

An Empirical Study to Evaluate the Performance of Synthetic Estimates of Substance Use in the National Survey on Drug Use and Health

Akhil K. Vaish,¹ Ralph E. Folsom,¹ Kathy Spagnola,¹
Neeraja Sathe,¹ Art Hughes²

¹RTI International, 3040 Cornwallis Road, RTP, NC 27709

²SAMHSA, Center for Behavioral Health Statistics and Quality,
Rockville, Maryland 20857

Abstract

Small area estimation methods are used to produce State and substate estimates of substance use and mental disorders using data from the National Survey on Drug Use and Health. Design-based estimates could be used as an alternative because they are less expensive than small area estimates and take less time to produce. Thus, it is important to determine how the small area estimates compare with their design-based counterparts in terms of accuracy and precision. A previous study demonstrated that small area estimates were generally more precise than design-based estimates while exhibiting only small levels of bias. In this paper, those results are extended by conducting an additional simulation study to evaluate the performance of synthetic estimates. These estimates are commonly produced for small areas where no sample data can be obtained, and this study aims to provide some guidance about the quality of such estimates.

Key Words: Small area estimation, design-based estimates, model-based estimates, synthetic estimates, National Survey on Drug Use and Health (NSDUH) small area estimates

1. Background

The National Survey on Drug Use and Health (NSDUH)¹ is the primary source of statistical information on the use of illicit drugs, alcohol, and tobacco by the U.S. civilian, noninstitutionalized population aged 12 or older. It 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 Center for Behavioral Health Statistics and Quality (CBHSQ, formerly the Office of Applied Studies [OAS]). Data collection and analysis are conducted under contract with RTI International.² NSDUH is an ongoing survey that administers a variety of questions on substance use and associated behaviors in a face-to-face setting via computer-assisted interviewing at the respondent's place of residence. It is a multistage area probability survey with a target sample of 67,500 persons nationwide. Prior to 1999, the NSDUH design employed a national probability sample that did not have a sufficient sample to produce State estimates. Beginning in 1999, the sample was expanded so that representative estimates

¹ Prior to 2002, the survey was known as the National Household Survey on Drug Abuse (NHSDA).

² RTI International is a trade name of Research Triangle Institute.

could be provided in each State and the District of Columbia. The current NSDUH design includes eight large sample States (California, Florida, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas) that collectively account for approximately 50 percent of the U.S. population. The annual target sample sizes are 3,600 for each of these large States and 900 for the remaining 42 States and the District of Columbia. The design oversamples youths aged 12 to 17 and young adults aged 18 to 25 so that each State's annual sample is approximately equally distributed among three major age groups: 12 to 17, 18 to 25, and 26 or older.

In previous validation studies, NSDUH State small area estimates were validated for annual sample sizes of about 900 (300 per age group) and 2-year pooled sample sizes of 1,800 (600 per age group) for the 12 or older age group (Wright, 2002a, 2002b, 2003a, 2003b). The results of these validation studies were presented to a panel of small area estimation (SAE) experts who recommended that single year NSDUH data will be sufficient to produce reliable age group-specific State estimates; however, single year data may not be sufficient to detect year-to-year changes in State prevalence rates for low prevalence outcome measures. The SAE panel members recommended producing State estimates using pooled 2 years of NSDUH data and estimating change as the difference between two consecutive 2-year moving averages.

It was inferred from the above-mentioned validation studies that if an area has a sample size of 300 respondents, then reliable point estimates could be produced. Based on this knowledge, for the first 1999 to 2001 substate report (OAS, 2005), the substate regions were formed by grouping contiguous counties (in some cases, census tracts) so that each of the substate regions had at least 300 respondents in the pooled 3 years of NSDUH data. The main objective was to produce substate region estimates for the 12 or older population. With the availability of data from subsequent NSDUHs, the scope of the substate reporting was expanded to include the production of age group-specific substate region estimates and change estimates between the two most recent substate region prevalence rates (e.g., changes between 2006 to 2008 vs. 2008 to 2010 substate prevalence rates). Moreover, the demand for substate region estimates has grown steadily because local area officials need such information for various purposes, such as treatment planning, intervention, and prevention. As a result, some substate region sample sizes are now becoming as small as 150 (50 per age group). To maintain the quality of substate region estimates, the age group-specific substate region estimates are subjected to a suppression rule that uses relative standard errors (RSEs) and effective sample size restrictions to suppress unreliable estimates. However, the accuracy (i.e., the mean squared error [MSE] that combines the squared bias and variance) of the substate region estimates depends heavily on model assumptions. Hughes, Vaish, Sathe, and Spagnola (2012) demonstrated that the substate region small area estimates were generally more precise than corresponding design-based estimates while exhibiting only small levels of bias.

