STATISTICS IN TRANSITION new series, March 2022
Vol. 23, No. 1, pp. 1–20, DOI 10.21307/stattrans-2022-001
Received – 01.01.2022; accepted – 24.02.2022

Estimation of mask effectiveness perception for small domains using multiple data sources

Aditi Sen[1], **Partha Lahiri**[2]

ABSTRACT

Understanding the impacts of pandemics on public health and related societal issues at granular levels is of great interest. COVID-19 is affecting everyone in the globe and mask wearing is one of the few precautions against it. To quantify people's perception of mask effectiveness and to prevent the spread of COVID-19 for small areas, we use Understanding America Study's (UAS) survey data on COVID-19 as our primary data source. Our data analysis shows that direct survey-weighted estimates for small areas could be highly unreliable. In this paper, we develop a synthetic estimation method to estimate proportions of perceived mask effectiveness for small areas using a logistic model that combines information from multiple data sources. We select our working model using an extensive data analysis facilitated by a new variable selection criterion for survey data and benchmarking ratios. We suggest a jackknife method to estimate the variance of our estimator. From our data analysis, it is evident that our proposed synthetic method outperforms the direct survey-weighted estimator with respect to commonly used evaluation measures.

Key words: cross-validation, jackknife, survey data, synthetic estimation.

1. Introduction

Mask effectiveness perception is a topic of great relevance during the COVID-19 pandemic with emergence of new variants, multiple waves and fluctuating infection r ates. In the United States, national estimates of mask effectiveness perception can be derived by weighted means or proportions from respondent level data from a national survey like the Understanding America Study. However, to draw conclusions for small areas (e.g., states) for which sample sizes are small, direct estimates are inappropriate and misleading with very low or high estimates and highly variable standard errors.

In this paper, we explore a synthetic estimation of the perception on mask effectiveness, i.e., proportion of people considering mask to be highly effective at the state level. The is an indirect method of borrowing strength from similar areas. A synthetic estimator is not area specific in the study variable of interest and can be applied to any probability and non-probability sample design. Such methods are often employed in practice for their simplicity and ability to produce estimates for areas with no sample from the sample survey. Moreover, when the survey does not provide any sample for many areas, a synthetic method may be appealing to public policy makers as the same estimation method is applied to all areas, irrespective of whether an area has sample or not. There is a widespread use of synthetic

[1]PhD Student, Applied Mathematics & Statistics, and Scientific Computation. E-mail: asen123@umd.edu
[2]Director and Professor, The Joint Program in Survey Methodology & Professor, Department of Mathematics, University of Maryland, College Park, MD 20742, USA. E-mail: plahiri@umd.edu

estimation in different small area applications; e.g. Ghosh (2020), Marker (1995) and others. A synthetic method uses explicit or implicit models to link several disparate databases in producing efficient estimates for small areas. Hansen et al. (1953) presented an early example of a regression method to produce synthetic estimates of the median number of radio stations heard during the day for over 500 counties of the United States. Stasny et al. (1991) developed a regression-synthetic method for estimating county acreage of wheat using a non-probability sample of farms along with auxiliary data on planted acreage and district indicators. Marker (1999) and Rao and Molina (2015) presented more examples of synthetic small area estimators based on regression models. For our problem, we combine UAS data with the census data and Covid Tracking Report data to develop synthetic estimates of mask effectiveness perception for the states.

In Section 2, we describe primary and supplementary data used in this paper. In Section 3, we evaluate performances of the state level direct survey-weighted estimates. The performance of the direct method is poor, which motivates synthetic estimation, described in Section 4. In this Section, we introduce a JACKKNIFE method to estimate variance of the synthetic estimator. We report main results from our data analysis in Section 5. In this Section, we introduce a new model selection criterion for complex survey data. Finally, we evaluate synthetic estimates by comparative analogy of plotting with direct estimates for a handful of states, some small like District of Columbia, Rhode Island, North Dakota and large states like New York, California, Florida. We conclude the paper by summarizing the utility of the methods described in the paper and discussing how they can be extended to any other binary, categorical or continuous variable from this survey or any other survey with little adjustments or modifications.

2. Data used

For this study, we will use the UAS as the primary data containing study variable on the perception of mask effectiveness and supplementary data containing information for building small domain modelling and estimation procedures.

2.1. The Primary Data: Understanding America Study (UAS)

The Understanding America Study (UAS), conducted by the University of Southern California (USC), is an internet panel of households representing the entire United States. A household is broadly defined as anyone living together with the person who signed up for participating in the UAS. Using members of the population-representative UAS panel, USC's Center for Economic and Social Research (CESR) launched the Understanding Coronavirus in America tracking survey on March 10, 2020. The survey provides useful information on attitudes, behaviours, including health care avoidance behaviour, mental health, personal finances around the novel coronavirus pandemic in the United States.

Initial requests were sent out to the UAS panel members in order to determine their willingness to participate in an ongoing Coronavirus of UAS surveys. Among 9,063 UAS panel members who responded to the initial request, 8,547 were found eligible to participate in the survey. On average until November, 2020 (wave 16) about six thousand respondents

Table 1: Understanding of America Survey (UAS) wave details

Wave Number	Wave Name	Time period	Sample size
1	UAS 230	March 10, 2020 - March 31, 2020	6,932
2	UAS 235	April 1, 2020 - April 28, 2020	5,478
3	UAS 240	April 15, 2020 - May 12, 2020	6,287
4	UAS 242	April 29, 2020 - May 26, 2020	6,403
5	UAS 244	May 13, 2020 - June 9, 2020	6,407
6	UAS 246	May 27, 2020 - June 23, 2020	6,408
7	UAS 248	June 10, 2020 - July 8, 2020	6,346
8	UAS 250	June 24, 2020 - July 22, 2020	6,077
9	UAS 252	July 8, 2020 - Aug 5, 2020	6,289
10	UAS 254	July 22, 2020 - Aug 19, 2020	6,371
11	UAS 256	Aug 5, 2020 - Sep 2, 2020	6,238
12	UAS 258	Aug 19, 2020 - Sep 16, 2020	6,284
13	UAS 260	Sep 2, 2020 - Sep 30, 2020	6,284
14	UAS 262	Sep 16, 2020 - Oct 14, 2020	6,129
15	UAS 264	Sep 30, 2020 - Oct 27, 2020	6,181
16	UAS 266	Oct 14, 2020 - Nov 11, 2020	6,181

participate in the surveys, as seen from sample size in Table 1. Beginning in March 2020, the first round was UAS 230, which fielded from March 10 to March 31, 2020, with most responses happening during the period of March 10-14, 2020. UAS 230 is the first round of the survey that includes questions specifically tailored to COVID-19. These questions were repeated in subsequent longitudinal waves. The survey is being conducted in multiple waves. As of November 11, 2020,there are 16 waves, as described in Table 1 with their time periods.

