Population Size Estimates for Men who Have Sex with Men and Persons who Inject Drugs

Alexandra M. Oster, Maya Sternberg, Amy Lansky, Dita Broz, Cyprian Wejnert, and Gabriela Paz-Bailey

ABSTRACT *Understanding geographic variation in the numbers of men who have sex with men (MSM) and persons who inject drugs (PWID) is critical to targeting and scaling up HIV prevention programs, but population size estimates are not available at generalizable sub-national levels. We analyzed 1999–2010 National Health and Nutrition Examination Survey data on persons aged 18–59 years. We estimated weighted prevalence of recent (past 12 month) male-male sex and injection drug use by urbanicity (the degree to which a geographic area is urban) and US census region and calculated population sizes. Large metro areas (population $\geq 1,000,000$) had higher prevalence of male-male sex (central areas, 4.4 % of men; fringe areas, 2.5 %) compared with medium/small metro areas (1.4 %) and nonmetro areas (1.1 %). Injection drug use did not vary by urbanicity and neither varied by census region. Three-quarters of MSM, but only half of PWID, resided in large metro areas. Two-thirds of MSM and two-thirds of PWID resided in the South and West. Efforts to reach MSM would benefit from being focused in large metro areas, while efforts to reach PWID should be delivered more broadly. These data allow for more effective allocation of funds for prevention programs.*

KEYWORDS *HIV infection, Men who have sex with men, Persons who inject drugs, Risk factors, Gay and bisexual men*

INTRODUCTION

Male-male sexual contact and injection drug use are common modes of HIV transmission in the USA.¹ Having accurate estimates of the size of these populations not only nationally but also in different geographic areas can aid in appropriate allocation of funds and determination of the reach of existing and future interventions, such as HIV testing efforts. Moreover, although HIV surveillance data provide estimates of the numbers of new cases falling into each of these categories (i.e., the numerator), calculation of disease rates is dependent on having estimates of the number of people in each of these populations (i.e., the denominator).

Oster, Lansky, Broz, Wejnert, and Paz-Bailey are with the Division of HIV/AIDS Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP), Centers for Disease Control and Prevention, Atlanta, GA, USA; Sternberg is with the Division of Laboratory Science, National Center for Environmental Health, Centers for Disease Control and Prevention, Atlanta, GA, USA.

Correspondence: Alexandra M. Oster, Division of HIV/AIDS Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention (NCHHSTP), Centers for Disease Control and Prevention, Atlanta, GA, USA. (E-mail: aoster@cdc.gov)

CDC recently published estimates of the population size of gay, bisexual, and other men who have sex with men (collectively referred to as MSM) in the USA². Although these estimates allow calculation of disease rates for HIV, syphilis, and other infections at the national level, extrapolating this estimate to different geographic areas is problematic, as the proportion of the population who are MSM may vary by smaller geographic unit. In 1994, Laumann et al. estimated the percentage of men identifying as gay or bisexual by place of residence (top 12 central cities, next 88 central cities, suburbs for each of the above groups, other urban areas, rural areas), based on data from 1992.³ To our knowledge, no other estimates of the prevalence of male-male sex by urbanicity (the degree to which a geographic area is urban) are publicly available. Lieb et al. estimated the percentage of men who have ever had sex with another man, by state and region.^{4,5} However, they extrapolated these estimates in part from Laumann's data on the percentage of persons who identify as gay or bisexual (rather than estimates of male-male sexual behavior) and did not present data for recent (e.g., past 12 month) sexual behavior.

Likewise, estimates of the size of the population of persons who inject drugs (PWID) have been published for the USA and for 92 large US metropolitan areas.⁶⁻⁸ These estimates are extremely useful for studies that are nationwide or are conducted in specific metropolitan areas, but they do not allow for aggregation of data across sub-national areas that also contain small cities or nonmetropolitan areas. We are not aware of published estimates of the overall prevalence of injection drug use by urbanicity, although estimates for prevalence of injection drug use by region and estimates for injection of specific drugs by urbanicity have been published.^{9,10}

The National Health and Nutrition Examination Study (NHANES) interviews a probability sample of the US population.¹¹ We used NHANES data to estimate the percent of the population that are MSM and PWID by urbanicity and region of the country.