In this paper, we extend the results of Hughes et al. (2012) by conducting an additional simulation study to evaluate the performance of synthetic estimates. About 2,220 counties are present in 3 years of pooled NSDUH data with $1 \leq n \leq 4,070$ and a median sample size = 32. In order to produce reliable design-based county-level estimates, many years (10 or more) of NSDUH data would have to be combined. This may not be an appealing option if estimates covering a more recent time period are needed. Even in the pooled 10 years of NSDUH data, many counties would still be missing in the sample. The second option would be to pool only enough NSDUH data to reflect current trends

and use SAE methodology. The estimates for the majority of the counties would be purely synthetic (i.e., they would be based on the model and the data from nonmissing counties). The quality of the synthetic estimates would depend on several factors, such as the model fit, leftover variation between counties after adjusting for fixed effects, and the similarity between the characteristics of the missing and nonmissing counties. This study aims to provide some guidance about the quality of synthetic estimates. Currently, county-level estimates using NSDUH data are not routinely produced because of the absence of any information about the MSE of county-level estimates. In Section 2, the simulation methodology used for judging the quality of synthetic estimates is presented, the results are discussed in Section 3, and conclusions and recommendations for future research are presented in Section 4.

2. Method

The main idea is to first create reliable benchmark estimates for comparing the synthetic estimates and the small area estimates with these reasonably "true values" for a variety of outcome measures (low, medium, and high prevalence rates). Within each benchmark area, 16 replicated random samples were selected. For ease of computation, these 16 replicated samples were modeled simultaneously by treating them as 16 pseudo substate areas. Small area estimates and associated synthetic estimates were produced for these pseudo substate areas and compared with the benchmark area-level estimates. We chose the following outcome measures from NSDUH:

- past month binge alcohol use (BNGALC) (~24.0 percent),
- past month use of cigarettes (CIGMON) (~22.0 percent),
- past year use of cocaine (COCYR) (~2.3 percent), and
- past month use of marijuana (MRJMON) (~9.4 percent),

where the percentages in parentheses show the national prevalence rate for the 12 or older age group.

For this study, pooled 2009 and 2010 NSDUH data were used. There are 48 State sampling regions (SSRs³) in each of the eight large States. Each of the 48 SSRs is expected to have eight segments⁴ in a single year of NSDUH data. Out of eight segments, about half of the segments are used again in the next year for sampling, and the other half of the segments are replaced with new segments. For example, if s_1, \dots, s_8 were the eight segments in one particular SSR (say, SSR1) in the 2009 NSDUH data, then in 2010, say, segments s_1, \dots, s_4 are kept, and s_5, \dots, s_8 were replaced by s_9, \dots, s_{12} . Hence, in 2010, SSR1 will have the s_1, \dots, s_4 and the s_9, \dots, s_{12} segments. So, in the pooled 2009 and 2010 NSDUH data, SSR1 will have s_1, \dots, s_{12} segments. Each of the single year NSDUH segments is expected to have on average 9.375 respondents. Because the s_1, \dots, s_4 segments occur twice in the pooled 2009 and 2010 NSDUH data, they will have in all about 75 ($4 \times 9.375 \times 2$) respondents, and the remaining 8 segments (s_5, \dots, s_{12}) will have

³ State sampling regions (SSRs) are contiguous geographic areas designed to yield approximately the same number of interviews. Within each SSR, 48 census tracts were selected with probability proportional to population size.