For each wave, eligible panel members are randomly assigned to respond on a specific day so that a full sample is invited to participate over a 14-day period. Respondents have 14 days to complete the survey but receive an extra monetary incentive for completing the survey on the day they are invited to participate. Thus, except for the first wave, the data collection period for each wave is four weeks with a two-week overlap between any two consecutive waves. Each wave data consists of, on an average, six thousand observations.

The UAS is sampled in batches, through address-based sampling. The batches are allocated for national estimation and also for special population estimation (Native Americans, California, and Los Angeles county). Essentially UAS is a multiple-frame survey with four frames: Nationally Representative Sample, Native Americans, Los Angeles (LA) County, and California. Table 2 shows the relationship between the batches and frames, but each batch draws from only one frame.

As of November 2020, there are 21 batches, the latest being added in August, 2020. Most batches use a two-stage probability sample design in which zip codes are drawn first and then households are drawn at random from the sampled zip codes (except for two small sub-groups that are simple random samples from lists). The National batches draw zip

Table 2: Relationship between Batches and Frames in the Understanding America Survey

Batch	Frame
1	U.S.
2,3	Native American
4	Los Angeles County young mothers
5 to 12	U.S.
13,14,18,19	Los Angeles County
15,16	California
17,20,21	U.S.

codes without replacement, but the Los Angeles County batches draw with replacement and do sometimes contain the same zip code in different batches.

The base weights account for the differential probability of sampling a zip-code and an address within it. The base weights are then adjusted for nonresponse. Finally, at the national level, the distribution of nonresponse adjusted weights is calibrated to that of the 2018 Current Population Survey (CPS) weights with respect to selected demographic variables. Weights are provided for all batches, except for batch 4, which comprises Los Angeles County young mothers, and non-Native American households in batches 2 and 3. Angrisani et al. (2019) describe the sampling and weighting for UAS in great detail.

The survey includes a national bi-weekly long-form questionnaire and a weekly Los Angeles County short-form questionnaire administered in each bi-weekly wave. The survey data contains information on different demographic variables such as age, race, sex, and Hispanic origin, education, marital status, work status, identifiers for the states and zip-codes, and various outcome variables affecting human lives (e.g., mental stress, personal finances, COVID-19 like symptoms, testing results, etc.) The data also contains base and final weights so survey-weighted direct estimates for different outcome variables of interest can be produced.

2.2. Supplementary Data

The COVID Tracking Project: Both national and state level data can be downloaded from https://covidtracking.com. We use the data as a source of state specific auxiliary variables in our models. The COVID Tracking Project collects and publishes testing data daily for the United States as a whole and also for states and territories. From this data we understand that for 50 states and the District of Columbia (DC) combined the total test count has been increasing fast with more than 1 million in April 2020 to close to total 200 million by end of November 2020. The daily test count also increased from around 180K in April 2020 to 1.5 million in November 2020. There are various state specific auxiliary variables that could be potentially predictive of the perception on mask effectiveness. They include COVID-19 daily total testing, total test results (positive/negative), death, recovery count (as obtained from Johns Hopkins data on coronavirus), hospitalization, ventilation, etc. With

the increase in tests (due to better supply of testing kits, increase in awareness, etc.) or increase in positive cases (due to mask mandate being relaxed, advent of a new variant, etc.), one may argue that people's perceptions of mask effectiveness may change. Thus for this study, we use the following auxiliary variables that could be potentially useful in explaining our outcome variable on perception of mask effectiveness:

(i) totalTestResults: total number of tests with positive or negative results,

(ii) positive: total number of positive tests.

To make the above two auxiliary variables comparable across 50 states and the District of Columbia, we have used appropriate scaling factors to create the following two auxiliary variables, which we have used in our modelling:

(i) Testing rate: (Total tests with positive or negative results)/(Total population of state),

(ii) Positivity rate: (Total positive tests)/ (Total tests with positive or negative results).

Population density data: We use population density estimates in our modelling. Population density estimates for US states in 2010 are obtained from the U.S. Census Bureau (2020). For this study, we have created a categorical variable from it with three levels as follows:

- low - when population density of a state is less than or equal to 101 people per square mile (1st quartile from 2010 census data), e.g. North Dakota, Wyoming, Alaska etc.,
- medium - when population density is greater than 101 but less than or equal to 231 people per square mile (median), e.g. Georgia, Michigan, Virginia, etc.,
- high - when population density is greater than 231 people per square mile, e.g. New York, California, District of Columbia, etc.

Democratic party affiliation: For a given state, we have created a binary variable, which takes on the value 1 if the Governor of the state is a Democrat and 0 otherwise. The information is prior to the 2020 election and obtained from Wikipedia (2020).

Region membership of the states: Since 1950, the United States Census Bureau defines four statistical regions, with nine divisions. Using information obtained from the Census Bureau (2010) we have marked each state as one of the four regions - Northeast, Midwest, South and West.

Census Bureau's Population Estimates Program (PEP): For our synthetic estimation method, we need population counts for different demographic groups in the 50 states and the District of Columbia. The Census Bureau releases various tables of population estimates. On June 2020, the Population Division of the U.S. Census Bureau released annual state

resident population estimates by age, sex, race, and Hispanic origin for the period April 1, 2010 to July 1, 2019. The Census Bureau essentially obtains these estimates using the 2010 decennial census as the base and updates by births, deaths, migration etc. available from the administrative records and others obtained from the American Community Survey (ACS) survey. We have used two data sources as follows:

1. SCPRC-EST2019-18+POP-RES: Estimates of the Resident Population Age 18 Years and Older for the US states from July 1, 2019 (released on December 2019), which can be directly used.

