Trajectories of and Disparities in HIV Prevalence among Black, White, and Hispanic/Latino Men who Have Sex with Men in 86 Large U.S. Metropolitan Statistical Areas, 1992–2013

Leslie D. Williams[1,*], **Ronald Stall**[2], **Barbara Tempalski**[3], **Kevin Jefferson**[4], **Justin Smith**[4], **Umed Ibragimov**[4], **H. Irene Hall**[5], **Anna Satcher Johnson**[5], **Guoshen Wang**[5], **David W. Purcell**[5], **Hannah L.F. Cooper**[4], **Samuel R. Friedman**[3,6]

[1] Division of Community Health Sciences, University of Illinois at Chicago School of Public Health, Chicago, IL

[2] University of Pittsburgh, Pittsburgh, PA

[3] Institute for Infectious Disease Research, National Development and Research Institutes, New York, NY

[4] Rollins School of Public Health, Emory University, Atlanta, GA

[5] Centers for Disease Control and Prevention, Atlanta, GA

[6] Department of Population Health, New York University School of Medicine, New York, NY

Abstract

The challenges of producing adequate estimates of HIV prevalence among men who have sex with men (MSM) are well known. Among them are accurately estimating MSM population size and obtaining HIV testing data from unbiased samples. Previous research has produced rigorous estimates of HIV prevalence among MSM in specific geographic locations (e.g., large cities with large populations of MSM), or for a broader range of locations, but only over a relatively short period of time (e.g., one year). No one, to our knowledge, has published annual estimates of HIV prevalence among MSM over an extended period of time and across a wide range of geographic

*Corresponding Author Leslie D. Williams, 1603 W. Taylor Street, Rm. 677 (MC 923), Chicago, IL 60612, lesliedw@uic.edu.

areas. This is an important gap in the literature, given that this information is needed to identify multi-level predictors of change over time in HIV prevalence among MSM and to help target resources to high-need areas - a national priority. This paper integrates data from numerous sources: Centers for Disease Control and Prevention's (CDC) National HIV Surveillance System and National HIV Prevention Monitoring and Evaluation data; estimates of 1992 MSM population size and HIV prevalence and incidence among MSM by Holmberg, 1997; and estimates of HIV among MSM from published literature using 1992–2013 data. It applies multilevel modeling to these data to estimate and validate trajectories of HIV prevalence among MSM from 1992–2013 for 86 of the largest metropolitan statistical areas (MSAs) in the United States. Our estimates indicate that, consistently, HIV prevalence among MSM increased during this time period in each MSA, from an across-MSA mean of 11% in 1992 to 20% in 2013 (with slightly smaller increases among MSAs with the initially-largest HIV burden among MSM; S.D. across all years = 3.5%). Our estimates by racial/ethnic subgroups of MSM suggest higher mean HIV prevalence among minority (Black and Hispanic/Latino) MSM than among white MSM across all years and geographic regions. The consistent increases found in HIV prevalence among all MSM are likely primarily attributable to decreases in mortality among HIV-positive MSM, and are likely secondarily attributable to increasing HIV incidence among racial/ethnic minority subpopulations of MSM. Future research is needed to confirm that these are in fact the factors driving the increases in HIV prevalence observed in our estimates. If so, without detracting from HIV prevention efforts targeting MSM, new healthcare initiatives may be needed which focus on targeted HIV prevention efforts among racial/ethnic minority MSM and on training healthcare providers to address cross-cutting health challenges of increased longevity among HIV-positive MSM populations.

Introduction

Since the beginning of the HIV epidemic, substantial scientific interest and public health resources have been devoted to estimating the prevalence of HIV infection among key high-risk populations, including men who have sex with men (MSM),[a] the population in the United States that suffers the greatest burden of disease attributable to HIV infection.[1,2] This long-standing interest is well justified. Prevalence is an important measure of the risks of further HIV-related complications and of future transmission within key populations, and thus provides a key metric to determine resources that should be devoted to prevention and care efforts for these groups.

Despite the importance of generating accurate estimates of HIV prevalence among MSM, there are numerous specific challenges in doing so. Among these challenges are the problems of enumerating a hidden population (i.e., MSM) and of reaching an un-biased sample for HIV testing across the wide variety of MSM sub-populations (e.g., by race/ethnicity and age). HIV testing data generally reflect an inherently biased sample, as people select into accessing this important healthcare service. In addition, people who test positive typically do not get retested in later years, but may move between geographic areas, which can lead to inaccuracy of prevalence estimates. Compounding these problems is the fact that

---

[a]"MSM" refers to and includes gay, bisexual, and other men who have sex with men.

HIV surveillance among MSM has not been entirely consistent across the vast geographic regions in which MSM reside or across the decades since the beginning of the HIV epidemic. Each of these problems raises significant challenges that hinder the production of accurate, setting-specific estimates of (and trajectories over time of) HIV prevalence among MSM.

Given these many challenges, it is unlikely that any one approach to estimating prevalence will yield a completely unbiased estimate of HIV prevalence among MSM. Thus, the most prudent strategy would be to use multiple approaches, with the aim of replicating estimates for HIV prevalence among MSM that are sufficiently accurate to guide HIV prevention and care policy. While previous research has produced good estimates of HIV prevalence among MSM in specific geographic locations (e.g., large cities with large populations of MSM[3,4,5]), or for a broader range of locations, but only for a short period of time (e.g., one year[6,7,8,9,10,11]), no previous studies have, to our knowledge, published annual estimates of HIV prevalence among MSM over an extended period of time and for a wide range of geographic areas. Yet such estimates would allow future research exploring setting-level predictors of change over time in HIV prevalence among MSM. Such estimates of HIV prevalence among MSM and such research on characteristics that predict change in HIV prevalence over time could be very useful to informing resource allocation and related policy decisions to adequately and efficiently address the health of MSM populations in the United States overall, and in specific settings.

Our study addresses the gap in the literature described above. Specifically, we present here one approach to generating estimates of prevalence of all HIV cases (both diagnosed and undiagnosed) among MSM in the United States. Our approach adds to the existing literature by using multiple sources of data to adjust for single-source biases in order to generate annual HIV prevalence estimates among MSM for a 22-year period of time for 86 of the largest metropolitan statistical areas (MSAs) in the United States. Additionally, we present annual MSA-level estimates of HIV prevalence among Black, White, and Hispanic/Latino subgroups of MSM, and examine variation in HIV prevalence among MSM both by geographic region and race.

Methods

Sample

The unit of analysis in this study is the MSA. The U.S. Census Bureau and Office of Management and Budget define an MSA as a set of contiguous counties that include one or more central cities of at least 50,000 people that collectively form a single, cohesive socioeconomic unit, defined by inter-county commuting patterns and socioeconomic integration.[12] MSAs were selected as the unit of analysis because data were readily available at this geographic level and because we posited that MSAs are meaningful epidemiologic units with which to study HIV among key populations and services designated for them.[13,14]

This study examines the largest MSAs in the United States over time, from 1992 to 2013. This time period was selected based on data availability and our aim of examining a long enough time period to be able to observe meaningful changes. MSAs (N=96) in the United

States that had a population greater than 500,000 in 1993 were eligible for inclusion in the present sample. Multiple imputation of missing data was conducted for tenable amounts of data missingness (see below). However, we were unable to compute estimates of HIV prevalence among MSM for 10 MSAs that had insurmountably large amounts of missing data (i.e., MSAs with six or more consecutive years of data missing on at least one data series used for estimates creation). As a result of such missingness, we removed the following 10 MSAs from our sample: Birmingham, AL; Boston, MA; Honolulu, HI; Little Rock, AR; Minneapolis, MN; Norfolk/Virginia Beach/Newport News, VA/NC; Richmond, VA; San Juan, PR; Springfield, MA; and Wichita, KS. The resulting sample of MSAs for which estimates of HIV prevalence among MSM were computed is therefore N = 86.

