Methods for Calculating State and National Prevalence Estimates: An Application of Estimates of Sexual Orientation and Gender Identity

Yangyang Deng[1], Ronaldo Iachan[1], Adam Lee[1], Marjorie Biel[1]
[1]ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850

Abstract
This research focuses on the methods for modeling estimates at the state level when data are available from a subset of states. We used the Sexual Orientation and Gender Identity (SOGI) optional module questions from the Behavioral Risk Factor Surveillance System (BRFSS) for 2016 to 2018 to develop models and provide estimates for all states. Models are validated against direct estimates where available. SOGI questions represent the most vigorous test of such a model in that a limited proportion of the sample who identify as transgender, bisexual and/or gay/lesbian. The process presented also provides a mechanism for imputation of responses where non-substantive answers are given (i.e. "do not know" or refusal to answer). The methodology is adaptable to other BRFSS optional models used by subsets of the states annually.

Key Words: BRFSS, Sexual Orientation, Gender Identity, Statistical Modeling, Imputation

1. Introduction

While there are several sources to obtain adult demographic data on race, ethnicity, sex, and income levels by state in the United States there is a relative paucity of sources for sexual orientation and gender identity (SOGI). One important source for race/ethnicity and sex estimates by state is the United States Census, a survey of the US population and 5 US territories conducted every 10 years (United States Census 2020). Lesbian, Gay, Bisexual, Trans (LGBT) advocacy groups have campaigned for the inclusion of questions asking about SOGI on the US Census but so far, their efforts have been unsuccessful, and the LGBT demographic has gone unmeasured (Services and Advocacy for GLBT Elders 2020). While the national Census does not ask about SOGI, there are other possible sources and methods to get estimates of the LGBT population in the United States. Other organizations have done surveys to estimate the LGBT population in the US. A 2017 Gallup poll and 2020 Williams Institute survey (LGBT Demographic Data Interactive 2019) found that 4.5% of adult Americans identified as LGBT and a 2016 Williams Institute survey found that 0.5% of the adult U.S. population identified as transgender. Our paper seeks to use data obtained from several recent cycles of the Behavioral Risk Factor Surveillance System (BRFSS) which included a SOGI survey module.

The BRFSS is the nation's premier system of health-related telephone surveys that collect state data about US residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. Established in 1984 with 15 states, BRFSS now collects data in all 50 states as well as the District of Columbia and participating territories. BRFSS completes more than 400,000 adult interviews each year nationally. The BRFSS sample is drawn by individual state health department, rather than being drawn as a single, national sample. The BRFSS is comprised of a core set of questions, which are adopted in standard form for all states. States may also select from standardized optional module on a number of health topics (Centers for Disease Control and Prevention 2017).

Data from the BRFSS have been used to model for Small Area Estimates (SAEs) in many studies (Guo et al. 2013, X. Zhang et al. 2014, Zhang et al. 2011). A recent publication reported a method of deriving county-level estimates from BRFSS state-level data (Pierannunzi et al. 2016). In most instances the large sample included in the BRFSS supports methods for creating sub-state prevalence estimates using state data or aggregating the BRFSS to a nationwide sample (Khalil and Crawford 2015) and then modeling sub-state areas (Song 2016, Li W 2009).

The literature provides instances where researchers have modeled from direct data from one geographic location that has sufficient direct observations to geographic areas where data are missing (National Cancer Institute 2017). Similar methods may be used to calculate estimates at the state level for questions within optional modules that have been asked only in a few states.

Since 2014, the BRFSS has used an optional model on sexual orientation and gender identity that states may choose to append to the core portion of the survey. Questions on sexual orientation and gender identity (SOGI) were included so researchers could use the data to compare responses from persons who identify as gay, lesbian, bisexual, and/or transgender with those of persons who do not identify themselves in these categories (Pierannunzi et al. 2017). The questions themselves are administered in two parts (Centers for Disease Control and Prevention 2017) as follows:

1. Do you consider yourself to be:
    1 Straight
    2 Lesbian or gay
    3 Bisexual
    4 Other
    7 Don't know/Not sure
    9 Refused

2. Do you consider yourself to be transgender?
    1 Yes, Transgender, male-to-female
    2 Yes, Transgender, female to male
    3 Yes, Transgender, gender nonconforming
    4 No
    7 Don't know/not sure
    9 Refused

SOGI questions used in the BRFSS optional module were developed by a group of survey professionals within the US Department of Health and Human Services (Institute of Medicine 2013). The questions are similar to those proposed by the Williams Group (Herman 2014). A number of other SOGI question formats have been proposed and are used on other surveys (Federal Interagency Working Group 2016). A total of 19 states participated in the optional module in 2014; 22 states used the module in 2015; 25 states participated in 2016, 28 states participated in 2017 and 2018. The list of states participating in the module by year is provided in Table 1.

