

Estimating the Population Size of Men Who Have Sex with Men in the United States to Obtain HIV and Syphilis Rates[§]

David W. Purcell^{*,1}, Christopher H. Johnson¹, Amy Lansky¹, Joseph Prejean¹, Renee Stein¹, Paul Denning¹, Zaneta Gau¹, Hillard Weinstock², John Su² and Nicole Crepaz¹

¹Division of HIV/AIDS Prevention, National Center for HIV, Viral Hepatitis, STD, and TB Prevention, Center for Disease Control and Prevention, Atlanta, GA, USA

²Division of STD Prevention, National Center for HIV, Viral Hepatitis, STD and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA, USA

Abstract: *Background:* CDC has not previously calculated disease rates for men who have sex with men (MSM) because there is no single comprehensive source of data on population size. To inform prevention planning, CDC developed a national population size estimate for MSM to calculate disease metrics for HIV and syphilis.

Methods: We conducted a systematic literature search and identified seven surveys that provided data on same-sex behavior in nationally representative samples. Data were pooled by three recall periods and combined using meta-analytic procedures. We applied the proportion of men reporting same-sex behavior in the past 5 years to U.S. census data to produce a population size estimate. We then calculated three disease metrics using CDC HIV and STD surveillance data and rate ratios comparing MSM to other men and to women.

Results: Estimates of the proportion of men who engaged in same-sex behavior differed by recall period: past year = 2.9% (95%CI, 2.6–3.2); past five years = 3.9% (3.5–4.4); ever = 6.9% (5.1–8.6). Rates on all 3 disease metrics were much higher among MSM than among either other men or women (38 to 109 times as high).

Conclusions: Estimating the population size for MSM allowed us to calculate rates for disease metrics and to develop rate ratios showing dramatically higher rates among MSM than among other men or women. These data greatly improve our understanding of the disproportionate impact of these diseases among MSM in the U.S. and help with prevention planning.

Keywords: Prevalence of same-sex behavior among men, MSM population size, HIV rates, syphilis rates.

INTRODUCTION

Since the beginning of the HIV epidemic in the United States, men have accounted for the largest proportion of cases, and the large majority of those men have been gay, bisexual, and other men who have sex with men (collectively referred to here as MSM). MSM accounted for 53% of estimated new HIV infections in the United States in 2006 and MSM who also injected drugs accounted for an additional 4% of estimated new infections [1]. Moreover, MSM comprise the only risk group with increasing numbers of new HIV infections [1]. Similarly, the number of cases of primary and secondary syphilis has grown rapidly among men in the past decade, and the majority of cases have been among MSM [2].

The disparity in number of HIV and syphilis cases among MSM compared to their population size has been difficult to quantify. While U.S. census data currently allow the Centers for Disease Control and Prevention (CDC) to calculate disease rates by age, sex at birth, and racial/ethnic groups [2, 3], there are no census data for the number of MSM in the United States. Estimation of a population size for MSM will allow for the calculation of disease rates among MSM to quantify the disparate impact of various diseases among MSM and to better guide allocation of prevention resources. The estimation of population size for MSM will also allow researchers to compare rates across regions (both within and outside the U.S.) and examine trends over time (since population sizes can change over time).

Research by Kinsey and his colleagues [4] led to the notion that approximately 10% of U.S. men are gay or bisexual. Later reviews lowered this estimate by approximately half [5]. In 1990, the U.S. census added questions about same-sex household partners, but these data cannot lead to a national estimate because people who are not partnered were uncounted [6]. Lieb and his colleagues [7] summarized a wide variety of methods that have been used to estimate the size of the MSM population in specific

*Address correspondence to this author at the Division of HIV/AIDS Prevention, National Center for HIV, Viral Hepatitis, STD, and TB Prevention, Center for Disease Control and Prevention, Atlanta, GA, USA; Tel: 404-639-1934; Fax: 404-639-1950; E-mail: dpurcell@cdc.gov

[§]Note: The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

cities, states, or for the whole United States. In their work, Lieb and his colleagues [7] provided statewide estimates ranging from 3.3% (North Dakota) to 13.2% (District of Columbia) with an overall estimate that MSM were 6.4% of the male population (7.1 million men). These estimates were derived using data from two national surveys that assessed whether men reported ever engaging in same-sex behavior [8, 9]. Lieb and colleagues [10] used their MSM population estimates to show that rates of HIV for MSM in 17 southern states were 36 times as high compared to other men in the south, with rates among MSM highest for African Americans in all states and lowest among white MSM in all but one state.

While there have been a variety of methods and estimates of the national population size of MSM, CDC was interested in developing a single estimate to use in interpreting HIV and sexually transmitted diseases (STD) surveillance data. Combining the findings from the existing literature into one population size estimate can be challenging because 1) sampling frames and recall periods differ, 2) measurements focus on different domains such as sexual behavior, sexual identity, sexual attraction, or having a relationship with a man, and 3) stigma may affect reporting rates on sensitive topics such as same-sex behaviors or identity based on whether the survey was self administered or involved speaking to an interviewer [11]. However, a quantitative synthesis (i.e., meta-analysis) of relevant and appropriate surveys would allow CDC to develop a national estimate of MSM population size to use to better understand the burden of HIV and STDs in the United States.