Measures

We estimated HIV prevalence among MSM using 1) MSA-level data on HIV among MSM from several data series from the Centers for Disease Control and Prevention (CDC); 2) previous MSA-level estimates of MSM population size and of HIV prevalence and incidence among MSM for 1992, created by Holmberg[6]; and 3) available research estimates of HIV prevalence among MSM from relevant published literature. In all CDC data sources, MSM were defined as 15- to 64-year-old men who reported having had any sexual contact with men (and are therefore comprised of people who identify with various sexual orientations). Likewise, data from published literature were eligible for inclusion if they described men having any sexual contact with other men, regardless of sexual orientation. Holmberg's[6] estimates defined MSM as men who had sex with other men in the past 12 months. Table 1 provides additional details about each of these sources of data. The CDC data series are the most comprehensive data available on rates of HIV among MSM, but each alone is subject to various limitations related to reporting accuracy and completeness (e.g., missing data points). By 1) using multiple CDC data series together, 2) anchoring trajectories by adjusting our model using a plausible and well-reasoned set of estimates of the characteristics of each MSA's HIV burden among MSM in 1992, at baseline (i.e., the Holmberg estimates), and 3) incorporating additional available data on HIV prevalence among MSM not captured by CDC (from the extant published literature), we aim to adjust the data series from CDC and reduce error in all available data to the highest degree possible.

Computing Annual Estimates of HIV Prevalence among MSM, 1992–2013

Multilevel models were used to create an equation that regressed estimates of HIV prevalence among MSM from previous literature on all other included data sources on HIV among MSM. The following multilevel equation, with time nested within MSAs, summarizes the estimation model:

$$Y = b_0 + b_1 x_{1jk} + \Sigma\, b_c x_{cjk} + \eta_{0k} + \alpha_{0jk}\,, \qquad \text{Equation 1}$$

with $\eta_{0k} : N(0, \sigma^2_{\eta0})$, $\alpha_{0ijk} : N(0, \sigma^2_{\alpha0})$, independently of one another, where j indexes MSAs, k indexes years, and c indexes all covariates (i.e., seven predictor variables, listed below); where η_{0k} reflects variability across MSAs and α_{0jk} reflects variability across years;

where x indexes independent variables (listed below) and where Y = "literature-based estimates."

The "literature-based estimates" outcome was based on a literature review of all studies computing HIV prevalence for MSM samples in our MSAs from data collected between 1992 and 2013. This outcome variable included 2008 and 2011 HIV estimates for MSM from the National HIV Behavioral Surveillance (NHBS) study (see Table 1). Where there were two or more research estimates for a given MSA in a given year, the mean was taken.

The independent variables in this equation used to predict the "literature-based estimates" outcome were: 1) the number of annual reported deaths among people with HIV ever classified as stage 3 (AIDS) from the CDC's National HIV Surveillance System (NHSS), divided by Holmberg's 1992 estimate of MSM population size; 2) the number of MSM living with AIDS (PLWA) per 10k population (also from the CDC's NHSS); 3) Holmberg's 1992 estimate of HIV prevalence among MSM; 4) Holmberg's 1992 estimate of HIV incidence among MSM; 5) Holmberg's 1992 estimate of the population size of MSM; 6) percent of MSM who were HIV positive in National HIV Prevention Monitoring and Evaluation (NHM&E) Testing data; 7) number of NHM&E testing events per 10k population; and 8) time.

To compute the HIV prevalence estimates, the multilevel equation described above was applied (using PROC MIXED in SAS version 9.3) only to data from the subset of 18 MSAs that had a reasonable amount of data from research estimates (including NHBS estimates), where a "reasonable" amount was defined as having at least one research estimate from early years (before 1997), at least one research estimate between 1997 and 2003, and at least one research estimate in later years (after 2003). The coefficients from this model analyzing the subset of 18 MSAs were then entered into the equation and used to compute estimates for all 86 MSAs.

Multiple imputation was conducted in SAS to address missing data in all key variables. Multiple imputation was completed using the PROC MI procedure, which relies upon single-chain regression-based algorithms to impute missing values using all relevant observed information.[15] To estimate imputed values, all variables used in the equation described above were included in the imputation model. Fifteen separate datasets were imputed, as extant literature suggests that more than five imputations are likely to be necessary[16,17] to provide unbiased parameter estimates.[18] The mean of the fifteen imputed values for each missing data point was used for all following analyses.

We observed during descriptive examination of NHM&E data that very low denominators (i.e., too few individuals tested in a given MSA and year) produced year-by-year NHM&E-computed percentages of MSM who tested positive that were highly unstable. Therefore, in the NHM&E data, if the number of individuals tested in a given MSA and year was less than 40% of the number of individuals tested the previous year in that MSA, or if the number of individuals tested was less than 30 in a given MSA and year, the NHM&E data points for that MSA and year were deleted. The 37 cases (1.9%) we deleted for this reason - out of our

total 1,892 (86 MSAs x 22 years) - were left missing among our final set of estimates, resulting in a final N of 1,855 estimated data points.

Sensitivity Analyses

Sensitivity analyses were conducted in order to assess the adequacy of the fit of our estimation model and to determine whether different specifications of the estimation model (i.e., the inclusion of different subsets of predictor variables in the estimation model) would result in better model fit. These analyses used the same methods described above, but with different versions of the multilevel regression model. Specifically, models were tested in which each predictor variable was removed (one variable removed per model) to compare model fit using Akaike's Information Criterion (AIC). Each of these models used for sensitivity analyses achieved poorer model fit (i.e., higher AIC) except for the model with Homberg's 1992 estimate of HIV incidence removed, which achieved the same model fit (i.e., the same AIC) as the original model (Equation 1). These sensitivity analyses therefore demonstrated that all variables in the model improved model fit (except for HIV incidence, which had no effect on model fit). We decided to retain the HIV incidence predictor in our final model to produce our estimates in order to remain consistent with the methods we have used to compute estimates of HIV prevalence among other key populations.[19]

Validation Analyses

We assessed convergent validity by computing bivariate correlations between our new 1992–2013 estimates and 1) 2006–2013 data from IMS Health20,[b] on the number of patients prescribed ARV medications in each MSA (not specific to MSM) per 10k population; and 2008–2013 data from CDC's National HIV Surveillance System on 2) MSM living with HIV (PLWH) per 10k population; 3) new HIV diagnoses among MSM per 10k population; and 4) estimates of HIV prevalence among MSM for 2012 published by Rosenberg et al.[21] Listwise deletion was applied to all correlation analyses, such that only the years of our estimates were analyzed against a given validator for the years for which data on that validator were available.

Estimates for Racial/Ethnic Subgroups

We used multiplier methods to apportion our estimates for each MSA and year by racial/ethnic subgroups for Black, White, and Hispanic/Latino populations. We used an equation defined by Lieb et al.[22] that applied the proportion of HIV-positive MSM in the NHM&E testing data who belonged to each of these racial/ethnic groups in a given MSA and year to our estimate of HIV prevalence among all MSM in that MSA and year. Specifically, we used the following equation to produce our racial subgroup estimates:

---

[b]IMS Health is a healthcare information and technology company in the United States. "IMS" does not currently seem to be an abbreviation for a longer name. IMS Health recently merged with another company and is now part of the company "IQVIA." However, as we obtained data from IMS Health before this occurred, we refer to these data using the IMS Health name.

Subgroup HIV prevalence =
$$\left(\frac{\%\ HIV\ positive\ among\ subgroup\ MSM\ in\ NHM\&E\ data}{\%\ HIV\ positive\ among\ all\ MSM\ in\ NHM\&E\ data}\right) X\ HIV$$
prevalence among all MSM

Equation 2

We could not compute estimates by racial/ethnic subgroups for all MSAs due to large amounts of missing or suppressed NHM&E data by race/ethnicity. The CDC suppressed all data points for which the value was less than 5 (i.e., for which there were less than 5 individuals of a certain race/ethnicity and risk group in a given MSA and year who were positive or who were tested). We operationalized a "too-great" amount of missingness if we could not compute "percent HIV positive" for a specific subgroup in a given MSA for 6 or more consecutive years because of missingness, as this amount of missing data would not be amenable to multiple imputation or other interpolation methods and would therefore make us unable to produce a reasonable trajectory for that subgroup in that MSA. If we were not able to compute estimates for at least two subgroups, we did not compute them for any subgroups, since the purpose of producing these subgroup estimates was to make comparisons and evaluate health disparities. Based on this rule of thumb for missingness, we were unable to compute any racial/ethnic subgroup estimates for 11 MSAs (Charleston-North Charleston, SC; El Paso, TX; Greensboro-Winston-Salem-High Point, NC; Harrisburg-Lebanon-Carlisle, PA; Knoxville, TN; Raleigh-Durham-Chapel Hill, NC; Scranton-Wilkes-Barre-Hazleton, PA; Seattle-Bellevue-Everett, WA; Syracuse, NY; Tacoma, WA; and Youngstown-Warren, OH); were unable to produce Hispanic/Latino subgroup estimates for an additional 24 MSAs (see Table 2), and were unable to produce Black subgroup estimates for an additional 9 MSAs. This resulted in a total N of 42 MSAs for which we were able to compute estimates for all three racial/ethnic subgroups, 24 MSAs for which we were able to compute estimates for Black and White subgroups only, and 9 MSAs for which we were able to compute estimates for Hispanic/Latino and White subgroups only.