2. SC-EST2019-ALLDATA5: Estimates of population by Age, Sex, Race, and Hispanic Origin – 5 race groups (5 race alone or in combination groups). This data needs to be adjusted by filtering out 18+ population (with "AGE") for the above-mentioned domains (using variables "RACE" for white and rest as other race and "ORIGIN" for Hispanic or Non-Hispanic). Sex is not used, although present in the data and hence set to value 0 for all. The domain wise populations are then adjusted with a factor (i.e., multiplying with domain wise population/total state population) so that the sum of all the domains is equal to the total state level estimate mentioned before.

3. Direct estimation

For the mask effectiveness perception problem, we focus on the following question from survey questionnaire: *How effective is wearing a face mask such as the one shown here for keeping you safe from coronavirus?* This is a categorical variable with five possible answers: *(i) Extremely Ineffective, (ii) Somewhat Ineffective, (iii) Somewhat Effective, (iv) Extremely Effective, and (v) Unsure.* The answer choices of respondents have been used to create a binary variable that takes on the value 1 is taken if mask is considered to be Extremely Effective by respondent and 0 otherwise. Using this binary variable the direct estimate works really well at overall national level with low standard error.

The survey data contains respondents residing in 50 states and DC, but naturally they are not evenly distributed. For larger states like California or Florida, there is a sizable volume in the sample of even as high as 2000 respondents and for smaller states like Delaware or Wyoming, there is very little representation of even 3 or 4 respondents. In such scenarios, direct survey-weighted estimates are highly misleading. For example, we see for the first three waves 0% of people in Wyoming think mask is extremely effective, which happens because all the respondents in the sample take the value 0 for binary response variable perceived mask effectiveness. Hence this is not a good method to draw a conclusion for the whole population of the states.

We observe extremely variable standard error (SE) or margin of error (ME). Estimated SE, or equivalently, estimated ME for a state depends on the sample size and the value of estimated proportion. For a state with small sample size, say less than 12, SE is either 0 or very high. From computations of direct estimates, from multiple waves we see that for Rhode Island, a state whose contribution in the wave is small with 2 or 3 respondents, estimated SE in the first few waves (1 and 2) is 0%. We obtain 0 SE when all the observations are the same. In the case of Rhode Island the cause is latter. But as soon as we have a mix

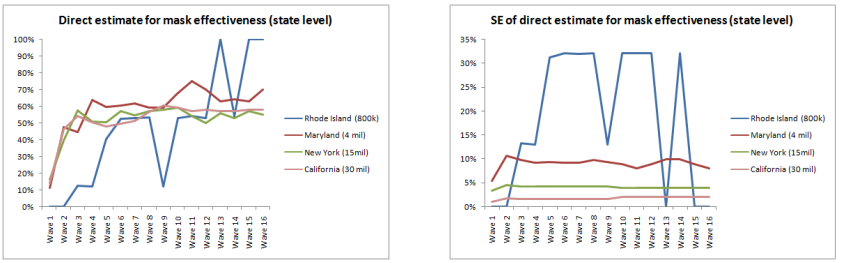

Figure 1: Direct estimates of perceived mask effectiveness and associated standard errors for four selected states.

of 0s and 1s, SE becomes very high, as high as even 30% from wave 5 onwards to wave 9 for Rhode Island.

Figure 1 displays erratic behaviour of direct estimates and standard errors for four states with varying population sizes (as estimated from the Census Bureau's PEP data)- one with high population (California - estimated adult population of 30 million from PEP), one with medium population (New York - estimated adult population of 15 million from PEP) – one with small population (Maryland - estimated adult population of 4 million from PEP) and one with very small population (Rhode Island - estimated adult population of 800k from PEP). The curves for Rhode Island are very unstable whereas, those for New York and California are quite stable. These SE estimates are thus surely very unstable or unreliable and typically, in public opinion polls margin of errors (2SE) is targeted at a low level such as 3% or 4%. Figure 1 for Rhode Island demonstrates high variability in state estimates for smaller states.

Along with high variability a demonstration of high bias in the direct state estimates can also be observed. Since we do not know the truth for perceived mask effectiveness, we cannot demonstrate bias properties for perceived mask effectiveness. But we can say if we consider another outcome variable for which "truth" is known from the PEP data, we can at least partially justify our claim. Using Figure 2 we show that UAS estimates of proportions of people falling in the four demographic groups or domains we considered do not match up with PEP data for states, but they more or less match at the national level. For large states like California, the difference between PEP estimate of the percentage of adult population and UAS direct wave estimate is negligible. This is similar for medium sized states like Maryland and New York, but for small states like Rhode Island and North Dakota, referring to Figure 2, we see the percentages vary significantly with even 0% or no contribution in some domains.

4. Synthetic method

For developing the synthetic estimation of perceived mask effectiveness for small areas, i.e., at state level, we first define the following notations and then derive a formula for the estimator from a logistic regression model. Let y_k denote the value of outcome (or