Table 1: States participating in the SOGI optional module by year

Year	Participated States
2014	Delaware, Hawaii, Idaho, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maryland, Minnesota, Montana, Nevada, New York, Ohio, Pennsylvania, Vermont, Virginia, Wisconsin, Wyoming
2015	**Colorado**, **Connecticut**, Delaware, **Georgia**, Hawaii, Idaho, **Illinois**, Indiana, Iowa, Kansas, Maryland, **Massachusetts** , Minnesota, **Missouri** , Nevada, New York , Ohio, Pennsylvania, **Texas**, Virginia, **West Virginia**, Wisconsin
2016	**California**, Connecticut, Delaware, Georgia, Hawaii, Idaho, Illinois, Indiana, Iowa, **Kentucky**, **Louisiana**, Massachusetts, Minnesota, **Mississippi**, Missouri, Nevada, New York, Ohio, Pennsylvania, **Rhode Island**, Texas, **Vermont**, Virginia, **Washington**, Wisconsin
2017	California, Connecticut, Delaware, **Florida**, Georgia, **Guam**, Hawaii, Illinois, Indiana, Iowa, Louisiana, Massachusetts, Minnesota, Mississippi, **Montana**, Nevada, New York, **North Carolina**, Ohio, **Oklahoma**, Pennsylvania, Rhode Island, **South Carolina**, Texas, Vermont, Virginia, Washington, Wisconsin

2018	**Arizona**, Connecticut, Florida, Guam, Hawaii, **Idaho**, Illinois, **Kansas**, Louisiana, **Maine**, Maryland, Minnesota, Mississippi, **Missouri**, Montana, Nevada, New York, North Carolina, Ohio, Oklahoma, Pennsylvania, Rhode Island, South Carolina, **Tennessee**, Texas, Vermont, Washington, **West Virginia**, Wisconsin

***States in bold are those first participated SOGI modules.**

The state-level sample allows for direct estimates for each state participating in the model. Direct state prevalence estimates, however, cannot be calculated for the states that did not participate in any given year. A method is needed to model prevalence estimates where no data were collected. The model-based methods described in this article build on published methods to model estimates from one geographic area with sufficient direct observations to other areas. Such models are usually applied to achieve small-area estimates. In this instance, however, state estimates from direct observations are used to model estimates in other states. In the process, we are also able to generate national estimates of the prevalence of SOGI.
Overall, here are three main research questions this paper aims to address:

1) Can we estimate prevalence of those identifying as LGBT in those states which have not opted for the SOGI module? Can we estimate the LGBT population at a national level?
2) What characteristics best predict LGBT estimates at the state level?
3) How do different models work for identifying those who are LGBT?

2. Methods

The estimates calculated herein are based on predictive models for each of the dichotomous outcomes produced by the SOGI module in the BRFSS. Model-based estimates for states not using the SOGI module are produced by predicting the outcome for each respondent in the state using a wide range of predictors. As a result, national estimates are also made possible. Data from the 2016-2018 BRFSS were combined resulting in a sample size of 1,352,637. Various modeling methods were experimented to identify characteristics associated with LGBT identification.

During the analysis, we had to contend with Don't Know (DK) responses and refusals to answer the questions for each of the outcomes. Our framework considered that the prevalence of each outcome in these categories (DK's or Refusals) was greater than for the population as a whole. We confirmed this premise by profiling the respondents in each of the DK/Refusal categories for the three outcomes along the dimensions defined by the predictors. A comparison with the profiles of respondents in each outcome group (e.g., the gay/lesbian group) confirmed that DK's and refusals were much more similar to these groups than to other respondents or to the population as a whole. As a result of this comparative analysis, we imputed responses in the DK-Refused at higher rates than random imputation would suggest. Specifically, we imputed 5% of the DK-Refusals to each of the gay/lesbian and bisexual categories. For the gender identity question, 2% of the DK-refusal responses were imputed as transgender.