Results

Trends in HIV Prevalence among MSM, 1992–2013

Figure 1 depicts the mean and median estimated annual HIV prevalence among MSM, across all 86 MSAs, from 1992–2013. (Plots of change over time in estimated HIV prevalence among MSM for each MSA, individually, are available in the online supplement for this article, in Figure S1.) Across all MSAs and years, mean HIV prevalence among MSM was 14.8% (S.D. = 3.5%), with a range of 6.5–32.2%. The mean (across all 86 MSAs) and median estimated HIV prevalence among MSM both increased steadily from 1992 (approximately 11%) to 2013 (approximately 20%). A trend of increasing prevalence occurred in most individual MSAs as well, with the exception that MSAs with the largest HIV burden among MSM at baseline (e.g., New York, San Francisco, Miami, Fort Lauderdale) maintained a relatively steady HIV prevalence among MSM during this time period, with only a small increase in prevalence. The average (across all years) annual standard deviation for MSA-level HIV prevalence among MSM was 2.07%.

Validation of Estimates

Our validation analyses revealed that our estimates of HIV prevalence among MSM were most strongly correlated ($r = 0.51$; $p = .06$) with the 2012 estimates of (diagnosed and undiagnosed) HIV prevalence among MSM published by Rosenberg et al.[21] This moderately high correlation is noteworthy particularly because it was based on only N = 14 MSAs, since Rosenberg et al. produced estimates for 17 of the 96 largest MSAs eligible for the present sample, but three of these were eliminated from our sample due to missing data. Our estimates of HIV prevalence among MSM were also significantly correlated with new HIV diagnoses among MSM per 10k population ($r = 0.26$; $p < .0005$) and with MSM living with HIV per 10k population ($r = 0.25$; $p < .0005$). These positive, significant correlations are in the expected direction and provide evidence of the validity of our estimates of HIV prevalence among MSM. Our estimates of HIV prevalence among MSM were also positively, significantly correlated with number of patients prescribed ARV medications per 10k population ($r = 0.14$; $p < .0005$), despite the fact that our data on ARV prescriptions were not disaggregated by risk group (i.e., are not specific to MSM).

Estimates for Racial/Ethnic Subgroups

Figure 2 depicts 1992–2013 trajectories of estimated HIV prevalence among subgroups of MSM by race/ethnicity, for the 42 MSAs for which we were able to compute such estimates for all three (Black, white, and Hispanic/Latino) groups. Figure 3 depicts a) trajectories of estimated HIV prevalence among only Black and white subgroups, for the 66 MSAs for which computing estimates for these two subgroups was possible; and b) trajectories of estimated HIV prevalence among only Hispanic/Latino and white subgroups, for the 51 MSAs for which computing estimates for these two subgroups was possible.

Both Figures 2 and 3 depict trajectories for HIV prevalence that are higher in all years among Black MSM ($M = 31.31\%$; $S.D. = 16.16$) than among Hispanic/Latino MSM ($M = 19.81\%$; $S.D. = 12.34$), and higher in all years among Hispanic/Latino MSM than among white MSM ($M = 12.21\%$; $S.D. = 6.02$), with the difference between estimated prevalence among Black MSM and that among white MSM increasing over time (from a difference of 12.39% in 1992 to a difference of 21.46% in 2013), as the slope of the trajectory for HIV prevalence among Black MSM appears to be steeper than that of the trajectories for the other racial/ethnic groups.

Figure 4 depicts the results of our analyses of HIV prevalence among MSM by U.S. Census-defined geographic regions. These results suggest that among all MSM, estimated HIV prevalence was higher on average in the Southern region of the United States than it was in the Western region of the United States in all years. In most years, estimated HIV prevalence among MSM was higher in the South than all other regions and was lower in the West than all other regions (although there were a few years during which this varied). The results of a one-way ANOVA ($F(3, 1851) = 51.91$; $p < .0005$) and corresponding post hoc analyses suggest that estimated prevalence in the South was significantly higher than in all other geographic regions ($M = 15.94\%$; $S.D. = 3.75$) and that estimated prevalence in the West was significantly lower than in all other geographic regions ($M = 13.39\%$; $S.D. = 3.05$).

Figure 5 and Table 3 present the results of our analyses of HIV prevalence among MSM by both geographic region and racial/ethnic subgroup. The differences between racial/ethnic subgroups described above held constant within each geographic region, separately (Figure 5). However, when examined by racial/ethnic subgroup, the direction of regional differences varied (Table 3). Among White MSM, estimated HIV prevalence was significantly higher in the South than in all other regions ($F_{(3, 1515)} = 4.56$; $p = .003$). However, among Black MSM ($F_{(3, 1307)} = 25.19$; $p < .0005$) and among Hispanic/Latino MSM ($F_{(3, 987)} = 17.98$; $p < .0005$), estimated HIV prevalence was significantly higher in both the Northeast and Midwest regions than in the South or West.

Discussion

This study extends prior work[24; 1–11] by generating HIV prevalence estimates among MSM for 86 of the largest U.S. MSAs over two decades. The prevalence estimates reported in this paper present a remarkably unified picture of increases in the prevalence rates for HIV infection among MSM from 1992–2013 across the largest MSAs in the United States (with the exception of a brief period of estimated decrease in mean prevalence between 1995 and 1998). At the end of the study period, the mean prevalence of HIV infection among MSM reached approximately 20%, almost twice the estimated 11% at baseline. The scope and persistence of these rises in prevalence rates over this two-decade period indicate the need to ensure that HIV prevention and care among MSM communities will remain a central focus in HIV-related public health efforts for many years to come.

Our estimates provide a more holistic picture of the trajectories of HIV prevalence among MSM in the United States than has been available from previous research. They integrate multiple data series from multiple sources in order to reduce biases present in single sources of data. They furthermore present trajectories of HIV prevalence among MSM across a longer period of time than have previous studies, and for a larger number of geographic regions than have most extant studies producing estimates of HIV prevalence among MSM (e.g.,[1–11, 6, 24]). They will therefore allow for future studies that explore MSA-level predictors of variation in HIV prevalence among MSM between MSAs and over time. These estimates, in combination with such future research on setting-level predictors of HIV prevalence among MSM, should help to inform HIV prevention and care resource allocation and related policy.

Increasing HIV prevalence rates among MSM are likely driven by a combination of increased survival among HIV positive men due to improved treatment options and ongoing new HIV infections. (Conversely, the brief period of estimated decreases in HIV prevalence between 1995 and 1998 may suggest that during this time, deaths among people with HIV were outpacing new diagnoses.) While increasing longevity among HIV-positive people is a most welcome development, this evolution presents new challenges for the HIV treatment and policy fields around the care of an increasingly older population with HIV, such as increased costs[23] and other challenges related to co-morbidities that are more common in older populations.[24] The scale of the rise in HIV prevalence shown in the data presented here strongly suggests that HIV burden will become increasingly centered among aging HIV positive MSM.[25,26,27] If so, HIV care will increasingly become an amalgam of treating the

diseases associated with aging, possible long term consequences of ARV treatments, and the treatment of psychosocial conditions (e.g., depression, internalized stigma) that constitute serious health challenges for MSM , with a focus on addressing the combined needs associated with each of these health challenges. Few care providers in the field have had training in treating the cross-cutting health problems of geriatric HIV positive populations. Finding ways to address these evolving challenges will be essential to helping long-term HIV-surviving (and therefore aging) MSM achieve a healthy and functional later life. Future research should examine changes in HIV prevalence over time among specific age groups of MSM, and variation in these trajectories between various settings, in order to identify settings with the largest increases in HIV prevalence among older MSM and to inform resource allocation towards addressing the unique needs of this population.

