Distribution Agreement

Signature:

Adam Stephen Vaughan	Date

Connecting Race, Place and HIV Prevalence:
A County-Level Analysis

By

Adam Vaughan
Master of Public Health

Global Epidemiology

Patrick S. Sullivan, DVM, PhD
Committee Chair

R. Luke Shouse, MD, MPH
Committee Member

Connecting Race, Place and HIV Prevalence:
A County-Level Analysis

By

Adam S. Vaughan

Bachelor of Engineering
Vanderbilt University
1997

Master of Science
University of Texas Southwestern Medical Center
1999

Thesis Committee Chair: Patrick S. Sullivan, DVM, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Epidemiology
2012

Abstract

Connecting Race, Place and HIV Prevalence:
A County-Level Analysis
By Adam S. Vaughan

Background: Within the United States, disparities in HIV prevalence among the three major racial/ethnic groups (white, blacks and Hispanics) are striking. The pattern of disparities, with blacks and whites at the extremes and Hispanics in the middle, extends across risk groups and geographies. In an attempt to explain these disparities, recent analyses have found statistically identical HIV prevalence rates among heterosexuals in impoverished urban settings.

Objective: This analysis aimed to use publicly available data to further explore associations between poverty, race and HIV in the United States.

Methods: We used national HIV surveillance data and other publicly available data to examine the interaction between race and poverty on county level HIV prevalence by urbanization. A negative binomial model, stratified on urbanization, was constructed with race-specific county-level HIV prevalence rate as the outcome and race, poverty and other social determinants of health as covariates. Prevalence rate ratios (PRR) were calculated for black and Hispanic populations, relative to white populations, at each level of urbanization and for various levels of poverty.

Results: The results indicated that poverty is associated with HIV prevalence only in major metropolitan areas. In non-urban areas, poverty was not associated with HIV prevalence. High levels of poverty eliminated or attenuated racial/ethnic disparities only in counties belonging to major metropolitan areas. Additionally, the relationship between HIV prevalence and poverty varied by race, with blacks having a stronger association that either whites or Hispanics.

Discussion: High poverty may explain HIV prevalence rate disparities in major metropolitan counties. However, outside of these counties, other factors, including sexual networks, must be examined to explain racial/ethnic rate disparities. These results indicate that future HIV prevention efforts must consider both location and race.

Connecting Race, Place and HIV Prevalence:
A County-Level Analysis

By

Adam S. Vaughan

Bachelor of Engineering
Vanderbilt University
1997

Master of Science
University of Texas Southwestern Medical Center
1999

Thesis Committee Chair: Patrick S. Sullivan, DVM, PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Global Epidemiology
2012

Acknowledgements

This thesis was possible only through the support – technical, conceptual and emotional – from a number of people:

Patrick Sullivan, for his confidence that this analysis was possible, unique and relevant;

Luke Shouse, for helping me to consider how prevention follows from analysis;

Eli Rosenberg, for spending the past year discussing the details of this analysis (and the many variations that preceded it) and for always allowing a quick question to morph into a thirty minute discussion;

My family and friends, for supporting me over the past two years as I pursued what I've always wanted to do – and never suggesting that I was crazy for doing it;

And, lastly, my dad who, after struggling with a Styrofoam head, sheets of acrylic and routed pieces of plywood during my first Master's degree, never received the acknowledgement that he deserved.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

INTRODUCTION

Within the United States, disparities in HIV prevalence among the three major racial/ethnic groups (white, blacks and Hispanics) are striking. At the end of 2009, 43% of people living with an HIV diagnosis were black, 35% white and 19% Hispanic (1). At the same time, blacks comprised only 13% of the population and Hispanics comprised 16% of the population (2). Disparities were even starker relative to population size. At the end of 2009, the estimated HIV prevalence rate in blacks was 952 per 100,000, a rate approaching the limit of a generalized epidemic (1). Among Hispanics, the prevalence rate was 320 per 100,000. Compared to whites (with a prevalence rate of 144 per 100,000), blacks were 6.6 times more likely to be living with HIV and Hispanics were 2.2 times more likely.

This pattern of disparities, with blacks and whites at the extremes and Hispanics in the middle, extends across risk groups and geographies. In 2009, among men living with HIV who were infected through male-to-male sexual contact, 29.2% were black and 20.1% were Hispanic (1). Similarly, 63.2% of men who contracted HIV through heterosexual contact were black and 21.2% were Hispanic. Black females comprised 59.0% of all women living with HIV, while Hispanic females were 19.3% (3). Geographic disparities also exist, with the burden of the HIV epidemic focused in America's cities. Metropolitan areas with populations greater than 500,000 persons accounted for 81.4% of all prevalent HIV cases in 2009 and had a prevalence rate 3.2 times that of nonmetropolitan areas (1).

A number of mechanisms, primarily structural and social factors, have been proposed to explain these stark racial/ethnic differences (4). Theoretical models have demonstrated that the disruption of minority social structures caused by institutionalized racism would disperse HIV away from urban centers (5). Other structural factors such as racism, oppression and lack of trust in government may hinder prevention outreach. Social constructs, such as homophobia and HIV stigma, and decreased healthcare resources may discourage open discussion of risk behaviors and limit HIV testing and treatment. Additionally, black men are incarcerated at much greater rates than white men and, during incarceration, are exposed to an environment in which HIV rates may be as high as 1.8% (6). Poverty, which connects all of these factors, has therefore frequently been suggested as a candidate for strong associations with racial/ethnic health disparities, including HIV.

However, potential relationships between social determinants and HIV are based on theory and plausibility, rather than a rigorous epidemiologic treatment. A thorough study of these associations must address the United States' complex relationships between race, poverty and geography. In 2009, the proportion of black and Hispanic Americans living in poverty was roughly twice the percent of white Americans living in poverty (7). This relationship is intensified in rural areas, which are, on average, poorer than urban areas. The most rural counties are the poorest and, with decreasing levels of urbanization, higher proportions of all races/ethnicities live in poverty as compared to urban areas.

Prior analyses of national surveillance and national survey data have focused on HIV's association with poverty in urban areas among specific populations. Studies have found no disparities in HIV prevalence rates among heterosexuals in urban settings when races live under the same impoverished conditions (8,9). Analysis of National HIV Behavioral Surveillance Survey (NHBS) data showed that, among heterosexuals living in urban areas with high AIDS prevalence, HIV prevalence was 2.3 times higher in individuals living below the poverty line (9). After adjusting for this association with poverty, no racial differences in prevalence rates were observed. More recent analysis of surveillance data confirmed the complex associations between social determinants of health and AIDS, with area-based measures of socioeconomic variables confounding associations with AIDS diagnosis rates (10). Similarly, the elimination or attenuation of racial/ethnic disparities in high poverty, urban areas have also been observed for chronic diseases and other health outcomes (11).

In order to address these sociodemographic complexities, we used publicly available data to examine the interaction between race and poverty on county level HIV prevalence by urbanization. We hypothesized that black/white and Hispanic/white disparities would be eliminated in all strata of urbanization after controlling for poverty. The results indicated that poverty is associated with HIV prevalence only in major metropolitan areas and that, in these counties, high poverty may explain HIV prevalence rate disparities. Outside of these urban areas, poverty was not associated with HIV prevalence and, therefore, other factors must be examined to explain racial/ethnic rate disparities.

METHODS

To evaluate the associations of poverty and race with HIV prevalence at different levels of urbanization, we used publicly available data to create a statistical model of county-level, race-specific HIV rates as a function of social determinants of health. County-level, race-specific case counts of persons aged ≥13 years living with a diagnosis of HIV at the end of 2008 were obtained from national HIV surveillance data (Centers for Disease Control and Prevention (CDC), presented through AIDSVu.org) (12). Race-specific population estimates for each county in 2008 were obtained from the U.S. Census Bureau for use as rate denominators. For a given county, white, black and Hispanic populations were included if: the number of prevalent race-specific cases was ≥five, the race-specific population was at least 1,000 persons, the state authorized CDC's release of data and the county was represented in the data by the white population and either the Black or Hispanic population. The former two criteria were implemented to protect the confidentiality of persons included in the case data.