The goals of this paper are to 1) meta-analyze population-based surveys to estimate the proportion of men in the United States who are MSM and estimate proportion of MSM by race/ethnicity and age if possible; 2) use these estimates to calculate disease metrics for MSM using national surveillance data for HIV and for primary and secondary (P&S) syphilis; and 3) calculate rate ratios to directly compare rates for MSM to rates in other men and women and to compare rates among MSM by race/ethnicity and age. The National HIV/AIDS Strategy [12] highlights MSM as a high priority population that must be the focus of prevention resources to reduce HIV infections in the United States and adds urgency to our desire to better describe and understand the disproportionate impact of HIV on MSM. An estimate of MSM population size and disease rates can help to guide resource allocation and programmatic efforts and establish benchmarks for monitoring progress.

METHODS

Meta-Analysis

Study Selection

We developed a search strategy to identify relevant reports published or in press between 1993 and 2008. The focus on more recent years was due to changes in social norms and acceptance of homosexual behavior that were expected to affect reporting of same-sex behavior. A CDC librarian conducted systematic searches of five electronic databases (Medline, CINAHL, PsycInfo, Sociological Abstracts, and Cochrane) by cross-referencing multiple search terms (i.e., keywords and each database's index

terms) in three domains: measurement descriptors (measurement, prevalence, data collection, assessment, epidemiologic measures, population), male-male sexual behavior (homosexuality, gay, male-to-male sexual contact, anal sex, men who have sex with men) and geography (limited to samples from some or all of the U.S.). We specifically kept the search terms broad in order to not miss any relevant references that might provide either data or different methods for estimating population size for MSM. These initial searches yielded 2862 references (see Fig. 1).

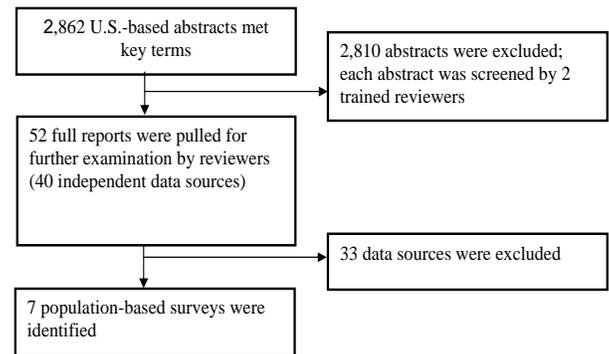


Fig. (1). Survey Selection Process for Meta-Analysis.

Each abstract was screened by two trained reviewers based on explicit inclusion and exclusion criteria. Studies were initially included if they reported prevalence of same-sex sexual behavior (sex with a man; male sex partners), sexual identity/orientation, or sexual attraction. To avoid overestimating the MSM population size, we excluded studies that recruited populations likely to have a high proportion of MSM (e.g., HIV-infected persons, STD clinics, homeless youth) and studies that specifically recruited MSM samples. Conference abstracts or dissertations were screened but not included, as they provided too few details for calculating the estimate.

The initial review of abstracts eliminated 2810 references and identified 52 potentially relevant references representing 40 unique studies with independent data sets for full-report coding. We then excluded school-based samples (including university students) that may have over or underestimated MSM population size and those that only measured sexual orientation or attraction but did not include a same-sex behavioral measure because our focus was on estimating the proportion of men engaging in same-sex behaviors that put them at risk for sexual acquisition of HIV and other STDs. We included only national, probability-based surveys. If multiple publications reported data from the same study, findings from the most comprehensive report were used to avoid overlap. These criteria were applied to the 40 studies and 7 met our criteria and were included in our meta-analyses (see Table 1). These surveys included questions on gender of sex partners (e.g., “Have your sex partners in the last 12 months been: exclusively male, exclusively female, or both male and female” for which we included “exclusively male” and “both male and female” in the population estimate; or “What have you done sexually with a man since you reached puberty” with responses options including oral and anal sex). To examine the extent to which the proportion of men reporting same-sex behavior varied by behavioral recall period, we calculated multiple population

Table 1. Description of Eligible Studies for Meta-Analysis on Population Size of MSM in the U.S. Population

| Study Name | Population Surveyed | Sampling Method | Data Collection Period for MSM Behavior | Data Collection Method | Behavioral Recall Period | MSM Questions |
|---|---|--|---|---|--|---|
| General Social Survey (GSS) [15] | National household survey of the general U.S. population of non-institutionalized English-speaking persons aged ≥18 years. Spanish-speaking persons included since 2006. | Probability sample of households in the U.S. One individual in each household completed the survey. | 1988–2008 for sex with a male in the past year; 1991–2008 for sex with a male in the past 5 years | Face-to-face, interviewer administered survey (paper in 2000; computer-assisted interview [CAI] since 2002). Self-administered paper questionnaire for sexual behavior. | Past year; past 5 years | Gender of sex partners |
| National Health and Nutrition Examination Surveys (NHANES) [16] | National household survey of the general U.S. population of non-institutionalized civilians aged ≥12 years. Sexual behavior questions only asked of persons aged 17–59 years. | Complex stratified, multistage cluster sample | 1999–2008 | Interviewer administered, in person CAI. Computerized self-administered questionnaire for sexual behavior. | Past year; lifetime | Number of male sex partners (past year, ever) where “sex” defined as oral or anal sex |
| National Household Survey on Drug Abuse (now National Household Survey on Drug Use and Health) [17] | National household survey of the general U.S. population of non-institutionalized civilians aged ≥12 years. Sexual behavior questions asked only of persons aged 18–59 years. | Complex stratified, multistage cluster sample | 1996 | In person interviewer-administered survey by computer-assisted interview (CAI); audio computer-assisted self-interview (A-CASI) for sexual behavior | Past year | Gender of sex partners |
| National Health and Social Life Survey (NHLS) [8, 18] | General U.S. population of English speakers aged 18–59 years living in households | Probability sampling | 1992 | In person interviewer-administered paper survey; self-administered paper questionnaire for sexual behavior | Past year; past 5 years; lifetime (since age 13 years) | Any same-gender sex where sex defined as oral or anal sex |
| National STD and Behavior Measurement Experiment (NSBME) [14] | National telephone survey of the general U.S. population with a residential phone number and an English-speaking respondent aged 18–45 years. | Telephone survey of a probability sample (list-assisted RDD) of men aged 18–45 residing in U.S. households with a working landline telephone | Sept. 1999 – April 2000 | Telephone survey; phone numbers were randomly assigned to the interviewer-administered computer-assisted telephone interview or telephone CASI. | Past year; past 5 years; lifetime | Sex with a man involving genital area/ penis contact |
| National Survey of Family Growth (NSFG) [9] | National household survey of the general U.S. population aged 15–44 years. Oversampling of teenagers and black and Hispanic adults | A nationally representative multistage area probability sample drawn from 121 areas of the U.S. | Cycle 6: Mar. 2002 – Mar. 2003 | In person interviewer administered by CAI; A-CASI for sexual behavior questions. | Past year; lifetime | Same sex sexual contact, where sexual contact defined as oral or anal sex |
| Project HOPE International Survey of AIDS-Risk Behaviors [13] | General U.S. population of non-institutionalized persons aged 16–50 years. Residents of Alaska and Hawaii were excluded. | Stratified, cluster sample of households in the U.S. (strata were geographic region and metropolitan versus nonmetropol. residence). | 1988 | In person interviewer administered paper survey; self-administered paper questionnaire for sexual behavior | Past 5 years | Oral or anal sex with a man. |