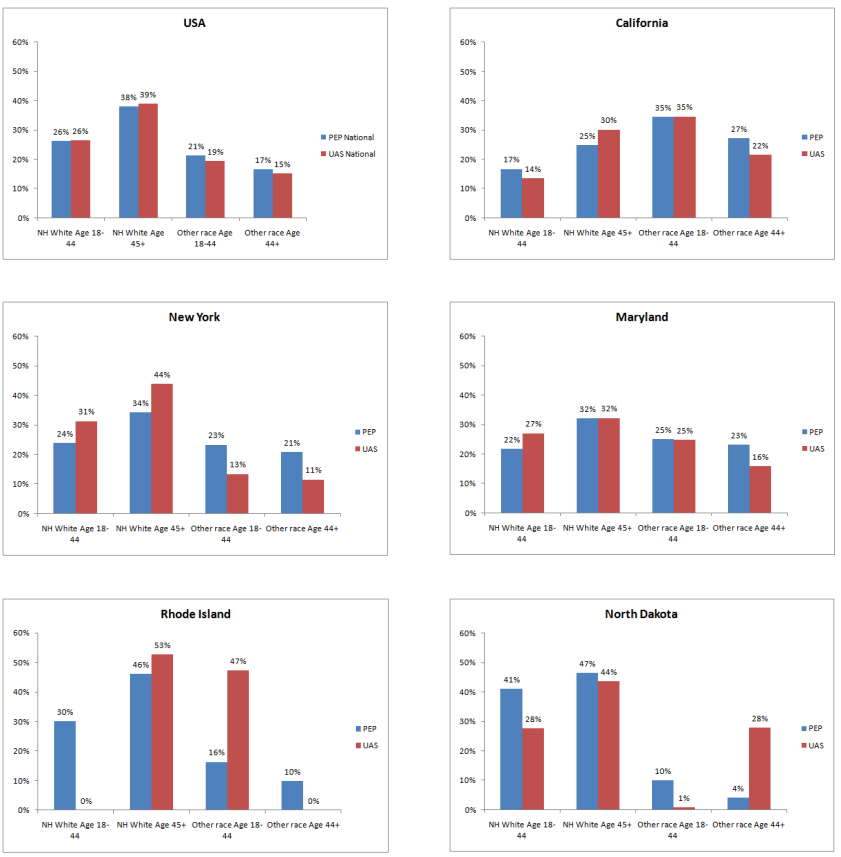

Figure 2: PEP vs. UAS estimates of 4 domains

dependent) variable for the kth respondent ($k = 1, \cdots, n$), where n denotes the number of respondents in a given wave (say, wave 2 covering April 1-April 28, 2020) of the UAS survey. The outcome variable is binary as defined by $y_k = 1$ if respondent k considers mask wearing to be extremely effective. Let $x_k = (x_{k1}, \cdots, x_{kp})'$ denote the value of a vector of auxiliary variables (same as independent variables or predictor variables or covariates) for respondent k. We have focused on the following two criteria for selecting the auxiliary variables for the unit level logistic regression model: (i) auxiliary variables should have good explanatory power in explaining the outcome variable of interest; (ii) total or mean of these auxiliary variables for the population should be available from a big data such as a large survey, administrative records or decennial census. Let N_i and N_{gi} be the population size of the adult (18+) and the gth group in state i, respectively. As discussed previously in the data section N_{gi} and N_i values are obtained from the US Census Bureau. Let y_{gik} be the value of the outcome variable for kth respondent in state i for the gth group ($g = 1, \cdots, G$; $i = 1, \cdots, m$; $k = 1, \cdots, N_{ig}$). Here we have $m = 51$ (50 states and DC) small areas. Let z_i be a vector of state specific auxiliary variables. For the estimation of mask-effectiveness variable for the 50 states and the District of Columbia, we write the population model as:

$$\text{Level 1: } y_{gik}|\theta_{gi} \overset{ind}{\sim} f(.;\theta_{gi}), \quad \text{Level 2: } h(\theta_{gi}) = x_g'\beta + z_i'\gamma, \tag{1}$$

for $k = 1, \cdots, N_{gi}$, $g = 1, \cdots, G; i = 1, \cdots, m$, where $f(.;\theta_{gi})$ is a suitable distribution with parameter θ_{gi} (here for binary variable this is a Bernoulli distribution with success probability θ_{gi}); $h(\theta_{gi})$ is a suitable known link function (for this application, we take logit link); β and γ are unknown parameters to be estimated using UAS micro data, i.e., at the respondent or unit level using survey weights.

We estimate population mean for state i by: $\hat{\bar{Y}}_i^{syn} = \sum_{g=1}^{G}(N_{gi}/N_i)\hat{\theta}_{gi} = \sum_{g=1}^{G}(N_{gi}/N_i)$ $h^{-1}(x_g'\hat{\beta} + z_i'\hat{\gamma})$, where h^{-1} is the inverse function of h; $\hat{\beta}$ and $\hat{\gamma}$ are survey-weighted estimators of β and γ, respectively. If $h(\cdot)$ is a logit function, we have $\hat{\bar{Y}}_i^{syn} = \sum_{g=1}^{G}(N_{gi}/N_i)\hat{\theta}_{gi} = \sum_{g=1}^{G}(N_{gi}/N_i)\exp(x_g'\hat{\beta} + z_i'\hat{\gamma})\left[1 + \exp(x_g'\hat{\beta} + z_i'\hat{\gamma})\right]^{-1}$.

We propose a jackknife method to estimate the variance of the proposed synthetic estimator. We obtain jth jackknife resample by deleting all survey observations in batch j. Thus we have $m = 20$ jackknife resamples from wave 14 onwards because there are 20 batches in total, whereas earlier for waves 1 to 13 there were in total 19 batches in each wave data, the latest addition being "21 MSG 2020/08 Nat. Rep. Batch 11" in August 2020 and LA County Young mothers is not present in any of the waves. For each jackknife resample, we recompute replicate synthetic estimate using (1). We will get m such replicate estimates, say, $\hat{\bar{Y}}_{i(-j)}^{syn}$ ($j = 1, \cdots, m$). We can then estimate the variance of $\hat{\bar{Y}}_i^{syn}$ by

$$v(\hat{\bar{Y}}_i^{syn}) = \frac{m-1}{m}\sum_{j=1}^{m}\left(\hat{\bar{Y}}_{i(-j)}^{syn} - \frac{1}{m}\sum_{j=1}^{m}\hat{\bar{Y}}_{i(-j)}^{syn}\right)^2. \tag{2}$$

Table 3: Direct national estimates of perceived mask effectiveness (associated standard errors) for selected demographic groups and first five waves.