A stratified negative binomial model was constructed with race-specific HIV case counts as the outcome and poverty and race as the exposures of interest. Race-specific poverty was included as a continuous predictor representing the percent of the race living in poverty within the county. Race was included in the model as two indicator variables for black and Hispanic, with white as the reference. The model accounted for potential differences in the association between poverty and race using interaction terms. The model accounted for potential differences in the associations between poverty and race by urbanization by stratifying on National Center for Health Statistics (NCHS) urbanization level (13). This categorization includes six levels, where the two highest levels represent core and fringe counties of metropolitan areas with population ≥1 million and the lowest represents purely rural, noncore counties.

Additional continuous predictors were included in the model as potential confounders due to previously documented associations with HIV prevalence. All data were publicly available (Table 1). These predictors were: race-specific educational attainment and county-level estimates of drug use, men who have sex with men (MSM) population, health insurance, correctional population and per capita health expenditures (6,14).

The model was constructed by first evaluating all potential covariates for collinearity (Collinearity SAS macro obtained from David Kleinbaum, Emory University, Atlanta, GA). For any given variable, collinearity was defined as having a condition index (CI) greater than 30 and a variance decomposition proportion (VDP) greater than 0.5. Covariates were singly removed from the model until collinearity was no longer present.

Confounding by covariates was evaluated using an all-possible subsets approach (15). Poverty, race and the corresponding interaction terms were treated as exposures and, therefore, could not be eliminated from the model during this process. Estimates of black and Hispanic prevalence rate ratios (PRR) at each level of urbanization were calculated at 10, 20 and 30% poverty for models representing all possible subsets of potential confounding variables. Therefore, for each model, 36 PRR estimates were compared to 36 PRRs from the fully adjusted (gold standard) model. PRRs within 10% of the gold standard PRRs were not considered to be confounded.

Models that were not confounded were then evaluated for precision through examination of PRR confidence interval width.

Following evaluation of collinearity and confounding and creation of the final model, PRRs for blacks and Hispanics (compared to whites) at each level of urbanization were calculated. Model-based PRRs were calculated using the same poverty level for both races and were evaluated at 10, 20 and 30% poverty. These values were based on the United States Federal definition of a poverty area (greater than 20% of the population with incomes below the poverty line) and the range of race-specific poverty in the data (16). Pairwise comparisons of model based PRRs between strata of urbanization were evaluated using Wald tests.

All analyses were completed using SAS v9.2 (Cary, NC). This analysis used summarized county-level data and was therefore exempt from IRB requirements.

RESULTS

Model selection

The outcome variable (number of race specific prevalent HIV cases) was overdispersed (mean=258, variance=1053310). Therefore, the negative binomial was selected as the most appropriate distribution for the model.

The full model containing all possible covariates was first evaluated for collinearity. Education was found to be highly collinear with poverty and was eliminated from the model. Variables remaining in the subsequent model were not collinear. This model (predicting race-specific HIV prevalence rates by race-specific poverty, race, poverty-race interaction, health insurance coverage, healthcare expenditures, MSM population, correctional population and drug use) was then defined as the fully adjusted, gold standard model.

Using the all possible subsets approach, 30 models were evaluated for confounding. This assessment found that no model adequately controlled for confounding as compared to the gold standard model (Table 2). The individual removal of variables representing MSM population, health insurance coverage and drug use from the gold standard model uniformly resulted in confounding toward the null. Removal of the variable representing correctional populations from the gold standard model resulted in confounding in both directions relative to the null. Removal of the variable representing healthcare expenditures from the gold standard model resulted in confounding away from the null.

Model precision was not assessed since the gold standard was the only model producing unconfounded results.

The final model produced adjusted estimates of the PRR according to Equation 1. This model allowed the PRR to vary with both race and poverty, and thus permitted an examination of the effect of race of HIV prevalence while controlling for both poverty and urbanization.

(1) $PRR_i = \beta_{ib}BLACK + \beta_{ih}HISP + \beta_{ipb}POV \cdot BLACK + \beta_{iph}POV \cdot HISP$

where i represented the six strata of NCHS urbanization, β_i were the regression coefficients for the given exposures, BLACK and HISP were indicator variables for black and Hispanic populations, respectively, and POV was race-specific poverty.

Model results

A total of 2,865 county-level race-specific HIV prevalence rates were included in this analysis, representing 1,150 counties and 37 states (Figure 1). Counties from each of the six levels of urbanization were included (Table 3). Each county in the most urban (i.e. large central metro) categorization contributed data for each of the three race groups. Most counties contributed black prevalent HIV cases to the analysis. Compared to white and black percentages, a smaller percentage of counties contributed Hispanic prevalent HIV cases. With decreasing levels of urbanization, a smaller proportion of counties contributed data for each race group. Counties in the two most urban categories (large central metro and large fringe metro) comprised 76.4% of all prevalent HIV cases, 76.4% of white cases, 74.3% of black cases and 82.8% of Hispanic cases.

Disparities in poverty rates existed among the three racial/ethnic groups (Figure 2). While the ranges of county-level poverty rates for all three racial/ethnic populations overlapped, the interquartile ranges (IQR) for white populations did not overlap the IQRs for black and Hispanic populations.

Generally, crude prevalence rates for blacks, whites and Hispanics decreased with increasing urbanization (Figure 3). However, crude Hispanic prevalence rates rose in the most rural classification. The crude black prevalence rate was four times the crude white prevalence rate in large central metro counties, and between seven and eight times the crude white rate for all other levels of urbanization. Hispanic/white PRRs were lower than black/white PRRs, showing little variation in all but the most rural (noncore) counties.

In the final model, race-specific poverty was a significant predictor of HIV prevalence in large central metro and large fringe metro counties (i.e. counties belonging to metropolitan areas with population $\geq$1 million) (Table 4). Poverty was not a significant predictor of HIV prevalence for other levels of urbanization. The indicator variable representing black populations was a significant predictor of HIV prevalence at all levels of urbanization. However, the black-poverty interaction term was significant only for the two most urban classifications. The indicator variable for Hispanic populations was not significant for large central metro counties, but was significant for all other strata of urbanization. However, the poverty-Hispanic interaction term was not significant at any level of urbanization.

Black/white and Hispanic/white model-based, adjusted PRRs exhibited similar patterns within strata of urbanization (Table 5). In the most urban (large central metro) counties, the black/white PRR was statistically greater than 1.0 at 10% poverty (3.3, 95% CI: (2.3, 4.8)), but not statistically different from 1.0 at poverty levels of 20% (1.6, 95% CI: (0.8, 3.0)) and 30% (0.8, 95% CI: (0.2, 2.3)) (Figure 4). For Hispanics in these counties, the adjusted model-based Hispanic/white PRR was not statistically different from 1.0 at poverty levels of 10% (0.7, 95% CI: (0.6, 1.0)), 20% (0.6, 95% CI: (0.3, 1.1)) or 30% (0.4, 95% CI: (0.1, 1.4)) (Figure 5).

In large fringe metro counties, the black/white PRR significantly decreased with increasing poverty, but remained significantly over 1.0. Within the same strata of urbanization, Hispanic/white PRRs followed the same statistically significant trend, although the PRR at 30% poverty was not statistically different from 1.0 (1.6, 95% CI: (0.8, 3.5)).

For all remaining levels of urbanization at race specific poverty levels between 10 and 30%, poverty did not significantly modify the association between race and HIV prevalence for both blacks and Hispanics relative to whites. All PRRs within these levels of urbanization were statistically greater than 1.0.

Black/white and Hispanic/white PRRs showed similar patterns of significance between levels of urbanization. Black/white PRRs evaluated at 20% poverty were significantly different between large central metro counties and all other levels of urbanization and between large fringe metro counties and all other levels of urbanization (Figure 4). No other pairwise comparisons of black/white PRRs between strata of urbanization were significant.

