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

Ronaldo Iachan[1], Yangyang Deng[1], Carol Pierannunzi[2]
[1]ICF, 530 Gaither Road, Suite 500, Rockville, MD 20850
[2]Centers for Disease Control and Prevention, 1600 Clifton Road NE, Atlanta, GA 30333

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 2014 to 2016 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 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

The Behavioral Risk Factor Surveillance System (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 states, 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 in 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, therefore, 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.

2. Methods

The estimates calculated herein are based on multivariate logistic regression models and multilevel logistic regression 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. The population for the analysis included all 50 states and the District of Columbia for the years 2014 to 2016. In phase I analysis, logistic regression was used for 2014 and 2015 year data (2016 data was not available at the time phase I was performed).In phase II analysis, multilevel model was used by adding state-level laws as levels for 2014-2016 data.

2.1 Phase I: Logistic Regression Model

The logistic regression models were fit to the weighted survey data for each cycle using SAS procedures for survey data (Proc Survey Logistic). The models included a range of predictors selected in bivariate analyses for the different outcomes: gay/lesbian, bisexual and transgender. (The bivariate tables are not presented in this manuscript; they were included in part in Pierannunzi et al. 2017).

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 transsexual question, 2% of the DK-refusal responses were imputed as transsexual.

2.2 Phase II: Multilevel logistic regression model

Multilevel logistic regression models are an extension of logistical models which are appropriate for data organized at more than one level (i.e., nested data). In the phase II multilevel models , state-level indices related to sexual minority and transgender laws were computed and used, to incorporate state effects on LGBT prevalence. Similar to phase I, the imptation and bivariate analyses were performed priot to modeling. We defined two different indices to reflect state-level laws in the general sexual minority and in the more specific transgender domains.

The sexual minority index is created in each state as the sum of four separate (0-1) indicators defined as follows for the different laws in this area:Indicator 1: States don't have hate crime laws specifically protecting LGBT; Indicator 2: States don't have non-discrimination employment laws protecting LGBT; Indicator3: States don't have laws prohibiting establishments from discriminating against LGBT customers; Indicator 4: States don't have laws making same-sex marriage legal.

The sexual minority law index indicator is the sum of these founr indices (Index 1 to Index 4). Any indicator equals to 1 suggests a law that is harmful to sexual minorities, so that larger values of the overall sum index suggest a state environment that is negative towards the LGBT population.
Following are the definitions for state-level The transgender law index is computed similarly for each state as the sum of four different indicators for relevant laws: Indicator 5: States don't have laws protecting youths from conversion therapy; Indicator 6: States don't have explicit bans on excluding trans individuals from receiving health insurance coverage; Indicator 7: States don't have laws for gender-neutral single-occupancy restrooms; Indicator 8: States have laws prohibiting transgender people from receiving documents reflecting their gender

identity.

Similar to the sexual minority index, larger values of the overall sum index suggest a state environment that is negative towards the transgender population.

3. Results

This section presents the model predictions for selected states for the phase I and phase II analyses.

3.1 Phase I

Table 2 presents the odds ratio estimates for the weighted logistic models for the three dependent variables of interest (non-significant predictors are shown as "NA"). The table shows that ever having an HIV test and education have the largest positive effects on the gay/lesbian prevalence (3.3 and 1.7). , marital status and children presence have the most negative impact on gay/lesbian (0.271 and 0.417). For the second dependent variable, those respondents with non-good mental status and those taking an HIV test are more likely to be bisexual (1.9 and 1.7); those who are married and male are less likely to be bisexual (0.4 and 0.6). Lastly, the odds for transgender are higher for males and Hispanics, (1.9 and 1.4), and lower for those with higher education and for drinkers (0.6 and 0.7).