Direct Estimate	Wave 1	Wave 2	Wave 3	Wave 4	Wave 5
Overall National	14% (0.6%)	41% (0.9%)	47% (0.9%)	46% (0.9%)	44% (0.9%)
NH White Age(18-44)	12% (1.1%)	33% (1.7%)	39% (1.6%)	37% (1.6%)	33% (1.6%)
NH White Age(45+)	11% (0.7%)	40% (1.2%)	46% (1.1%)	46% (1.1%)	44% (1.1%)
Other race Age(18-44)	20% (1.7%)	48% (2.6%)	54% (2.3%)	50% (2.3%)	49% (2.3%)
Other race Age(45+)	17% (1.7%)	47% (2.7%)	58% (2.5%)	58% (2.5%)	58% (2.4%)

5. Data analysis

At national level in order to understand the broader question on the identification of demographic factors influencing effectiveness perceptions certain domains or groups are created based on race-ethnicity x age. These four groups are Non Hispanic White Age 18-44, Non Hispanic White Age 45+, Other race Age 18-44 and Other race Age 45+. Considerable variation among these groups is observed across multiple waves with all standard errors (SE) from direct estimates around 2%, after which it is chosen for further estimation study. The direct survey-weighted estimates at the national level as well as domain level from waves 1 to 5 are provided in Table 3 along with the standard errors in parenthesis; see also Figure 3. We observe that the overall national estimate and the domain NH White Age(45+) behave similarly (e.g, 46% and 44% for waves 4 and 5, respectively). The Other Race Age (45+) domain has the highest perception of mask effectiveness (e.g., 58% for waves 4 and 5), whereas the domain NH White Age (18-44) has the least value of such estimate (e.g., 37% and 33% at wave 4 and 5, respectively). Thus this breakdown of the population into domains can be used further for modelling. We have used R **survey** package to compute such estimates with the weights of respondents as provided in the wave data. We refer to the papers by Lumley (2004, 2010, 2020) for understanding the R **survey** package.

From the aforementioned observations, it is clear that direct estimates are not stable even at the state level. The synthetic estimators essentially would borrow strength from other states through implicit or explicit models and combine information from the sample survey, various administrative/census records, or previous surveys. Synthetic estimators are highly effective and appealing in small area estimation. Referring to synthetic estimation methods explained in Lahiri and Pramanik (2019), we employ a unit level logistic model with respondent level characteristics like the age x race/ethnicity along with state level auxiliary variables such as regional identifier (e.g., Northeast, Midwest, South or West), party affiliation of state governor or DC mayor (Democratic or Republic) and even the state level COVID-19 testing or positivity rate. Thus we have combined the data in UAS survey with

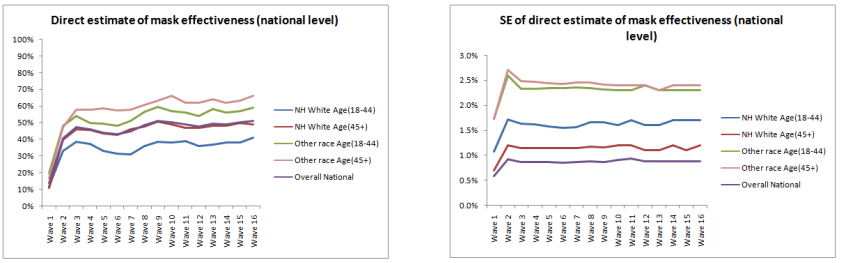

Figure 3: National direct wave estimates of perceived mask effectiveness and associated standard error direct estimates; overall estimates as well as estimates for four groups are provided.

the US Census Bureau data and Covid Tracking Project data to derive state level synthetic estimates of population means and totals for the variable of interest.

5.1. Variable Selection

For all the 16 waves, we first fit the full model, i.e., the model with all auxiliary variables listed earlier. Table 4 displays significant auxiliary variables in all the waves. We then concentrate our focus on the models given in Table 5. These are logistic regression models for the indicator response variable perceived mask effectiveness with different combinations of auxiliary variables. In every case, we use R survey package to run weighted logistic regression with quasi-Bernoulli family, where weights are the final post-stratification weights as provided by UAS and design is defined with such weights and no strata or cluster.

The full model, i.e., M1, is our starting point. True values of some of the coefficients of M1 may be zero; if the sample size is large, those coefficients will be estimated at near zero. But, if we keep too many covariates in a model, the estimates may be subject to high variability (and thereby we may lose some predictive power if we select a model with a lot of covariates.)

We now explore the possibility of reducing the number of auxiliary variables from M1. There are a large number of possible models so we proceed systematically. To this end, we fit M1 for all the 16 waves. Table 4 reports significant auxiliary variables for each of the 16 waves. In all the models, we include intercept (whether or not it is significant). Using information in Table 4, we create Table 5, which lists a number of competing models with less number of model parameters. We now explain why we want to consider models M1-M7 for further comparison.

All the auxiliary variables except for the democratic party affiliation appear in at least one wave. Thus, a natural question is what happens if we drop the democratic party affiliation from M1, which motivates keeping M2 for further investigation. Positivity rate is significant only in wave 5. This suggests inclusion of model M3 for further investigation. The factors NH Whites (18-44), NH Whites (45+), Other Race (18-44), population density are all significant for 5 waves: 6,7, 12, 14, and 16. So the model M7 seems to be a natural choice. We then consider models M3-6. Note that, in addition to NH Whites (18-44), NH

Table 4: Significant covariates in Model 1 for different waves; from R package output of significance code and p-value pairs to be interpreted as '***' for [0, 0.001], '**' for (0.001, 0.01], '*' for (0.01, 0.05], '•' for (0.05, 0.1], ' ' for (0.1, 1]

Wave	intercept	NH White Age(18-44) (indicator)	NH White Age(44+) (indicator)	Other race Age(18-44) (indicator)	testing rate	positivity rate	population density (categorical)	region Northeast (indicator)	region Midwest (indicator)	region South (indicator)	Democratic party (indicator)
1	***	*	***						•	**	
2		***	*		•						
3		***	***				***				
4		***	***	*			***			**	
5	*	***	***	*		**	**	•			
6		***	***	*			***				
7		***	***	*			**				
8		***	***				***		*		
9		***	***		•		***				
10		***	***	*	*		**				
11		***	***	•			***		•	•	
12		***	***	*			***				
13		***	***	•	*		**			**	
14		***	***	•			***				
15		***	***	•					•		
16		***	***	*			***				

Table 5: A list of competing models

Model	intercept	NH White Age(18-44) (indicator)	NH White Age(44+) (indicator)	Other race Age(18-44) (indicator)	testing rate	positivity rate	population density (categorical)	region Northeast (indicator)	region Midwest (indicator)	region South (indicator)	Democratic party (indicator)
M1	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
M2	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
M3	✓	✓	✓	✓	✓		✓	✓	✓	✓	
M4	✓	✓	✓	✓	✓		✓			✓	
M5	✓	✓	✓	✓			✓			✓	
M6	✓	✓	✓	✓	✓		✓				
M7	✓	✓	✓	✓			✓				

Whites (45+), Other Race (18-44), population density, each of these three models includes an additional auxiliary variable significant in at least one wave. For example, M4 includes an additional auxiliary variable testing rate because all M4 coefficients are significant in wave 10.