Hispanic/white PRRs evaluated at 20% poverty in the most urban classification were significantly different from all other levels of urbanization (Figure 5). Hispanic/white PRRs in large fringe metro and medium metro classifications were significantly different from the remaining levels of urbanization. Hispanic/white PRRs in the most rural (noncore) counties were significantly different from all other classifications of urbanization.

DISCUSSION

Using population based HIV surveillance data, we investigated the relationships between HIV prevalence, race and poverty across levels of urbanization. We observed that, for both blacks and Hispanics, poverty was significantly associated with HIV prevalence only in counties within major metropolitan areas (population ≥1 million). Poverty was not associated with HIV prevalence outside of these major metropolitan areas. Additionally, we observed that the relationship between poverty and HIV prevalence within strata of urbanization differed for blacks and Hispanics.

The association between individual level poverty and HIV is well known, especially in urban areas (17–25). Our analysis found that adjusting for poverty eliminates or attenuates racial/ethnic HIV prevalence disparities in counties belonging to major metropolitan areas. This finding supports and extends prior analyses which found similar results among heterosexuals from select cities (8,9). Our analysis expands previous reports by including prevalent HIV cases resulting from all modes of transmission from a larger number of cities and by accounting for factors, such as health insurance coverage and healthcare utilization, which could be on the causal pathway between poverty and HIV prevalence. The inclusion of these factors in the current analysis suggests that the association between HIV and poverty is not mediated solely by these factors.

Furthermore, our analysis expands on prior results by including non-urban areas. The association we observed between poverty, race and HIV prevalence in the most urban counties was not detected in counties at lower levels of urbanization. In these non-metropolitan counties, we observed that, despite adjusting for poverty, severe racial/ethnic disparities in HIV prevalence remained.

This finding, that controlling for poverty eliminates racial/ethnic HIV disparities in major metropolitan counties, but has no association in non-urban counties, may indicate that the drivers of disparities differ by level of urbanization. That is, where urban HIV disparities appear to be driven by poverty, poverty does not drive non-urban HIV. Differences in access to healthcare have been suggested as a primary motivator for this rural/urban difference. Residents of rural areas have been reported to be medically underserved and, as a result, have reduced access to care and treatment (26). However, this analysis attempted to control for these factors by including health insurance coverage and healthcare expenditures as covariates in the model. Since the inclusion of these variables did not attenuate disparities, other factors distinguishing the HIV epidemic in urban and rural America must be explored.

The connection between population movement and sexual networks may represent a factor explaining the lack of an association between HIV prevalence and poverty in non-urban America. Migration within the United States from urban to rural has been associated with increased HIV prevalence. A study of rural HIV in the Southeastern United States found that one in five HIV-positive individuals were likely infected in urban areas, but that two in three had lived in urban areas prior to diagnosis (27). Additionally, immigration from Latin America to the rural United States has been associated with increased HIV prevalence. Among Hispanic men who immigrated to rural communities, cultural differences, job instability and physical, cultural and social isolation have been associated with increased risk HIV risk behaviors (28,29).

In both cases, this movement of individuals to rural areas may create geographically overlapping sexual networks, with high prevalence areas (both urban and international) increasing HIV prevalence among racial/ethnic minorities. Additionally, given the low population density, sexual networks within rural America also exhibit extreme concurrency (30). Among racial/ethnic minorities, this low population density is compounded by socioeconomic factors, such as a low ratio of men to women (due to incarceration), discrimination and economic oppression (31). If these highly concurrent rural sexual networks are then linked to high prevalence urban sexual networks, racial/ethnic disparities could result.

We also found that the association between poverty and HIV in black populations differs from that in Hispanic populations. The associations between poverty and HIV prevalence were not statistically different for white and Hispanic populations at all levels of urbanization. However, the association between poverty and HIV prevalence was stronger for black populations than for white populations only in major metropolitan areas. Given that we controlled for healthcare system factors, this difference in the effect of poverty may indicate structural differences (such as racism or stigma) between these three racial/ethnic groups.

Overall, this analysis presents a novel picture of the HIV epidemic in the United States. By focusing on counties at all levels of urbanization from across the country, we have elaborated on the association between poverty, race and HIV. These findings suggest an approach to prevention that differs based on level of urbanization and is focused on poverty in urban areas and race in non-urban areas.

Within major metropolitan areas (which include over 80% of all HIV cases), the relationships between poverty and HIV prevalence are clear. Structural prevention efforts focused on the urban poor, regardless of race, may decrease the number of cases and mitigate or eliminate disparities in major metropolitan areas. These prevention efforts may directly target poverty (such as microfinance) or may attempt to increase HIV awareness among impoverished populations. Although untested in the United States, microfinance programs have shown promise in reducing HIV risk factors (32).

Major urban areas are currently the focus of national prevention strategies. The current National HIV/AIDS Strategy calls for greater prevention "in communities where HIV is most heavily concentrated"(33). This recommendation has been advanced through the Enhanced Comprehensive HIV Prevention Planning (ECHPP) project which targets HIV prevention in the twelve U.S. cities with the highest AIDS prevalence (34).

However, although non-metropolitan counties contribute less than 20% of national HIV cases, disparities are greater in these counties than in the most urban counties and are not associated with poverty. This suggests that a focus on disparity reduction may necessitate an additional prevention strategy for non-urban areas of the country, with interventions focused more on decreasing risk factors among racial and ethnic minorities – whether through intervention at the individual or social level. These interventions could be based on reducing the size of sexual networks or on reducing stigma and racism in these areas.

Limitations

Limitations of this analysis are those typical of ecologic analyses, specifically the ecological fallacy. We assumed that county-level data, including poverty, represented HIV positive individuals within the county. In high density counties, this assumption may be more valid than in less dense counties. In urban areas, the percent of the population living in poverty represents large numbers individuals living both above and below the poverty line. As a result, HIV positive individuals were pulled from a large population that may share similar characteristics. However, in rural areas, impoverished individuals (who could be at increased risk of HIV) may live in counties with low poverty rates and, consequently, population values may not reflect the individual. Therefore, ecologic bias could result in failure to detect a difference in more rural areas, but no bias in areas with large populations. Although this analysis found differences between urban and non-urban areas, the estimates for non-urban areas were similar for a wide range of populations. Therefore, even in non-urban areas, we believe that the ecologic bias is minimal.

This analysis may also have been affected by selection bias. We attempted to account for potential selection bias by requiring the inclusion a county's white population in order for the county's black or Hispanic population to be included. Excluded counties had either a low population (<1,000) of a low number of cases (<5). These criteria led to the inclusion of fewer race-specific populations with increasing levels of urbanization. Over 85% of counties contributed both black and white populations to the analysis across all levels of urbanization. However, outside of core metro counties, the percent of counties contributing Hispanic populations ranged from 17 to 72%. As a result, estimates for blacks and whites may be less subject to selection bias, but estimates for Hispanics may be biased. Since low case counts were suppressed, this likely increased the observed association (i.e. biased the results away from the null).

While this analysis used all available data, this small sample size, especially for Hispanic populations living in non-core counties, could have limited the power to detect associations between race and poverty. This small sample size would likely lead to estimates away from the null, since high prevalence populations were included.

Selection bias was also possible due to the distribution of the counties included in the analysis. The demographics of the United States, with high populations on both coasts, and the epidemiology of the HIV epidemic, led to a non-uniform distribution of counties across the country. The West was severely underrepresented in the analysis. Previous studies have observed the regional variation in HIV prevalence, with the epidemic concentrated in the Northeast and South and with higher prevalence in these two regions (8,9,26). Most included counties in this analysis were from these two regions. Therefore, generalization of our results may be better restricted to the South and Northeast; generalization of results to the West and Midwest should be made with caution.

Finally, this analysis assumes independence of rates within and between counties. Spatial correlation of rates between counties was accounted for by using the NCHS Urban-Rural Classification Scheme for Counties. In this categorization, the definitions of classes of urbanization partially account for proximity to major metropolitan areas and, therefore, may adequately account for spatial autocorrelations. Additionally, since race-specific rates within counties as associated with similar county services, sexual networks and sociodemographic factors, intra-county rates may be correlated. Both of these possible sources of bias likely biased estimates away from the null, since correlation could have accounted for some of the prevalence disparities.