Preliminary variable selection identified a set of individual level variables as potential predictors of LGBT identification, through excluding variables with high missingness and skip patterns. These predictors were tested along with state level indices summarizing the acceptance of LGBT according to state laws.

The LGBT Welcome Index was created using the Human Rights Campaign's (HRC) and Equality Federation Institute's (EFI) State Equality Indexes from 2016, 2017, and 2018. These annual reports provide a breakdown of the number of "good" and "bad" laws present, which are supportive or harmful of the LGBT population, introduced, and passed in each state related to the LGBT community. Laws within each of the good or bad categories are split up into broad categories describing the purpose of the law. When creating the LGBT Welcome Index one positive point was given to each state with the presence of a "good" law. Good laws are broken up into the following broad categories in the reports: those which prohibit discrimination in employment, housing, public accomodations and education, laws that specifically protect against bullying based on sexual orientation and gender identity, laws which allow for second-parent adoption for same-sex couples, laws which ban insurance exclusions for transgender-specific healthcare needs, laws that provide transgender-inclusive health benefits for state employees, laws and policies that allow gender marker changes on driver's licenese and/or birth certificates, laws which allow hate crimes based on sexual orientation or gender identity to be addressed and laws which protect youth from conversion therapy. One negative point is given to each state for the presence of "bad" laws which are broken up into broad categories as well: laws that prevent school districts from specifically protecting LGBT students against bullying, laws which prevent schools from including LGBT topics in the curriculum, laws which criminalize behaviors that have a low or negligible risk of HIV transmission, laws that allow for transgender exlusions in Medicaid, and laws that prevent gender marker changes on identifying documents. All points were added together within each of the years to assign an overall LGBT welcome index predictor to each state. States with a higher score are more welcoming to LGBT people and states with a lower index are less welcoming.

Regarding the modeling procedure, we split 60% of our data as a training set, 20% as a validation set and 20% as a test set. We used training and validation sets to build and tune the models, then used the test data to select the best performing model setting for the prediction. The modeling methods included multilevel logistic regression, random forest, and LASSO regression. Machine Learning algorithms generally are good at handling data that are multi-dimensional; The multilevel logistic regression allowed for generalization of linear models that vary at more than one level and the accurate variance calculation relative to the clustering of data by state; LASSO regression, as any regularization method, can avoid overfitting and can be applied for feature selection purpose (Heinze, Georg et al 2018). For multilevel logistic regression, we used two set of predictors which were selected using backward method and LASSO method, as another method experiment aspect. Specific to these models, variance could be observed at the individual or state level.

The four dependent variables used in the models were binary indicators of lesbian/gay identification, bisexual identification, transgender identification, and overall LGBT identification. The indices of state laws were included as random effects in multilevel logistic regression and as regular predictors in machine learning and LASSO. Our initial analyses of the BRFSS data led to 41 potential predictors ranging from demographics and health care access variables to chronic health condition and health behavior variables.

3. Results

This section presents the model predictions for selected states for the analysis.

We applied the three criteria of mean square area (MSE, area under the ROC curve (AUC), and how close predictions are to the true state-level prevalence, which we called mean state error. We wanted to choose the model with the largest AUC, the smallest MSE, and the smallest mean state error. Table 2 shows the model selection results for the four dependent variables of gay/lesbian, bisexual, transgender, and LGBT; "X" refers to the model option with the best performance. Generally, multilevel logistic regression performs better than random forest and LASSO regression and looking at the variable selection method LASSO is performing better than backwards selection for all dependent variables except for overall LGBT identity. The transgender variable was an exception because MLM did not converge and thus regular logistic regression methods were used.

Table 2. Model Selection

Model	Gay/Lesbian	Bisexual	Transgender	LGBT
MLM- backward variable selection				X
MLM-LASSO variable selection	X	X	X*	
LASSO regression				
Random forest				

*Note: MLM did not converge for transgender, so regular logistic regression methods were used.

Now we will look more closely at mean state error. The mean state error was the average difference between the true and predicted state estimates. For each dependent variable we produced the weighted state-level estimates using predictive probabilities and true values to represent the predicted and true state-level estimates. The scatter plots in Figures 1-4 show how closely the true and predicted state level estimates align to each other. The closer the true and predicted values are, the closer the trend line is to the diagonal line and the better the model performs. As we can see, the MSE is smaller for the gay/lesbian model than the bisexual and the transgender model seems to perform the worst.