Table 2: Odds Ratio Estimates of Phase I Models

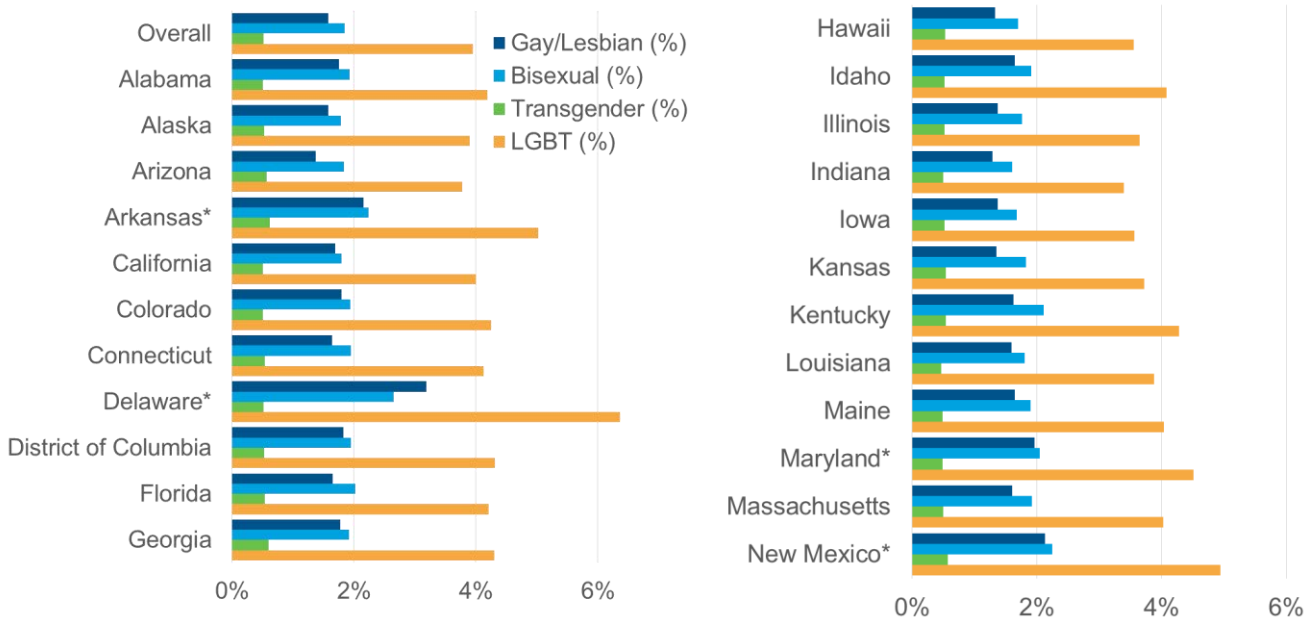
DEPENDENT VARIABLES	GAY/LESBIAN			BISEXUAL			TRANSGENDER		
INDEPENDENT VARIABLES	**Estimate**	**95% Confidence Limits**		**Estimate**	**95% Confidence Limits**		**Estimate**	**95% Confidence Limits**	
GENDER: MALE VS FEMALE	1.551	1.387	1.735	**0.625**	0.554	0.706	**1.887**	1.518	2.345
RACE: WHITE VS NON-WHITE	1.224	1.06	1.413	NA	NA	NA	0.694	0.535	0.9
ETHNICITY: HISPANIC VS NON-HISPANIC	1.296	1.069	1.572	NA	NA	NA	**1.422**	0.997	2.029
EDUCATION: COLLEGE GRADUATE VS BELOW COLLEGE	**1.711**	1.532	1.91	NA	NA	NA	**0.584**	0.456	0.748
MARITAL STATUS: MARRIED VS NON-MARRIED	**0.271**	0.235	0.311	**0.432**	0.379	0.493	NA	NA	NA
RENT OR OWN: RENT VS OWN HOUSE	NA	NA	NA	1.395	1.229	1.582	1.285	0.996	1.658
EMPLOYMENT: EMPLOYED VS NOT EMPLOYED	NA	NA	NA	0.901	0.798	1.016	0.802	0.649	0.992
MENTAL STATUS: NOT GOOD VS GOOD	1.234	1.101	1.382	**1.909**	1.688	2.159	1.276	1.021	1.596
HIV TEST: HAS TEST VS NO TEST	**3.333**	2.956	3.759	**1.697**	1.503	1.915	NA	NA	NA
OBESE: VS OBESE VS NOT OBESE	NA	NA	NA	1.124	0.992	1.273	NA	NA	NA
SMOKING STATUS: SMOKER VS NON-SMOKER	1.144	0.998	1.312	1.155	1.007	1.324	NA	NA	NA

DRINKING STATUS: DRINKER VS NON-DRINKER	1.246	1.111	1.398	1.178	1.041	1.334	**0.689**	0.558	0.851
CHILDREN PRESENCE: HAS CHILD VS HAS NO CHILD	**0.417**	0.357	0.487	NA	NA	NA	NA	NA	NA

Figure 1 presents weighted model estimates for years 2014 and 2015 for the prevalence of gay/lesbian, bisexual and transgender individuals in each state and nationally. Nationally, the prevalence rates are nearly 4% (3.95%) for the combined gay/lesbian/bisexual categories: 1.58% for gay/lesbian and 1.85% for bisexual. The prevalence for transgender is 0.52%. The states have highest prevalence rates for overall LGBT are Delaware (6.36%) and Arkansas (5.02%). The full list of states with model predictions will be presented in Appendix.

In order to validate the estimates, comparisons were made for the model-predicted prevalence rates with the survey estimates for those states where direct survey data are available. For gay/lesbian, 83% of states with direct SOGI survey data have model predictions which fall into the 95% confidence intervals for the probability sample estimates. The corresponding rates are 82% for the bisexual model predictions and 79% for the transgender model predictions. Appendix A includes the full list of states for these comparisons.

Figure 1: Weighted model estimates for years 2014-2015



3.2 Phase II

We then developed weighted logistic multilevel models which incorporate the state-level effects associated with the laws in the LBGT domain. Table 3 presents the odds ratio estimates for these models for all three dependent variables. For gay and lesbian, those who are not married and have taken an HIV test have higher prevalence (3.3and 3.3); in contrast, those have children and lower education tend to have lower prevalence (0.4 and 0.6). For bisexual, the results show those have not with non-good mental status and not married are more likely to be bisexual (2.1 and 2.3); those who are not obese are less likely to be bisexual (0.9). Lastly, the transgender odds are higher for those respondents with less education (1.9), and lower for females and drinkers (0.6 and 0.7).