estimates if a study provided more than one relevant behavioral recall period. For example, if a study reported data for the past year and past five years, we calculated two population estimates for this study and examined these data separately.

Obtaining Data for Meta-Analysis

For each study we obtained the proportion of men who reported same-sex behaviors or partners and its variance (standard error) using the data presented in the publications or by analyzing the original survey data. We estimated the

proportion of MSM for each of 3 recall periods most commonly used in the surveys: ever (i.e., lifetime), past 5 years, and past year. The population-level proportion estimates and variances were taken directly from publications for the National Survey of Family Growth (NSFG) [9], Project HOPE International Survey of AIDS-Risk Behaviors (Project HOPE) [13], the National STD and Behavior Measurement Experiment (NSBME) [14], and the National Health and Social Life Survey (NHSLs) (for lifetime data only) [8]. For the General Social Survey (GSS) [15], NHANES (1999-2008) [16], the National Household Survey on Drug Abuse (NHSDA) [17], and the NHSLs (for past year and past 5 year data only) [18], we analyzed or asked the original researchers to analyze the original survey data using SUDAAN [19] to take the sample design of each survey into account in estimating the proportion of MSM and variance.

We applied a meta-analytic method that has recently been extended to survey data [20]. For each specific recall period, we first multiplied each prevalence estimate by a weight inversely proportional to its variance, summed the weighted prevalence estimates across studies, and then divided by the sum of the weights. Because the surveys included in our analysis used questions that were worded differently, employed different sampling designs to reach their target populations, and were conducted over a period of several years, we used Rao's method [20] to account for the variability of estimates by adding a corresponding between-studies variance term before deriving the overall prevalence estimate. We examined the heterogeneity of prevalence estimates across surveys using the Q statistic as described in Rao [20] and Higgins' I^2 index [21]. Values of the I^2 index around 25, 50, and 75 typically correspond to low, medium, and high heterogeneity, respectively. The aggregated findings for the overall proportion of MSM were based on random-effects models, which provide a more conservative estimate of variance and generate more accurate inferences about a population of studies beyond those included in this review.

For calculating rates, we chose to use the overall past-5-year estimate as a conservative, mid-range estimate of population size of MSM. The past year estimate is the smallest, and it may exclude MSM who have not been sexually active recently. The "ever" estimate is the largest and may include some men who only had incidental same-sex contact many years ago or once. We considered the past-5-year estimate the best available recall period for MSM behavior to use to estimate the proportion of MSM in the total population (of men and overall) because we wanted to compare calculated disease rates for MSM to those often reported in CDC surveillance data based on US census categories.

We planned to conduct a second set of meta-analyses to estimate the MSM prevalence by racial/ethnic group and by age categories because we assumed that there might be differences by race or age. Four of the 7 eligible surveys had some data by race/ethnicity and age (NSFG, NHANES, NHSDA, and NHSLs). Three of these surveys had a recall period of "ever" and one had a recall period of past year. However, due to the small number of surveys with race- and

age-specific data and small samples sizes of MSM when stratified by race/ethnicity, proportion estimates calculated in preliminary analyses were very imprecise with large associated confidence intervals (data not shown). It also would have been necessary to adjust the data to be parallel with the overall past-5-year estimate. Because of our inability to estimate race-specific population sizes of MSM, we applied the overall past-5-year estimate of the proportion of MSM to the race/ethnicity and age-specific distributions from U.S. census data. Categories included three mutually exclusive racial/ethnic groups (black/African American [non-Hispanic], Hispanic/Latino, and white [non-Hispanic]) and 5 age groups (13-24, 25-34, 35-44, 45-54, and >55 years).

All meta-analytic calculations per Rao's method were carried out in Microsoft Excel® 2007 spreadsheets and verified in SAS® Version 9.2. The Comprehensive Meta-Analysis software (version 2) [22] which incorporates assumptions appropriate for synthesizing results from observational studies and clinical trials, was used to verify results and look for comparable patterns in the data.