Figure 1. Gay/Lesbian State-level Prevalence (%)

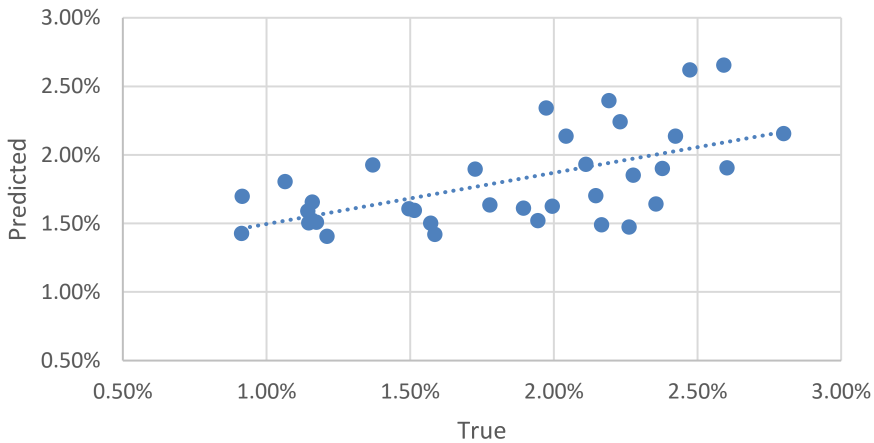
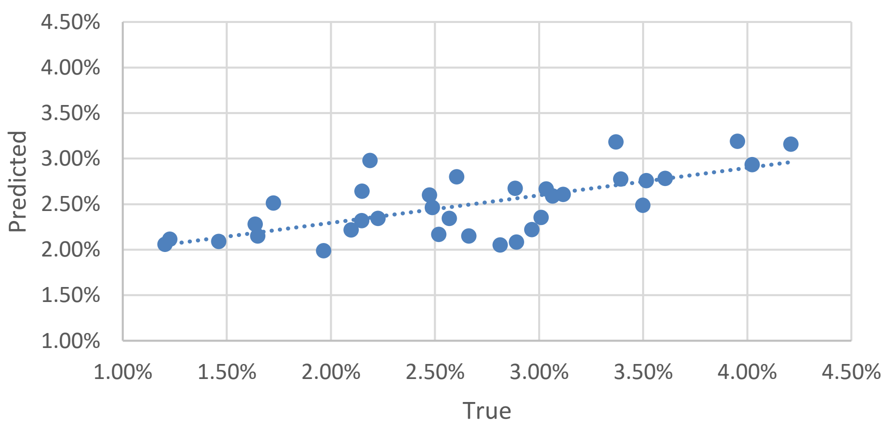
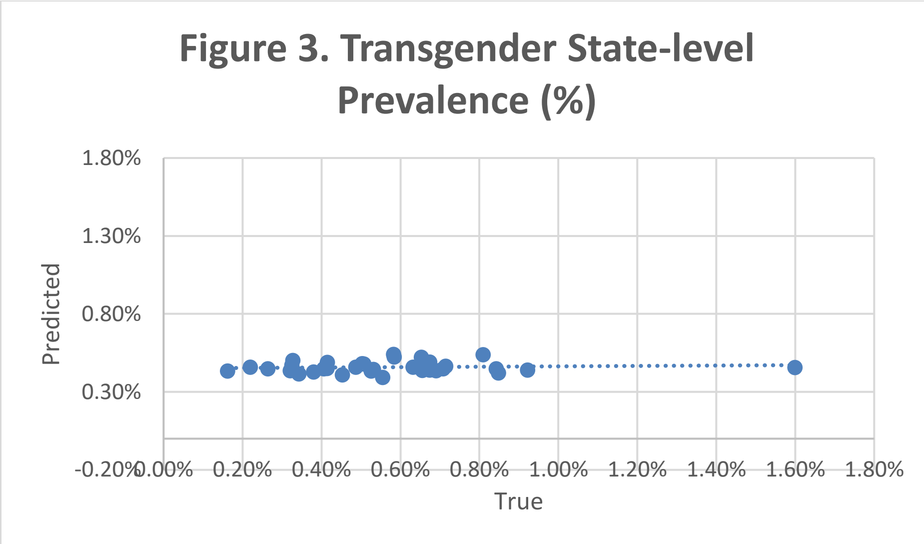
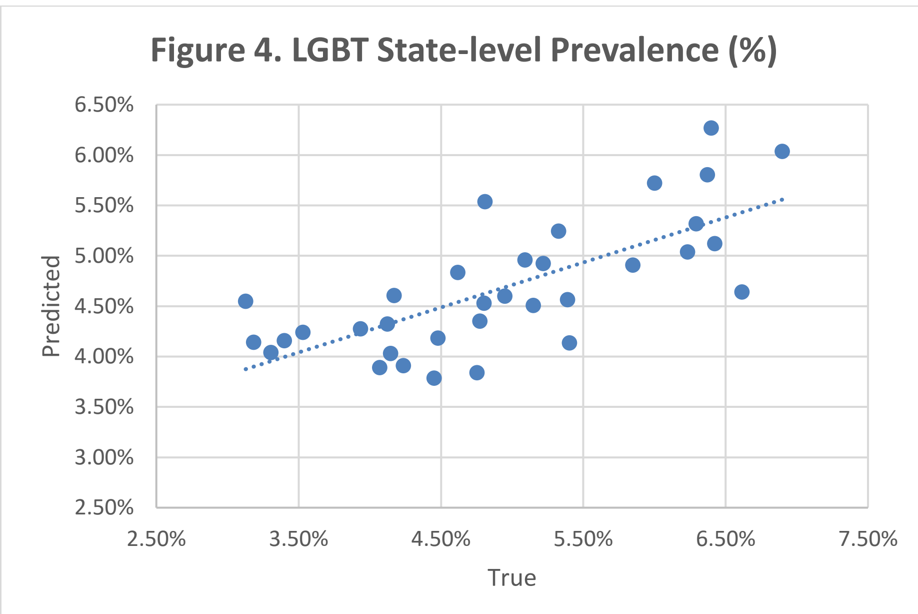