Another factor which is likely contributing to increasing trajectories of HIV prevalence among MSM is high HIV incidence rates among Black and Hispanic/Latino MSM. For example, HIV incidence among Black MSM is estimated to be almost twice as high as incidence among MSM overall.[28] This staggering health disparity among subpopulations of minority MSM will continue to contribute to increasing HIV prevalence among MSM overall.

Indeed, consistent with prior research,[e.g., 29] our racial/ethnic subgroup estimates suggest that there are large disparities in HIV prevalence between minority MSM and white MSM. HIV prevalence was found to be much higher on average among Black MSM than among white MSM in all years. HIV prevalence among Hispanic/Latino MSM was also found to be higher on average in all years compared to HIV prevalence among white MSM, although this disparity was smaller than that between Black and white subgroups of MSM. Our analyses of geographic variation found that these racial/ethnic disparities in HIV prevalence among MSM were similar in each of the four regions of the U.S. defined by the Census (South, West, Northeast, and Midwest).

In examining geographic differences in HIV prevalence among MSM, we found that estimated HIV prevalence among all MSM was highest in the South, on average. This difference appeared to be driven by white MSM, as this regional pattern only held true in our subgroup estimates for white MSM. Among Black and Hispanic/Latino subgroups of MSM, estimated HIV prevalence was higher in the Northeast and Midwest than in the South or West, on average. Although racial/ethnic disparities in HIV prevalence among MSM were found to be largest in the Northeast on average, they appear to be increasing over time in the South, as the trajectory of HIV prevalence among Black MSM appears to be increasing most steeply in that region (from a difference between Black MSM and white MSM of 9.79% in 1992 to a difference of 18.53% in 2012). However, mean estimates and trajectories for racial/ethnic subgroups by region are likely to be biased by the fact that we were only able to compute such estimates for a subset of MSAs due to missing data, as described above.

Nonetheless, the above findings suggest that there are marked disparities in HIV prevalence between racial/ethnic subgroups of MSM in all regions of the United States. Continued and increased efforts and resources are therefore needed for building comprehensive preventive and treatment-based programs and services that target racial/ethnic minority subpopulations

(e.g., programs connecting Black and Hispanic/Latino MSM to PrEP or to ART). Special attention might be warranted in the South and Northeast to address larger (in the Northeast) and/or rapidly growing (in the South) disparities.

Limitations

Although our use of multiple data sources in our estimation model was a major methodological strength in that it allowed us to adjust for single-source biases, it also introduced some limitations. Using many data sources collected by many entities over a period of 22 years could possibly result in a substantial degree of systematic measurement error, in part because some data series modified operational definitions of constructs and/or changed other data collection and reporting methodologies over the 22-year period. Furthermore, even when such specific changes are known, their exact influence on trends in the data series is not always easily understood. For example, NMH&E testing changed its data collection requirements between 2007 and 2008, when a new "Expanded Testing Program (ETP)" was introduced that aimed to target disproportionately affected populations and populations with lower testing rates for HIV testing.[30] However, our examination of the data between these two years did not reveal noticeable changes in the number of MSM tested or in the number of those tested who were HIV-positive; and furthermore, descriptive analysis of the proportion of cases that came from the "ETP" and those that did not showed dramatically different patterns over time between individual MSAs, suggesting that MSAs did not all implement this program in a similar manner. Another potential source of bias is the sampling procedures used by NHBS (which sampled MSM from bars or dance clubs, thereby possibly capturing a higher risk subpopulation)[31] or by any of the studies we incorporated into our literature-based estimates outcome variable. However, combining all available estimates should help to reduce such potential bias.

An additional important limitation is that most of our data sources contained missing data. Though we addressed data missingness with multiple imputation, this is an imperfect process that could have introduced bias to our estimates if, for example, our imputation model was poorly specified. Our descriptions of mean estimates of HIV prevalence among racial/ethnic subgroups of MSM are subject to additional potential bias as a result of our removing MSAs with large amounts of missing data, as discussed above. This removal of some MSAs is especially likely to have affected our analyses by both racial/ethnicity and geographic region, as MSAs with large amounts of missing data among racial/ethnic minority MSM were disproportionately in the South or West. Our use of NHM&E Testing data to apportion our estimates by race/ethnicity is also a limitation, since people who receive HIV testing are a biased sample, as discussed above. Additionally, all of our estimates are potentially limited insofar as they were created using only the secondary data that were available to us. While we used 7 different datasets representing the publicly available data of which we are aware, using additional or alternative data could alter the estimates reported herein. Finally, it is always possible that we may have not specified the best estimation model, although the sensitivity analyses described above suggest that we specified the best model possible with the data series available.

Conclusions

In summary, the data presented here provide a holistic picture of changes in HIV prevalence among MSM in the largest MSAs in the United States across a two-decade time period. Specifically, mean (across all 86 MSAs) estimated HIV prevalence among MSM increased steadily from approximately 11% in 1992 to approximately 20% in 2013. This trend of steadily increasing prevalence was estimated for most individual MSAs as well. As ARV treatment has become both more effective and accessible, and the lifespan of HIV-positive MSM has increased, HIV prevalence has increased consistently, and almost uniformly, across large metropolitan areas in various regions of the country. Increasing HIV prevalence and increased lifespan of HIV-positive MSM has created new challenges related to adequately addressing the healthcare needs of this population. In addition, it is likely that very high HIV incidence rates among racial/ethnic minority MSM contribute substantially to increasing HIV prevalence rates among MSM. Our estimates for racial/ethnic subgroups of MSM found that HIV prevalence was markedly higher in all years and all geographic regions among Black MSM than among white MSM, and was markedly higher on average among Hispanic/Latino MSM than among white MSM (although the latter disparity is smaller). Continued and increased efforts and resources are needed that target racial/ethnic minority MSM in order to address these disparities. Future research is needed that estimates trajectories of HIV prevalence among MSM by age group, in order to ascertain the degree to which the increases in HIV prevalence found in the present study among MSM overall are indeed driven specifically by increases in HIV prevalence among older MSM. If so, new healthcare initiatives may be needed that focus on training health providers to address the cross-cutting health challenges of expanding comprehensive prevention (e.g., PrEP) and treatment services among racial/ethnic minority populations and of treating aging HIV-positive populations, as the proportion of MSM who are living with HIV continues to grow.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments:

References

1. Centers for Disease Control and Prevention. HIV Surveillance Report, 2017; vol. 29. http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html. Published 11 2018.

2. Centers for Disease Control and Prevention. Estimated HIV incidence and prevalence in the United States, 2010–2015. HIV Surveillance Supplemental Report 2018;23(No. 1). http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html. Published 3 2018.

3. Koblin Beryl A.; Torian Lucia V.; Guilin Vince; Ren Leigh; MacKellar Duncan A.; Valleroy Linda A. "High prevalence of HIV infection among young men who have sex with men in New York City" AIDS: 8 18th, 2000 - Volume 14 - Issue 12 - p 1793–1800. [PubMed: 10985317]

4. Osmond D, Pollack L, Paul J, Catania J, "Changes in Prevalence of HIV Infection and Sexual Risk Behavior in Men Who Have Sex With Men in San Francisco: 1997–2002"; AJPH, 2007, Vol 97, No. 9, 1677–83.