Disease Rates and Rate Ratios

HIV and AIDS Data

We analyzed data on adults and adolescents (ages 13 or older at diagnosis) with a diagnosis of HIV infection (regardless of stage of disease) in 2008 and reported to CDC through June 2009. Cases of HIV infection were reported from 37 states with confidential, name-based reporting since at least January 2005 [3]. Estimated diagnoses of HIV infection are based on numbers of diagnoses in a given year adjusted for delays in reporting of cases and for missing risk factor information, but not for incomplete reporting [3, 23]. We estimated the number of persons living with a diagnosis of HIV for those persons diagnosed with HIV (regardless of stage of disease at diagnosis) and include persons diagnosed through 2007 and living at the end of 2007 [3]. The estimated numbers of persons living with HIV are based on diagnoses and deaths which were adjusted for delays in reporting of cases and of deaths and for missing risk factor information [3, 23].

Primary and Secondary Syphilis Data

Reports of P&S syphilis represent recently acquired infection, and therefore represent valid indicators of incidence [24]. Data from interviews with patients with P&S syphilis are reported by health departments to CDC *via* the National Electronic Telecommunications System for Surveillance (NETSS). We reviewed NETSS data from states where sex of sex partner (male, female, or both) was reported for 70% or more of male cases of primary and secondary syphilis 13 years of age and older during 2008. The 70% threshold represented the best balance between including male cases of P&S syphilis while gathering the most complete epidemiologic data for those cases. For states meeting this threshold of reporting, we reviewed information on sex, sex of sex partner, age, race and ethnicity. Thirty-nine states and Washington, D.C. met inclusion criteria, yielding 12,110 reports of P&S syphilis that accounted for 89.7% of all reported P&S syphilis in the U.S. during 2008.

U.S. Census Data and Numbers of MSM, Other Men, and Women

To match ages used in HIV and syphilis surveillance data, we used U.S. census data to calculate numbers of men and women ages 13 and older [25]. For men, we then divided the population into the number of MSM and number of other men. To do this, we calculated number of MSM by multiplying our newly derived past-5-year estimate of the proportion of MSM among men (and its 95% confidence interval) by the 2008 population estimate of men aged 13 and older. The number of MSM was then subtracted from the total of all men aged 13 and older to compute the population size of other men.

Calculating Rates and Rate Ratios

We used the estimated number of MSM and other men and the U.S. census count for women as denominators to calculate rates per 100,000 persons for 3 disease metrics (estimated diagnoses of HIV infection, estimated number of people living with a diagnosis of HIV infection, cases of primary or secondary syphilis). Rates were calculated by dividing the estimated number of cases in a population group by the number of people in that population group. Rates were first calculated for MSM, other men, and women. Among MSM, rates were further calculated for race/ethnicity and age groups. For all groups and subgroups except for women, estimated rates are reported with upper and lower confidence intervals to reflect the uncertainty related to the calculation of the proportion of MSM in the population. As

the estimate of the number of MSM increased, the estimate of the number of other men decreased, meaning that rates changed inversely for MSM and for other men.

Finally, we calculated rate ratios (and confidence intervals where relevant) to directly compare rates between different groups. First, we calculated the rate ratios between MSM and other men and between MSM and women by dividing the rate for MSM by the rate for other men and for women. Next we calculated rate ratios within MSM by race/ethnicity and age. For race/ethnicity, we used white MSM as the reference group and compared their rates with the rates for black/African American MSM and Hispanic/Latino MSM. For age, we used the youngest age group of MSM as the reference and compared their rates to the 4 older age groups of MSM. There are no confidence intervals for rate ratios within MSM for race/ethnicity or age because the proportion of men in each subgroup does not change as the proportion of MSM in the population changes.

RESULTS

Proportion of the Population Estimated to be MSM

Table 2 shows the estimated proportion of males who were MSM, by recall period, for each of the relevant population-based surveys and the combined estimates for each recall period. The aggregated estimate from the 4 surveys with a “past 5 year” recall period is 3.9% (95% CI = 3.5%, 4.4%). This estimate falls between the combined estimate from the 6 surveys with a “past year” recall period (2.9%, 95% CI = 2.6%,

Table 2. Estimated Proportion of Males Who are MSM for Individual Studies and Combined by Meta-analysis, by Behavioral Recall Period

| Study Name | Time Point | Recall Period | Estimated Prevalence | Lower 95% CI | Upper 95% CI |
|--------------------------|------------|---------------------|----------------------|--------------|--------------|
| GSS | 1988–2008 | past year | 3.2% | 2.8% | 3.7% |
| NHANES | 1999–2008 | past year | 2.9% | 2.1% | 3.6% |
| NHSDA | 1996 | past year | 2.3% | 1.6% | 3.1% |
| NHSLS | 1992 | past year | 3.5% | 2.1% | 4.9% |
| NSBME | 1999–2000 | past year | 2.6% | 0.8% | 4.4% |
| NSFG | 2002–2003 | past year | 2.9% | 2.2% | 3.6% |
| Combined estimate | | past year | 2.9% | 2.6% | 3.2% |
| GSS | 1991–2008 | past 5 years | 3.8% | 3.3% | 4.3% |
| NHSLS | 1992 | past 5 years | 4.1% | 2.7% | 5.5% |
| NSBME | 1999–2000 | past 5 years | 4.3% | 1.8% | 6.8% |
| Project Hope | 1988 | past 5 years | 6.2% | 3.5% | 9.0% |
| Combined estimate | | past 5 years | 3.9% | 3.5% | 4.4% |
| NHANES | 1999–2008 | Ever | 5.2% | 4.4% | 6.1% |
| NHSLS | 1992 | Ever | 9.1% | 7.6% | 10.6% |
| NSBME | 1999–2000 | Ever | 8.2% | 4.9% | 11.5% |
| NSFG | 2002–2003 | Ever | 6.0% | 5.0% | 7.0% |
| Combined estimate | | Ever | 6.9% | 5.1% | 8.6% |

Notes:

GSS: General Social Survey; NHANES: National Health and Nutrition Examination Survey; NHSDA: National Household Survey on Drug Abuse; NHSLS: National Health and Social Life Survey; NSBME: National STD and Behavior Measurement Experiment; NSFG: National Survey of Family Growth; Project Hope: Project Hope International Survey of AIDS-Risk Behaviors.