Future Directions

Our analysis found differences in the associations between race and poverty that varied by race and level of urbanization. Although we proposed possible mechanisms for these distinctions, additional research is needed. From the urban/non-urban perspective, research should focus on sexual networks and, more specifically, on race-specific assortativity, concurrency and overlap between rural and urban networks. Given that poverty drives disparities in the urban epidemic, but has no association in the non-urban epidemic, these individual level analyses may help to elucidate one driver of racial/ethnic disparities.

As a continuation of county-level analyses, spatial analyses of these data would account for geographic correlation of HIV. By incorporating distances between urban and rural counties, clusters of high prevalence could be detected, thus suggesting the geographic range of sexual networks and the potential influence of urban HIV on non-urban rates. Spatial analysis would also implicitly account for correlations between races with counties, thus providing more precise estimates of association.

CONCLUSION

This analysis focused on three keys factors in the practice of public health: place, race and poverty. In doing so, it reinforced the complexities of understanding relationships among these factors, but also identified nuances in associations between race, poverty, and HIV prevalence. We have observed, to the best of our knowledge for the first time, county-level differences in racial/ethnic disparities by level of urbanization. We found that, while poverty drives disparities in the most urban counties, different drivers must exist to explain the high disparities outside of major metropolitan areas. Additional research should focus on defining these drivers in order to further target prevention efforts. These potential interventions must be specific to the level of urbanization, with urban interventions addressing poverty and its associated social structures and non-urban interventions potentially addressing sexual networks and risk behaviors among racial/ethnic minorities. Specific urban interventions could encompass microfinance or stigma reduction, while non-urban interventions could focus on individual risk factors. While this analysis cannot identify the potential efficacy of specific interventions, it can inform their development by stating the necessity to consider both race and place in HIV prevention in the United States.

REFERENCES

1. Centers for Disease Control and Prevention. HIV Surveillance Report: Diagnoses of HIV Infection and AIDS in the United States and Dependent Areas, 2010. 2012.
2. U.S. Census Bureau. Section 1: Population. In: *Statistical Abstract of the United States: 2011*. Washington, DC: 2010:1-62.
3. United States Department of Agriculture Economic Research Service. Rural Poverty at a Glance. 2004.
4. Centers for Disease Control and Prevention. A Heightened National Response to the HIV/AIDS Crisis Among African Americans. Atlanta, GA: 2007.
5. Wallace R, Wallace D. U.S. Apartheid and the Spread of AIDS to the Suburbs: A Multi-City Analysis of the Political Economy of Spatial Epidemic Threshold. *Social Science & Medicine*. 1995;41(3):333-345.
6. Bureau of Justice Statistics. HIV in Prisons, 2004. 2006.
7. U.S. Census Bureau. Section 13: Income, Expenditures, Poverty, and Wealth. In: *Statistical Abstract of the United States: 2012*. Washington, DC: 2010:431-470.
8. Denning P, Dinenno E. Communities in Crisis: Is There a Generalized HIV Epidemic in Impoverished Urban Areas of the United States? In: *XVIII International AIDS Conference*. Vienna: 2010:1.
9. Centers for Disease Control and Prevention. Characteristics Associated with HIV Infection Among Heterosexuals in Urban Areas with High AIDS Prevalence - 24 Cities, United States, 2006-2007. *Morbidity and mortality weekly report*. 2011;60(31):1045-9.

10. Song R, Hall HI, Harrison KM, et al. Identifying the impact of social determinants of health on disease rates using correlation analysis of area-based summary information. *Public health reports*. 2011;126(Suppl 3):70-80.

11. Laveist T, Pollack K, Thorpe R, et al. Place, not race: disparities dissipate in southwest Baltimore when blacks and whites live under similar conditions. *Health affairs*. 2011;30(10):1880-7.

12. Emory University. AIDSVu. (AIDSVu.org). (Accessed January 9, 2011)

13. Ingram D, Franco S. NCHS Urban-Rural Classification Scheme for Counties. *Vital Health Statistics*. 2012;2(154).

14. Andrulis DP. Access to Care Is the Centerpiece in the Elimination of Socioeconomic Disparities in Health. *Annals of Internal Medicine*. 1998;129(5):412-416.

15. Kleinbaum DG, Klein M. Modeling Strategy for Assessing Interaction and Confounding. In: *Logistic Regression: A Self-Learning Text*. New York: Springer; 2010:203-239.

16. U.S. Census Bureau. Poverty Areas. 1995;(http://www.census.gov/population/socdemo/statbriefs/povarea.html). (Accessed December 11, 2011)

17. Arnold MP, Raymond HF, Fisher HR, et al. Socioeconomic position and HIV risk-relevant behavior among lower-income heterosexuals in San Francisco. *AIDS and behavior*. 2011;15(2):499-506.

18. Ellerbrock TV, Chamblee S, Bush TJ, et al. Human immunodeficiency virus infection in a rural community in the United States. *American journal of epidemiology*. 2004;160(6):582-8.

19. Fife D, Mode C. AIDS Prevalence by Income Group in Philadelphia. *Journal of acquired immune deficiency syndrome*. 1992;5(11):1111-1115.

20. Hixson BA, Omer SB, del Rio C, et al. Spatial clustering of HIV prevalence in Atlanta, Georgia and population characteristics associated with case concentrations. *Journal of urban health*. 2011;88(1):129-41.

21. Krieger N, Chen JT, Waterman PD, et al. Painting a truer picture of US socioeconomic and racial/ethnic health inequalities: the Public Health Disparities Geocoding Project. *American journal of public health*. 2005;95(2):312-23.

22. Krueger LE, Wood RW, Diehr PH, et al. Poverty and HIV seropositivity: the poor are more likely to be infected. *AIDS*. 1990;4(8):811-814.

23. Simon PA, Hu DJ, Diaz T, et al. Income and AIDS Rates in Los Angeles County. *AIDS*. 1995;9(3):281-284.

24. Wiewel EW, Hanna DB, Begier EM, et al. High HIV prevalence and diagnosis rates in New York City black men. *Journal of community health*. 2011;36(1):141-9.

25. Zierler S, Krieger N, Tang Y, et al. Economic deprivation and AIDS incidence in Massachusetts. *American journal of public health*. 2000;90(7):1064-73.

26. Hall HI, Li J, McKenna MT. HIV in predominantly rural areas of the United States. *Journal of rural health*. 2005;21(3):245-53.

27. Lansky A, Nakashima A, Diaz T, et al. Human immunodeficiency virus infection in rural areas and small cities of the southeast: contributions of migration and behavior. *Journal of rural health*. 2000;16(1):20-30.

28. Winett L, Harvey SM, Branch M, et al. Immigrant Latino men in rural communities in the Northwest: social environment and HIV/STI risk. *Culture, health & sexuality*. 2011;13(6):643-56.

29. Apostolopoulos Y, Sonmez S, Kronenfeld J, et al. STI/HIV risks for Mexican migrant laborers: exploratory ethnographies. *Journal of immigrant and minority health*. 2006;8(3):291-302.

30. Stratford D, Ellerbrock TV, Chamblee S. Social organization of sexual-economic networks and the persistence of HIV in a rural area in the USA. *Culture, health & sexuality*. 2012;9(2):121-35.

31. Adimora AA, Schoenbach VJ, Doherty IA. HIV and African Americans in the southern United States: sexual networks and social context. *Sexually transmitted diseases*. 2006;33(Suppl 7):S39-45.

32. Stratford D, Mizuno Y, Williams K, et al. Addressing poverty as risk for disease: recommendations from CDC's consultation on microenterprise as HIV prevention. *Public Health Reports*. 2008;123(1):9-20.