5. Sifikis J, Hylton J, Flynn C, Solomon L, MacKellar D, Valleroy L; "Prevalence of HIV Infection and Prior HIV Testing among Young Men Who have Sex with Men. The Baltimore Young Men's Survey", AIDS and Behavior, 2010, Volume 14, Issue 4, pp 904–912. [PubMed: 17968648]

6. Holmberg SD. The estimated prevalence and incidence of HIV in 96 large US metropolitan areas. American Journal of Public Health. 1996;86(5):642–654. [PubMed: 8629714]

7. Harawa N Greenland S, Bingham T, *; Johnson D, Cochran S, Cunningham W, Celentano D, Koblin B,, LaLota M, MacKellar D McFarland W Shehan D, Stoyanoff S, *; Thiede H, Torian L, Valleroy L; #"" Associations of Race/Ethnicity With HIV Prevalence and HIV-Related Behaviors Among Young Men Who Have Sex With Men in 7 Urban Centers in the United States", JAIDS Journal of Acquired Immune Deficiency Syndromes: 2004 - Volume 35 - Issue 5 - p 526–536. [PubMed: 15021318]

8. Valleroy L; . MacKellar D;Karon J, Rosen D, McFarland W; Shehan DA, Stoyanoff S,; LaLota M, Celentano D, ; . Koblin B, ; Thiede H , Katz M; Torian L; Janssen R; for the Young Men's Survey Study Group HIV Prevalence and Associated Risks in Young Men Who Have Sex With Men; JAMA. 2000;284(2):198–204. doi:10.1001/jama.284.2.198. [PubMed: 10889593]

9. Smith A ; Le B ; Finlayson T ; Oster A ; DiNenno E; Prevalence and awareness of HIV infection among men who have sex with men - 21 cities, United States, 2008., Morbidity and Mortality Weekly Report 2010 Vol.59 No.37 pp.1201–1207. [PubMed: 20864920]

10. Catania JA, Osmond D, Stall RD, Pollack L, Paul JP, Blower S, Binson D, Canchola JA, Mills TC, Fisher L, Choi KH, Porco T, Turner C, Blair J, Henne J, Bye LL, and Coates TJ "The continuing HIV epidemic among men who have sex with men." Am J Public Health. 2001 6; 91(6): 907–914. [PubMed: 11392933]

11. Lieb Spencer,, Prejean Joseph, Thompson Daniel R., Fallon Stephen J., Cooper Hannah, Gates Gary J., Liberti Thomas M., Friedman Samuel R., Malow Robert M.; "HIV Prevalence Rates Among Men Who Have Sex with Men in the Southern United States: Population-Based Estimates by Race/Ethnicity"; AIDS and Behavior, 2011, Volume 15, Issue 3, pp 596–606. [PubMed: 20872062]

12. Office of Management and Budget. Standards for defining metropolitan and micropolitan statistical areas. Federal Register 2000; 65:8228–8238.

13. Friedman SR, Tempalski B, Cooper H, Perlis T, Keem M, Friedman R, & Flom PL (2004). Estimating numbers of injecting drug users in metropolitan areas for structural analyses of community vulnerability and for assessing relative degrees of service provision for injecting drug users. Journal of Urban Health, 81(3), 377–400. [PubMed: 15273263]

14. Friedman SR, Tempalski B, Brady JE, Friedman JJ, Cooper HL, Flom PL, … & Des Jarlais DC (2007). Predictors of the degree of drug treatment coverage for injection drug users in 94 metropolitan areas in the United States of America. International Journal of Drug Policy, 18(6), 475–485. [PubMed: 18061873]

15. SAS Institute, Inc. (2015). SAS/STAT User's Guide: The MI Procedure. SAS Institute, Inc.: Cary, NC.

16. Rubin DB (1987). Multiple Imputation for Nonresponse in Surveys. New York:Wiley.

17. Schafer JL, & Olsen MK (1998). Multiple imputation for multivariate missing-data problems: a data analyst's perspective. Multivariate Behavioral Research. Special Issue: Innovative Methods for Prevention Research, 33, 545–571.

18. Graham JW, Olchowski AE, & Gilreath TD (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prevention Science, 8, 206–213. [PubMed: 17549635]

19. Williams LD, Tempalski B, Ibragimov U, Stall R, Satcher A, Wang G, Cooper HLF, & Friedman SR (2020). Regional variation in trends over time in HIV prevalence among people who inject drugs in 89 large U.S. metropolitan statistical areas, 1992–2013. Annals of Epidemiology.

20. IMS Health (2015). IMS LifeLink Longitudinal Prescription Claims (LRx) Database. IMS HEALTH Incorporated.

21. Rosenberg ES, Grey JA, Sanchez TH, & Sullivan PS (2016). Rates of prevalent HIV infection, prevalent diagnoses, and new diagnoses among men who have sex with men in US states, metropolitan statistical areas, and counties, 2012–2013. JMIR Public Health and Surveillance, 2(1).

22. Lieb S, Fallon SJ, Friedman SR, Thompson DR, Gates GJ, Liberti TM, & Malow RM (2011). Statewide estimation of racial/ethnic populations of men who have sex with men in the US. Public health reports, 126(1), 60–72.

23. Krentz HB, & Gill MJ (2015). Increased costs of HIV care associated with aging in an HIV-infected population. HIV medicine, 16(1), 38–47. [PubMed: 25105798]

24. Balderson BH, Grothaus L, Harrison RG, McCoy K, Mahoney C, & Catz S (2012). Chronic illness burden and quality of life in an aging HIV population. AIDS Care, 25(4), 10.1080/09540121.2012.712669

25. Samji Hasina, Cescon Angela, Hogg Robert S., Modur Sharada P., Althoff Keri N., Buchacz Kate, Burchell Ann N., Cohen Mardge, Gebo Kelly A., John Gill M "Closing the Gap: Increases in Life Expectancy among Treated HIV-Positive Individuals in the United States and Canada". PLOS; Published: 12 18, 2013, 10.1371/journal.pone.0081355.

26. Murray John M., McDonald Ann M., Law Matthew G. "Rapidly ageing HIV epidemic among men who have sex with men in Australia"Sexual Health 6(1) 83–86 10.1071/SH08063.

27. The Antiretroviral Therapy Treatment Collaboration "Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies", Lancet HIV 2017, 4, e349–356. [PubMed: 28501495]

28. Matthews DD, Herrick AL, Coulter RW, Friedman MR, Mills TC, Eaton LA, … & POWER Study Team. (2016). Running backwards: Consequences of current HIV incidence rates for the next generation of black MSM in the United States. AIDS and Behavior, 20(1), 7–16. [PubMed: 26267251]

29. Hall HI, Byers RH, Ling Q, & Espinoza L (2007). Racial/ethnic and age disparities in HIV prevalence and disease progression among men who have sex with men in the United States. American journal of public health, 97(6), 1060–1066. [PubMed: 17463370]

30. Centers for Disease Control and Prevention. Expanded Testing Program Initiative (ETP) data 2008–2013. Division of HIV/AIDS Prevention. National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention.