TABLE 3: Odds Ratio Estimates of Phase II Models

DEPENDENT VARIABLES	GAY/LESBIAN			BISEXUAL			TRANSGENDER		
INDEPENDENT VARIABLES	**Estimate**	**95% Confidence Limits**		**Estimate**	**95% Confidence Limits**		**Estimate**	**95% Confidence Limits**	
GENDER: FEMALE VS MALE	0.662	0.659	0.664	1.483	1.477	1.489	**0.65**	0.646	0.653
RACE: NON-WHITE VS WHITE	0.826	0.822	0.83	NA	NA	NA	1.586	1.577	1.594
ETHNICITY: NON-HISPANIC VS HISPANIC	0.842	0.837	0.847	NA	NA	NA	0.768	0.763	0.773
EDUCATION: BELOW COLLEGE VS COLLEGE GRADUATE	**0.61**	0.608	0.613	1.09	1.085	1.095	**1.965**	1.951	1.98
MARITAL STATUS: NON-MARRIED VS MARRIED	**3.285**	3.268	3.302	**2.268**	2.258	2.278	1.052	1.047	1.058
RENT OR OWN: RENT VS OWN HOUSE	1.043	1.039	1.048	1.411	1.405	1.416	NA	NA	NA
EMPLOYMENT: NOT EMPLOYED VS EMPLOYED	0.937	0.933	0.942	1.058	1.054	1.063	1.1	1.094	1.105
MENTAL STATUS: NOT GOOD VS GOOD	1.288	1.282	1.293	**2.083**	2.075	2.091	1.364	1.358	1.371
AID TEST: HAS TEST VS NO TEST	**3.269**	3.255	3.284	1.725	1.718	1.732	1.046	1.04	1.051
OBESE: NOT OBESE VS OBESE	0.997	0.992	1.001	**0.879**	0.875	0.882	1.172	1.166	1.178
SMOKING STATUS: SMOKER VS NON-SMOKER	1.13	1.125	1.136	1.133	1.128	1.138	1.137	1.13	1.144
DRINKING STATUS: DRINKER VS NON-DRINKER	1.264	1.258	1.269	1.205	1.201	1.21	**0.678**	0.674	0.681
CHILDREN PRESENCE: HAS CHILD VS HAS NO CHILD	**0.43**	0.428	0.432	1.062	1.057	1.066	0.832	0.828	0.837

Besides the fixed effects shown in Table 3, random effects are also taken account in multilevel models. State-level law indices are fit as random effects, and the results are presented in Table 4.

For gay and lesbian, the same sex law indicator 0 has a significant difference at .1 level from other levels (P-value: 0.525), which raises the intercept term by 0.1043, and this implies a positive relationship between sexual minority laws and the gay and lesbian prevalence. The results are similar for the bisexual outcome variable. For the transgender variable, there is a negative relationship between state-laws and prevalence.

TABLE 4: Random Effects for Multilevel Models

SOLUTION FOR RANDOM EFFECTS			
SUBJECT	**Estimate**	**Std Err Pred**	**Pr > |t|**
GAY/LESBIAN			
SAME SEX LAW INDICATOR = 0	0.1043	0.05381	**0.0525**
SAME SEX LAW INDICATOR = 3	-0.02938	0.05381	0.5852
SAME SEX LAW INDICATOR = 4	-0.07497	0.05382	0.1636
BISEXUAL			
SAME SEX LAW INDICATOR = 0	0.07828	0.03943	**0.0471**
SAME SEX LAW INDICATOR = 3	-0.03104	0.03943	0.4321
SAME SEX LAW INDICATOR = 4	-0.04724	0.03943	0.2309
TRANSGENDER			
TRANSGENDER LAW INDICATOR = 0	-0.2201	0.07850	**0.0051**
TRANSGENDER LAW INDICATOR = 1	-0.02956	0.07848	0.7065
TRANSGENDER LAW INDICATOR = 2	-0.00926	0.07846	0.9060
TRANSGENDER LAW INDICATOR = 3	-0.01227	0.07846	0.8758
TRANSGENDER LAW INDICATOR = 4	0.2711	0.07848	**0.0006**