Figure 2. Bisexual State-level Prevalence (%)







After selecting the best model settings for each dependent variable, we used the training, validation, and test sets to select the final covariates in order to get the most prediction power. We then used the final model to predict for all the states plus DC whether they had an LGBT module. Table 3 shows how the variables were selected in the models. The bolded variables were selected into all 4 models and are considered main contributors to predicting LGBT identification status. These variables lifetime depressive disorder diagnosis, marital status, education level, employment status, income level, ever been diagnosed with some sort of arthritis by a doctor, race/ethnicity, age, had at least one drink of alcohol in the past 30 days, and whether or not they live in a private residence.

Table 3. Covariate selection

LABEL	Gay/Lesbian	Bisexual	Transgender	LGBT
Ever had a depressive disorder	X	X	X	X
Marital Status	X	X	X	X
Education Level	X	X	X	X
Employment Status	X	X	X	X
Income Level	X	X	X	X
Ever had a doctor diagnose them as having some form of arthritis	X	X	X	X
Race/ethnicity	X	X	X	X
Age	X	X	X	X
Had at least one drink of alcohol in the past 30 days	X	X	X	X
Live in a private residence or not	X	X	X	X
Have Health Care Professionals	X	X	X	
Could Not See Doctor Because of Cost	X	X	X	
Number of Days Physical Health Not Good	X	X	X	
Number of Days Mental Health Not Good	X	X	X	
Always or Nearly Always Wear Seat Belts	X	X	X	
Number of Adults in household	X	X	X	
Difficulty Dressing or Bathing	X	X		X
Ever had asthma	X	X		X
Smoking Status	X	X		X
have ever been tested for HIV	X	X		X

Table 4 shows more selected variables. The variables in red were selected for only one model, mostly for the transgender model. These variables include lifetime cancer diagnosis of any type, general health status, presence of any health care coverage, length of time since last routine coverage, lifetime diagnosis of chronic obstructive pulmonary disease, emphysema, or chronic bronchitis, lifetime diagnosis of diabetes, and lifetime diagnosis of coronary heart disease (CHD) or myocardial infarction (MI).