33. Office of National AIDS Policy. National HIV/AIDS Strategy for the United States. Washington, DC: 2010.

34. Centers for Disease Control and Prevention. Enhanced Comprehensive HIV Prevention Planning (ECHPP) Project. (http://www.cdc.gov/hiv/strategy/echpp/pdf/echpp-factsheet.pdf). (Accessed , 2012)

TABLES AND FIGURES

Table 1 Predictors and data sources of HIV prevalence rates used in the model

Variable (Variable Name)	Data Source	Comments
Poverty (POV)	2005-2009 American Community Survey (ACS) 5-Year Estimates	Percent of the race living in poverty.
Educational attainment (ED)	2005-2009 American Community Survey (ACS) 5-Year Estimates	Percent of the race age ≥25 having earned at least a high school diploma or equivalent.
Urbanization (URB)	2006 National Center for Health Statistics (NCHS) Urban-Rural Classification Scheme for Counties	This categorization defines counties based on size and proximity to metropolitan areas. The six categories are: central counties of metro areas of ≥1 million population (large central metro), fringe counties of metro areas of ≥1 million population (large fringe metro), counties in metro areas of 250,000-999,999 population (medium metro), counties in metro areas of 50,000-249,999 population (small metro), micropolitan counties and noncore counties.
Drug use (DU)	2006-2008 National Surveys on Drug Use and Health, Substate Estimates from the Substance Abuse and Mental Health Services Administration (SAMHSA)	Percent of population aged 12 or older who used an illicit drug other than marijuana in past month. Each county within the defined substate region was assigned the same value.
Population of men who have sex with men (MSM)	2005-2009 American Community Survey (ACS) 5-Year Estimates	Percent of male-male unmarried partner households.
Health insurance coverage (HI)	2008 Small Area Health Insurance Estimates	Percent of the population under age 65 lacking health insurance.
Correctional population (CORR)	2010 United States Decennial Census	Percent of the county population in federal or state prisons
Healthcare expenditures (EXP)	2007 Health Resources and Services Administration (HRSA) Area Resource File	Per capita hospital expenditures.

Table 2 Results of confounding assessment

Model (Variables dropped from gold standard)	Number of confounded PRR estimates	
	Toward the null	Away from the null
A (CORR)	6	4
B (MSM)	10	0
C (DU)	12	0
D (HI)	10	2
E (EXP)	0	16
F (CORR, MSM)	14	2
G (CORR, DU)	18	2
H (CORR, HI)	13	5
I (CORR, EXP)	1	13
J (MSM, DU)	18	0
K (MSM, HI)	22	0
L (MSM, EXP)	1	13
M (DU, HI)	17	0
N (DU, EXP)	0	7
O (HI, EXP)	4	7
P (CORR, MSM, DU)	20	2
R (CORR, MSM, EXP)	1	9
S (MSM, DU, HI)	23	0
T (MSM, DU, EXP)	9	6
U (DU, HI, EXP)	11	5
V (CORR, DU, HI)	18	4
W (CORR, DU, EXP)	2	6
X (CORR, HI, EXP)	2	7
Y (MSM, HI, EXP)	11	8
Z (CORR, MSM, DU, HI)	24	2
AA (CORR, MSM, DU, EXP)	8	4
BB (CORR, DU, HI, EXP)	10	6
CC (MSM, DU, HI, EXP)	12	5
DD (CORR, MSM, HI, EXP)	6	6
EE (CORR, MSM, HI, EXP, DU)	11	5

Table 3 Distribution of counties, race groups and prevalent HIV cases by level of urbanization

	Total		**White**		**Black**		**Hispanic**	
Urbanization	**Number of Counties (%)**	**Number of Cases (%)**	**Number of Counties (%)**	**Number of Cases (%)**	**Number of Counties (%)**	**Number of Cases (%)**	**Number of Counties (%)**	**Number of Cases (%)**
Large central metro	57 (5)	434,979 (58.8)	57 (100)	141,582 (32.5)	57 (100)	193,337 (44.4)	57 (100)	100,060 (23)
Large fringe metro	234 (20.3)	130,408 (17.6)	234 (100)	48,775 (37.4)	223 (95.3)	63,245 (48.5)	165 (70.5)	18,413 (14.1)
Medium metro	203 (17.7)	103,901 (14)	203 (100)	40,045 (38.5)	196 (96.6)	46,494 (44.7)	147 (72.4)	17,362 (16.7)
Small metro	197 (17.1)	33,318 (4.5)	197 (100)	13,935 (41.8)	177 (89.8)	15,677 (47.1)	128 (65)	3,706 (11.1)
Micropolitan	290 (25.2)	27,749 (3.8)	290 (100)	9,748 (35.1)	248 (85.5)	15,103 (54.4)	127 (43.8)	2,898 (10.4)
Noncore	169 (14.7)	9,227 (1.2)	169 (100)	2,285 (24.8)	162 (95.9)	6,357 (68.9)	28 (16.6)	585 (6.3)
Total (%)	1,150 (100)	739,582 (100)	1,150 (100)	256,370 (34.7)	1,063 (92.4)	340,213 (46)	652 (56.7)	143,024 (19.3)

Table 4 Model coefficients for exposures of interest

Urbanization	**Poverty Coefficient (95% CI)**		**Black Coefficient (95% CI)**		**Hispanic Coefficient (95% CI)**		**Poverty-Black Interaction Coefficient (95% CI)**		**Poverty-Hispanic Interaction Coefficient (95% CI)**	
Large central metro	0.07	(0.02, 0.13)	1.94	(1.25, 2.63)	-0.04	(-0.64, 0.57)	-0.07	(-0.13, -0.02)	-0.03	(-0.08, 0.03)
Large fringe metro	0.05	(0.01, 0.08)	2.08	(1.78, 2.37)	0.79	(0.45, 1.13)	-0.04	(-0.07, 0)	-0.01	(-0.04, 0.02)
Medium metro	-0.01	(-0.04, 0.02)	1.94	(1.51, 2.37)	0.48	(0.05, 0.91)	0.00	(-0.03, 0.04)	0.03	(0, 0.06)
Small metro	0.00	(-0.02, 0.03)	2.26	(1.80, 2.71)	1.32	(0.84, 1.81)	-0.01	(-0.04, 0.02)	0.00	(-0.03, 0.02)
Micropolitan	0.00	(-0.02, 0.02)	2.30	(1.85, 2.74)	1.12	(0.61, 1.63)	0.00	(-0.03, 0.02)	0.02	(-0.01, 0.04)
Noncore	-0.01	(-0.03, 0.02)	2.41	(1.95, 2.87)	1.62	(0.84, 2.39)	0.00	(-0.03, 0.02)	0.02	(-0.01, 0.05)

Table 5 Black/white and Hispanic/white model-based PRRs by strata of urbanization

	Black/White PRR (95% CI)						**Hispanic/White PRR (95% CI)**					
Urbanization	**10% Poverty**		**20% Poverty**		**30% Poverty**		**10% Poverty**		**20% Poverty**		**30% Poverty**	
Large central metro	3.3	(2.3, 4.8)	1.6	(0.8, 3)	0.8	(0.2, 2.3)	0.7	(0.6, 1)	0.6	(0.3, 1.1)	0.4	(0.1, 1.4)
Large fringe metro	5.5	(4.6, 6.7)	3.9	(2.5, 5.9)	2.7	(1.3, 5.6)	2.0	(1.6, 2.4)	1.8	(1.2, 2.8)	1.6	(0.8, 3.5)
Medium metro	7.3	(5.9, 9.1)	7.7	(5.4, 10.9)	8.1	(4.3, 15.4)	2.2	(1.8, 2.8)	3.1	(2.1, 4.4)	4.2	(2.2, 8.1)
Small metro	8.6	(6.6, 11.3)	7.8	(5.9, 10.3)	7.0	(4.4, 11.3)	3.6	(2.7, 4.8)	3.5	(2.7, 4.6)	3.4	(2.1, 5.5)
Micropolitan	9.5	(7.4, 12.2)	9.1	(7.1, 11.6)	8.7	(5.6, 13.4)	3.6	(2.6, 4.8)	4.2	(3.2, 5.4)	4.9	(3.1, 7.6)
Noncore	10.7	(8.4, 13.6)	10.3	(8.1, 13.1)	9.9	(6.2, 15.8)	6.1	(3.7, 10.1)	7.4	(5.1, 10.8)	9.0	(5.4, 15.2)