31. Oster AM, Johnson CH, Le BC, Balaji AB, Finlayson TJ, Lansky A, Mermin J, Valleroy L, MacKellar D, Behel S, & Paz-Bailey G (2014). Trends in HIV prevalence and HIV testing among young MSM: five United States cities, 1994–2011. AIDS and Behavior, 18(3), 237–247.
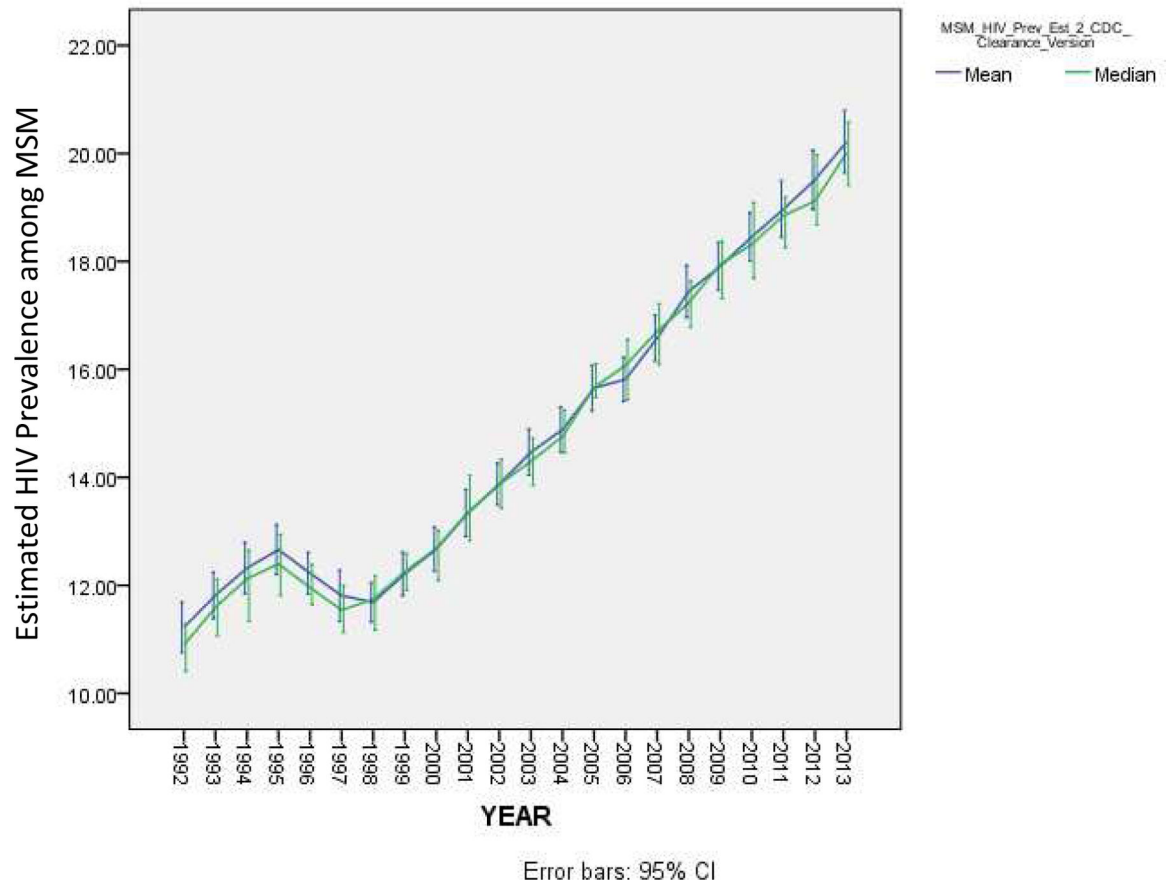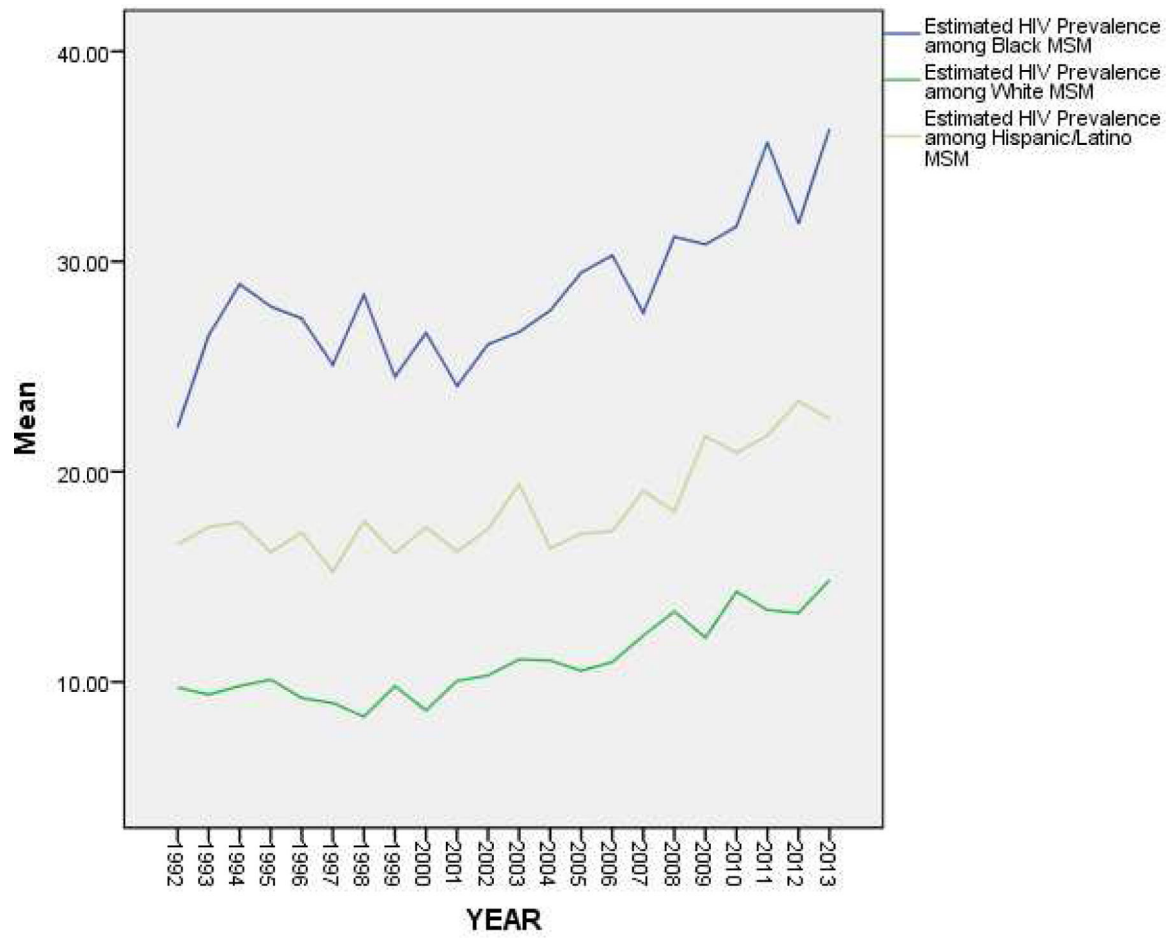
Figure 1.
Mean Trajectory of HIV Prevalence among Men who Have Sex with Men across 86
Metropolitan Statistical Areas, 1992–2013

