Identifying Sexual Orientation Health Disparities in Adolescents: Analysis of Pooled Data From the Youth Risk Behavior Survey, 2005 and 2007

We studied sexual orientation disparities in health outcomes among US adolescents by pooling multiple Youth Risk Behavior Survey (YRBS) data sets from 2005 and 2007 for 14 jurisdictions. Here we describe the methodology for pooling and analyzing these data sets.

Sexual orientation–related items assessed sexual orientation identity, gender of sexual contacts, sexual attractions, and harassment regarding sexual orientation. Wording of items varied across jurisdictions, so we created parallel variables and composite sexual minority variables.

We used a variety of statistical approaches to address issues with the analysis of pooled data and to meet the aims of individual articles, which focused on a range of health outcomes and behaviors related to cancer, substance use, sexual health, mental health, violence, and injury. (*Am J Public Health.* 2014;104:211–217. doi:10.2105/AJPH.2013.301748)

Brian Mustanski, PhD, Aimee Van Wagenen, PhD, Michelle Birkett, PhD, Sandra Eyster, PhD, and Heather L. Corliss, PhD

RESEARCH ON THE HEALTH OF lesbian, gay, and bisexual (LGB) youths has primarily come from nonprobability samples.[1,2] Such studies have been crucial for identifying health issues, their developmental course, and risk and protective factors, but their designs are less suited to describing health disparities. Their primary limitation is the inability to ensure that the LGB and heterosexual youths are drawn from the same or even comparable populations. When sexual orientation questions are included, probability-based sampling approaches can ameliorate this problem because individuals are sampled from a known population (e.g., students in schools). However, until recently very few large federal and state health surveillance surveys included sexual orientation items.[1]

Even when sexual orientation items are included in population health studies, the low prevalence of LGB identities and same-sex sexual behaviors often leads to too few individuals represented in the cells of interest. Small numbers of LGB individuals prevent analysis of sexual orientation subgroups (e.g., lesbian–gay vs bisexual) or comparisons of effects across other key social characteristics such as age, race, and gender. This is problematic because evidence shows heterogeneity in the health of LGB subgroups. For example, a review of multiple school-based samples found bisexuals to have higher risk for suicidality than heterosexuals, but results were mixed for gay and lesbian youths.[3] Very few studies have looked at the intersections of sexual orientation and other sociodemographic characteristics, such as race.[1]

When large health surveys measure sexual orientation, they frequently use a single item that assesses either sexual orientation identity or the gender of past sexual partners.[4] Such single items fail to capture the multiple dimensions of sexual orientation—including attractions, behaviors, and identity—that may not align with one another, particularly among youths.[4–6] The relationship between these dimensions and various health outcomes may also differ. For example, one study found that LGB sexual orientation identity was associated with increased mood and anxiety disorders, but that women reporting only same-sex partners had the lowest rates of most disorders.[3] Therefore, population-based studies that assess more than 1 component of sexual orientation are at a considerable advantage in understanding its relationship with health outcomes. The set of articles in this special issue extend the literature by focusing on sexual orientation disparities in several health domains through analysis of data from population-based samples.

DATA SOURCE

Data came from the Youth Risk Behavior Surveillance System, operated by the Centers for Disease Control and Prevention (CDC). Since 1990, the system has monitored health-related behaviors and outcomes that contribute to the leading causes of death, disability, and social problems among youths and adults.[7] A component is the Youth Risk Behavior Survey (YRBS), which is conducted every other year both nationally and jurisdictionally in 47 states and more than 20 territorial, tribal, and local regions. Jurisdictional surveys are administered by departments of health and education with assistance from the CDC. Because the national YRBS did not include questions on sexual orientation, our pooled project used data from the jurisdictional surveys that elected to include questions about sexual orientation.

Each state and local school-based YRBS uses a 2-stage, cluster sampling design to produce representative samples of students in grades 9 to 12 in its jurisdiction.[7] All but a few jurisdictions survey only public schools, and each local sample incorporates only schools in the funded school district. In most jurisdictions, in the first sampling stage, schools are selected with probability proportional to school enrollment size. In the second sampling stage, intact classes of a required subject or intact classes during a required period are selected randomly. All students in sampled classes are eligible to participate. Response

rates vary across jurisdiction. For example, in 2003, school response rates ranged from 67% to 100%, student response rates ranged from 60% to 94%, and overall response rates ranged from 60% to 90%.[7] The CDC provides jurisdictions with the questionnaire for a particular year, and sites may modify the questionnaire. They can add and remove questions, but two thirds of the questions from the CDC-provided YRBS questionnaire must remain unchanged.[7] More information about the YRBS procedures can be found in Brener et al.[7]