To select one out of the seven models listed in Table 5, we apply a cross-validation leave-one-state-out method. We now describe the method. We leave out the entire UAS survey data on the outcome variable y_i (e.g., perceived mask effectiveness) for state i and predict the vector of outcome variables for all sampled units of the leave out state using x_g for the sampled unit and z_{-i} for the leave out state. Let $f(y_i|y_{-i})$ denote the joint density of y_i, all the observations in state i, conditional on the data from the rest of the states, say y_{-i}. For the Bernoulli distribution of y_i for state i, using independence, we have for known model parameters β and γ:

$$\begin{aligned}
\log f(y_i|y_{-i};\beta,\gamma) &= \sum_{g=1}^{G}\sum_{k=1}^{n_{gi}}\left[y_{gik}\log\theta_{gi}+(n_{gi}-y_{gik})\log(1-\theta_{gi})\right] \\
&= \sum_{g=1}^{G}\sum_{k=1}^{n_{gi}}\left[y_{gik}\log\left(\frac{\theta_{gi}}{1-\theta_{gi}}\right)+n_{gi}\log(1-\theta_{gi})\right] \\
&= \sum_{g=1}^{G}\sum_{k=1}^{n_{gi}}\left[y_{gik}(x_g'\beta+z_i'\gamma)-n_{gi}\log\left(1+\exp(x_g'\beta+z_i'\gamma)\right)\right].
\end{aligned}$$

Using data from the rest of states, i.e., $y_{(-i)}$ we get survey-weighted estimates β and γ and plug in the above expression. Let these estimates be $\hat{\beta}_{w,(-i)}$ and $\hat{\gamma}_{w;(-i)}$. We then define our model selection criterion as:

$$C=\sum_{i=1}^{m}\sum_{g=1}^{G}\sum_{k=1}^{n_{gi}}w_{gik}\left[y_{gik}(x_g'\hat{\beta}_{w,(-i)}+z_i'\hat{\gamma}_{w;(-i)})-n_{gi}\log\left(1+\exp(x_g'\hat{\beta}_{w,(-i)}+z_i'\hat{\gamma}_{w;(-i)})\right)\right].$$

For each of the models in Table 5, we compute C model selection measures for all the waves (wave 1-16). In Table 6, we report the quantiles (minimum, first quartile, median, third quartile, maximum) and mean of C values (over the 16 waves) for each model in Table 5. We divide C value from each model by the sample size of the wave to scale down the numbers for ease of comparison. The C values are all negative, as these are logarithm of fractions. For every state, iteratively regressions are run and regression estimates are obtained, which are used in the formula. Using Table 6, we select M2 as the best performing model because this model produces maximum value of all descriptive statistics reported in Table 6.

5.2. Synthetic estimation of the perception of mask effectiveness for the states

In this section, we consider benchmarked synthetic estimates, which are obtained from the synthetic estimates after a ratio adjustment. These benchmarked synthetic estimates, when appropriately aggregated over the 50 states and the District of Columbia, yield the national direct estimate. We compare both synthetic and benchmarked synthetic estimates

Table 6: Cross validation leave one state out statistic for all models

Model	0%	25%	50%	75%	100%	Mean
M1	-89	-83	-75	-68	-16	-71
M2	-79	-55	-10	-8	-7	-28
M3	-92	-86	-83	-78	-20	-78
M4	-92	-86	-82	-77	-21	-78
M5	-91	-84	-80	-76	-19	-76
M6	-68	-59	-55	-51	-17	-54
M7	-90	-82	-76	-69	-15	-72

with the corresponding direct sample survey estimates (i.e., weighted proportion of people from UAS who believe mask is extremely effective) for the 50 states and the District of Columbia. This gives us an idea about the magnitude of biases in the synthetic and benchmarked synthetic estimates because direct estimates, though unreliable, are unbiased or approximately so. In Figure 4, we have 6 plots corresponding to 6 states (3 with small population - District of Columbia, North Dakota, Rhode Island, and 3 with large population - California, New York, Florida) of point estimates (direct and benchmarked synthetic) vs waves, which display time series trends from wave 1 to wave 16. The direct estimates of perceived mask effectiveness for small states could be unreasonable. For example, for the District of Columbia, direct estimates are 0% for both waves 1 and 2. On the other hand, benchmarked synthetic estimates 18% and 39% are more reasonable for these two waves – they are more in line with the national estimates for such waves. Similarly, for Rhode Island unreasonable 100% perceived mask effectiveness direct estimates for waves 13 and 15 have also been modified to more reasonable benchmarked synthetic estimates.

Figure 5 displays standard errors of direct and benchmarked synthetic estimates. The proposed jackknife method is used to compute standard errors for benchmarked synthetic estimates. We denote standard errors of direct and benchmarked synthetic estimates by SE and STD, respectively. If we focus on the error graphs, the values from direct estimates get as high as 32% for small states (i.e. one with low contribution to overall sample size). Using benchmarked synthetic estimates at the state level, the error has reduced to almost 6 times with as low as 6% standard error from the jackknife method. For larger states like New York and Florida, the errors reduce using benchmarked synthetic estimate, although not to a great extent. For the state contributing most to the sample size, California, the standard errors are more or less similar.