MATERIALS AND METHODS

Sampling and Recruitment

NHANES is a series of cross-sectional surveys designed to assess the health and nutritional status of adults and children in the USA¹¹. Conducted periodically during 1971-1994, NHANES has been conducted continuously since 1999. It examines a nationally representative sample of the civilian, noninstitutionalized population by surveying about 5000 persons from 15 counties each year, with data released in 2-year cycles. Participants are chosen using a complex, stratified, multistage probability sampling design. The portion of the interview regarding sexual and drug use behaviors is administered in a private room using audio computer-assisted self-interview (ACASI). NHANES activities were reviewed and approved by the Research Ethics Review Board of the National Center for Health Statistics, and participants provide written informed consent for participation.

Study Populations

We first calculated the prevalence of several HIV risk behaviors. To ensure adequate sample size, we included data from 1999-2010. We included adults aged 18-59 years in the analysis and examined prevalence of the following: (1) male-male sex-ever (i.e., among men, having ever had oral or anal sex with another man), (2)

male-male sex-past 12 months (i.e., among men, having had oral or anal sex with another man during the past 12 months), (3) injection drug use-ever (i.e., ever injecting drugs not prescribed by a doctor), and (4) injection drug use-past 12 months (i.e., injecting drugs not prescribed by a doctor during the past 12 months). Prevalence estimates for male-male sex are presented with two different denominators: men only and all persons. Persons who reported both male-male sex and injection drug use were eligible to be included in both the male-male sex and injection drug use categories. The small numbers of persons reporting both of those behaviors did not permit analysis as a separate category.¹²

Urbanicity

To determine urbanicity, we used the 2006 CDC National Center for Health Statistics (NCHS) Urban-Rural Classification Scheme for Counties, which classifies all US counties and county-equivalents into one of six urbanization levels—four for metropolitan counties and two for nonmetropolitan counties (Fig. 1).¹³ Large metropolitan areas (those with population ≥ 1 million) are divided into large central metro counties, which correspond to counties that contain the urban core of a city, and large fringe metro counties, which are the more suburban counties. Due to the small sample size, there was a risk of confidentiality disclosure that necessitated collapsing medium and small metro counties into one category (“medium/small metro”) and all nonmetropolitan counties into another category (“nonmetro”). Table 1 shows the distribution of the US population by urbanicity.

Region

We also stratified by region, classifying each US state into one of four regions as designated by the US Census Bureau (available at http://www.census.gov/geo/maps-data/maps/pdfs/reference/us_regdiv.pdf): Northeast, Midwest, South, and West (Fig. 1). Table 1 shows the distribution of the US population by region.

Analysis

We calculated weighted percentages and confidence intervals overall and stratified by urbanicity and by US census region. Data were analyzed using the statistical software packages SAS (SAS Institute Inc., 2002) callable SUDAAN (SUDAAN Release 10.0, 2008). SUDAAN uses sample weights and calculates variance estimates that account for the complex survey design. All estimates were weighted to account for selection probabilities and non-response. All standard error estimates were calculated using the Taylor series (linearization) method within SUDAAN. Confidence intervals for the prevalence were calculated using a logit transformation. The degrees of freedom for variance estimation were based on subtracting the number of strata from the number of primary sampling units. *P* values for comparisons by urbanicity or region were based on a test for independence using an *F*-statistic derived from a Wald chi-square for categorical variables. Estimates with relative standard error (RSE) greater than 30 % have been flagged and should be interpreted with caution. No estimates had RSE greater than 50 %.

Because these analyses used geographic variables that are not available in the public data set, analyses were conducted at the NCHS Research Data Center in Atlanta, GA (<http://www.cdc.gov/rdc/>).

FIG. 1 Urbanicity level and census region, United States.