Beginning in October of 2008, we requested YRBS data files from jurisdictions that we knew or suspected included sexual orientation–related measures. We did not discover a centralized, publicly accessible, complete source of information about the inclusion of sexual orientation questions in the YRBS by jurisdiction and year. Therefore, we consulted several sources to determine which data files to request: the Web site http://www.LGBTData.com, a PubMed literature review, and a list of which YRBS sites asked about sexual orientation provided by a health scientist at the CDC's Division of Adolescent and School Health (L. Whittle, personal communication, 2009). For access to data from some jurisdictions, the CDC directed us to contact state YRBS coordinators, who reviewed our data request and provided approval for the CDC to distribute data files to us. We concluded data requests on December 31, 2009. Washington, DC, was the only jurisdiction from which we requested data that did not grant access before the start of 2010. One inherent limitation of the data obtained is that it came from more progressive and urban jurisdictions; thus these data may not be representative of adolescents living in rural or less progressive regions.

CODING OF SEXUAL ORIENTATION ITEMS

Sexual orientation–related measures included in the YRBS varied across jurisdictions and year of data collection. Measures fell into 4 broad categories: gender of sexual contacts (opposite sex, same sex, both sexes, none), sexual orientation identity (heterosexual, lesbian–gay, bisexual, unsure), gender of sexual attraction (opposite sex, same sex, both sexes, no one), and sexual orientation–related harassment. Table 1 outlines jurisdictions that included sexual orientation–related measures, by category.

Sexual Orientation Identity

Nine of 14 jurisdictions included a question about sexual orientation identity. In the most common variation (1a), students were asked, "Which of the following best describes you?" Response options were "Heterosexual (straight)," "Gay or lesbian," "Bisexual," and "Not sure." Variation 1b was identical to 1a except that the second response option read, "Homosexual (gay or lesbian)." Variation 1c was identical to 1a except that "None of the above" was an additional response option. We pooled these variations into a single sexual orientation identity variable, with respondents who checked "None of the above" in variation 1c set to missing. Table 2 provides the unweighted number and weighted proportion (with 95% confidence intervals [CIs]) of participants in each jurisdiction, collapsed across years of administration, who endorsed each sexual orientation identity. Of participants administered the sexual orientation identity item, 3.2% did not respond.

Sexual Contacts and Attraction

Twelve of 14 jurisdictions included a question about the gender of sexual contacts. In the most common variation (1a), students were asked, "During your life, with whom have you had sexual contact?" Response options were "I have never had sexual contact," "Females," "Males," and "Females and males." Variation 1b asked, "During your life, the person(s) with whom you have had sexual contact is (are) . . ."; options were "I have not had sexual contact with anyone," "Female(s)," "Male (s)," and "Female(s) and male(s)." Variation 1c offered the same response options in the singular, but asked, "The person(s) with whom you have had sexual contact during your life is (are): . . ." In several jurisdictions, the survey asked students about sexual intercourse rather than sexual contact. This variation (2) asked, "During your life, with whom have you had sexual intercourse?" Options were "I have never had sexual intercourse," "Females," "Males," "Females and males." We used these questions and the gender of respondents to create the following categories: no reported sexual partners, exclusively same-sex sexual partners, same- and opposite-sex sexual partners, and exclusively opposite-sex partners. Table 3 provides the unweighted number and weighted proportion (with 95% CIs) of participants in each jurisdiction, collapsed across years of administration, who reported each pattern of no, same-sex, both-sex, and opposite-sex sexual partners. Of participants administered the gender of sexual contacts item, 3.9% did not respond.

Attraction was measured in only 1 jurisdiction (Hawaii), with the item (variation 1), "Who are you sexually attracted to?" Response options were "Males," "Females," "Both males and females," and "I am not sexually attracted to anyone." We used these questions and the gender of respondents to create the following categories: attracted to opposite sex, attracted to same sex, attracted to both sexes, not attracted to anyone. Weighted proportions, collapsed across years of administration, showed that most respondents indicated being attracted to the opposite sex (84.7%; 95% CI = 82.6%, 86.5%), followed by those who indicated not being sexually attracted to anyone (6.6%; 95% CI = 5.3%, 8.2%), those attracted to the same sex (4.1%; 95% CI = 3.1%, 5.4%), and those with bisexual attractions (4.6%; 95% CI = 3.9%, 5.5%). Of participants administered the gender of sexual attraction item, 1.8% did not respond.