For the chosen model M2, we create a state level comparative diagram of benchmarked synthetic estimates with direct estimates in Figure 6 using data from wave 16. As at the state level, the synthetic estimates and the corresponding benchmarked synthetic estimates are really close, we have not plotted synthetic estimates for ease of viewing. We observe that our synthetic estimates are much more stable than the corresponding direct estimates. The states arranged in increasing order of population sizes show that the issue of highly variable state level direct estimates for the smaller states has been mitigated by the synthetic method.

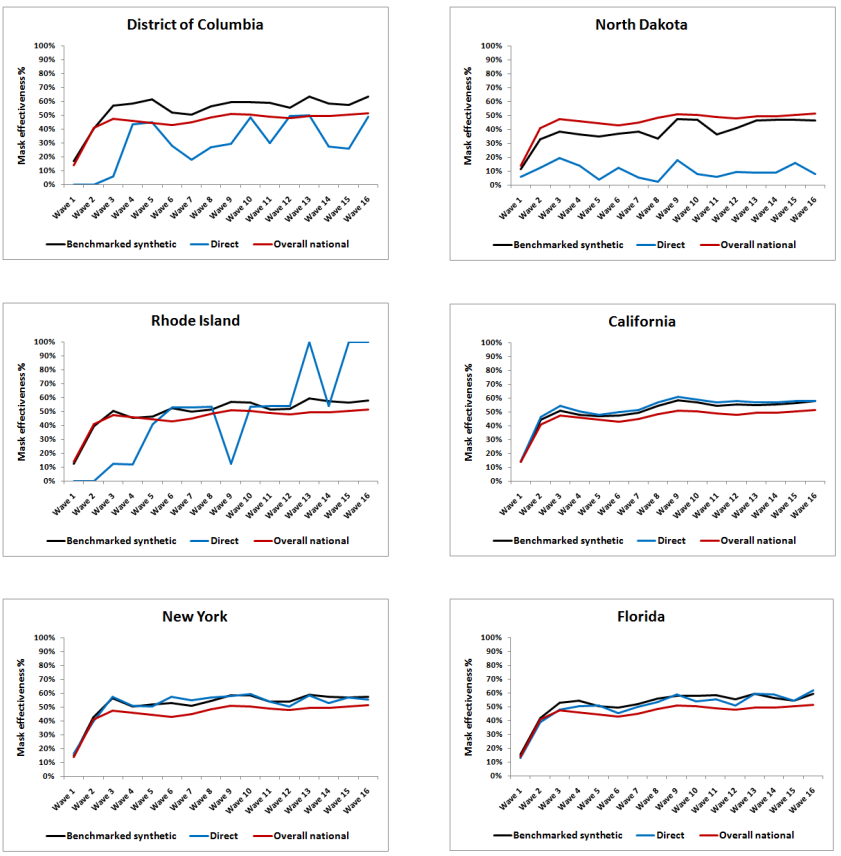

Figure 4: Time series trend of direct and benchmarked synthetic estimate for 6 sample states (3 small, 3 large)

Figure 5: Time series trend of SE of direct and benchmarked synthetic estimate for 6 sample states (3 small, 3 large)

For largely populated states as well as for small ones, the benchmarked synthetic estimates are doing a good job of estimating the proportion of the response variable. We next check the robustness of the synthetic estimator in terms of variance through the jackknife method.

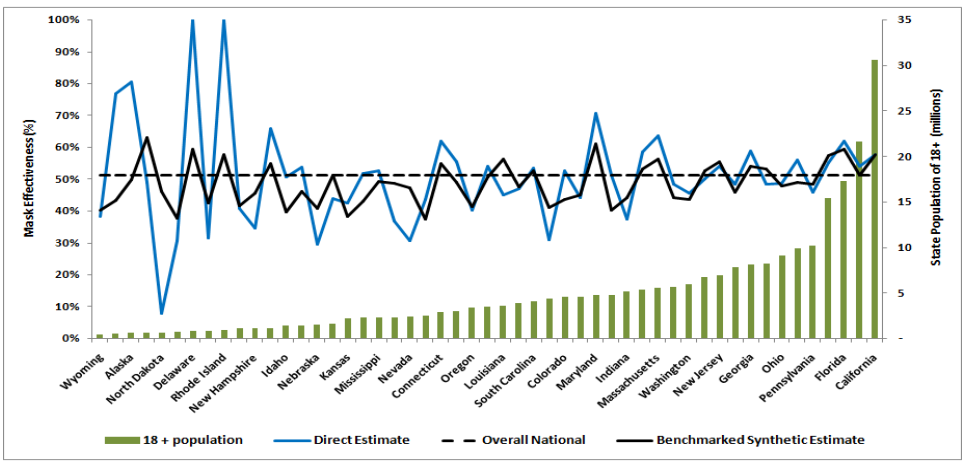

Figure 6: State level comparison of direct and synthetic estimates from M1 for wave 16

We fitted M2 for wave 16 data and obtained jackknife estimates of variances and hence standard errors at the state level. We provide a comparative view with the SE from direct estimates at the state level. The two graphs in Figure 7 are based on the wave 16 data. In the x-axis, states are arranged in increasing order of sample sizes. In the first graph, the y-axis is the ratio of direct estimate (survey-weighted) and synthetic estimate. In the second graph, the y-axis is the ratio of STD and SE, where SE is the standard error of direct estimate coming right from UAS (treating states as domains) and STD is the jackknife standard error of benchmarked synthetic estimate. For states with small sample sizes (e.g. Rhode Island, Wyoming), we see a lot of differences between the survey-weighted direct estimates and the synthetic estimates. For states with large sample sizes (e.g., California), the ratio is approaching to 1 (as plotted by the straight line) as the auxiliary variables used to construct the synthetic estimator are reasonable. We observe that all the jackknife estimates are much smaller than direct estimates and we conclude that the model is a fair one at estimating the perceived mask effectiveness at state level.