⁴ Within sampled census tracts, adjacent census blocks were combined to form the second-stage sampling units or area segments. One area segment was selected within each sampled census tract with probability proportional to population size.

about 75 (8×9.375) respondents. Hence, SSR1 will have about 150 respondents in the pooled 2009 and 2010 NSDUH data, with one third of the segments (four segments) containing on average 18.75 respondents per segment and two thirds of the segments (eight segments) containing 9.375 respondents per segment. Across 48 SSRs, there will be about $150 \times 48 = 7,200$ respondents in each of the eight large States in the pooled 2009 and 2010 NSDUH data.

Each of the eight large States was partitioned into two parts with $n = \sim 3,600$ resulting in $8 \times 2 = 16$ benchmark areas, and their design-based estimates were treated as "true value." It is as if each of the large States has two substate areas or counties. Part 1 of each of the eight large States was formed to have lower average prevalence rates than Part 2 across the four outcome measures. This was done to simulate real-life situations where counties within States have varying prevalence rates (i.e., some are higher and some are lower than the State prevalence rates), and some of the county-level rates are similar to the State prevalence rates. For example, to create two benchmark areas in California, design-based estimates for 48 SSRs were produced and ranked (1 to 48) for four outcome measures. Each of the 48 SSRs was again ranked based on its average rank over four outcome measures. The first 24 SSRs were grouped to form Part 1 of California, and the remaining 24 SSRs were grouped to form Part 2. The process was repeated for the remaining seven large sampling States. Table 1 shows the overall State and its Part 1 and Part 2 design-based prevalence rates for four outcome measures for the 12 or older age group.

Table 1: Overall State and Its Part 1 and Part 2 Prevalence Rates

<i>State</i>	<i>Outcome Measure</i>	<i>Overall (%)</i>	<i>Part 1 (%)</i>	<i>Part 2 (%)</i>
California	BNGALC	22.6	19.2	26.3
California	CIGMON	18.2	15.4	21.4
California	COCYR	2.1	1.2	3.1
California	MRJMON	8.2	5.9	10.7
Florida	BNGALC	20.7	16.7	24.8
Florida	CIGMON	22.6	19.2	26.1
Florida	COCYR	1.9	0.8	2.9
Florida	MRJMON	6.2	5.5	7.1
Illinois	BNGALC	25.9	22.3	29.8
Illinois	CIGMON	24.5	19.9	29.3
Illinois	COCYR	1.7	0.7	2.7
Illinois	MRJMON	6.4	4.3	8.7
Michigan	BNGALC	27.0	23.4	31.1
Michigan	CIGMON	26.3	23.1	29.8
Michigan	COCYR	1.6	0.9	2.4
Michigan	MRJMON	8.6	6.7	10.9
New York	BNGALC	23.8	19.1	28.6
New York	CIGMON	21.5	17.8	25.3
New York	COCYR	2.2	1.2	3.2
New York	MRJMON	7.7	5.4	10.1
Ohio	BNGALC	23.6	20.6	26.8
Ohio	CIGMON	25.8	23.0	28.6
Ohio	COCYR	1.4	0.9	2.0
Ohio	MRJMON	6.0	4.3	7.7

(continued)

Table 1: Overall State and Its Part 1 and Part 2 Prevalence Rates (continued)

<i>State</i>	<i>Outcome Measure</i>	<i>Overall (%)</i>	<i>Part 1 (%)</i>	<i>Part 2 (%)</i>
Pennsylvania	BNGALC	24.4	21.1	28.2
Pennsylvania	CIGMON	24.5	19.5	30.3
Pennsylvania	COCYR	1.7	0.9	2.7
Pennsylvania	MRJMON	5.5	3.6	7.8
Texas	BNGALC	24.2	22.6	25.9
Texas	CIGMON	22.1	18.9	25.6
Texas	COCYR	1.8	1.2	2.5
Texas	MRJMON	5.3	3.8	6.9