Sexual Orientation–Related Harassment

Five of 14 jurisdictions included a question about harassment arising from perceived sexual orientation. The most common variation (1a) asked students, "During the past 12 months, how many times have you been harassed because someone thought you were gay, lesbian, or bisexual?" Response options were zero, 1, 2 or 3, 4 or 5, 6 or 7, 8 or 9, 10 or 11, and 12 or more times. In variation 1b, a jurisdiction asked the exact same question, but included "on school property" in the question.

Variation 2 asked, "During the past 30 days, have you been the victim of a verbal slur because of your gender or sexual orientation?" Response options were yes

TABLE 1—Sexual Orientation-related Measures by Year and Jurisdiction: Youth Risk Behavior Survey, United States, 2005 and 2007

Jurisdiction	Sexual Orientation Identity, Variation[b]	Gender of Sexual Contacts, Variation[a]	Gender of Sexual Attraction, Variation	Sexual Orientation Harassment, Variation[c]
2005				
Boston, MA	1a	1b		
Chicago, IL	1a	1a		
Connecticut				1b
Delaware	1c	2		
Hawaii			1	1a
Maine		1b		3
Massachusetts	1a	1b		
New York City, NY	1a	1a		
San Diego, CA		2		1a
San Francisco, CA	1a			2
Vermont	1a	2		
2007				
Boston, MA	1a	1a		
Chicago, IL	1a	1a		1a
Connecticut		1a		1b
Delaware	1b	2		
Hawaii			1	1a
Maine	1a	1a		3
Massachusetts	1a	1a		
New York City, NY	1a	1a		
Rhode Island	1a	1a		
San Diego, CA				1a
San Francisco, CA	1a			4
Vermont	1a	2		
Wisconsin		1a		
Milwaukee, WI		1a		

Note. Only Hawaii asked for gender of persons respondents were sexually attracted to.
[a]All variations asked for gender of lifetime sexual contacts, with slightly different wording in questions and responses.
[b]Response options for all variations were heterosexual or straight; homosexual, gay, or lesbian; bisexual; and not sure, with slight variations in wording; 1c also had none of the above.
[c]Variation 1a asked about harassment because of sexual orientation in the past 12 months, with number responses; 1b asked about harassment on school property. Variation 2 asked about gender or sexual orientation verbal slurs in the past 30 days; variation 3 asked about offensive comments or attacks because of perceived sexual orientation at school or traveling to or from school; both had yes or no responses. Variation 4 asked how many times respondents heard harassing statements in the past 12 months.

and no. Variation 3 asked, "Has anyone ever made offensive comments or attacked you because of your perceived sexual orientation at school or on their way to or from school?" Response options were yes and no. Variation 4 asked, "During the past 12 months, how many times did you hear other students make harassing statements based on sexual orientation, such as faggot, dyke, and 'that's so gay'?" and offered the same response options as variation 1a. Because wording varied so significantly, we did not construct a common pooled variable.

CODING OF RACE/ETHNICITY ITEMS

The measurement of race/ethnicity varied across years and by jurisdiction, particularly in the response options for Asian participants. In 2005, the survey measured race/ethnicity with a single item that asked students to check all options that applied. In 9 of 14 jurisdictions, the questionnaire asked, "How do you describe yourself?" Response options for all jurisdictions were American Indian or Alaska Native, Asian, Black or African American, Hispanic or Latino, Native Hawaiian/other Pacific Islander, and White. The CDC provides a standard recode of these items that creates 2 additional multiracial categories: multiple Hispanic and multiple non-Hispanic.[8] In 2007, the survey measured race/ethnicity with 2 items: "Are you Hispanic or Latino?" and "What is your race? (Select one or more responses.)" In 9 jurisdictions, the options were American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or other Pacific Islander, and White. The CDC provides a standard recode of the items to create 8 racial/ethnic categories that correspond to the 2005 categories.[9]

Across years, 5 jurisdictions included other Asian, Native Hawaiian/other Pacific Islander, and non-White options. We recoded the following to Asian: Southeast Asian American (e.g., Cambodian, Vietnamese, Laotian, Thai), Asian American (e.g., Chinese, Japanese, Korean, East Indian), Filipino, Indochinese, other Asian, Chinese, Japanese, and other Asian or Pacific Islander. One jurisdiction offered the response option other–non-White, which we recoded to multiple non-Hispanic. One jurisdiction included the option Native Hawaiian/part Hawaiian, which we recoded to Native Hawaiian/other Pacific Islander.