We define Benchmark Ratio (BR) as the ratio of the overall direct national estimate to the synthetic estimate (aggregated at the national level). The synthetic estimates, which are obtained at the state level, are aggregated by multiplying by the ratio of the adult state population to the overall US adult population estimate and then adding up. The closer the value of BR is to 1 the better is the model. We see from Table 7 that BR is close to 1 for all waves, using which we compute the Benchmarked or BR synthetic estimate.

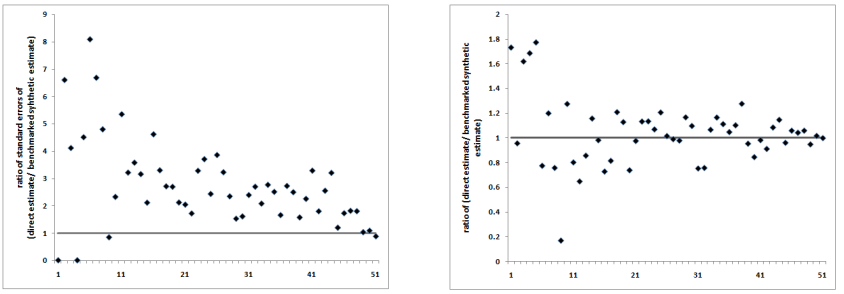

Figure 7: Comparison of direct with benchmarked synthetic estimates through the ratio SE/STD and ratio of estimates from M1 for wave 16; domains arranged in increasing order of sample size.

Table 7: Benchmarking ratios and national synthetic and benchmarked synthetic estimates for last five waves; synthetic estimates are based on Model 1.

Model	0%	25%	50%	75%	100%	Mean
Benchmarking Ratio	0.98	0.98	0.99	0.99	1.00	0.99
Synthetic	14.08%	45.22%	48.63%	50.43%	51.94%	46.00%
Benchmarked Synthetic	13.86%	44.62%	48.02%	49.66%	51.22%	45.38%

6. Conclusion

The method of estimating population means or totals for the states of USA explained in the paper provides sensible and numerically sound estimates and the model selection with all standard error of estimates within 2%. We noticed high variability of synthetic estimates at the state level estimation. We further note that while direct UAS estimates are designed to produce approximately unbiased estimates at the national level, they are subject to biases for the state level estimation. Biases in the direct proportion estimates at the state level may arise from the fact that they are essentially ratio estimates since the state sample sizes are random and expected sample sizes are small for most states. Moreover, the UAS weights are not calibrated at the state level.

From our investigation, we found that synthetic estimates improve on UAS direct estimates in terms of variance reduction, especially for the small states. But since synthetic estimates are derived using a working model, they are subject to biases when working model is not reasonable. However, we observe that the benchmarking ratios for all waves are consistently around 1 showing lack of evidence for bias. Our benchmarked synthetic estimates are close to the synthetic estimates because the benchmarking ratios are close to 1. None-theless by benchmarking synthetic estimates we achieve data consistency and it is reasonable to expect to reduce biases as well. We add that it is possible to reduce biases at the state level by benchmarking the synthetic estimates to the UAS direct estimates for a group of states

(e.g., benchmarking with a division). This may be needed for other synthetic estimation problems.

Infection rates has declined declining in most parts of the USA. However, with differential vaccination hesitancy rates across the US states and emergence of new COVID-19 variants, identification of granular level mask effectiveness perception rates may remain an important problem in the US. While we wait to reach herd immunity through aggressive vaccination program, good control of the spread of COVID-19 and its different variants in different parts of the world is essential. Thus, it will be of interest to understand mask effectiveness perception rates in communities throughout the world, especially where infection rates are high. Not only COVID-19, but for other infectious diseases, mask effectiveness perception is likely to stay relevant. While we illustrate the proposed synthetic methodology for state level estimation of perceived mask effectiveness, the methodology is general and can be applied to granular sub-state levels with no sample from the primary survey data. Moreover, similar synthetic methodology can be developed in the future to estimate granular level proportions related to personal finance, mental health, etc.

7. Acknowledgements

The project described in this paper relies on data from survey(s) administered by the Understanding America Study, which is maintained by the Center for Economic and Social Research (CESR) at the University of Southern California. The content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of USC or UAS. The collection of the UAS COVID-19 tracking data is supported in part by the Bill & Melinda Gates Foundation and by grant U01AG054580 from the National Institute on Aging. The second author's research was partially supported by the U.S. National Science Foundation Grant SES-1758808.

References

Angrisani, M., A. Kapteyn, E. Meijer, and H.-W. Saw, (2019). Sampling and weighting the Understanding America Study, Working Paper, No. 2019-004, Los Angeles, CA: University of Southern California, Center for Economic and Social Research.

Census Bureau, (2010). Census Regions and Divisions of the United States.

Ghosh, M., (2020). Small area estimation: Its evolution in five decades. Statistics in Transition New Series, Special Issue on Statistical Data Integration, pp. 1–22.

Hansen, M., W. Hurwitz, and W. Madow, (1953). Sample Survey Methods and Theory, Vol. 1. Wiley-Interscience.

Lahiri, P. and S. Pramanik, (2019). Estimation of average design-based mean squared error of synthetic small area estimators, Austrian Journal of Statistics, 48, pp. 43–57.

Lumley, T., (2004). Analysis of complex survey samples, Journal of Statistical Software, 9(1), pp. 1–19, R package verson 2.2.

Lumley, T., (2010). Complex Surveys: A Guide to Analysis Using R: A Guide to Analysis Using R. John Wiley and Sons.

Lumley, T., (2020). Survey: analysis of complex survey samples, R package version 4.0.

Marker, D., (1995). Small Area Estimation: A Bayesian Perspective. Phd thesis, University of Michigan, Ann Arbor, MI.

Marker, D., (1999). Organization of small area estimators using a generalized linear regression framework, Journal of Offcial Statistics, 15, pp. 1–24.

Rao, J. N. K., I. Molina, (2015). Small Area Estimation, 2nd Edition, Wiley.

Stasny, E., P. Goel, and D. Rumsey, (1991). County estimates of wheat production, Survey Methodology, 17 (2), pp. 211–225.

U.S. Census Bureau, (2020). Census Bureau Population Estimates. Wikipedia, (2020). Political party strength in U.S. states.