After creating 16 benchmark areas, we created 16 pseudo substate areas with $n = 225$ (approximately) in each of the 16 benchmark areas, resulting in $16 \times 16 = 256$ pseudo substate areas. For this purpose, we randomly partitioned each benchmark area's sample of segments into 16 replicated subsamples. Note that Part 1 and Part 2 of each large State contains 24 SSRs, and each of the SSRs contains 12 segments. Hence, there are $24 \times 12 = 288$ segments in each of the two parts. These 288 segments were randomly partitioned into 16 groups, which are referred to as pseudo substate areas. Each of the 16 pseudo substate areas had about 18 segments. As mentioned earlier, about one third of the segments (6 segments) contain on average 18.75 respondents per segment, and two thirds of the segments (12 segments) contain 9.375 respondents per segment, resulting in about $6 \times 18.75 + 12 \times 9.375 = 225$ respondents in each of the 16 pseudo substate areas. From the previous two validation studies, $n = 225$ appears to be sufficient to produce reliable small area estimates as compared with the corresponding design-based estimates.

For producing small area estimates, the age group-specific (12 to 17, 18 to 25, 26 or older) respondent-level weights in each of the 256 pseudo substate areas were post-stratified to the corresponding 1/16 of the benchmark area-level analysis weight totals. Small area estimates were produced for past month binge alcohol use (BNGALC), past month cigarette use (CIGMON), past year cocaine use (COCYR), and past month marijuana use (MRJMON) using a simpler version of the current NSDUH-SAE models⁵ (i.e., only age group-specific pseudo substate area-level random effects were included in the logistic mixed models along with the current fixed effects predictors). Note that, in the current NSDUH-SAE models, age group-specific State random effects and within State-level (grouped three SSRs or substate region) random effects are fitted. Specifically, the following logistic mixed model was fitted:

$$\log[\pi_{aijk}(1 - \pi_{aijk})] = x'_{aijk} \beta_a + \eta_{bj},$$

where π_{aijk} is the probability of engaging in the behavior of interest (e.g., using marijuana in the past month) for person- k belonging to age group- a ($1 = 12$ to 17 , $2 = 18$ to 25 , $3 = 26$ to 34 , and $4 = 35$ or older) in pseudo substate area- j ($j = 1, \dots, 16$) of benchmark area- i ($i = 1, \dots, 16$). Let x_{aijk} denote a $p_a \times 1$ vector of auxiliary (predictor) variables associated with age group- a and β_a denote the associated vector of regression parameters. The age group-specific vectors of auxiliary variables are defined for every

⁵ For details, see the "2010-2011 National Survey on Drug Use and Health: Guide to State Tables and Summary of Small Area Estimation Methodology" at <http://www.samhsa.gov/data/NSDUH/2k11State/NSDUHsaeMethodology2011.htm>.

block group in the Nation and also include person-level demographic variables, such as race/ethnicity and gender. Because of small sample sizes for the 26 to 34 age group at the pseudo substate area level, the 26 to 34 age group was combined with the 35 or older age group and the pseudo substate area-level random effects (η_{bj}) were fitted for the pooled age group- b ($1 = 12$ to 17 , $2 = 18$ to 25 , and $3 = 26$ or older). The vectors of the pseudo substate area-level random effects $\eta'_{bj} = (\eta_{1j}, \eta_{2j}, \eta_{3j})$ are assumed to have a multivariate normal distribution $\eta_{bj} \sim MNV(0, W)$, where W is a 3×3 variance-covariance matrix. The model parameters (β_a , η_{bj} , and W) were estimated using the survey-weighted hierarchical Bayes SAE methodology of Folsom, Shah, and Vaish (1999).

Using the fitted model, small area estimates and synthetic estimates for the 256 pseudo substate areas for four outcome measures were produced. The corresponding design-based estimates (weighted averages) were also produced.

The small area estimates for group- a ($1 = 12$ to 17 , $2 = 18$ to 25 , $3 = 26$ to 34 , and $4 = 35$ or older), pseudo substate area- j ($j = 1, \dots, 16$) within the benchmark area- i ($i = 1, \dots, 16$) are constructed by first forming the population-level averages of the person- k level predicted values given in equation (1) for each Markov Chain Monte Carlo (MCMC) sample; these population-level averages are then further averaged over the MCMC samples to get the required estimate. The person- k level predicted values for each MCMC sample is given by the following:

$$P_{aijk}(SAE) = \frac{\exp(X'_{aijk} \hat{\beta}_a + Z'_{aijk} \hat{\eta}_{aj})}{1 + \exp(X'_{aijk} \hat{\beta}_a + Z'_{aijk} \hat{\eta}_{aj})}, \quad (1)$$