MEASUREMENT OF SEXUAL ORIENTATION

Several challenges arose in defining sexual orientation. First, surveys administered up to 3 separate measures. These dimensions were not conceptually equivalent and did not necessarily align; for instance, some heterosexually identified respondents reported same-sex sexual partners. The overlap of these dimensions across

TABLE 2—Sexual Orientation Identity by Jurisdiction: Youth Risk Behavior Survey, United States, 2005 and 2007

Jurisdiction	Heterosexual		Gay or Lesbian		Bisexual		Unsure		Total, No.	Sexual Majority, No. (%)	Sexual Minority,[a] No. (%)
	No.	% (95% CI)	No.	% (95% CI)	No.	% (95% CI)	No.	% (95% CI)			
Boston, MA	3051	94.2 (93.1, 95.2)	38	1.1 (0.8, 1.6)	91	2.9 (2.2, 3.8)	58	1.8 (1.3, 2.4)	3238	3264 (91.5)	297 (8.5)
Chicago, IL	1697	90.9 (89.1, 92.4)	45	2.3 (1.5, 3.5)	66	3.6 (2.6, 4.8)	57	3.3 (2.2, 4.8)	1865	1852 (89.4)	208 (10.6)
Connecticut	…	…	…	…	…	…	…	…	…	4170 (96.2)	158 (3.8)
Delaware	4890	94.2 (93.4, 94.9)	56	1.1 (0.8, 1.5)	189	3.4 (2.9, 4.0)	67	1.3 (1.0, 1.7)	5202	4983 (93.5)	361 (6.5)
Hawaii	…	…	…	…	…	…	…	…	…	2622 (91.5)	231 (8.5)
Maine	1241	94.5 (92.5, 96.0)	11	0.8 (0.4, 1.5)	36	3.0 (1.9, 4.5)	32	1.7 (1.2, 2.5)	1320	2479 (92.5)	220 (7.5)
Massachusetts	6095	93.8 (93.1, 94.4)	89	1.3 (1.0, 1.6)	225	3.2 (2.8, 3.8)	117	1.7 (1.3, 2.1)	6526	6036 (91.0)	617 (9.0)
Milwaukee, WI	…	…	…	…	…	…	…	…	…	1721 (91.6)	171 (8.4)
New York City, NY	15 117	92.0 (91.3, 92.8)	222	1.1 (0.9, 1.4)	648	3.7 (3.2, 4.1)	459	3.1 (2.7, 3.7)	16 446	15 470 (89.9)	1750 (10.1)
Rhode Island	1954	90.1 (88.4, 91.5)	47	1.9 (1.3, 2.7)	123	5.4 (4.2, 6.8)	55	2.7 (1.9, 3.8)	2179	1907 (86.6)	303 (13.4)
San Diego, CA	…	…	…	…	…	…	…	…	…	3165 (97.8)	72 (2.2)
San Francisco, CA	4357	89.8 (88.6, 90.9)	80	1.5 (1.1, 2.0)	176	3.7 (3.2, 4.4)	229	4.9 (4.1, 5.9)	4842	4521 (90.2)	485 (9.8)
Vermont	16 293	93.3 (92.7, 93.9)	185	0.9 (0.7, 1.2)	584	3.1 (2.8, 3.5)	516	2.7 (2.3, 3.0)	17 578	16 379 (92.7)	1416 (7.3)
Wisconsin	…	…	…	…	…	…	…	…	…	1983 (94.8)	111 (5.2)

Note. CI = confidence interval. Percentages are weighted and adjusted; samples sizes are unweighted.
[a]Weighted proportion of individuals who had same-sex contacts; identified as gay, lesbian, bisexual, or unsure; or reported same-sex attractions.

demographic groups (age, gender, race) is the focus of an article in this special issue.[10] Furthermore, measures of sexual orientation varied by jurisdiction and by year. Consequently, defining sexual orientation by using a single dimension of identity, behavior, or attraction required limiting analysis to only jurisdictions that included that measure. Specifically, only Hawaii asked for gender of sexual attraction, and surveys for 96.3% of the pooled sample did not include this question. Surveys administered to 20.5% of the pooled sample did not ask about sexual orientation identity, and 15.3% of the sample did not receive a measure of gender of sexual contacts. One conceptually appealing approach to measurement of sexual orientation would be to construct a joint multidimensional measure of sexual orientation identity and gender of sexual contacts from both of these variables. However, this approach would also result in a large degree of missingness (27.0% of the pooled sample).