Figure 2: Selected Final Model Prediction for Phase II

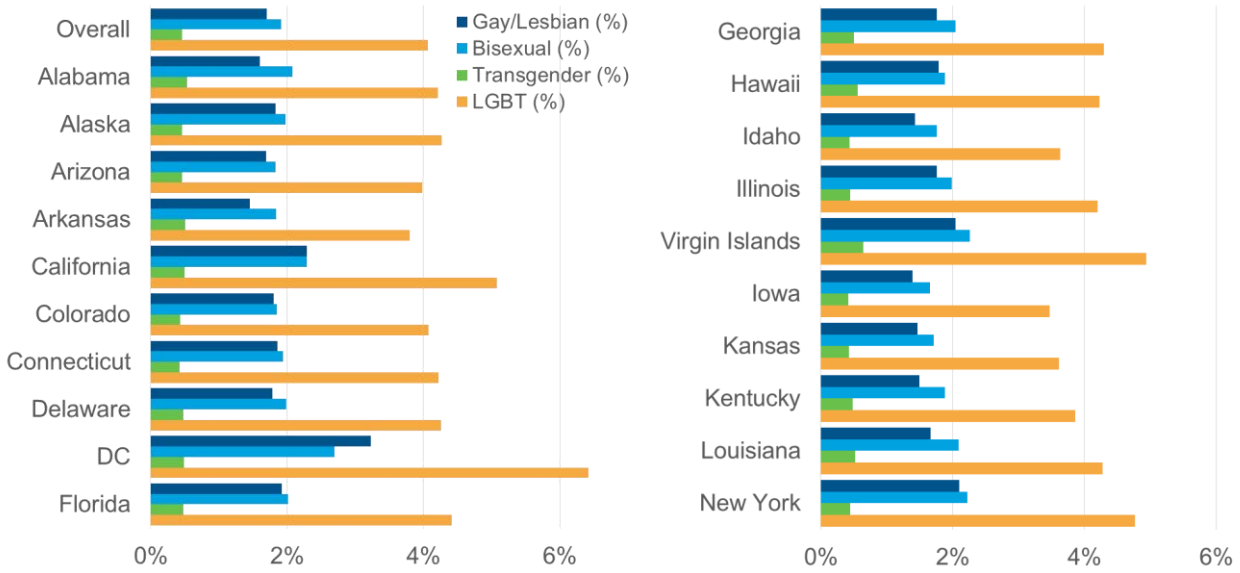


Figure 2 presents weighted model estimates for years 2014 to 2016 for the prevalence of gay/lesbian, bisexual and transgender individuals in each state and nationally. Nationally, the prevalence rates are over 4% (4.07%) for the combined gay/lesbian/bisexual categories: 1.70% for gay/lesbian and 1.91% for bisexual. The prevalence for transgender is 0.46%. The states with highest prevalence rates for overall LGBT are DC (6.42%) and California (5.07%). Appendix A provides h the full set of state-level model predictions.

In order to validate the estimates, we compared the model-predicted prevalence rates with the survey estimates for those states where direct survey data. For gay/lesbian, there are 78% of such states' model predictions fall into the 95% confidence intervals for the probability sample estimates, while the counterpart rates are 60% for bisexual has 60% and 81% for transgender. Compare to the phase I models, multilevel models in phase II increase

performance for transgender, while decrease for gay/lesbian and bisexual. Appendix A also includes the full set of comparisons.

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. This may not be a factor if the modeling approach were to be used for estimates of Body Mass Index (BMI), diabetes, or other health indicators found in the BRFSS. 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.

Future research will consider alternative imputation procedures for missing values as well as DNK and refusal answers. We will also consider one overall LGBT indicator as a new dependent variable to decrease the effect of a small minority and to best account for the effects of laws that affect this broader population as a whole.

References

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

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.

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.

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/.

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.

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.

Appendix A

Table A-1: Phase I Validity Testing for Model Predictions

Phase I: Survey estimates vs Model Predictions - 2014 and 2015					
	SOGI Variable	**Survey Estimates**	**CI Lower**	**CI Upper**	**Model Prediction**
Overall	Gay/Lesbian	1.62%	1.54%	1.70%	1.56%
	Bisexual	2.05%	1.93%	2.17%	1.86%
	transgender	0.57%	0.51%	0.63%	0.50%
Colorado	Gay/Lesbian	2.02%	1.55%	2.49%	1.71%
	Bisexual	2.26%	1.71%	2.81%	1.85%
	transgender	0.27%	0.13%	0.41%	0.50%

Connecticut	Gay/Lesbian	1.95%	1.54%	2.36%	1.83%
	Bisexual	2.14%	1.67%	2.61%	2.02%
	transgender	0.28%	0.10%	0.46%	0.48%
Delaware	Gay/Lesbian	2.45%	1.94%	2.96%	1.66%
	Bisexual	2.36%	1.75%	2.97%	1.95%
	transgender	0.80%	0.41%	1.19%	0.53%
Georgia	Gay/Lesbian	1.61%	1.06%	2.16%	1.62%
	Bisexual	1.66%	0.97%	2.35%	1.98%
	transgender	0.56%	0.21%	0.91%	0.54%
Hawaii	Gay/Lesbian	1.67%	1.38%	1.96%	1.72%
	Bisexual	1.67%	1.36%	1.98%	1.88%
	transgender	0.71%	0.51%	0.91%	0.60%
Idaho	Gay/Lesbian	0.94%	0.61%	1.27%	1.33%
	Bisexual	2.02%	1.57%	2.47%	1.73%
	transgender	0.54%	0.32%	0.76%	0.50%
Illinois	Gay/Lesbian	1.41%	0.98%	1.84%	1.62%
	Bisexual	2.12%	1.53%	2.71%	1.96%
	transgender	0.54%	0.27%	0.81%	0.51%
Indiana	Gay/Lesbian	1.63%	1.32%	1.94%	1.38%
	Bisexual	2.24%	1.77%	2.71%	1.82%
	transgender	0.58%	0.40%	0.76%	0.50%
Iowa	Gay/Lesbian	1.05%	0.72%	1.38%	1.24%
	Bisexual	1.68%	1.19%	2.17%	1.63%
	transgender	0.30%	0.10%	0.50%	0.46%
Kansas	Gay/Lesbian	1.28%	1.10%	1.46%	1.38%
	Bisexual	1.80%	1.58%	2.02%	1.69%
	transgender	0.48%	0.38%	0.58%	0.48%
Kentucky	Gay/Lesbian	1.61%	1.10%	2.12%	1.33%
	Bisexual	1.28%	0.91%	1.65%	1.83%
	transgender	0.54%	0.30%	0.78%	0.51%
Louisiana	Gay/Lesbian	1.47%	1.06%	1.88%	1.62%
	Bisexual	1.73%	1.22%	2.24%	2.11%
	transgender	0.50%	0.28%	0.72%	0.54%
Maryland	Gay/Lesbian	1.75%	1.40%	2.10%	1.63%
	Bisexual	2.24%	1.73%	2.75%	1.89%
	transgender	0.46%	0.28%	0.64%	0.47%
Massachusetts	Gay/Lesbian	2.19%	1.78%	2.60%	2.05%
	Bisexual	2.49%	2.02%	2.96%	2.17%
	transgender	0.35%	0.17%	0.53%	0.50%
Minnesota	Gay/Lesbian	1.47%	1.29%	1.65%	1.47%
	Bisexual	1.96%	1.74%	2.18%	1.72%
	transgender	0.62%	0.50%	0.74%	0.45%
Missouri	Gay/Lesbian	2.01%	1.42%	2.60%	1.45%
	Bisexual	1.88%	1.17%	2.59%	1.88%
	transgender	0.68%	0.25%	1.11%	0.52%