3.2%) and the combined estimate from the 4 surveys with an “ever” recall period (6.9%, 95% CI = 5.1%, 8.6%). Q statistics and I² indicate heterogeneity of results across surveys with an “ever” recall period (Q₃ = 22.2, P <.01, I² = 86.5). However, sensitivity tests did not reveal any individual survey that exerted disproportionate influence on the overall estimate. The heterogeneity tests for surveys with a “past 5 years” recall period (Q₃ = 3.1, P =.38, I² = 2.3) and with a “past year” recall period (Q₅ = 5.3, P =.38, I² = 5.9) were not significant indicating lack of heterogeneity.

Table 3. Population Size of Men Aged 13 and Older in the United States, 2008 (N = 122,852,862), and Number of MSM and Other Men Using Past-Five-Year Proportion Estimate of MSM from Meta-Analysis

| MSM Proportion | MSM Population Size | Other Men Proportion | Other Men Population Size |
|----------------|---------------------|----------------------|---------------------------|
| 3.5% | 4,299,850 | 96.5% | 118,553,012 |
| 3.9% | 4,791,262 | 96.1% | 118,061,600 |
| 4.4% | 5,405,526 | 95.6% | 117,447,336 |

Table 3 shows the population size of men aged 13 and older and the number of MSM and other men using the 3.9% proportion estimate (and 95% confidence interval). We estimate that in 2008, 4,791,262 men in the U.S. were MSM, with a range from 4,299,850 to 5,405,526.

Disease Rates and Rate Ratios

Rates of estimated diagnoses of HIV infection, rates of people living with HIV, and rates of P&S syphilis were higher among MSM than among other men or women

(Tables 4-6). The rate ratios indicate disparities between MSM and other men and MSM and women. Comparing MSM to other men, the estimated rate of diagnoses of HIV infection in 2008 was 59 to 75 times as high (Table 4), the estimated rate of MSM living with a diagnosis of HIV infection was 38 to 48 times as high (Table 5), and the P&S syphilis diagnosis rate was 63 to 79 times as high (Table 6). Rate ratios comparing MSM and women were similar to those comparing MSM and other men for the HIV metrics (Tables 4-5). For P&S syphilis, the rate ratio was higher for MSM compared to women than for MSM compared to other men (Table 6).

Among MSM, black/African Americans had the highest disease rates while whites had the lowest rates, regardless of epidemiologic measure. By age group, the highest estimated rate of diagnoses of HIV infection occurred among MSM aged 25 to 44. Estimated rates of MSM living with a diagnosis of HIV infection were highest among men aged 35 to 54. The estimated rate of P&S syphilis among MSM was highest among men aged 25 to 44.

Rate ratios indicate substantial disparities between MSM by racial/ethnic groups for all disease metrics. Across the 3 disease metrics, rates for black/African American MSM were 3.8 to 6 times as high as those for white MSM. Rates for Hispanic/Latino MSM were 1.8 to 2.7 times as high as those for white MSM (Tables 4-6).

DISCUSSION

Using meta-analysis with behavioral data from population-based surveys, we found that MSM comprised 3.9% (3.5% - 4.4%) of the U.S. adult male population. This translates into approximately 2.0% of the overall population

Table 4. Estimated Number, Rate Per 100,000 Persons, and Rate Ratios for Diagnoses of HIV Infection in 2008, 37 States with Confidential Name-Based HIV Infection Reporting, for MSM, Other Men, Women, and MSM by Race/Ethnicity and Age

| | Estimated Number of HIV Diagnoses | Estimate rate per 100,000 Population (95% CI) | Rate Ratio (95% CI)* |
|---|-----------------------------------|---|----------------------|
| MSM and Other Men | | | |
| Other Men | 8,286 | 10.1 (10.0 – 10.1) | ref |
| MSM | 22,469 | 672 (596 – 749) | 67 (59 – 75) |
| MSM and Women | | | |
| Women# | 10,332 | 11.5 | ref |
| MSM | 22,469 | 672 (596 – 749) | 58 (52 – 65) |
| MSM Race/Ethnicity | | | |
| White | 8,478 | 363 (322 – 405) | ref |
| Hispanic | 4,201 | 980 (869 – 1,092) | 2.7 |
| Black/ African American | 9,133 | 2,165 (1,919 – 2,412) | 6.0 |
| MSM Age at Diagnosis (in Years): | | | |
| 13–24 | 5,032 | 713 (632 – 795) | ref |
| 25–34 | 6,380 | 1,126 (998 – 1,255) | 1.6 |
| 35–44 | 6,041 | 1,048 (929 – 1,168) | 1.5 |
| 45–54 | 3,709 | 624 (553 – 696) | 0.9 |
| 55+ | 1,306 | 145 (129 – 162) | 0.2 |

*There are no confidence intervals for rate ratios for MSM race/ethnicity or age because the proportion of men in each subgroup does not change as the proportion of MSM in the population changes.
 #There are no confidence intervals for women because the rate calculation uses U.S. census count, not an estimated number as with MSM and other men.