Clearly, any definition of sexual orientation would have certain weaknesses. Therefore, after much deliberation, study authors across the entire special issue decided to define sexual orientation in different ways, depending on the focus of the specific article. First, because our goal was to give a broad epidemiological view of the health of sexual minority adolescents in the United States, we created a pooled sexual orientation variable that allowed for inclusion of as many jurisdictions as possible in the analysis sample. We created a binary sexual minority status variable to identify individuals who endorsed a minority sexual orientation on any of the 3 dimensions measured (gay or lesbian, bisexual, or unsure sexual orientation identity; exclusive same sex or same and opposite sex in gender of sexual contacts; and attracted to same sex or attracted to both sexes in gender of sexual attraction). Although we acknowledge that constructing the sexual minority status variable out of 3 measures obscures important variability across dimensions of sexual orientation, combining measures allowed us to use the entire analytic sample to report data on the overall population of sexual minority adolescents. Because of the difficulties in interpreting this variable however, we used sexual minority status only as a first step in analysis to provide an overview of disparities in broad strokes. Some authors began by reporting findings by sexual minority status and next analyzed 1 dimension of sexual orientation (i.e., sexual orientation identity or gender of sexual contacts) or 2 dimensions (i.e., sexual orientation identity and gender of sexual contacts), depending on the outcome under study and the most relevant dimensions of sexual orientation.

It is important to note that the use of gender of sexual contacts as an indicator of sexual orientation is complicated, particularly in an adolescent sample, because it limits the measure of sexual orientation to only students who have had sexual contact. Developmental differences in abstinence, as well as differences by gender (adolescent girls report more abstinence), race/ethnicity (adolescent Whites and Asians report more abstinence), and sexual orientation identity (adolescents who identify as heterosexual report more abstinence) mean that high school students reporting same-sex behavior likely differ in important ways from high school students who report same-sex attraction or a gay or lesbian identity. Researchers must

TABLE 3—Gender of Sexual Partners by Jurisdiction: Youth Risk Behavior Survey, United States, 2005 and 2007

Jurisdiction	No Reported Sexual Partners		Exclusively Same-Sex Partners		Same- and Opposite-Sex Partners		Exclusively Opposite-Sex Partners				
	No.	% (95% CI)	No.	% (95% CI)	No.	% (95% CI)	No.	% (95% CI)	Total, No.	Sexual Majority, No. (%)	Sexual Minority,[a] No. (%)
Boston, MA	1202	35.4 (33.0, 37.9)	92	3.0 (2.3, 3.9)	89	3.0 (2.4, 3.7)	1760	58.6 (56.1, 61.1)	3143	3264 (91.5)	297 (8.5)
Chicago, IL	656	35.7 (32.4, 39.1)	52	2.8 (2.1, 3.7)	58	3.2 (2.1, 4.9)	1099	58.4 (54.7, 62.0)	1865	1852 (89.4)	208 (10.6)
Connecticut	823	39.4 (35.9, 43.0)	48	2.6 (2.0, 3.4)	110	5.2 (4.1, 6.5)	1031	52.8 (49.8, 55.8)	2012	4170 (96.2)	158 (3.8)
Delaware	2210	43.2 (40.9, 45.6)	45	0.9 (0.7, 1.2)	149	2.6 (2.1, 3.2)	2698	53.3 (51.0, 55.6)	5102	4983 (93.5)	361 (6.5)
Hawaii	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	–	2622 (91.5)	231 (8.5)
Maine	1163	41.7 (38.5, 44.9)	74	2.6 (2.0, 3.4)	101	3.4 (2.8, 4.2)	1326	52.3 (49.2, 55.5)	2664	2479 (92.5)	220 (7.5)
Massachusetts	2609	40.1 (37.9, 42.2)	142	2.3 (1.9, 2.7)	255	3.9 (3.4, 4.5)	3361	53.8 (51.7, 55.9)	6367	6036 (91.0)	617 (9.0)
Milwaukee, WI	604	32.0 (29.0, 35.2)	69	3.8 (2.8, 5.1)	102	4.8 (3.7, 6.1)	1077	59.4 (56.4, 62.4)	1852	1721 (91.6)	171 (8.4)
New York City, NY	6555	43.2 (40.6, 45.7)	394	2.2 (1.8, 2.6)	631	3.8 (3.1, 4.6)	8761	50.9 (48.7, 53.0)	16 341	15 470 (89.9)	1750 (10.1)
Rhode Island	825	36.7 (33.6, 39.8)	72	3.1 (2.4, 4.0)	115	5.3 (4.1, 6.8)	1162	55.0 (51.7, 58.2)	2174	1907 (86.6)	303 (13.4)
San Diego, CA	948	58.3 (54.4, 62.2)	32	2.1 (1.4, 3.2)	40	2.4 (1.7, 3.3)	624	37.1 (33.4, 41.0)	1644	3165 (97.8)	72 (2.2)
San Francisco, CA	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	4521 (90.2)	485 (9.8)
Vermont	11 322	63.2 (60.5, 65.8)	123	0.7 (0.5, 0.9)	380	1.9 (1.7, 2.1)	5698	34.2 (31.8, 36.8)	17 523	16 379 (92.7)	1416 (7.3)
Wisconsin	787	37.6 (34.1, 41.2)	36	1.8 (1.3, 2.5)	75	3.5 (2.7, 4.4)	1181	57.1 (53.8, 60.3)	2079	1983 (94.8)	111 (5.2)