Montana	Gay/Lesbian	0.96%	0.59%	1.33%	1.46%
	Bisexual	1.56%	1.05%	2.07%	1.76%
	transgender	0.48%	0.23%	0.73%	0.48%
Nevada	Gay/Lesbian	2.29%	1.62%	2.96%	1.80%
	Bisexual	2.50%	1.89%	3.11%	2.08%
	transgender	0.46%	0.19%	0.73%	0.54%
New York	Gay/Lesbian	1.99%	1.72%	2.26%	2.04%
	Bisexual	2.41%	2.04%	2.78%	2.25%
	transgender	0.62%	0.44%	0.80%	0.54%
Ohio	Gay/Lesbian	1.24%	1.00%	1.48%	1.40%
	Bisexual	2.18%	1.77%	2.59%	1.85%
	transgender	0.73%	0.49%	0.97%	0.50%
Pennsylvania	Gay/Lesbian	1.44%	1.17%	1.71%	1.56%
	Bisexual	1.67%	1.36%	1.98%	1.92%
	transgender	0.47%	0.31%	0.63%	0.49%
Texas	Gay/Lesbian	1.39%	1.00%	1.78%	1.59%
	Bisexual	2.18%	1.65%	2.71%	1.84%
	transgender	0.70%	0.39%	1.01%	0.55%
Vermont	Gay/Lesbian	2.07%	1.66%	2.48%	1.73%
	Bisexual	2.51%	2.00%	3.02%	1.88%
	transgender	0.67%	0.40%	0.94%	0.43%
Virginia	Gay/Lesbian	1.75%	1.46%	2.04%	1.64%
	Bisexual	1.76%	1.45%	2.07%	1.87%
	transgender	0.64%	0.46%	0.82%	0.49%
West Virginia	Gay/Lesbian	1.05%	0.74%	1.36%	1.60%
	Bisexual	1.09%	0.78%	1.40%	1.92%
	transgender	0.83%	0.54%	1.12%	0.53%
Wisconsin	Gay/Lesbian	1.48%	1.15%	1.81%	1.47%
	Bisexual	1.91%	1.54%	2.28%	1.85%
	transgender	0.41%	0.17%	0.65%	0.47%
Wyoming	Gay/Lesbian	1.01%	0.52%	1.50%	1.34%
	Bisexual	1.99%	1.17%	2.81%	1.60%
	transgender	0.27%	0.09%	0.45%	0.47%

Table A-2: Phase I Model Predictions for All States in US

Weighted Estimates for Gay/Lesbian, Bisexual and Transgender - National and All States, 2014 and 2015		
	SOGI Variables	**Model Predictions**
Overall	Gay/Lesbian	1.58%
	Bisexual	1.85%
	Transgender	0.52%
Alabama	Gay/Lesbian	1.75%
	Bisexual	1.93%

	Transgender	0.51%
Alaska	Gay/Lesbian	1.58%
	Bisexual	1.79%
	Transgender	0.53%
Arizona	Gay/Lesbian	1.37%
	Bisexual	1.84%
	Transgender	0.57%
Arkansas	Gay/Lesbian	2.16%
	Bisexual	2.24%
	Transgender	0.62%
California	Gay/Lesbian	1.69%
	Bisexual	1.80%
	Transgender	0.51%
Colorado	Gay/Lesbian	1.80%
	Bisexual	1.94%
	Transgender	0.51%
Connecticut	Gay/Lesbian	1.64%
	Bisexual	1.95%
	Transgender	0.54%
Delaware	Gay/Lesbian	3.19%
	Bisexual	2.65%
	Transgender	0.52%
District of Columbia	Gay/Lesbian	1.83%
	Bisexual	1.95%
	Transgender	0.53%
Florida	Gay/Lesbian	1.65%
	Bisexual	2.02%
	Transgender	0.54%
Georgia	Gay/Lesbian	1.78%
	Bisexual	1.92%
	Transgender	0.60%
Hawaii	Gay/Lesbian	1.33%
	Bisexual	1.70%
	Transgender	0.53%
Idaho	Gay/Lesbian	1.65%
	Bisexual	1.91%
	Transgender	0.52%
Illinois	Gay/Lesbian	1.37%
	Bisexual	1.76%
	Transgender	0.52%
Indiana	Gay/Lesbian	1.29%
	Bisexual	1.61%

	Transgender	0.50%
Iowa	Gay/Lesbian	1.37%
	Bisexual	1.68%
	Transgender	0.52%
Kansas	Gay/Lesbian	1.35%
	Bisexual	1.83%
	Transgender	0.54%
Kentucky	Gay/Lesbian	1.63%
	Bisexual	2.11%
	Transgender	0.54%
Louisiana	Gay/Lesbian	1.60%
	Bisexual	1.81%
	Transgender	0.47%
Maine	Gay/Lesbian	1.65%
	Bisexual	1.90%
	Transgender	0.49%
Maryland	Gay/Lesbian	1.97%
	Bisexual	2.05%
	Transgender	0.49%
Massachusetts	Gay/Lesbian	1.61%
	Bisexual	1.92%
	Transgender	0.50%
Michigan	Gay/Lesbian	1.49%
	Bisexual	1.70%
	Transgender	0.47%
Minnesota	Gay/Lesbian	1.45%
	Bisexual	1.98%
	Transgender	0.59%
Mississippi	Gay/Lesbian	1.46%
	Bisexual	1.86%
	Transgender	0.54%
Missouri	Gay/Lesbian	1.49%
	Bisexual	1.72%
	Transgender	0.50%
Montana	Gay/Lesbian	1.22%
	Bisexual	1.58%
	Transgender	0.50%
Nebraska	Gay/Lesbian	1.81%
	Bisexual	2.01%
	Transgender	0.55%
Nevada	Gay/Lesbian	1.55%
	Bisexual	1.73%