Table 5. Estimated Number, Rate Per 100,000 Persons, and Rate Ratios for MSM Living with a Diagnosis of HIV Infection, year-End 2007, 37 States with Confidential Name-Based HIV Infection Reporting, for MSM, Other Men, Women, and MSM by Race/Ethnicity and Age

| | Estimated Number of MSM Living with HIV | Estimate rate per 100,000 Population (95% CI) | Rate Ratio (95% CI)* |
|---|---|---|------------------------------|
| MSM and Other Men | | | |
| Other Men | 152,468 | 187 | (186 – 188) Ref |
| MSM | 265,330 | 7,929 | (7,116 – 8,951) 42 (38 – 48) |
| MSM and Women | | | |
| Women# | 153,814 | 173 | Ref |
| MSM | 265,330 | 7,929 | (7,116 – 8,951) 46 (41 – 52) |
| MSM Race/Ethnicity | | | |
| White | 128,290 | 5,523 | (4,896 – 6,155) Ref |
| Hispanic | 43,908 | 10,601 | (9,396 – 11,812) 1.9 |
| Black/ African American | 87,446 | 21,018 | (18,629 – 23,420) 3.8 |
| MSM Age at the End of 2007 (in years): | | | |
| 13–24 | 10,836 | 1,538 | (1,363 – 1,713) Ref |
| 25–34 | 41,376 | 7,412 | (6,570 – 8,260) 4.8 |
| 35–44 | 96,284 | 16,514 | (14,637 – 18,401) 10.7 |
| 45–54 | 82,532 | 14,071 | (12,472 – 15,679) 9.2 |
| 55+ | 34,302 | 3,925 | (3,479 – 4,374) 2.6 |

*There are no confidence intervals for rate ratios for MSM race/ethnicity or age because the proportion of men in each subgroup does not change as the proportion of MSM in the population changes.

#There are no confidence intervals for women because the rate calculation uses U.S. census count, not an estimated number as with MSM and other men.

or 4,791,262 MSM in the United States (95%CI = 4,299,850 to 5,405,526) in 2008. It is important to have these population size estimates to calculate disease rates and other metrics and to compare disease burden on a standard scale. Compared to other men and women, for the two HIV disease metrics, MSM rates were 38 to 75 times as high, and for P&S syphilis, rates were 63 to 109 as high. Data in the U.S. show that MSM make up approximately half of all persons living with HIV infection (whether diagnosed or undiagnosed) [26] and that in 2009, MSM accounted for 61% of new infections [27]. However, the rates calculated here indicate that this disease burden, relative to the size of the MSM population, is of tremendous magnitude. The disparity in rates for MSM compared to other men and women exceeds disparities between race/ethnicity groups reported in HIV surveillance data [3] or between MSM race/ethnicity groups reported here. Quantifying these disparities can help to better focus the National HIV/AIDS Strategy and allocate resources on the most disproportionately affected groups [12]. Our data also quantify well known disparities among MSM, showing that African American men are the most affected group within the already disproportionately affected MSM group. Not surprisingly, our estimates for racial/ethnic disparities between MSM are very similar to those calculated by Lieb and colleagues [7] because these relative disparities between groups of MSM are not affected by using different estimates of MSM population size.

The estimated proportion of MSM that we calculated for 3 different recall periods can be used by others to calculate

rates for other diseases or measures (e.g., past year estimate for HIV incidence rate). Using meta-analysis to calculate an estimate gives us a more robust and stable proportion of MSM than selecting a population size estimate from one particular study and is especially useful when applying these estimates to a variety of surveillance data.

It is important to reiterate that our estimates are based on male-male sexual behavior. We did not use data on sexual identity or attraction, and thus, our estimate is not analogous to the number of self-identified gay and bisexual men in the United States. Because we focused on disease rates, we were more focused on behavior that puts people at risk for infection when we estimated our population size.

Despite the strengths of our methods, a few limitations regarding these data should be noted. First, a meta-analysis rests on the strengths of the underlying studies — and while the study designs are quite robust in the 7 national surveys that met our inclusion criteria, they are hampered by small sample sizes of MSM. Because MSM are a small proportion of the population, obtaining a large enough overall sample so that there are adequate numbers of MSM is difficult in such studies without huge samples sizes. This issue is compounded when we are interested in understanding more about subgroups of MSM such as racial/ethnic or age groups. MSM are seldom over-sampled in population-based surveys to improve representativeness and representation. In addition, under-reporting of same-sex behavior due to perceived stigma could result in an underestimate of population size; however, this bias should be mitigated to

Table 6. Number, Rate, and Rate Ratio for All MSM, Other Men, and Women and for MSM by Race/Ethnicity and Age for Primary and Secondary (P&S) Syphilis Diagnoses, 2008, 39 States*

| | Number of P&S Syphilis Diagnoses | Estimate Rate Per 100,000 Population (95% CI) | Rate Ratio (95% CI)* |
|---------------------------|----------------------------------|---|----------------------|
| MSM and Other Men | | | |
| Other Men† | 2,294 | 2.2 (2.2 – 2.2) | ref |
| MSM† | 6,606 | 154 (136 – 172) | 71 (63 – 79) |
| MSM and Women | | | |
| Women# | 1,821 | 1.6 | ref |
| MSM† | 6,606 | 154 (136 – 172) | 96 (85 – 109) |
| MSM Race/Ethnicity | | | |
| White | 2,699 | 78 (69 – 87) | ref |
| Hispanic | 1,234 | 141 (125 – 158) | 1.8 |
| Black/African American | 2,248 | 353 (313 – 393) | 4.5 |
| MSM Age: | | | |
| 13–24 | 1,524 | 166 (147 – 185) | Ref |
| 25–34 | 2,001 | 277 (245 – 308) | 1.7 |
| 35–44 | 1,920 | 258 (229 – 287) | 1.6 |
| 45–54 | 941 | 123 (109 – 137) | 0.7 |
| 55+ | 219 | 19 (17 – 21) | 0.1 |

*There are no confidence intervals for rate ratios for MSM race/ethnicity or age because the proportion of men in each subgroup does not change as the proportion of MSM in the population changes.