Figure 1 Race groups included in the analysis, by county

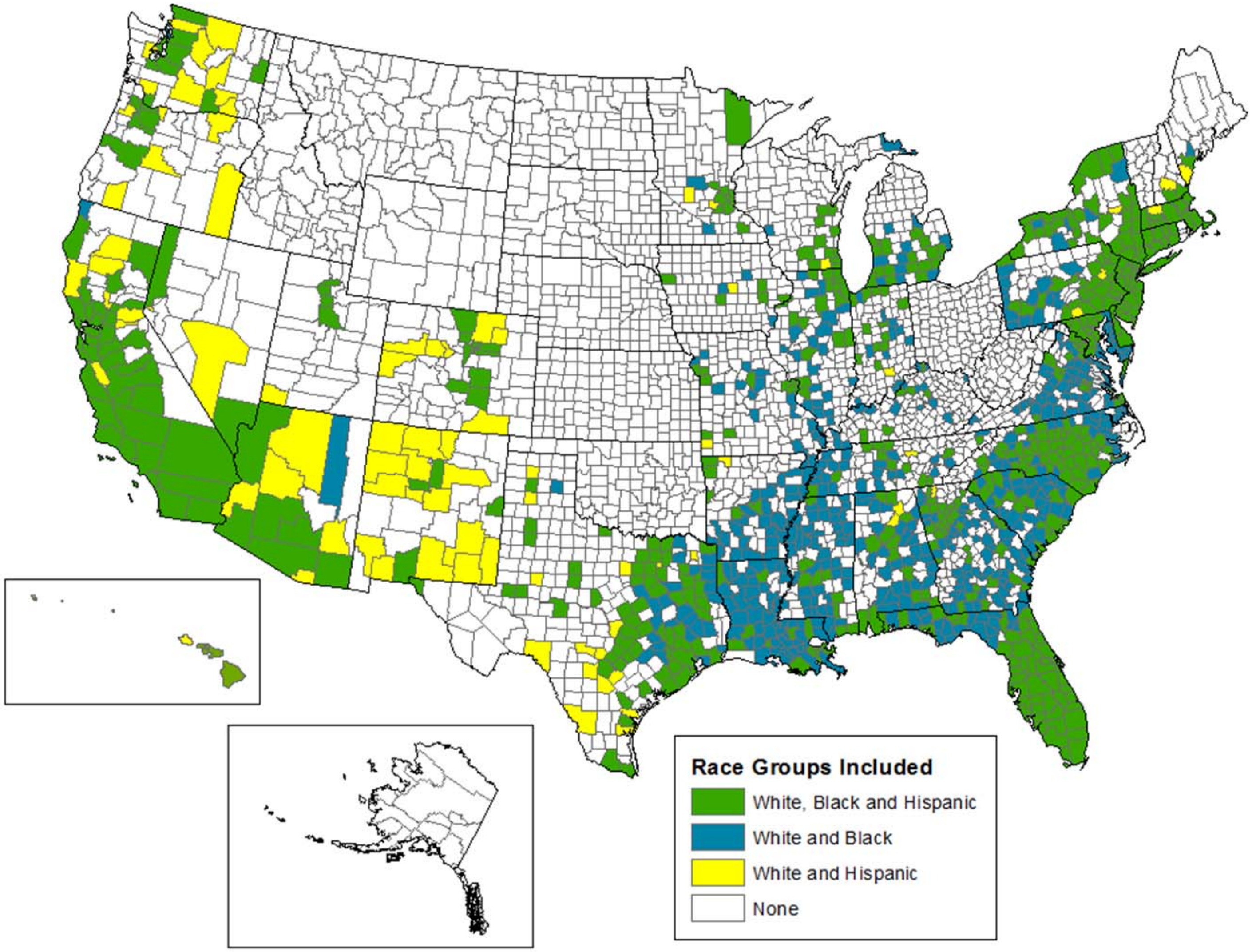


Figure 2 County-level poverty by race

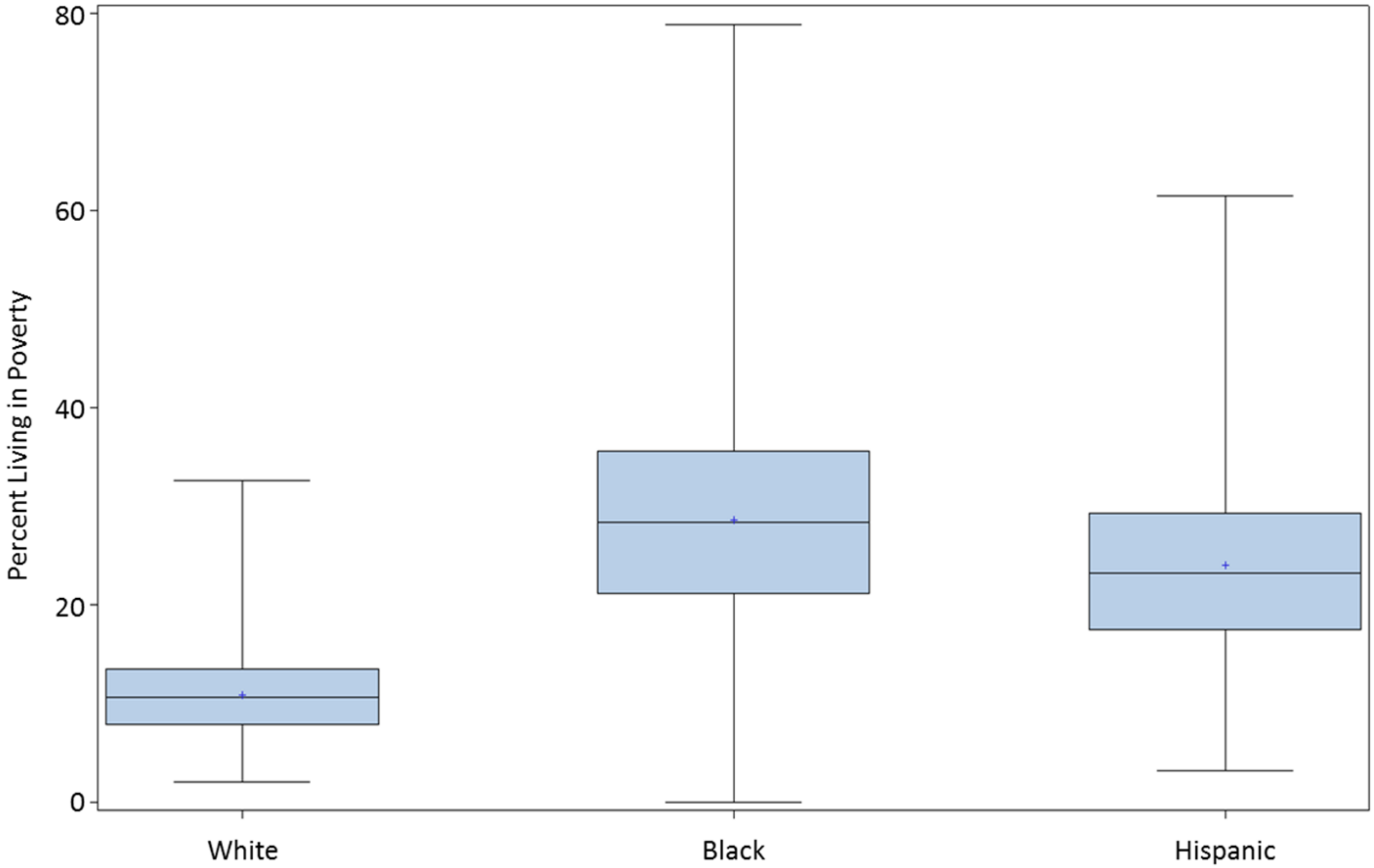


Figure 3 Crude county-level prevalence rates and PRRs by race and urbanization, 2008