	Transgender	0.46%
New Hampshire	Gay/Lesbian	1.74%
	Bisexual	1.98%
	Transgender	0.52%
New Jersey	Gay/Lesbian	1.63%
	Bisexual	1.83%
	Transgender	0.60%
New Mexico	Gay/Lesbian	2.13%
	Bisexual	2.25%
	Transgender	0.57%
New York	Gay/Lesbian	1.70%
	Bisexual	2.00%
	Transgender	0.57%
North Carolina	Gay/Lesbian	1.25%
	Bisexual	1.55%
	Transgender	0.48%
North Dakota	Gay/Lesbian	1.40%
	Bisexual	1.81%
	Transgender	0.53%
Ohio	Gay/Lesbian	1.30%
	Bisexual	1.73%
	Transgender	0.55%
Oklahoma	Gay/Lesbian	1.83%
	Bisexual	1.96%
	Transgender	0.50%
Oregon	Gay/Lesbian	1.60%
	Bisexual	1.88%
	Transgender	0.50%
Pennsylvania	Gay/Lesbian	1.76%
	Bisexual	1.96%
	Transgender	0.50%
Rhode Island	Gay/Lesbian	1.51%
	Bisexual	1.92%
	Transgender	0.56%
South Carolina	Gay/Lesbian	1.28%
	Bisexual	1.64%
	Transgender	0.50%
South Dakota	Gay/Lesbian	1.51%
	Bisexual	1.93%
	Transgender	0.55%
Tennessee	Gay/Lesbian	1.59%
	Bisexual	1.83%

	Transgender	0.57%
Texas	Gay/Lesbian	1.13%
	Bisexual	1.53%
	Transgender	0.56%
Utah	Gay/Lesbian	1.75%
	Bisexual	1.82%
	Transgender	0.46%
Vermont	Gay/Lesbian	1.68%
	Bisexual	1.87%
	Transgender	0.50%
Virginia	Gay/Lesbian	1.71%
	Bisexual	1.86%
	Transgender	0.50%
Washington	Gay/Lesbian	1.51%
	Bisexual	1.83%
	Transgender	0.56%
West Virginia	Gay/Lesbian	1.49%
	Bisexual	1.81%
	Transgender	0.50%
Wisconsin	Gay/Lesbian	1.35%
	Bisexual	1.59%
	Transgender	0.50%
Wyoming	Gay/Lesbian	1.31%
	Bisexual	1.53%
	Transgender	0.49%

Table A-3: Phase II Validity Testing for Model Predictions

Phase II: Survey estimates vs Model Predictions - 2014 to 2016					
State	**SOGI Variable**	**Survey Estimate**	**CI Lower**	**CI Upper**	**Model Prediction**
overall	Gay/Lesbian	1.77%	1.69%	1.85%	1.69%
	Bisexual	2.18%	2.09%	2.26%	1.91%
	transgender	0.53%	0.48%	0.57%	0.45%
California	Gay/Lesbian	2.41%	1.98%	2.83%	2.22%
	Bisexual	1.95%	1.60%	2.30%	2.26%
	transgender	0.42%	0.19%	0.65%	0.50%
Colorado	Gay/Lesbian	2.01%	1.54%	2.48%	1.79%
	Bisexual	2.20%	1.66%	2.74%	1.86%
	transgender	0.28%	0.14%	0.42%	0.44%
Connecticut	Gay/Lesbian	1.80%	1.53%	2.07%	1.86%
	Bisexual	2.42%	2.05%	2.78%	1.95%

