Estimated Numbers and Percentages of Past Year Cigarette Users (2006 NSDUH)

	Using current weight				*Using recalibrated weight*			
Domain	*Esti-mated number*	*SE*	*Esti-mated percent-age*	*SE*	*Esti-mated number*	*SE*	*Esti-mated percent-age*	*SE*
Age/Age Group								
12-17	4,315	88	17.0	0.35	4,270[b]	85	16.8[b]	0.33
12	119	14	3.1	0.36	125[b]	15	3.1	0.37
13	288	22	6.9	0.52	283	21	6.8[b]	0.51
14	547	29	12.8	0.65	545	28	12.8	0.67
15	845	40	18.6	0.79	803[b]	34	18.7	0.79
16	1,143	50	26.4	1.00	1,132	41	26.0	0.94
17	1,374	50	32.8	0.98	1,382	41	32.8	0.97
18 to 25	15,390	158	47.0	0.48	15,430	156	47.1	0.48
18	1,933	74	42.1	1.13	1,715[b]	46	41.6[b]	1.13
19	1,960	76	46.9	1.25	1,904[b]	51	47.1	1.26
20	1,904	74	47.3	1.41	1,932[a]	55	47.6	1.35
21	2,112	79	49.2	1.19	1,992[b]	49	49.0	1.19
22	1,950	69	49.4	1.32	1,994[b]	54	49.6	1.34
23	1,945	74	49.3	1.30	2,015[b]	53	49.3	1.30
24	1,804	66	46.0	1.26	1,906[b]	52	46.1	1.27
25	1,783	67	46.5	1.23	1,971[b]	51	46.9	1.22
26 to 34	13,879	270	39.5	0.77	13,814	267	39.3	0.76
35 to 49	20,890	410	32.0	0.63	20,938	412	32.1	0.63
50 to 64	13,251	430	25.6	0.83	13,233	430	25.5	0.83
65 or Older	3,951	288	11.1	0.81	3,909	287	11.0	0.81
Race/Ethnicity								
Hispanic	9,166	288	27.4	0.86	9,028[a]	283	27.0[a]	0.85
Non-Hispanic White	51,475	674	30.3	0.40	51,502	675	30.3	0.40
Non-Hispanic Black	8,140	290	27.6	0.98	8,184	291	27.7	0.98
Non-Hispanic Others	2,896	192	21.7	1.44	2,880	189	22.1	1.45
Gender								
Male	38,802	559	32.5	0.47	38,695	556	32.4	0.47
Female	32,875	529	26.0	0.42	32,900	527	26.0	0.42

[a] Difference in the estimated number/percentage produced from the two sets of weights is significant at $\alpha = 0.05$.

[b] Difference in the estimated number/percentage produced from the two sets of weights is significant at $\alpha = 0.01$.

Appendix 3: Estimated Numbers and Percentages of Past Year Marijuana Users (2006 NSDUH)

	Using current weight				*Using recalibrated weight*			
Domain	*Estimated number*	*SE*	*Estimated percentage*	*SE*	*Estimated number*	*SE*	*Estimated percentage*	*SE*
Age/Age Group								
12 to 17	3,348	78	13.2	0.31	3,319[a]	78	13.1[a]	0.31
12	40	10	1.0	0.27	42[b]	11	1.0	0.27
13	136	15	3.3	0.36	131	15	3.1	0.35
14	363	24	8.5	0.55	366	24	8.6	0.57
15	723	37	15.9	0.74	681[b]	31	15.8	0.73
16	974	43	22.5	0.87	975	38	22.4	0.88
17	1,112	45	26.6	0.93	1,124[a]	40	26.7	0.94
18 to 25	9,169	149	28.0	0.46	9,090[b]	148	27.8[b]	0.45
18	1,433	67	31.2	1.11	1,279[b]	45	31.0	1.10
19	1,309	63	31.3	1.14	1,269[b]	45	31.4	1.12
20	1,295	68	32.1	1.29	1,303	51	32.1	1.25
21	1,283	64	29.9	1.14	1,211[b]	47	29.8	1.15
22	1,096	56	27.7	1.16	1,112[a]	47	27.7	1.18
23	1,001	52	25.4	1.05	1,033[b]	44	25.3	1.07
24	933	50	23.8	1.09	986[b]	46	23.8	1.10
25	820	48	21.4	1.10	897[b]	46	21.3	1.10
26 to 34	5,018	211	14.3	0.60	5,039	212	14.4	0.60
35 to 49	5,570	262	8.5	0.40	5,593	260	8.6	0.40
50 to 64	2,099	192	4.1	0.37	2,103	192	4.1	0.37
65 or Older	174	75	0.5	0.21	169	73	0.5	0.20
Race/Ethnicity								
Hispanic	2,684	149	8.0	0.45	2,636	144	7.9	0.43
Non-Hispanic White	18,150	397	10.7	0.23	18,146	399	10.7	0.23
Non-Hispanic Black	3,694	177	12.5	0.60	3,690	174	12.5	0.59
Non-Hispanic Others	849	81	6.4	0.61	842	80	6.4	0.61
Gender								
Male	15,592	364	13.1	0.30	15,571	362	13.0	0.30
Female	9,785	254	7.7	0.20	9,742	255	7.7	0.20

[a] Difference in the estimated number/percentage produced from the two sets of weights is significant at $\alpha = 0.05$.

[b] Difference in the estimated number/percentage produced from the two sets of weights is significant at $\alpha = 0.01$.