#There are no confidence intervals for women because the rate calculation uses U.S. census count, not an estimated number as with MSM and other men.

† Sex of sex partner data were available for 8,900 (86.5%) of 10,289 men in these 39 states during 2008.

some extent by the private self-report format for sexual behavior used in all the surveys included in the meta-analysis. Given the possible factors that affected the data in the original studies, it is important to consider our analysis, which is essentially a secondary analysis, in this context. Our rates should always be interpreted in the context of the confidence intervals we presented and the range of likely values of the proportion of MSM in the population, as well as a range of values for disease rates and rate ratios. Another caution is that the age ranges of MSM reported in the surveys did not match the age ranges of men in the two surveillance systems reported here or in U.S. census data. When selecting age ranges from census data, we chose to match them to the age ranges reported in the surveillance data which was easier because of the diversity of ages of men in the surveys. In our analyses, we did not account for these age-range differences between the surveys and our other data. We also did not examine trends in age distribution over time in the population which also could potentially affect our results. Finally, our choice to use the estimate based on behavioral data from “past 5 years” does not correspond to the HIV surveillance definition of MSM, which is same-sex behavior since 1977. However, as noted previously, we judged this population estimate to be more robust than measures for the other two time periods. Because we have presented our other two estimates here, other researchers or program staff can use these higher or lower estimates to determine rates or rate ratios if those estimates better fit their needs.

In sum, the meta-analysis results presented here represent one method for making population size estimates for MSM. Future estimates may be improved or refined by considering methods other than meta-analysis [28], including data from younger or more recent cohorts, and recruiting larger samples from which to calculate age- and race/ethnicity-specific estimates. Although national estimates are useful for monitoring the epidemic on a broad scale, local jurisdictions would benefit from using appropriate methods to develop local estimates of MSM population size to monitor their local epidemics [28]. The population estimates of the number of MSM (as well as rates and ratios) can be updated annually based on U.S. census population figures, although they are unlikely to change dramatically from year to year. In contrast, the meta-analysis estimate of the population proportion of MSM is not expected to be updated frequently, as meta-analyses are best updated when a significant amount of new data are available and national samples focused on MSM behavior are limited in number. Trends from population-based surveys will be monitored as part of CDC’s program of HIV behavioral surveillance [29] and the MSM population size estimate meta-analysis can be updated as new data accumulate.

While these rates will help our public health efforts, going forward, CDC will report them separately from their surveillance reports. To date, disease rate information for population subgroups assessed by the U.S. census, such as men and women and racial and ethnic groups, has been reported in HIV and STD surveillance reports. Because the

MSM population estimate was derived from sources other than direct U.S. census population figures, disease rates for MSM and for persons in other transmission categories will be reported separately from the annual HIV Surveillance Report [3] and STD Surveillance Report [2]. Regular reporting of these data will help CDC and local areas monitor the progress of disease prevention.

Because meta-analysis quantitatively combines estimates from multiple studies, it provides a strong basis for a national estimate of the proportion of MSM in the U.S. Our analyses are complementary to other recent publications on this topic which provide guidance so that local areas or states can develop their own estimate of the size of local MSM populations to guide program planning [28]. When taken together, these local or state estimates and our national estimate of population size can be used to provide a broader understanding of the HIV epidemic and the surge in syphilis among those most at risk for transmission and acquisition of these infectious diseases. The MSM meta-analysis estimate and the disease rate calculations provide important tools for monitoring and characterizing the HIV and syphilis epidemics in the United States and they can be helpful in implementing and monitoring the National HIV/AIDS Strategy and syphilis elimination efforts. These data — like all surveillance data — will establish their utility as they are translated into action for public health purposes such as resource allocation, program design, and evaluation of policies and programs. Resources must be directed to the populations at highest risk and to strategies that are cost-effective. The best available data must be used to guide decision-making at the national, state, and local levels [30]. The MSM population size estimate and resulting rates and rate ratios are important additions to the data we use to make these critical decisions.

ACKNOWLEDGEMENT

Declared none.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflicts of interest.

NOTE

*References marked with an asterisk were included in the meta-analyses.

REFERENCES

- [1] Hall HI, Song R, Rhodes P, *et al.* Estimation of HIV incidence—United States. *JAMA* 2008; 300: 520-9.
- [2] Centers for Disease Control and Prevention. Sexually Transmitted Disease Surveillance, 2009. Atlanta: U.S. Department of Health and Human Services; 2010. Available at: <http://www.cdc.gov/std/stats09/default.htm> [Accessed March 20, 2011].
- [3] Centers for Disease Control and Prevention. HIV/AIDS Surveillance Report, 2008. vol. 20. Atlanta: U.S. Department of Health and Human Services; 2010. Available at: <http://www.cdc.gov/hiv/topics/surveillance/resources/reports/>. [Accessed February 1, 2011].
- [4] Kinsey AC, Pomeroy WB, Martin CD. Sexual behavior in the human male. Philadelphia: W. B. Saunders 1948.
- [5] Diamond M. Homosexuality and bisexuality in different populations. *Arch Sex Behav* 1993; 22: 291-310.
- [6] Black D, Gates GJ, Sanders S, Taylor L. Demographics of the gay and lesbian population in the United States: evidence from available systematic data sources. *Demography* 2000; 37: 139-54.