	transgender	0.51%	0.33%	0.69%	0.42%
Delaware	Gay/Lesbian	2.15%	1.78%	2.52%	1.79%
	Bisexual	2.27%	1.82%	2.73%	1.99%
	transgender	0.62%	0.35%	0.90%	0.48%
Georgia	Gay/Lesbian	2.08%	1.61%	2.54%	1.71%
	Bisexual	1.93%	1.45%	2.40%	1.99%
	transgender	0.61%	0.34%	0.88%	0.51%
Hawaii	Gay/Lesbian	1.90%	1.63%	2.17%	1.78%
	Bisexual	1.91%	1.62%	2.19%	1.88%
	transgender	0.69%	0.50%	0.87%	0.56%
Idaho	Gay/Lesbian	1.02%	0.73%	1.31%	1.42%
	Bisexual	2.20%	1.81%	2.60%	1.75%
	transgender	0.43%	0.27%	0.59%	0.44%
Illinois	Gay/Lesbian	1.72%	1.37%	2.06%	1.73%
	Bisexual	2.15%	1.71%	2.59%	2.00%
	transgender	0.48%	0.29%	0.66%	0.46%
Indiana	Gay/Lesbian	1.63%	1.37%	1.90%	1.47%
	Bisexual	2.21%	1.86%	2.56%	1.84%
	transgender	0.55%	0.41%	0.69%	0.46%
Iowa	Gay/Lesbian	0.96%	0.74%	1.19%	1.35%
	Bisexual	1.96%	1.57%	2.35%	1.64%
	transgender	0.34%	0.20%	0.49%	0.42%
Kansas	Gay/Lesbian	1.30%	1.11%	1.48%	1.46%
	Bisexual	1.78%	1.57%	1.99%	1.69%
	transgender	0.48%	0.38%	0.58%	0.43%
Kentucky	Gay/Lesbian	1.62%	1.28%	1.95%	1.49%
	Bisexual	1.76%	1.45%	2.06%	1.86%
	transgender	0.55%	0.39%	0.70%	0.48%
Louisiana	Gay/Lesbian	1.65%	1.21%	2.10%	1.66%
	Bisexual	1.63%	1.23%	2.04%	2.09%
	transgender	0.49%	0.31%	0.67%	0.51%
Maryland	Gay/Lesbian	1.70%	1.36%	2.03%	1.71%
	Bisexual	2.36%	1.83%	2.88%	1.88%
	transgender	0.47%	0.29%	0.64%	0.44%
Massachusetts	Gay/Lesbian	2.43%	2.08%	2.77%	2.19%
	Bisexual	3.30%	2.86%	3.73%	2.16%
	transgender	0.38%	0.22%	0.53%	0.43%
Minnesota	Gay/Lesbian	1.54%	1.39%	1.69%	1.59%
	Bisexual	2.07%	1.88%	2.25%	1.75%
	transgender	0.53%	0.45%	0.62%	0.40%
Mississippi	Gay/Lesbian	0.80%	0.45%	1.16%	1.58%
	Bisexual	1.01%	0.54%	1.47%	2.03%
	transgender	0.81%	0.40%	1.22%	0.55%
Missouri	Gay/Lesbian	1.45%	1.11%	1.78%	1.55%
	Bisexual	2.31%	1.71%	2.91%	1.89%

	transgender	0.42%	0.20%	0.64%	0.48%
Montana	Gay/Lesbian	1.02%	0.63%	1.40%	1.54%
	Bisexual	1.57%	1.06%	2.09%	1.77%
	transgender	0.45%	0.20%	0.71%	0.44%
Nevada	Gay/Lesbian	2.17%	1.70%	2.64%	1.94%
	Bisexual	2.61%	2.11%	3.10%	2.13%
	transgender	0.45%	0.25%	0.66%	0.49%
New York	Gay/Lesbian	2.19%	1.96%	2.41%	2.05%
	Bisexual	2.49%	2.22%	2.76%	2.20%
	transgender	0.60%	0.46%	0.74%	0.45%
Ohio	Gay/Lesbian	1.36%	1.16%	1.57%	1.50%
	Bisexual	2.19%	1.88%	2.50%	1.87%
	transgender	0.73%	0.55%	0.92%	0.46%
Pennsylvania	Gay/Lesbian	1.45%	1.22%	1.68%	1.69%
	Bisexual	1.92%	1.62%	2.22%	1.97%
	transgender	0.42%	0.31%	0.53%	0.45%
Rhode Island	Gay/Lesbian	1.80%	1.30%	2.29%	1.97%
	Bisexual	2.84%	2.04%	3.63%	2.12%
	transgender	0.77%	0.31%	1.23%	0.42%
Texas	Gay/Lesbian	1.68%	1.31%	2.05%	1.66%
	Bisexual	2.11%	1.74%	2.49%	1.83%
	transgender	0.63%	0.37%	0.90%	0.47%
Vermont	Gay/Lesbian	1.91%	1.58%	2.23%	1.89%
	Bisexual	2.86%	2.41%	3.31%	1.93%
	transgender	0.56%	0.36%	0.75%	0.39%
Virginia	Gay/Lesbian	1.71%	1.48%	1.94%	1.76%
	Bisexual	1.90%	1.64%	2.16%	1.91%
	transgender	0.56%	0.42%	0.69%	0.46%
Washington	Gay/Lesbian	1.88%	1.56%	2.20%	1.84%
	Bisexual	3.58%	3.09%	4.07%	1.95%
	transgender	0.51%	0.34%	0.69%	0.42%
West Virginia	Gay/Lesbian	1.08%	0.75%	1.41%	1.69%
	Bisexual	1.07%	0.76%	1.38%	1.95%
	transgender	0.85%	0.55%	1.14%	0.49%
Wisconsin	Gay/Lesbian	1.35%	1.08%	1.61%	1.56%
	Bisexual	2.24%	1.87%	2.61%	1.87%
	transgender	0.38%	0.20%	0.55%	0.43%
Wyoming	Gay/Lesbian	1.04%	0.55%	1.52%	1.41%
	Bisexual	1.91%	1.09%	2.73%	1.58%
	transgender	0.28%	0.11%	0.46%	0.43%

Table A-4: Phase I Model Predictions for All States in US

Weighted Estimates for Gay/Lesbian, Bisexual and Transgender - National and All States, 2014 to 2016