where X_{aijk} denotes the population-level vector of covariates, $\hat{\beta}_a$ denotes the vector of the estimated regression parameters, $Z'_{aijk} = (z_{1ijk}, z_{2ijk}, z_{3ijk}, z_{4ijk})$ denotes a 4×1 vector of indicator values (1/0), where $z_{aijk} = 1$ (or 0) if person- k belongs (or does not belong) to age group- a , and $\hat{\eta}_{aj}$ denotes the estimated random effects vector for pseudo substate area- j , where $\hat{\eta}_{4j} = \hat{\eta}_{3j}$ because a common random effect was fitted to age groups 3 and 4.

If estimates are desired for areas where there are no sample data, then synthetic estimates are produced. The population-level averages of the person- k level predicted values given in equation (2) are formed for each MCMC sample and then averaged over the MCMC samples to produce the synthetic estimates for the age group and substate area of interest. The person- k level predicted values for each MCMC sample is given by the following:

$$P_{aijk}(SYE) = \frac{\exp(X'_{aijk} \hat{\beta}_a + Z'_{aijk} \varepsilon)}{1 + \exp(X'_{aijk} \hat{\beta}_a + Z'_{aijk} \varepsilon)}, \quad (2)$$

where $\varepsilon \sim MNV(0, \hat{W})$. Note that $\hat{\beta}_a$ and \hat{W} are estimated using the data from other areas in the sample and it is assumed that X'_{aijk} is known for the missing areas. Because fixed effects parameters ($\hat{\beta}_a$) are estimated using all of the data in the sample, there is very little variation between the fixed parts of the synthetic estimates for the 16 pseudo substate areas. To obtain an estimate of the model mean squared prediction error (MSPE) for the synthetic estimates, we randomly generated random effects using a multivariate normal distribution $\varepsilon \sim MNV(0, \hat{W})$. Note that after averaging over the MCMC samples, the synthetic estimates effectively reduce to just the fixed part of the population average because the simulated random effects (ε) average to zero over the MCMC samples. The addition of the zero mean random effects (ε) at each MCMC sample is there to estimate the squared bias part of the MSE.

The main difference between small area estimates and the synthetic estimates is the ability to estimate area-specific random effects (η_{bj}) using associated area-level data. When there are no data from an area, random effects are simulated by randomly generating $\varepsilon \sim MNV(0, \hat{W})$ to produce proper MSPE. For example, if 100 areas are not present in the sample, 100 random vectors (ε) are simulated from the $MNV(0, \hat{W})$ distribution for each MCMC sample and used in equation (2) to produce synthetic estimates.

3. Results

Recall that there are 16 benchmark areas (2 in each of the 8 large States) and that for each benchmark area 16 pseudo substate areas were created; hence, there are $16 \times 16 = 256$ pseudo substate areas. A total of four age group-specific (12 or older, 12 to 17, 18 to 25, and 26 or older) small area estimates for BNGALC, CIGMON, COCYR, and MRJMON were produced for each of the 256 pseudo substate areas. Hence, 4,096 (256 pseudo substate areas \times 4 age groups \times 4 outcome measures) small area estimates, synthetic estimates, and the corresponding design-based estimates were produced. These estimates were compared with the corresponding benchmarked estimates, which are design-based estimates for benchmark areas. In our simulation study, we estimated the random effects for each of the 256 pseudo substate areas and used equation (1) to produce small area estimates. Later, we simulated 256 random vectors (ε) using $MNV(0, \hat{W})$ and used equation (2) to produce corresponding synthetic estimates. Note that between the small area estimates and the synthetic estimates, $X'_{aijk}\hat{\beta}_a$ remains the same, which allowed us to specifically see the impact of random effects on the estimates. For comparing estimates with the benchmarked estimates, we defined the following measures:

Let $\hat{\theta}_{ij}$ denote an estimate (small area estimate, synthetic estimate, or design-based estimate) of θ_i (true value or benchmarked estimate) for the pseudo substate area- j ($j = 1$ to 16) within benchmark area- i ($i = 1$ to 16), then relative absolute bias (RAB) and relative root mean squared error (RRMSE) of $\hat{\theta}_{ij}$ is defined as

$$RAB(\hat{\theta}_i) = 100 \times \frac{abs(\bar{\theta}_i - \theta_i)}{\theta_i} \text{ and}$$

$$RRMSE(\hat{\theta}_i) = 100 \times \frac{\sqrt{\frac{1}{16} \sum_{j=1}^{16} (\hat{\theta}_{ij} - \theta_i)^2}}{\theta_i},$$

respectively, where $\bar{\theta}_i = (\sum_{j=1}^{16} \hat{\theta}_{ij} \div 16)$. Additionally, design-based 95 percent confidence

interval (CI) widths and synthetic estimate (or SYE) CI widths were compared with the small area estimate Bayes confidence interval (BCI) (or prediction interval) widths by using the following measures:

$$WDB_SAE = \frac{1}{16} \sum_{j=1}^{16} \frac{\text{Design-based CI width for the } j\text{th pseudo substate area}}{\text{SAE BCI width for the } j\text{th pseudo substate area}} \text{ and}$$

$$WSYE_SAE = \frac{1}{16} \sum_{j=1}^{16} \frac{\text{SYE CI width for the } j\text{th pseudo substate area}}{\text{SAE BCI width for the } j\text{th pseudo substate area}}.$$

In some cases, a design-based CI did not exist because some of the pseudo substate area design-based estimates were 0; such cases were excluded when calculating the ratio of widths.

Table 2 shows the RAB (percentage) and RRMSE (percentage) for 4 outcome measures averaged over 16 benchmark areas. Age group-specific (12 to 17, 18 to 25, and 26 or older) RABs and RRMSEs were further averaged and shown in Table 2 along with the 12 or older RABs and RRMSEs. Similarly, Table 3 shows the corresponding averages for ratio of 95 percent design-based CI width with SAE prediction interval width and averages for the ratio of synthetic estimate CI width with SAE prediction interval width.

Table 2: Average Relative Absolute Bias and Relative Root Mean Squared Error

Outcome	Age Group ¹	Relative Absolute Bias (%)			Relative Root Mean Squared Error (%)		
		DBE	SAE	SYE	DBE	SAE	SYE
BNGALC	0	1.1	10.3	15.0	18.6	12.8	15.1
	1, 2, 3	1.7	7.9	11.1	29.0	10.4	11.2
CIGMON	0	1.0	10.0	14.6	19.5	12.9	14.7
	1, 2, 3	1.5	8.7	13.0	29.6	13.0	13.1
COCYR	0	3.5	45.2	90.8	67.6	52.1	91.2
	1, 2, 3	5.4	49.8	90.6	107.4	55.7	90.8
MRJMON	0	2.6	18.1	36.8	35.1	25.0	36.9
	1, 2, 3	3.0	16.9	34.1	47.3	24.3	34.3

DBE = design-based estimate; SAE = small area estimate; SYE = synthetic estimate.

¹Age Group: 0 = 12 or older; 1 = 12 to 17; 2 = 18 to 25; and 3 = 26 or older.

Table 3: Average Ratio of Widths of 95 Percent Confidence Intervals

<i>Outcome</i>	<i>Age Group¹</i>	<i>Ratio of Widths of 95 Percent Confidence Intervals</i>	
		<i>DBE CI/SAE BCI</i>	<i>SYE CI/SAE BCI</i>
BNGALC	0	1.9	2.6
	1, 2, 3	2.6	2.4
CIGMON	0	2.0	2.6
	1, 2, 3	2.3	2.6
COCYR	0	1.9	3.5
	1, 2, 3	4.5	3.0
MRJMON	0	1.8	3.3
	1, 2, 3	2.4	3.1

BCI = Bayes confidence interval; CI = confidence interval; DBE = design-based estimate; SAE = small area estimate; SYE = synthetic estimate.

¹Age Group: 0 = 12 or older; 1 = 12 to 17; 2 = 18 to 25; and 3 = 26 or older.