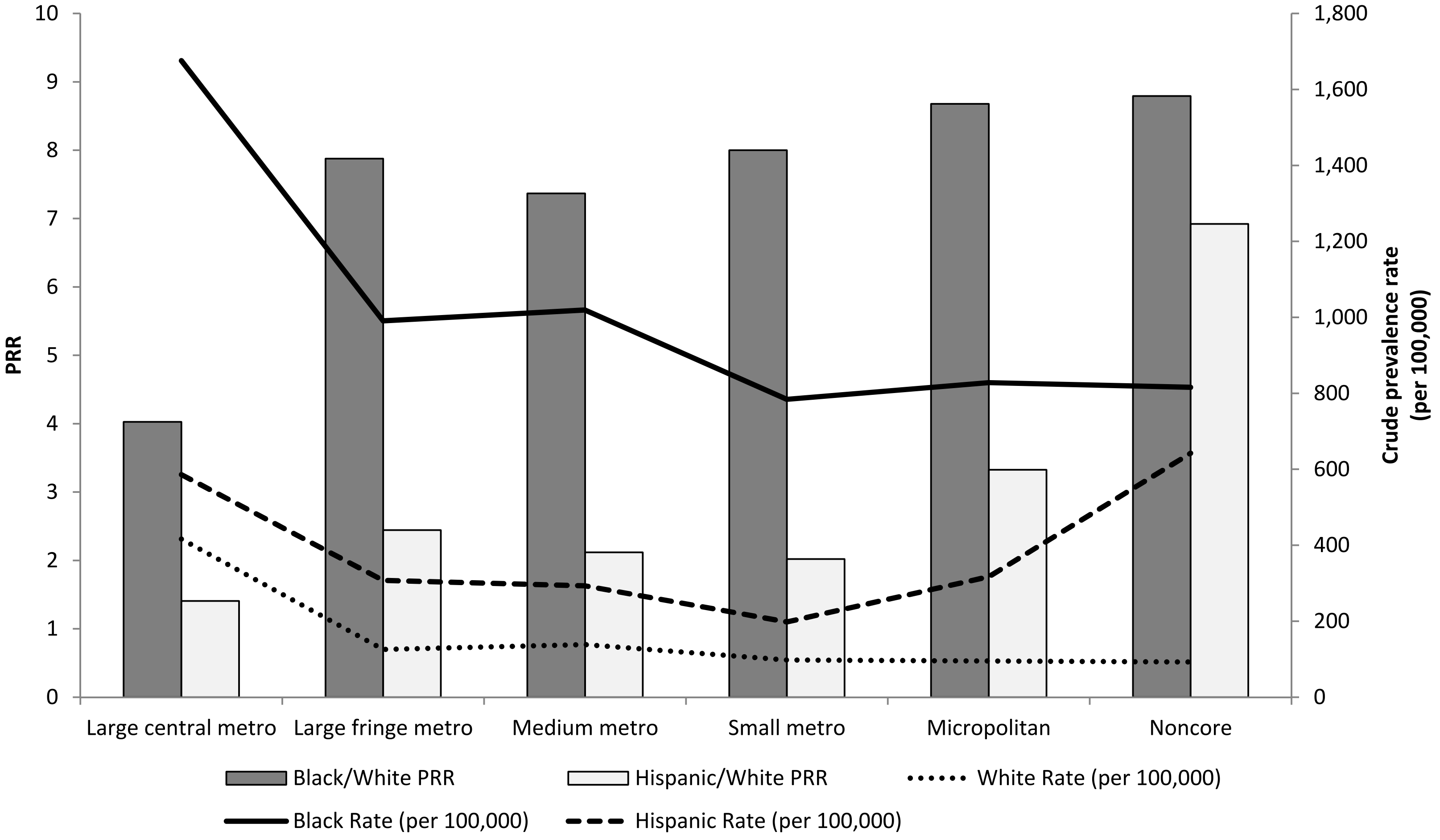


Figure 4 Black HIV prevalence rate ratios by urbanization, 2008

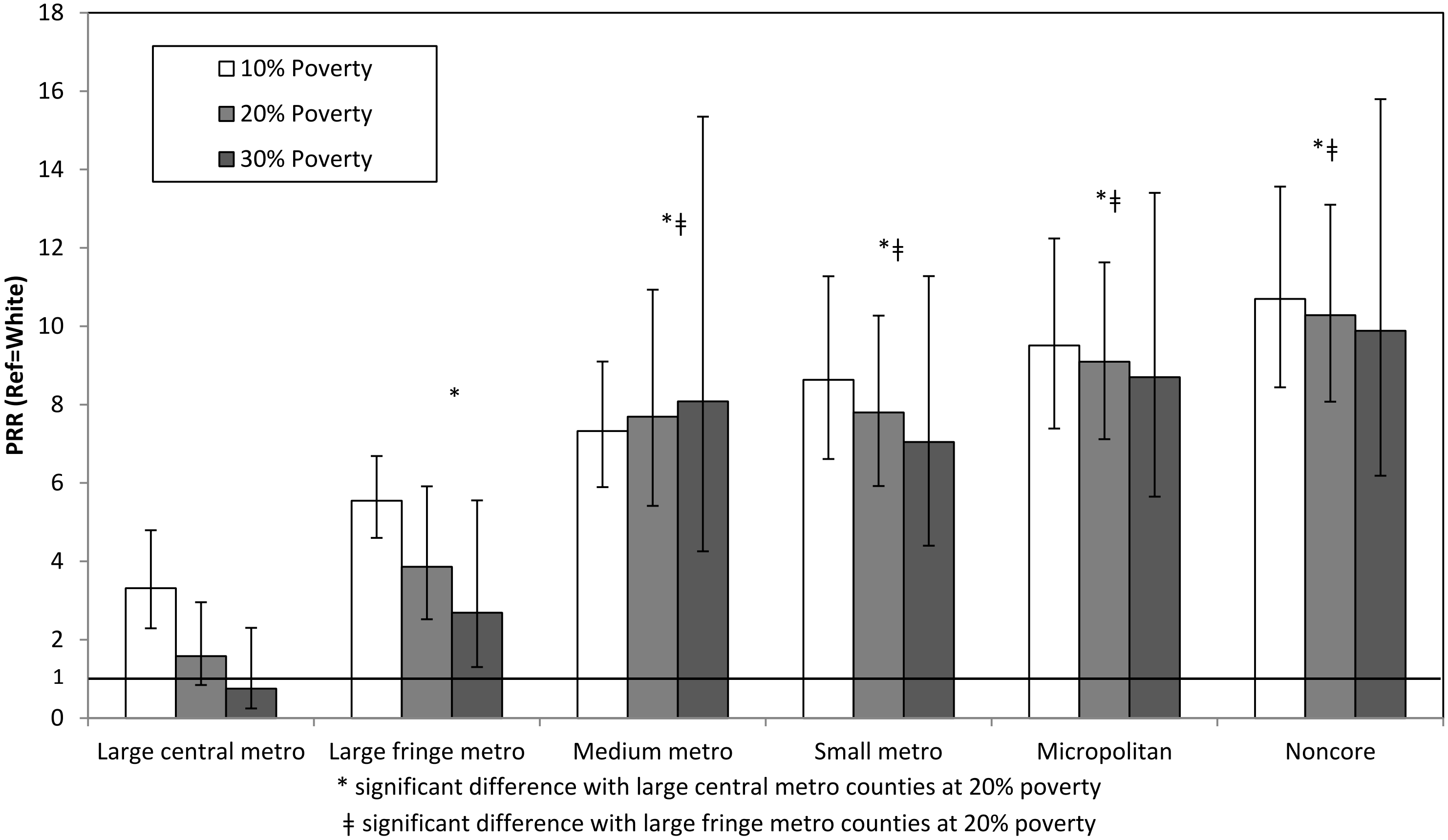
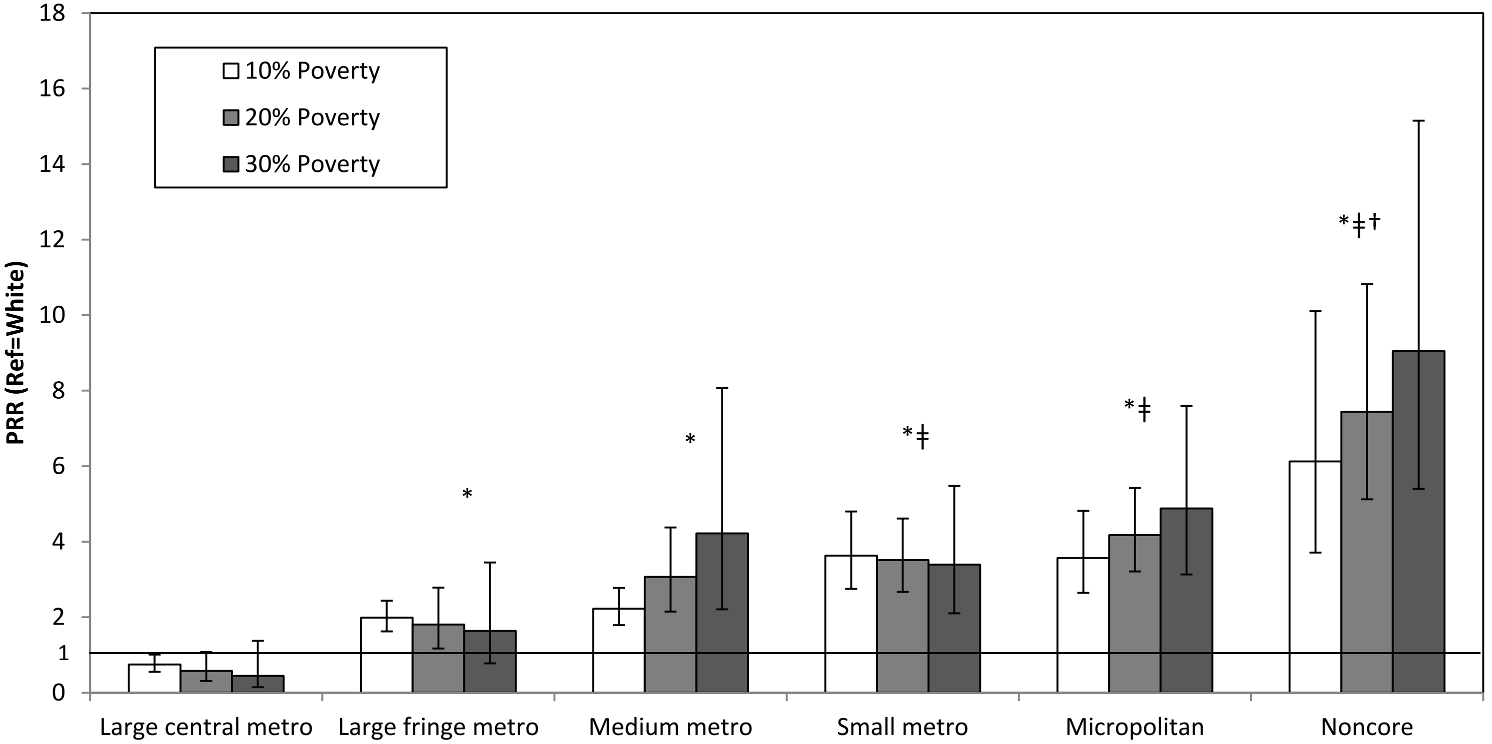