Table 4. Covariate selection

LABEL	Gay/Lesbian	Bisexual	Transgender	LGBT
Are You A Veteran	X		X	X
Own or Rent Home		X	X	X
Difficulty Concentrating or Remembering		X	X	X
Difficulty Doing Errands Alone		X	X	X
body mass index (BMI)		X	X	X
landline/cellphone respondent		X	X	X
Did physical activity or exercise during the past 30 days	X	X		
number of children in household	X		X	
Adult flu shot/spray past 12 months	X			X
Difficulty Walking or Climbing Stairs		X		X
Ever had any type of cancer	X			
General Health Status		X		
Have any health care coverage			X	
Length of time since last routine checkup			X	
Ever had chronic obstructive pulmonary disease, emphysema or chronic bronchitis			X	
Ever had diabetes			X	
Ever had coronary heart disease (CHD) or myocardial infarction (MI)			X	

Random effects are also taken into account in multilevel models. State-level law indices are fit as random effects, and the results are presented in Tables 5-7. We collapsed the LGBT welcome index into five categories because using the original categories caused convergence issues. We also only tested random effects for the intercept term. For the gay/lesbian model, we can see that only the fourth level is significantly different from the baseline intercept term, and the estimate shows a higher prevalence at this level. For Bisexual and LGBT, we can see the first and last levels are significant, the lowest level declines and the highest-level increases in prevalence. Interestingly for LGBT, the third level is significant declines on LGBT prevalence, even though the welcome index is positive.

Table 5. Random effects for gay/lesbian model

Gay/lesbian: Solution for Random Effects					
Effect	**Subject**	**Estimate**	**Std Err Pred**	**Pr > |t|**	**State index**
Intercept	statelaw 1	-0.06704	0.0415	0.106	(-1,-2,-3,-4)
Intercept	statelaw 2	-0.01593	0.0401	0.691	0
Intercept	statelaw 3	-0.06031	0.041	0.142	(1 ~ 6)
Intercept	statelaw 4	**0.08006**	0.0385	**0.038**	(7 ~ 9)
Intercept	statelaw 5	0.06321	0.0381	0.097	(10 ~ 12)

Table 6. Random effects for bisexual model

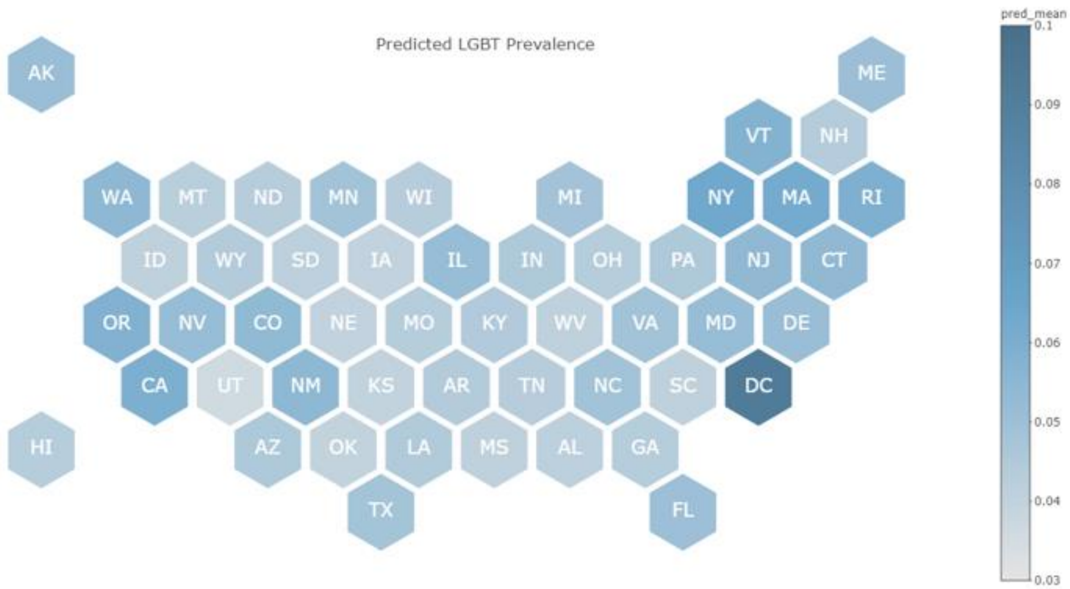
Bisexual: Solution for Random Effects					
Effect	**Subject**	**Estimate**	**Std Err Pred**	**Pr > |t|**	**State index**
Intercept	statelaw 1	**-0.1804**	0.0704	**0.01**	(-1,-2,-3,-4)
Intercept	statelaw 2	0.00892	0.0693	0.898	0
Intercept	statelaw 3	-0.1044	0.0702	0.137	(1 ~ 6)
Intercept	statelaw 4	0.121	0.0685	0.077	(7 ~ 9)
Intercept	statelaw 5	**0.155**	0.0683	**0.023**	(10 ~ 12)