Population Size Estimates

We used the prevalence estimates for each level of urbanicity and each census region to calculate the total number of MSM and PWID in each of these areas. To do so, we first used US census data from the Vintage 2011 file from the US Census Bureau¹⁴ to calculate the number of men and women aged 13 years or older in each geographic area (urbanicity level or census region) as of July 1, 2011. We then multiplied the weighted prevalence estimate from NHANES by the population estimate of interest (men for estimates of MSM and all persons for estimates of PWID) and rounded to the nearest thousand.

RESULTS

Prevalence of HIV Risk Behaviors by Urbanicity

In our analysis of 23,037 persons interviewed by NHANES during 1999–2010, we found that prevalence of male-male sex (both ever and past 12 months) varied significantly by urbanicity, with highest prevalence in the most urban areas and lowest prevalence in the least urban areas (Table 2). The relative difference was large. For example, the percentage of men reporting male-male sex in the past 12 months varied fourfold between large central metro areas and nonmetro areas (4.4 vs. 1.1 %). In contrast, the percentage of persons reporting injection drug use, either ever or in the past 12 months, did not vary significantly by urbanicity.

TABLE 1 Distribution of US population, aged 13 years or older, in 2011, by urbanicity and US census region

	No. of counties/ states	Male population	Female population	Total Population	% of US population
Urbanicity ^a		N	N	N	%
Large central metro	63	36,813,054	39,115,336	75,928,390	29
Large fringe metro	354	30,659,758	32,470,764	63,130,522	24
Medium/small metro	673	37,488,252	39,282,978	76,771,230	30
Nonmetro	2052	21,277,469	21,533,381	42,810,850	17
Total	3142	126,238,533	132,402,459	258,640,992	100
Region ^b					
Northeast	9	22,612,582	24,327,863	46,940,445	18
Midwest	12	27,242,994	28,529,728	55,772,722	22
South ^c	16	46,630,154	49,307,196	95,937,350	37
West	13	29,753,222	30,238,070	59,991,292	23
Total	50	126,238,952	132,402,857	258,641,809	100

Data from US Census bureau estimates for July 1, 2011.

^aCounties assigned to urbanicity levels in accordance with 2006 NCHS urban-rural classification scheme

^bStates assigned to regions in accordance with US Census Bureau determinations

^cAlso includes District of Columbia

Prevalence of HIV Risk Behaviors by Region

By region, there was no significant variation in prevalence of male-male sex (Table 2). Lifetime prevalence of injection drug use varied, with highest prevalence in the West (3.3 %), lowest prevalence in the Northeast (1.3 %), and intermediate prevalence in the Midwest (1.7 %) and South (2.4 %). However, although history of injection drug use in the past 12 months seemed to follow a similar pattern, the percentages reporting injection drug use were small and differences were not statistically significant.

Population Size Estimates by Urbanicity

We estimated that large central metro areas, which contain 29 % of the US population, are home to 43 % of the approximately 6 million men who had ever had sex with a man and 51 % of the 3 million men who had sex with a man in the past 12 months (Table 3 and Fig. 2). Meanwhile, nonmetro areas, which contain 17 % of the US population, are home to only 9 % of men who ever had sex with a man and 7 % of men who had sex with a man in the past 12 months. On the other hand, PWID are fairly equally distributed across levels of urbanicity.

Population Size Estimates by Region

By region, the South, which contains 37 % of the US population, houses 33–40 % of each risk group (Table 3). This percentage is highest for PWID. The West, which contains 23 % of the US population, houses 33 % of persons who ever injected drugs and 31 % of persons who injected drugs recently. The South and West together comprise the majority of MSM and more than two-thirds of PWID (Fig. 2b).