Note. CI = confidence interval. Percentages are weighted and adjusted; samples sizes are unweighted.
[a]Weighted proportion of individuals who had same-sex contacts; identified as gay, lesbian, bisexual, or unsure; or reported same-sex attractions.

fully think through these issues when deciding on the use of a particular measure of sexual orientation.

COMPLEX SAMPLING DESIGN AND DATA POOLING

Analysis of the YRBS data sets involved pooling observations across samples that varied across jurisdiction (city and state) as well as within the same jurisdiction across time (2005 and 2007). Pooling cross-sectional data has limitations. Caution is required in generalizing to a specific population (i.e., estimating the percentage of youths with a specific characteristic), because there is no longer a clear population of interest. Another consideration is that it is possible that the same individual responded to surveys in both years, and therefore the observations were not fully independent. However, previous studies have pooled data across years of YRBS,[11,12] and evidence suggests that this does not compromise validity, because rates of repeat participation across years are low.[13,14] Another important consideration is that the jurisdictions that asked questions about sexual orientation may not have been representative of the entire United States. Therefore, the primary methodological advantage of this approach is that the data came from large probability samples rather than convenience samples.

Descriptive and Inferential Statistics

For analyses of the prevalence of demographic characteristics, health behaviors, and health outcomes, we followed instructions provided by the CDC for analysis of YRBS data.[15] These instructions indicate that a variety of specialized (e.g., SUDAAN version 11.0, Research Triangle Institute, Research Triangle Park, NC) and general (e.g., SPSS version 21, SPSS, Inc, Chicago, IL) statistical packages with procedures for analyzing complex survey data produce the same point estimates and inconsequential differences in SEs and CIs when properly implemented to use the stratum and primary sampling unit. Stratum and primary sampling unit identify the sampling stratum and selected units within the stratum, respectively, and are used in complex sample variance estimation methods such as Taylor series linear approximation. These consistent findings across software packages allowed authors to estimate their descriptive statistics with any of the software options and associated procedures described in the CDC instructions.

The analytic approach for incorporating the complex sampling design into the estimation of inferential statistics was more complicated because several articles aimed to address questions that required statistical approaches that could not be implemented in the software or with the complex sampling data analytic procedures described by the CDC.[15] For example, one article sought to explore how school climate influenced suicidality, and therefore analyses needed to account for the nested nature of school climate measures within jurisdictions, which was accomplished by using the hierarchical linear modeling (HLM) software version 7,[16] which only allows for a single weighting variable.[17]

Failing to take into consideration design effects in these models would result in incorrectly computed SEs (usually involving underestimating the magnitude of the SE). Therefore, to estimate

complex models and inferential statistics, we considered 2 alternative approaches[18]: (1) computing SEs under the assumption of simple random sampling, followed by a postcomputation adjustment of each standard error to reflect the design effect, and (2) incorporating the design effect into the analytic weights, followed by simple random sampling SE estimation to yield complex sample design–adjusted SEs. We used the latter approach here.