	SOGI Variables	Model Predictions
Overall	Gay/Lesbian	1.70%
	Bisexual	1.91%
	Transgender	0.46%
Alabama	Gay/Lesbian	1.60%
	Bisexual	2.08%
	Transgender	0.53%
Alaska	Gay/Lesbian	1.83%
	Bisexual	1.98%
	Transgender	0.46%
Arizona	Gay/Lesbian	1.69%
	Bisexual	1.83%
	Transgender	0.46%
Arkansas	Gay/Lesbian	1.45%
	Bisexual	1.84%
	Transgender	0.51%
California	Gay/Lesbian	2.29%
	Bisexual	2.29%
	Transgender	0.50%
Colorado	Gay/Lesbian	1.80%
	Bisexual	1.85%
	Transgender	0.43%
Connecticut	Gay/Lesbian	1.86%
	Bisexual	1.94%
	Transgender	0.42%
Delaware	Gay/Lesbian	1.79%
	Bisexual	1.99%
	Transgender	0.48%
District of Columbia	Gay/Lesbian	3.23%
	Bisexual	2.70%
	Transgender	0.49%
Florida	Gay/Lesbian	1.92%
	Bisexual	2.02%
	Transgender	0.48%
Georgia	Gay/Lesbian	1.76%
	Bisexual	2.04%
	Transgender	0.50%
Hawaii	Gay/Lesbian	1.79%
	Bisexual	1.88%
	Transgender	0.56%
Idaho	Gay/Lesbian	1.43%
	Bisexual	1.76%

	Transgender	0.44%
Illinois	Gay/Lesbian	1.76%
	Bisexual	1.99%
	Transgender	0.45%
Indiana	Gay/Lesbian	1.49%
	Bisexual	1.85%
	Transgender	0.46%
Iowa	Gay/Lesbian	1.39%
	Bisexual	1.66%
	Transgender	0.42%
Kansas	Gay/Lesbian	1.47%
	Bisexual	1.71%
	Transgender	0.43%
Kentucky	Gay/Lesbian	1.50%
	Bisexual	1.88%
	Transgender	0.48%
Louisiana	Gay/Lesbian	1.67%
	Bisexual	2.09%
	Transgender	0.52%
Maine	Gay/Lesbian	1.73%
	Bisexual	1.89%
	Transgender	0.41%
Maryland	Gay/Lesbian	1.78%
	Bisexual	1.94%
	Transgender	0.44%
Massachusetts	Gay/Lesbian	2.12%
	Bisexual	2.13%
	Transgender	0.42%
Michigan	Gay/Lesbian	1.73%
	Bisexual	1.99%
	Transgender	0.45%
Minnesota	Gay/Lesbian	1.60%
	Bisexual	1.76%
	Transgender	0.41%
Mississippi	Gay/Lesbian	1.53%
	Bisexual	1.97%
	Transgender	0.55%
Missouri	Gay/Lesbian	1.57%
	Bisexual	1.90%
	Transgender	0.48%
Montana	Gay/Lesbian	1.60%
	Bisexual	1.77%

	Transgender	0.45%
Nebraska	Gay/Lesbian	1.33%
	Bisexual	1.65%
	Transgender	0.42%
Nevada	Gay/Lesbian	1.94%
	Bisexual	2.12%
	Transgender	0.49%
New Hampshire	Gay/Lesbian	1.68%
	Bisexual	1.82%
	Transgender	0.39%
New Jersey	Gay/Lesbian	1.84%
	Bisexual	2.01%
	Transgender	0.46%
New Mexico	Gay/Lesbian	1.72%
	Bisexual	1.86%
	Transgender	0.52%
New York	Gay/Lesbian	2.10%
	Bisexual	2.22%
	Transgender	0.45%
North Carolina	Gay/Lesbian	1.81%
	Bisexual	2.04%
	Transgender	0.51%
North Dakota	Gay/Lesbian	1.37%
	Bisexual	1.62%
	Transgender	0.41%
Ohio	Gay/Lesbian	1.52%
	Bisexual	1.88%
	Transgender	0.46%
Oklahoma	Gay/Lesbian	1.39%
	Bisexual	1.77%
	Transgender	0.49%
Oregon	Gay/Lesbian	1.95%
	Bisexual	2.10%
	Transgender	0.42%
Pennsylvania	Gay/Lesbian	1.75%
	Bisexual	2.00%
	Transgender	0.45%
Rhode Island	Gay/Lesbian	1.89%
	Bisexual	2.06%
	Transgender	0.42%
South Carolina	Gay/Lesbian	1.59%
	Bisexual	1.94%

	Transgender	0.51%
South Dakota	Gay/Lesbian	1.37%
	Bisexual	1.67%
	Transgender	0.43%
Tennessee	Gay/Lesbian	1.63%
	Bisexual	1.99%
	Transgender	0.49%
Texas	Gay/Lesbian	1.70%
	Bisexual	1.86%
	Transgender	0.47%
Utah	Gay/Lesbian	1.21%
	Bisexual	1.57%
	Transgender	0.46%
Vermont	Gay/Lesbian	1.90%
	Bisexual	1.92%
	Transgender	0.39%
Virginia	Gay/Lesbian	1.81%
	Bisexual	1.93%
	Transgender	0.45%
Washington	Gay/Lesbian	1.82%
	Bisexual	1.93%
	Transgender	0.42%
West Virginia	Gay/Lesbian	1.64%
	Bisexual	1.92%
	Transgender	0.48%
Wisconsin	Gay/Lesbian	1.59%
	Bisexual	1.89%
	Transgender	0.43%
Wyoming	Gay/Lesbian	1.47%
	Bisexual	1.63%
	Transgender	0.43%
Virgin Islands	Gay/Lesbian	2.04%
	Bisexual	2.26%
	Transgender	0.64%