Table 2 shows that the bias of the design-based estimate is less than or equal to the bias of the small area estimate, which is less than or equal to the bias of the synthetic estimate. Note that small area estimates are designed to accept some bias in order to minimize MSE, whereas the design-based estimates should be almost unbiased under repeated sampling. The synthetic estimates are supposed to exhibit more bias than small area estimates. Hence, these results are not unexpected. The bias for design-based estimates is small but not negligible. This may be due to the fact that only 16 pseudo substate areas (sample replicates) were created, which was not a sufficient number for the average of the associated design-based estimates to match closely the corresponding benchmarked estimates. The relatively small number of segments (about 18) per pseudo substate area may also lead to nonnegligible ratio estimation bias. From Table 2, we also notice that small area estimates have the least average RRMSEs. For COCYR, the performance of all three estimators deteriorates, with the small area estimates showing less deterioration than the others. The RAB and the RRMSE for the synthetic estimates are very close to each other, which is also an expected result because synthetic estimates are formed using the fixed parameter estimates that exhibit very little variance and therefore do not change across the 16 pseudo substate areas within a benchmark area.

It is clear from Table 3 that the small area estimate BCI widths are always smaller than the design-based CI widths and the synthetic estimate CI widths. It can be noted that the small area estimates are clearly performing much better than the design-based estimates and the synthetic estimates in terms of RAB, RRMSE, and the ratio of widths.

4. Conclusions and Recommendations

The synthetic estimates performed better than expected, especially for the higher prevalence outcome measures (BNGALC, CIGMON). Perhaps after adjusting for age, race, gender, and a few other covariates, not enough variability remained among our pseudo substate areas to benefit from the fitted substate area-level random effects. This needs to be investigated further. The CIs for synthetic estimates are much wider than expected. That is, the simulation MSEs are much smaller than the model-based MSEs for the synthetic estimates. This also needs to be investigated further.

Our limited simulation study does not fully account for real-life situations where substate areas within States generally exhibit wide-ranging characteristics. Intuitively, if the missing areas lie close to the fixed regression line, then their synthetic estimates will be

acceptable. Otherwise, they will be highly biased. A more comprehensive simulation study is needed to further evaluate the quality of the synthetic estimates.

Acknowledgment

The 2010-2011 NSDUH was funded by the Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality, under Contract No. HHSS283200800004C and conducted under RTI Project No. 0211838.

Disclaimer

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

References

- Folsom, R. E., Shah, B., & Vaish, A. (1999). Substance abuse in States: A methodological report on model based estimates from the 1994-1996 National Household Surveys on Drug Abuse. In *Proceedings of the 1999 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, Baltimore, MD* (pp. 371-375). Alexandria, VA: American Statistical Association.
- Hughes, A., Vaish, A., Sathe, N., & Spagnola, K. (2012). Evaluation of small area estimates of substance use in the National Survey on Drug Use and Health. In *Proceedings of the 2012 Joint Statistical Meetings, American Statistical Association, Survey Research Methods Section, San Diego, CA* (pp. 4661-4674). Alexandria, VA: American Statistical Association.
- Office of Applied Studies. (2005, May). *Substate estimates from the 1999-2001 National Surveys on Drug Use and Health*. Rockville, MD: Substance Abuse and Mental Health Services Administration.
- Wright, D. (2002a). *State estimates of substance use from the 2000 National Household Survey on Drug Abuse: Volume I. Findings* (HHS Publication No. SMA 02-3731, NHSDA Series H-15). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.
- Wright, D. (2002b). *State estimates of substance use from the 2000 National Household Survey on Drug Abuse: Volume II. Supplementary technical appendices* (HHS Publication No. SMA 02-3732, NHSDA Series H-16). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.
- Wright, D. (2003a). *State estimates of substance use from the 2001 National Household Survey on Drug Abuse: Volume I. Findings* (HHS Publication No. SMA 03-3775, NHSDA Series H-19). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.
- Wright, D. (2003b). *State estimates of substance use from the 2001 National Household Survey on Drug Abuse: Volume II. Individual State tables and technical appendices* (HHS Publication No. SMA 03-3826, NHSDA Series H-20). Rockville, MD: Substance Abuse and Mental Health Services Administration, Office of Applied Studies.