Design effects are a measure of the ratio of the SE that accounts for the complex sample to an SE computed under the assumption of simple random sampling. Within each sample, we computed the design effects for each variable by computing the SE of an estimate (1) assuming a simple random sample and (2) correctly accounting for the complex design through Taylor series linearization estimation. The ratio of these 2 SEs is the design effect. Averaging this value across all variables yields an average root design effect. We calculated all design effects within each sample. For example, we computed design adjustments and applied them separately for Boston, Massachusetts, 2005; Boston 2007; Chicago, Illinois, 2005; Chicago 2007; and so on. The average design effects ranged from 1.08 (New York City, 2005) to 2.69 (Vermont, 2005).

Once we determined the average design effect for each sample, we normalized the analytic weights provided by the CDC for each observation by multiplying the weight by the average sampling fraction (n/N); we then divided the normalized weight by the squared design effect, which altered the effective sample size by adjusting the weight downward as a function of the overall design effect. In the pooled data file, we could use this design-adjusted analytic weight to compute design-adjusted SE estimates as if the data reflected a simple random sample design. We then used the design-adjusted analytic weight variable when we fit models in HLM.

Statistical Models

Several investigators used HLM to estimate associations between variables and in some cases to explain differences between jurisdictions. In initial multilevel modeling, researchers included year of data collection at level 2 across outcomes, but they found that year did not explain a significant amount of variance. Therefore, in the interest of parsimony, they included only jurisdictional differences at level 2 in the reported multilevel models. For each health outcome, researchers calculated an intraclass correlation to determine the amount of variance between jurisdictions. Although several studies used HLM, most did not include explanatory variables at level 2 beyond jurisdiction.

To understand the implications of using HLM with design-adjusted analytic weights for estimating associations between variables, we compared the odds ratios and associated *P* values with those produced by SUDAAN when calculations followed CDC analysis recommendations.[15] These sensitivity analyses showed that the odds ratios and *P* values were negligibly different (differences in the hundredths for odds ratios; differences of hundredths or thousandths for *P* values, with HLM *P* values always more conservative). Therefore, depending on the aims of individual articles, to estimate associations between variables, authors either used CDC-recommended software and approaches or HLM with design-adjusted analytic weights.

CONCLUSIONS

Recently the CDC published a surveillance report on sexual orientation–related health disparities that analyzed YRBS data from 2001 to 2009.[19] We view the articles in this special issue as complementing and extending the surveillance data in that report in several ways. This issue's studies used a pooled data set that allowed for estimation of results across the represented jurisdictions. This approach enabled greater investigation of interactions between sexual minority status and other sociodemographic factors. These studies also provide in-depth discussion of a range of outcomes across physical, sexual, and mental health as well as social determinants of health, such as experiences of victimization. They make important contributions to the science of LGB youth health disparities. ■

About the Authors

Brian Mustanski and Michelle Birkett are with the Department of Medical Social Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL. Aimee Van Wagenen is with Fenway Health, Boston, MA. Sandra Eyster is with the American Institutes for Research, Washington, DC. Heather L. Corliss is with the Division of Adolescent Medicine, Boston Children's Hospital and Harvard Medical School, Boston, MA.

Correspondence should be sent to Brian Mustanski, PhD, Department of Medical Social Sciences, Northwestern University, 625 N Michigan Ave, Suite 2700, Chicago, IL 60611 (e-mail: brian@northwestern.edu). Reprints can be ordered at http://www.ajph.org by clicking the "Reprints" link.

This article was accepted October 21, 2013.

Contributors

B. Mustanski conceptualized the study and interpreted data. A. Van Wagenen coordinated data pooling. M. Birkett analyzed and interpreted data. S. Eyster and H. L. Corliss developed the analytic plan. All authors wrote and revised the article.

Acknowledgments

This project was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (grant R21HD051178) and by the IMPACT LGBT Health and Development Program, Northwestern University. Brian Mustanski was supported by a William T. Grant Scholars Award. Heather L. Corliss was supported by the National Institute on Drug Abuse (career development award DA23610).

Assistance from the Centers for Disease Control and Prevention (CDC) Division of Adolescent and School Health and the work of the state and local health and education departments who conduct the Youth Risk Behavior Surveillance System made the project possible. We thank Michael Cohen of the American Institutes for Research for his consultation on complex sampling analyses.

Note. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health, the CDC, the William T. Grant Foundation, or any of the agencies involved in collecting the data.

Human Participant Protection

Protocol approval was not necessary because we obtained de-identified data from secondary sources. We obtained data use agreements from the Vermont and Rhode Island departments of health, the only 2 that required these agreements for access to the data.