TABLE 2 Prevalence of male-male sex and injection drug use, by urbanicity and US census region, National Health and Nutrition Examination Survey, 1999–2010

		Ever had male-male sex (of men)	Ever had male-male sex (of all persons)	Had male-male sex, past 12 mo (of men)	Had male-male sex, past 12 mo (of all persons)	Ever injected drugs	Injected drugs, past 12 mo
		% (95 % CI)	% (95 % CI)	% (95 % CI)	% (95 % CI)	% (95 % CI)	% (95 % CI)
Overall	N=23037	4.7 (4.1–5.4)	2.2 (1.9–2.5)	2.5 (2.0–3.0)	1.2 (0.9–1.4)	2.3 (2.0–2.6)	0.3 (0.2–0.4)
Urbanicity ^a	<i>p</i> value	0.0004	0.0004	0.0002	0.0002	0.3	0.6
Large Central Metro	N=9060	7.0 (5.6–8.6)	3.2 (2.6–4.0)	4.4 (3.4–5.7)	2.0 (1.6–2.6)	2.0 (1.6–2.5)	0.3 (0.1–0.5) ^c
Large Fringe Metro	N=3916	4.6 (3.4–6.1)	2.2 (1.6–3.0)	2.5 (1.7–3.8)	1.2 (0.8–1.9)	2.2 (1.5–3.1)	0.3 (0.2–0.6) ^c
Medium/Small Metro	N=6820	3.8 (3.0–4.9)	1.8 (1.4–2.3)	1.4 (0.9–2.2)	0.6 (0.4–1.0)	2.2 (1.7–2.9)	0.3 (0.1–0.4)
Nonmetro	N=3233	2.6 (1.7–3.9)	1.2 (0.8–1.8)	1.1 (0.5–2.5) ^c	0.5 (0.2–1.2) ^c	3.0 (2.2–4.0)	0.5 (0.2–1.0) ^c
Region ^b	<i>p</i> value	0.4	0.4	0.5	0.5	0.0002	0.3
Northeast	N=3643	4.7 (3.3–6.6)	2.2 (1.5–3.1)	1.9 (1.0–3.6) ^c	0.9 (0.5–1.7) ^c	1.3 (0.9–1.7)	0.2 (0.1–0.4) ^c
Midwest	N=4571	4.6 (3.4–6.2)	2.2 (1.6–3.0)	2.1 (1.3–3.3)	1.0 (0.6–1.6)	1.7 (1.3–2.3)	0.3 (0.1–0.6) ^c
South	N=8809	4.2 (3.5–5.1)	2.0 (1.6–2.4)	2.6 (2.0–3.4)	1.2 (0.9–1.6)	2.4 (1.9–3.0)	0.3 (0.2–0.6)
West	N=6014	5.7 (4.4–7.4)	2.7 (2.1–3.5)	3.0 (2.0–4.6)	1.4 (0.9–2.2)	3.3 (2.6–4.3)	0.4 (0.2–0.8) ^c

^aCounties assigned to urbanicity levels in accordance with 2006 NCHS urban-rural classification scheme. Due to missing data, urbanicity levels do not sum to overall number

^bStates assigned to regions in accordance with US Census Bureau determinations

^cRelative standard error between 30 % and 50 %

TABLE 3 Population size estimates for men who have sex with men and persons who inject drugs, by urbanicity and US census region

	Ever had male-male sex		Had male-male sex, past 12 mo		Ever injected drugs		Injected drugs, past 12 mo	
	n	%	n	%	n	%	n	%
Overall ^a	5,933,000	100	3,156,000	100	5,949,000	100	776,000	100
Urbanicity ^b								
Large Central Metro	2,577,000	43	1,620,000	51	1,519,000	26	228,000	29
Large Fringe Metro	1,410,000	24	766,000	24	1,389,000	23	189,000	24
Medium/Small Metro	1,425,000	24	525,000	17	1,689,000	28	230,000	30
Nonmetro	553,000	9	234,000	7	1,284,000	22	214,000	28
Region ^c								
Northeast	1,063,000	18	430,000	14	610,000	10	94,000	12
Midwest	1,253,000	21	572,000	18	948,000	16	167,000	22
South	1,958,000	33	1,212,000	38	2,302,000	39	288,000	37
West	1,696,000	29	893,000	28	1,980,000	33	240,000	31

^aEach population estimate is calculated by multiplying the prevalence estimate for that geographic area (from analysis of data from the National Health and Nutrition Examination Survey) by the total population size for that geographic area (as determined from census data) and is rounded to the nearest thousand. Because the prevalence estimates are rounded, the individual estimates may not sum to the overall number or to 100 %

^bCounties assigned to urbanicity levels in accordance with 2006 NCHS urban-rural classification scheme. Due to missing data, urbanicity levels do not sum to overall number

^cStates assigned to regions in accordance with US Census Bureau determinations

FIG. 2 Population size of men who have sex with men (MSM) and persons who inject drugs (PWID), by (a) urbanicity and (b) U.S. census region.