- [7] Lieb S, Fallon SJ, Friedman SR, *et al.* Statewide estimation of racial/ethnic populations of men who have sex with men in the USA. *Public Health Rep* 2011; 126: 60-72.
- [8] Laumann EO, Gagnon JH, Michael RT, Michaels, S. The social organization of sexuality: sexual practices in the United States. Chicago: University of Chicago Press 1994.*.
- [9] Mosher WD, Chandra A, Jones J. Sexual behavior and selected health measures: Men and women 15 – 44 year of age, United States, 2002. Advance Data from Vital and Health Statistics; no 362. Hyattsville, MD: National Center for Health Statistics 2005*.
- [10] 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: 596-606.
- [11] Turner CF, Ku L, Rogers S, *et al.* Adolescent sexual behavior, drug use, and violence: Increased reporting with computer survey technology. *Science* 1998; 280: 867-73.
- [12] White House Office of National AIDS Policy. The National HIV/AIDS Strategy. Released July 13, 2010. Available at: <http://aids.gov/federal-resources/policies/national-hiv-aids-strategy/nhas-implementation.pdf>. [Accessed March 20, 2011].
- [13] Sell RL, Wells JA, Wypij D. The prevalence of homosexual behavior and attraction in the United States, the United Kingdom and France: Results of national population-based samples. *Arch Sex Behav* 2005; 24: 235-48.*
- [14] Villarroel MA, Turner CF, Eggleston E, *et al.* Same-gender sex in the United States: Impact of T-CASI on prevalence estimates. *Public Opin Q* 2005; 70: 166-96.*
- [15] Davis JA, Smith TW, Marsden PV. General Social Surveys, 1972-2008 [Cumulative File] [Computer file]. ICPSR25962-v2. Storrs, CT: Roper Center for Public Opinion Research, University of Connecticut/Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributors], 2010-02-08. doi:10.3886/ICPSR25962. Available at: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/25962/detail> [Accessed March 21, 2011].*
- [16] United States Department of Health and Human Services. Centers for Disease Control and Prevention. National Center for Health Statistics. National Health and Nutrition Examination Survey (NHANES), 2007-2008 [Computer file]. ICPSR25505-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2010-05-04. doi:10.3886/ICPSR25505. Each year available separately at: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/25505?archive=ICPSR&q=NHANES> [Accessed March 21, 2011].*
- [17] United States Department of Health and Human Services. Substance Abuse and Mental Health Services Administration. Office of Applied Studies. National Survey on Drug Use and Health, 2008 [Computer file]. ICPSR26701-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2009-12-16. doi:10.3886/ICPSR26701. Each year available separately at: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/26701/detail>. [Accessed March 21, 2011].*
- [18] Laumann EO, Gagnon JH, Michael RT, Michaels S. National Health and Social Life Survey, 1992: [United States] [Computer file]. ICPSR06647-v2. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2008-04-17. doi:10.3886/ICPSR06647. Available at: <http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/6647/detail>. [Accessed March 21, 2011].*
- [19] Research Triangle Institute. SUDAAN User's Manual, Release 9.0. Research Triangle Park, NC: Research Triangle Institute 2004.
- [20] Rao SR, Graubard BI, Schmid CH, *et al.* Meta-analysis of survey data: application to health services research. *Health Serv Outcomes Res Method* 2008; 8: 98-114.
- [21] Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; 327: 557-60.
- [22] Comprehensive Meta-Analysis [computer program]. Version 2. Englewood, NJ: Biostat Inc 2005.
- [23] Harrison KM, Kajese T, Hall HI, Song R. Risk factor redistribution of the national HIV/AIDS surveillance data: an alternative approach. *Public Health Rep* 2008; 123: 618-27.
- [24] Peterman TA, Kahn RH, Ciesielski CA, *et al.* Misclassification of the stages of syphilis: implications for surveillance. *Sex Transm Dis* 2005; 32: 144-9.

- [25] U.S. Census Bureau. Population estimates: entire dataset. July 1, 2008. Published August 6, 2009. Available at: <http://www.census.gov/popest/estimates.html>. [Accessed April 29, 2010].
- [26] Centers for Disease Control and Prevention. HIV surveillance – United States, 1981 -2008. *MMWR* 2011; 60: 689-93.
- [27] Prejean J, Song R, Hernandez A, *et al.* Estimated HIV Incidence in the United States, 2006-2009. *PLOS One* 2011; 6(8): e17502.
- [28] Lieb S, Thompson DR, Misra S, *et al.* for the Southern AIDS Coalition MSM Project Team. Estimating populations of men who have sex with men in the southern United States. *J Urban Health* 2009; 86: 887-901.
- [29] Lansky A, Drake A, DiNenno E, Lee C-W. HIV behavioral surveillance among the US general population. *Public Health Rep* 2007; 122(Suppl1): 24-31.
- [30] Centers for Disease Control and Prevention. HIV Prevention in the United States at a Critical Crossroads, August 2009. Available at: http://www.cdc.gov/hiv/resources/reports/hiv_prev_us.htm#PrevStatus [Accessed March 21, 2011].

Received: April 26, 2011

Revised: August 18, 2011

Accepted: September 20, 2011

© Purcell *et al.*; Licensee *Bentham Open*.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.