References

1. Committee on Lesbian Gay Bisexual and Transgender Health Issues and Research Gaps and Opportunities, Institute of Medicine. *The Health of Lesbian, Gay, Bisexual, and Transgender People: Building a Foundation for Better Understanding.* Washington, DC: National Academies Press; 2011.

2. Mustanski BS, Newcomb ME, Du Bois SN, Garcia SC, Grov C. HIV in young men who have sex with men: a review of epidemiology, risk and protective factors, and interventions. *J Sex Res.* 2011;48(2–3):218–253.

3. Saewyc EM, Skay CL, Hynds P, et al. Suicidal ideation and attempts in North American school-based surveys: are bisexual youth at increasing risk? *J LGBT Health Res.* 2007;3(2):25–36.

4. Saewyc EM, Bauer GR, Skay CL, et al. Measuring sexual orientation in

adolescent health surveys: evaluation of eight school-based surveys. *J Adolesc Health.* 2004;35(4):345.e1–e15.

5. Savin-Williams RC, Ream GL. Prevalence and stability of sexual orientation components during adolescence and young adulthood. *Arch Sex Behav.* 2007;36(3):385–394.

6. Mustanski B, Kuper L, Greene GJ. Development of sexual orientation and identity. In: Tolman DL, Diamond LM, eds. *Handbook of Sexuality and Psychology.* Washington, DC: American Psychological Association; 2013:597–628.

7. Brener ND, Kann L, Kinchen SA, et al. Methodology of the Youth Risk Behavior Surveillance System. *MMWR Recomm Rep.* 2004;53(RR-12):1–13.

8. Centers for Disease Control and Prevention. 2005 National YRBS data users manual. 2005. Available at: ftp://ftp.cdc.gov/pub/data/yrbs/2005/2005NationalYRBSDataUsersManual.pdf. Accessed March 29, 2012.

9. Centers for Disease Control and Prevention. 2007 National YRBS data users manual. 2007. Available at: ftp://ftp.cdc.gov/pub/data/yrbs/2007/2007_National_YRBS_Data_Users_Manual.pdf. Accessed March 29, 2012.

10. Mustanski B, Birkett M, Greene GJ, Rosario M, Bostwick W, Everett BG. The association between sexual orientation identity and behavior across race/ethnicity, sex, and age in a probability sample of high school students. *Am J Public Health.* 2014;104(2):237–244.

11. Kann L, Brener ND, Warren CW, Collins JL, Giovino GA. An assessment of the effect of data collection setting on the prevalence of health risk behaviors among adolescents. *J Adolesc Health.* 2002;31(4):327–335.

12. Goodenow C, Szalacha LA, Robin LE, Westheimer K. Dimensions of sexual orientation and HIV-related risk among adolescent females: evidence from a statewide survey. *Am J Public Health.* 2008;98(6):1051–1058.

13. Corliss HL, Goodenow CS, Nichols L, Austin SB. High burden of homelessness among sexual-minority adolescents: findings from a representative Massachusetts high school sample. *Am J Public Health.* 2011;101(9):1683–1689.

14. Goodenow C, Netherland J, Szalacha L. AIDS-related risk among adolescent males who have sex with males, females, or both: evidence from a statewide survey. *Am J Public Health.* 2002;92(2):203–210.

15. Centers for Disease Control and Prevention. *Software for Analysis of YRBS Data.* Atlanta, GA: US Dept of Health and Human Services; 2009.

16. Raudenbush SW, Bryk AS, Cheong YF, Congdon RT, du Toit M. *HLM 7: Hierarchical Linear and Nonlinear Modeling.* Lincolnwood, IL: Scientific Software International Inc; 2011.

17. Hatzenbuehler ML, Birkett M, VanWagenen A, Meyer IH. Protective school climates and reduced risk for suicide ideation in sexual minority youths. *Am J Public Health.* 2014;104(2):279–286.

18. Thomas SL, Heck RH. Analysis of large-scale secondary data in higher education research: potential perils associated with complex sampling designs. *Res Higher Educ.* 2001;42(5):517–540.

19. Kann L, Olsen EO, McManus T, et al. Sexual identity, sex of sexual contacts, and health-risk behaviors among students in grades 9–12—youth risk behavior surveillance, selected sites, United States, 2001–2009. *MMWR Surveill Summ.* 2011;60(7):1–133.