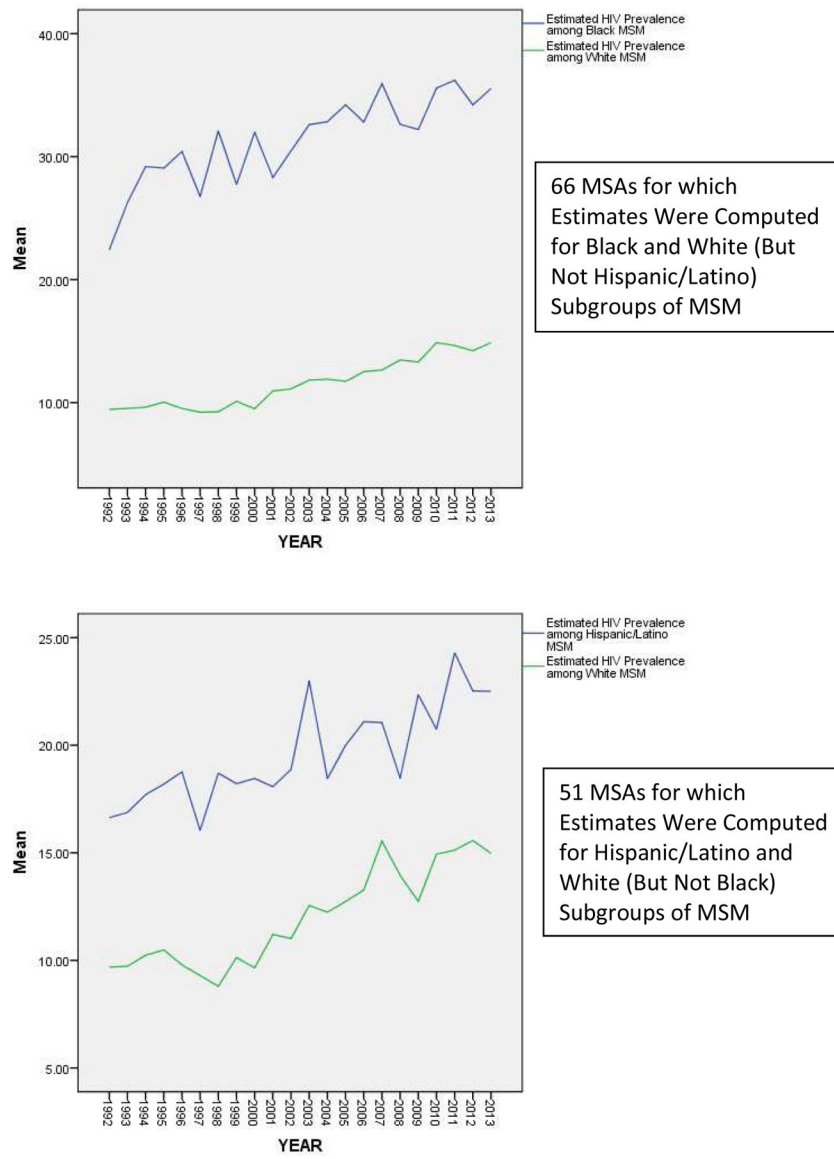**Figure 2.**
Estimated HIV Prevalence among Subgroups of MSM by Race/Ethnicity, Over Time (1992–2013) among 42 Metropolitan Statistical Areas
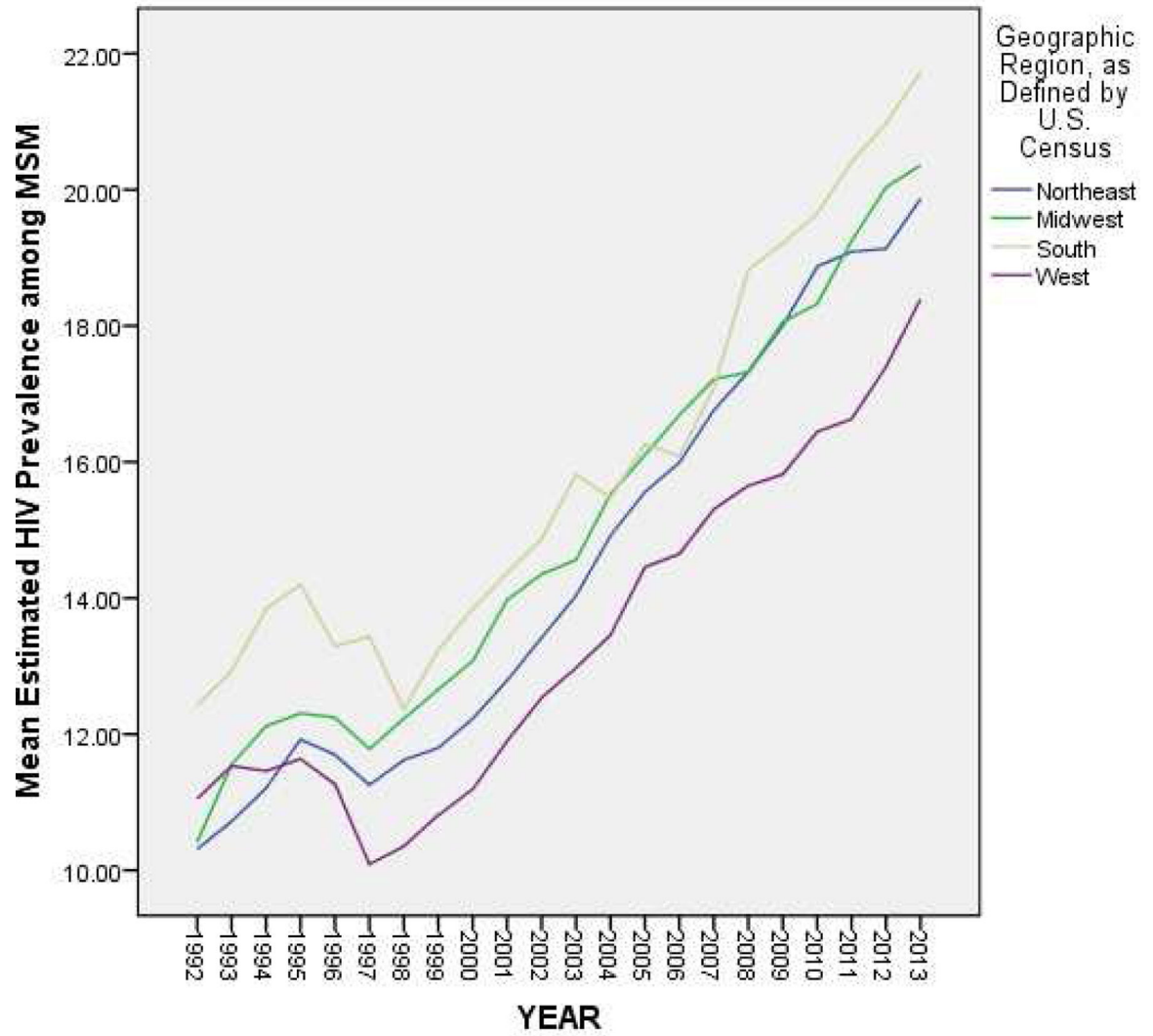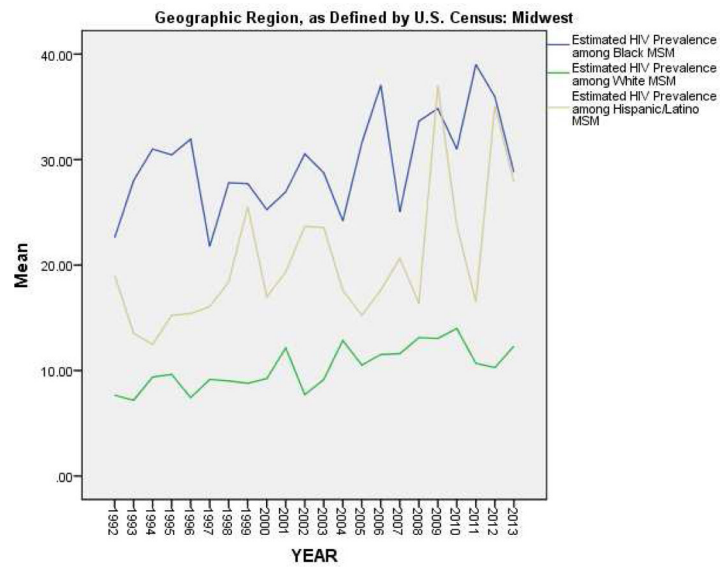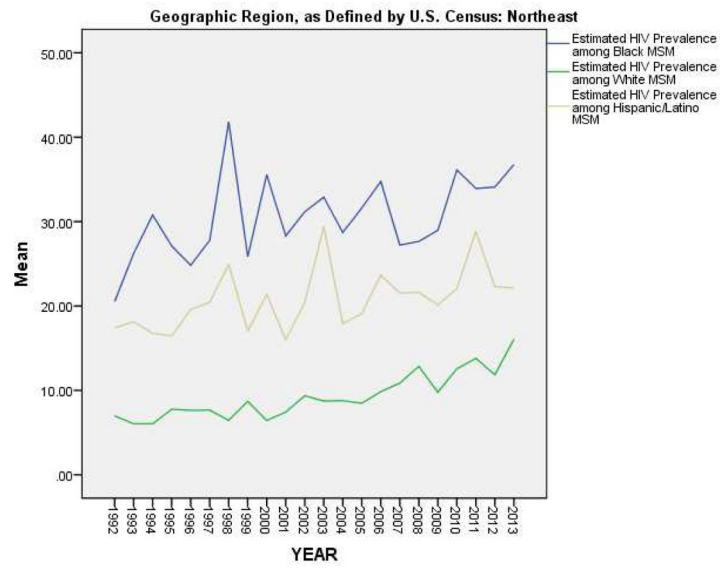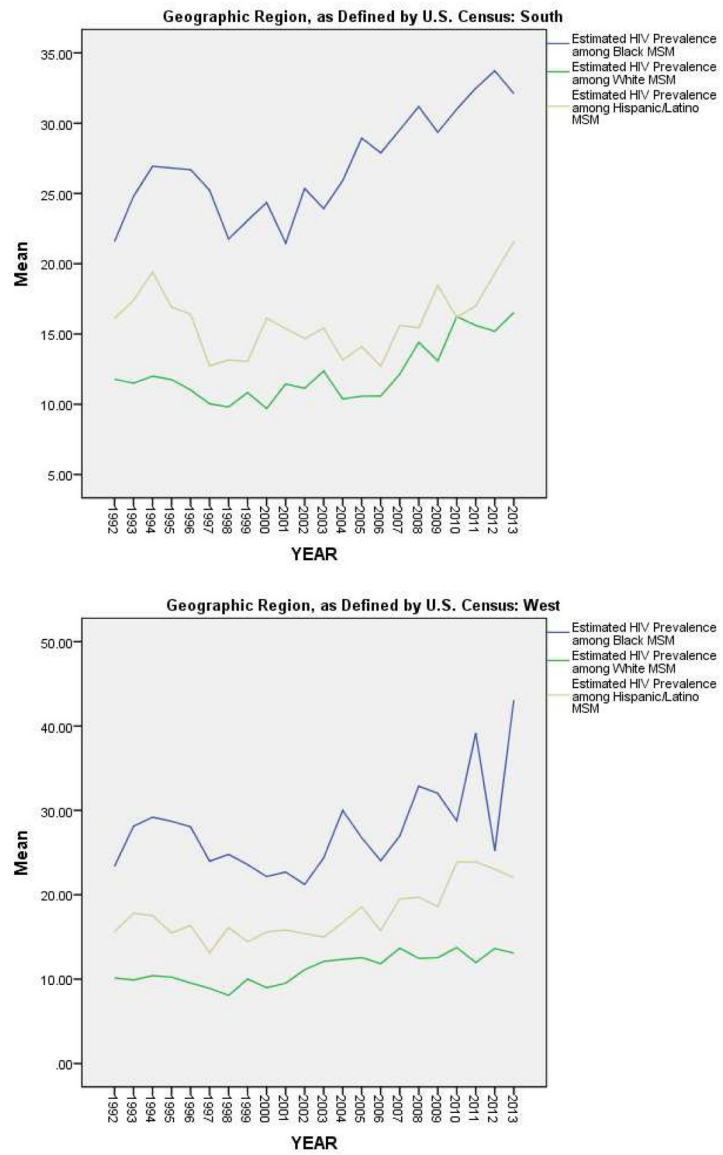
Figure 3.
Estimated HIV Prevalence among Subgroups of MSM by Race/Ethnicity, Over Time (1992–2013) among MSAs for which at Least Two Subgroups Were Computed