Table 7. Random effects for LGBT model

LGBT: Solution for Random Effects					
Effect	**Subject**	**Estimate**	**Std Err Pred**	**Pr > |t|**	**State index**
Intercept	statelaw 1	**-0.08473**	0.0431	**0.049**	(-1,-2,-3,-4)
Intercept	statelaw 2	-0.00229	0.0424	0.957	0
Intercept	statelaw 3	**-0.08603**	0.0433	**0.047**	(1 ~ 6)
Intercept	statelaw 4	0.07553	0.0419	0.071	(7 ~ 9)
Intercept	statelaw 5	**0.09752**	0.0418	**0.02**	(10 ~ 12)

In the final stage of analysis, we apply our models to all the states in US to produce predictions. Figure 5 show our final predictions. DC has the highest prevalence and MA, NY, and CA also have higher LGBT populations. Overall, the northeast region and the west coast have higher prevalence, and this pattern is somewhat similar to the distribution of LGBT welcoming laws.

Figure 5. Final model prediction for Overall LGBT (%)



4. Conclusions and Discussion

This study showed the feasibility of developing multivariate models to generate state estimates that borrow estimation power from states with module data. The estimation methods were validated by comparing with states with direct survey estimates and sufficient observations. The methodology also supported the computation of national estimates based on the incomplete mosaic of states with module data. While developed in the context of the BRFSS data for SOGI outcomes, the approach can be used for other BRFSS topics and/or for other national surveys based on state samples.

The model-based methodology developed for this study can be applied to any BRFSS modules that are used in a subset of states as long as the number of states exceeds a minimum (15–16 states) in order to provide sufficient observations. Although an exact number of observations is not specified herein, researchers will have to take care when the number of states is low and/or the number of observations is a substantial portion of the total number of observations. Researchers are urged to review the application of this method to other variables where the total number of persons who report the variable of interest is low. Our research was conducted with a variable in which less than 1% of the total number of observations responded that they were transgender. It is unlikely, therefore, that researchers will apply the method to an indicator with lower prevalence in the state-level population; however, as a general rule, researchers applying the method must ensure that the demographic and/or risk groups are of sufficient size and scope to represent the other states.

As with all research, we found some limitations in our approach. Given that we began with a demographic that represented a small portion of the population, we believe that some of the variability of our approach resulted from the low prevalence estimates of persons who are transgender. To solve the issue of small proportion prevalence, we experimented some case balancing techniques by under sampling the non-LGBT people. It did a much better job to predict the true positive at individual level, but significantly overestimated the state-level prevalence.

In reviewing our results and predictions, we noticed that transgender has different characteristic and geographic distribution than gay/lesbian and bisexual. In the future study, we may develop a separate state law index only for transgender. Instead of creating an overall LGBT dependent variable, we may consider combine gay/lesbian and bisexual as one dependent variable but not transgender.

Regarding to the validation method, we may consider develop a leave-one state our method. For example, in this data we have 34 states with SOGI module, we can randomly select 33 states to train the model and use 1 state to test the performance. It is similar as a state-level cross validation. We did not have the resources to run this test because of processing power limit.

We also acknowledge that the treatment of persons who refuse to answer and/or answered “do not know” to any of the questions included in the analyses is subjective. In future research we intend to delve deeper into these responses. It may be that respondents in fact “do not know” their sexual orientation and/or gender identity, or it may be that some respondents do not understand the questions themselves. Another potential response bias is that the surveys are based on self-reports, which may cause different understandings among different respondents regarding to the question description. In addition, we found that there is a relatively high proportion of DNK and refusal answers among non-English speaking respondents, which may be caused by the difference interpretation when the questionnaire is translation into another language such as Spanish.

References

Centers for Disease Control and Prevention. 2017. *Behavioral Risk Factor Surveillance System.* April 18. http://www.cdc.gov/brfss.

Conron, Keith J, and Shoshana K Goldberg. 2020. *Adult LGBT Population in the United States.* Fact sheet, The Williams Institute.

Flores, Andrew R, Jody L Herman, Gary J Gates, and Taylor N T Brown. 2016. *How Many Adults Identify as Transgender in the United States?* Report, The Williams Institute.

Guo, Jing, Thomas Land, Jean M Zotter, Xingyou Zhang, Erica Marshall, and Wenjun Li. 2013. "Small-area Estimation on Current Asthma Prevalence Among Adults in Massachusetts Using BRFSS Survey Data." *American Public Health Association.* Boston, Massachusetts: American Public Health Association. 283510. https://apha.confex.com/apha/141am/webprogram/Paper283510.html.