DISCUSSION

We used NHANES data to calculate the prevalence of behaviors that confer a high risk for HIV infection and calculated population size estimates by urbanicity of residence and region of the country. These nationally representative, population-based data demonstrate important geographic differences in the distribution of these population groups. Prevalence of male-male sex varied by urbanicity and not by region, while prevalence of lifetime injection drug use varied by region and not by urbanicity.

These data can be used in several ways. First, estimates of the prevalence of risk behaviors may be used by local jurisdictions to estimate the number of MSM or PWID in their jurisdiction. Although local data are preferable when available, in the absence of local data, our findings provide a basis for cities or for states in a given region to estimate the number of MSM or PWID. These estimates might also be useful to researchers and health departments as a denominator for calculating disease rates, given a certain number of incident or prevalent cases of HIV or STIs among these populations. Population size and disease rates can help to determine the scale required for interventions that aim to reach these populations. These data may also be useful on a national level, when determining when and how to focus prevention efforts in specific geographic areas or by providers interested in tailoring their testing and counseling behaviors to better address their local population.

We found that 76 % of MSM live in large metropolitan areas (population >1,000,000). These findings confirm what is known from HIV surveillance data, which indicate that 84 % of MSM living with HIV infection lived in a metropolitan statistical area (population >500,000) at the time of diagnosis.¹⁵ Both findings demonstrate the importance of focusing prevention efforts among MSM in large urban areas.

Additionally, we did not find variation in injection drug use by urbanicity, although we did find significant differences in lifetime injection drug use by region. National drug use surveys conducted in 1999–2005 demonstrated that prevalence of recent injection drug use did not vary by urbanicity (although the types of drugs used did vary by urbanicity) and that injection drug use did vary by region, with higher injection drug use in the West.^{9,10} Taken together, these findings provide support for ensuring that prevention reaches PWID in urban and rural areas alike.

Limitations

Due to small sample sizes of some populations, we had to combine data from multiple cycles and were not able to assess trends. Therefore, these estimates of differences by urbanicity and region represent an average across this time period. Additionally, although our prevalence estimates were based on data collected during 1999–2010, we used census data from 2011 to calculate population sizes. We opted to use more contemporary census data so estimates would be more relevant to readers, but these estimates may not be reflective of population sizes when the data were collected. Additionally, although our NHANES study population included persons aged 18–59 years, we extrapolated population size estimates to persons aged 13 years or older; the prevalence of these risk behaviors may be higher among persons aged 18–59 years than among all persons aged 13 years or older.

Small sample sizes of persons who recently injected drugs led to large relative standard errors for this population, suggesting unreliable estimates. Participants may have under-reported male-male sex and injection drug use; this may have affected the size of our risk groups. However, audio computer-assisted self-interview was used to minimize this bias.

NHANES captures the civilian, noninstitutionalized population; homeless and institutionalized populations are not represented, and this may particularly affect our estimate of number of PWID. NHANES is designed to be representative at the national level, but does not necessarily produce estimates that are representative by urbanicity level or by region, as interviews are conducted in only 15 counties each year.

Finally, we focused on two time periods for behaviors of interest—past 12 months and ever. Therefore, these data are not directly comparable to estimates that use behavior in the past 5 years^{2,8}. However, we believe that the past 12 month estimates are appropriate for identifying populations who continue to be at risk of HIV acquisition.

Estimates of male-male sex by urbanicity (9 % in urban areas, 4 % in suburban areas, and 1 % in rural areas) that have been used in other publications^{4,5} were based on data on sexual identity, rather than same-sex risk behavior. The original source of those estimates did provide prevalence of male-male sexual behavior by urbanicity that is similar to ours, although direct comparison is not possible due to differences in the urbanicity categories used³.