Figure 4.
Estimated Trajectories of HIV Prevalence among MSM by Geographic Region, Over Time (1992–2013)

Geographic Region, as Defined by U.S. Census: Northeast

Geographic Region, as Defined by U.S. Census: Midwest

Figure 5.
Estimated Trajectories of HIV Prevalence among MSM by Racial/Ethnic Subgroup, for Each Geographic Region, Over Time (1992–2013)

Table 1.

Description of Data Sources Used to Compute Estimates of HIV Prevalence among Men who Have Sex with Men

Database	Description and Characteristics
Counseling & Testing Services [*]	Individual reports of HIV tests at CDC-funded HIV testing sites (NHM&E) (1992–2013), including HIV Expanded Testing Program Initiative (ETP) in 2008–2013. Data are a count of total tests (and of positive tests) - not of unique individuals - and thus can count an individual more than once a year. Data reports are for MSM ages 15–64.
Deaths of people with HIV infection ever classified as stage 3 (AIDS)	Number of individuals with diagnosed HIV infection ever classified as AIDS who have died, as reported to CDC by state and local health departments through the CDC's National HIV Surveillance System. Deaths may have been due to any cause. These data report yearly number of such deaths among MSM for ages 15–64 for years 1992–2013.
PLWH ever classified as stage 3 (AIDS) [**]	Number of persons living with HIV infection ever classified as stage 3 (AIDS), as reported to CDC by state and local health departments through the CDC's National HIV Surveillance System. These data report yearly number of such cases among MSM, for ages 15–64 for years 1992–2013.
National HIV Behavioral Surveillance (NHBS) [***]	Data were collected as part of the CDC's National HIV Behavioral Surveillance (NHBS), a CDC-funded multi-city annual cross-sectional survey designed to characterize HIV prevalence, behavioral risks among high-risk populations and extent and nature of these populations' contact with HIV related services. Surveillance is conducted in rotating, annual cycles in three different populations at increased risk for HIV (MSM, people who inject drugs, and high risk heterosexuals). Data on MSM used in the present study were available for 2008 and 2011.
Holmberg Estimates	This series uses previously-published estimates by Holmberg from 1992 on HIV incidence and prevalence among MSM, ages 15–64. These estimates were created using multiple data sources: previously published studies, HIV cases reported to the CDC, and data from local clinics and testing sites. These estimates used methods and assumptions that are described in detail by Holmberg (1996). MSM were defined as men who had sex with other men in the last 12 months.
Population Size	U.S. Census Bureau's Population Estimates Program (PEP). Data for each MSAs are by race/ethnicity and ages 15–64 per 10,000 population for years 1990–2013.
Literature Estimates	Our research estimates were based on a review of published literature and conference abstracts, as well as web-based searches and inquiries of researchers to find HIV prevalence rate estimates among MSM in any of the 96 MSAs of interest. To be eligible, a study had to have been conducted during 1992–2013 and to have determined HIV serostatus through the testing of blood, urine, or saliva samples rather than through self-report. An additional inclusion criterion was that the study could not have been part of the CDC NHM&E HIV testing system (as we already captured NHM&E data directly and included it in the computation of our estimates). We identified eligible research-based estimates from 18 of 96 metropolitan areas totaling 132 data points over time. The annual number of research-based estimates was concentrated mainly within MSAs having substantial research institutions with an interest in drug using populations and tending to have highly populated central cities (i.e., New York City, Chicago, Baltimore, Los Angeles and San Francisco).

[*] Data were acquired through a special data request from CDC's HIV Counseling and Testing in Publicly Funded Sites; and HIV Expanded Testing Program Initiative (ETP).

[***] Data were acquired through a special data request to the CDC's National HIV Behavioral Surveillance.

Table 2.

MSAs which Were Excluded from Creation of Racial/ethnic Subgroups

	Black MSM	White MSM	Hispanic/Latino MSM
MSAs which Had to Be Excluded Due to Data Missingness	Albuquerque, NM; Allentown, PA; Bakersfield, CA; Charleston, SC; El Paso, TX; Fresno, CA; Greensboro, NC; Harrisburg, PA; Knoxville, TN; Providence, RI; Raleigh, NC; Salt Lake City, UT; Scranton, PA; Seattle, WA; Stockton-Lodi, CA; Syracuse, NY; Tacoma, WA; Tucson, AZ; Ventura, CA; Youngstown, OH	Charleston, SC El Paso, TX Greensboro, NC Harrisburg, PA Knoxville, TN Raleigh, NC Scranton, PA Seattle, WA Syracuse, NY Tacoma, WA Youngstown, OH	Akron, OH; Albany, NY; Ann Arbor, MI; Baltimore, MD; Buffalo, NY; Charleston, SC; Charlotte, NC; Cincinnati, OH; Columbus, OH; Dayton, OH; El Paso, TX; Gary, IN; Greensboro, NC; Greenville, SC; Harrisburg, PA; Indianapolis, OH; Jacksonville, FL; Knoxville, TN; Louisville, KY; Memphis, TN; Monmouth, NJ; Nashville, TN; Omaha, NE; Pittsburgh, PA; Raleigh, NC; Saint Louis, MO; Sarasota, FL; Scranton, PA; Seattle, WA; Syracuse, NY; Tacoma, WA; Toledo, OH; Tulsa, AZ; Wilmington, NE; Youngstown, OH

Table 3.

Results of One-way ANOVAs Assessing Differences between Geographic Regions for Each Racial/Ethnic Subgroup, Separately

		Mean (S.D.)	F (df)	p
Estimated HIV Prevalence among Black MSM	Northeast	30.67% (13.92)	4.39 (3, 802)	.004
	Midwest	30.07% (12.25)		
	South	27.06% (8.40)		
	West	27.83% (12.94)		
Estimated HIV Prevalence among White MSM	Northeast	9.29% (5.23)	16.73 (3, 802)	<.0005
	Midwest	10.47% (3.82)		
	South	12.33% (5.09)		
	West	11.16% (3.80)		
Estimated HIV Prevalence among Hispanic/Latino MSM	Northeast	20.81% (12.26)	11.76 (3, 802)	<.0005
	Midwest	20.80% (16.27)		
	South	16.10% (5.72)		
	West	17.75% (5.72)		