Heinze, Georg et al. 2018. " Variable selection - A review and recommendations for the practicing statistician." *Biometrical journal. Biometrische Zeitschrift* vol. 60,3 (2018): 431-449. doi:10.1016/j.amepre.2014.08.021.

Institute of Medicine. 2013. "Collecting Sexual Orientation and Gender Identity Data in Electronic Health Records: Workshop Summary." *Board on the Health of Select Populations.* Washington DC: National Academies Press. Accessed April 18, 2017. doi:https://doi.org/10.17226/18260.

J.L. Herman, editor. 2014. *Best Practices for asking questions to identify transgender and other gender minority respondents on population surveys.* Los Angeles: The Williams Institute.

Khalil GM, Crawford CAG. 2015. "A Bibliometric Analysis of U.S.-Based Research on the Behavioral Risk Factor Surveillance System." *American Journal of Preventive Medicine* 48 (1): 50-57. doi:10.1016/j.amepre.2014.08.021.

LGBT Demographic Data Interactive. 2019. *Los Angeles, CA: The Williams Institute, UCLA School of Law.* https://williamsinstitute.law.ucla.edu/visualization/lgbt-stats/?topic=LGBT#about-the-data

Li W, Kelsey JL, Zhang Z, Lemon SC, Mezgebu S, Boddie-Willia C, Reed GW. 2009. "Small-Area Estimation and Prioritizing Communities for Obesity Control in Massachusetts." *American Journal of Public Health* 99 (3): 511-519. doi:10.2105/AJPH.2008.137364.

National Cancer Institute. 2017. *Cancer Prevalence Statistics: Approaches to Estimation Using Cancer Registry Data.* Accessed April 21, 2017. https://surveillance.cancer.gov/prevalence/approaches.html.

National Cener for Biotechnology Information. 2017. *Collecting Sexual Orientation and Gender Identity Data in Electronic Health Records: Workshop Summary.* April 18. https://www.ncbi.nlm.nih.gov/books/NBK154077/.

National LGBTQ Task Force. n.d. *Queer the Census.* Accessed 2020. https://www.sageusa.org/counting-lgbt-communities-sage-and-the-2020-census/.

Newport, Frank. 2018. "In U.S., Estimate of LGBT Population Rises to 4.5%." *Gallup.* May 22. Accessed September 25, 2020. https://news.gallup.com/poll/234863/estimate-lgbt-population-rises.aspx.

Pierannunzi, C, F Xu, R C Wallace, W Garvin, K J Greenlund, W Bartoli, D Ford, P Eke, and G M Town. 2016. "A methodological approach to small area estimation for the Behavioral Risk Factor Surveillance System." *Preventing Chronic Disease.*

Song L, Mercer L, Wakefield J, Laurent A, Solet D. 2016. "Using Small-Area Estimation to Calculate the Prevalence of Smoking by Subcounty Geographic Areas in King County, Washington, Behavioral Risk Factor Surveillance System, 2009–2013." *Preventing Chronic Disease* 13.

United States Census Bureau. 2020. *About.* April 21. Accessed September 15, 2020. https://www.census.gov/programs-surveys/decennial-census/2020-census/about.html.

Zhang, X, J B Holt, H Lu, A G Wheaton, E S Ford, K J Greenlund, and J B Croft. 2014. "Multilevel regression and post stratification for small-area estimation of population health outcomes: a case study of chronic obstructive pulmonary disease prevalence using the Behavioral Risk Factor Surveillance System." *American Journal of Epidemiology* 179 (8): 1025-33. doi:10.1093/aje/kwu018.

Zhang, Xingyou, James B Holt, Ann G Wheaton, Earl S Ford, Kurt J Greenlund, and Janet B Croft. 2014. "Multilevel Regression and Poststratification for Small Area Estimation of Population Health Outcomes: A Case Study of Chronic Pulmonary Disease Prevalence Using the Behavioral Risk Factor Surveillance System." *American Journal of Epidemiology* 179 (8): 1025-1033. doi:10.1093/aje/kwu018.

Zhang, Z, L Zhang, A Penman, and W May. 2011. "Using small-area estimation to calculate county-level prevalence of obesity in Mississippi 2007-2009." *Preventing Chronic Disease* 8 (4): 1-11.