CONCLUSION

We found important differences in the prevalence of key HIV risk behaviors by urbanicity and region. Our results suggest that efforts to reach MSM could benefit from being focused in large central and fringe metro areas, while efforts to reach PWID could benefit from broader implementation. These nationally representative, population-based data are the most recent data available on prevalence of these HIV risk behaviors by urbanicity and region and can allow for more effective allocation of funds for prevention programs and determination of reach of existing programs.

ACKNOWLEDGMENTS

We would like to thank Kim Elmore for creating the map used in this publication and Jianmin Li for assistance with census data. We would also like to thank Ajay Yesupriya at the Research Data Center and the NHANES staff for their work in study design and data collection and preparation.

Disclaimer. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Source of Funding. The authors listed on this manuscript are federal employees working at the Centers for Disease Control and Prevention and have not obtained outside sources in the form of grants, equipment, drugs or any combination of these to complete this study.

REFERENCES

1. Centers for Disease Control and Prevention. *HIV Surveillance Report, 2011*. Atlanta, GA: CDC; 2013.
2. Purcell DW, Johnson CH, Lansky A, et al. Estimating the population size of men who have sex with men in the United States to obtain HIV and syphilis rates. *Open AIDS J*. 2012; 6: 98–107.
3. Laumann EO, Gagnon JH, Michael RT, Michaels S. *The Social Organization of Sexuality: sexual practices in the United States*. Chicago, IL: The University of Chicago Press; 1994.
4. Lieb S, Fallon SJ, Friedman SR, et al. Statewide estimation of racial/ethnic populations of men who have sex with men in the U.S. *Public Health Rep*. 2011; 126(1): 60–72.
5. Lieb S, Prejean J, Thompson DR, et al. HIV prevalence rates among men who have sex with men in the southern United States: population-based estimates by race/ethnicity. *AIDS Behav*. 2011; 15(3): 596–606.
6. Brady JE, Friedman SR, Cooper HL, Flom PL, Tempalski B, Gostnell K. Estimating the prevalence of injection drug users in the U.S. and in large U.S. metropolitan areas from 1992 to 2002. *J Urban Health*. 2008; 85(3): 323–351.
7. Tempalski B, Pouget ER, Cleland CM, et al. Trends in the population prevalence of people who inject drugs in US metropolitan areas 1992–2007. *PLoS One*. 2013; 8(6), e64789.
8. Lansky A, Finlayson T, Johnson C, et al. Estimating the number of persons who inject drugs in the United States by meta-analysis to calculate national rates of HIV and hepatitis C virus infections. *PLoS One*. 2014; 9(5), e97596.

9. Substance Abuse and Mental Health Services Administrations (SAMHSA). Injection Drug Use. *The NHSDA Report*. March 14, 2003.
10. Substance Abuse and Mental Health Services Administrations (SAMHSA). Demographic and Geographic Variations in Injection Drug Use. *The NSDUH Report*. July 19, 2007.
11. Zipf G, Chiappa M, Porter KS, Ostchega Y, Lewis BG, Dostal J. National health and nutrition examination survey: plan and operations, 1999–2010. *Vital Health Stat*. 2013; 1(56): 1–37.
12. Harris N, Johnson C, Sionean C, et al. Estimated percentages and characteristics of men who have sex with men and use injection drugs—United States, 1999–2011. *MMWR Morb Mortal Wkly Rep*. 2013; 62(37): 757–762.
13. Ingram DD, Franco SJ. NCHS urban-rural classification scheme for counties. *Vital Health Stat*. 2012; 2(154): 1–65.
14. U.S. Census Bureau. Population estimates [entire data set]. Available at: <http://www.census.gov/popest/data/>.
15. Centers for Disease Control and Prevention. Diagnosed HIV infection among adults and adolescents in metropolitan statistical areas—United States and Puerto Rico, 2010. *HIV Surveillance Supplemental Report*. 2013; 18(1): 1–87.