Figure 5 Hispanic HIV prevalence rate ratios by urbanization, 2008



* significant difference with large central metro counties at 20% poverty

‡ significant difference with large fringe metro counties at 20% poverty

† significant difference with medium metro, small metro and micropolitan counties at 20% poverty

APPENDIX A: SAS CODE

```
/**********************************************************************
* Create by Adam Vaughan, Fall 2011 and Spring 2012
* This SAS file imports various data sets and creates one final data
set for use with
* my thesis. This file does not perform any analysis.
* The final data file has one observation for each race in each county.
* This file imports CDC data reported as rates, not quintiles.
**********************************************************************/

libname a "h:\thesis\data\";

* Import the data file containing the FIPS county data;
PROC IMPORT OUT= fips
            DATAFILE=
"H:\Thesis\Data\dc_acs_2009_5yr_g00\dc_acs_2009_5yr_g00_geo.xls"
            DBMS=EXCEL REPLACE;
     RANGE="dc_acs_2009_5yr_g00_geo$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

/**********************************************************************
************************
* Income
**********************************************************************
***********************/

* Import files containing ACS 5 year estimates of per capita income and
median income, sex. These text files were downloaded
from factfinder.census.gov on 7/20/11 and then maniulated in Excel
prior to importing into SAS.;
PROC IMPORT OUT= WORK.income
            DATAFILE= "H:\Thesis\Data\Income_ACS_5YrEst.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
       guessingrows=3000;
RUN;

/**********************************************************************
************************
* Gini
**********************************************************************
***********************/
PROC IMPORT OUT= WORK.gini
            DATAFILE= "H:\Thesis\Data\Gini_ACS_5YR_2006-2010.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=2;
       guessingrows=3000;
RUN;
```

```
/**********************************************************************
************************
* Education, Poverty, Demographics and MSM data
***********************************************************************
***********************/

* Import other files containing demographics, poverty, education and
MSM data stratified by age, race, sex. These text files were downloaded
from factfinder.census.gov on 7/20/11 and then maniulated in Excel
prior to importing into SAS.;
PROC IMPORT OUT= WORK.data1
            DATAFILE=
"H:\Thesis\Data\dc_acs_2009_5yr_g00\dc_acs_2009_5yr_g00__data1.xls"
            DBMS=EXCEL REPLACE;
     RANGE="dc_acs_2009_5yr_g00__data1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

PROC IMPORT OUT= WORK.data2
            DATAFILE=
"H:\Thesis\Data\dc_acs_2009_5yr_g00\dc_acs_2009_5yr_g00__data2.xls"
            DBMS=EXCEL REPLACE;
     RANGE="dc_acs_2009_5yr_g00__data2$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

PROC IMPORT OUT= WORK.data3
            DATAFILE=
"H:\Thesis\Data\dc_acs_2009_5yr_g00\dc_acs_2009_5yr_g00__data3.xls"
            DBMS=EXCEL REPLACE;
     RANGE="dc_acs_2009_5yr_g00__data3$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

PROC IMPORT OUT= WORK.data4
            DATAFILE=
"H:\Thesis\Data\dc_acs_2009_5yr_g00\dc_acs_2009_5yr_g00__data4.xls"
            DBMS=EXCEL REPLACE;
     RANGE="dc_acs_2009_5yr_g00__data4$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;
```

```
PROC IMPORT OUT= WORK.data5
            DATAFILE=
"H:\Thesis\Data\dc_acs_2009_5yr_g00\dc_acs_2009_5yr_g00__data5.xls"
            DBMS=EXCEL REPLACE;
     RANGE="dc_acs_2009_5yr_g00__data5$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

PROC IMPORT OUT= WORK.data6
            DATAFILE=
"H:\Thesis\Data\dc_acs_2009_5yr_g00\dc_acs_2009_5yr_g00__data6.xls"
            DBMS=EXCEL REPLACE;
     RANGE="dc_acs_2009_5yr_g00__data6$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

PROC IMPORT OUT= WORK.data7
            DATAFILE=
"H:\Thesis\Data\dc_acs_2009_5yr_g00\dc_acs_2009_5yr_g00__data7.xls"
            DBMS=EXCEL REPLACE;
     RANGE="dc_acs_2009_5yr_g00__data7$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

PROC IMPORT OUT= WORK.data8
            DATAFILE=
"H:\Thesis\Data\dc_acs_2009_5yr_g00\dc_acs_2009_5yr_g00__data8.xls"
            DBMS=EXCEL REPLACE;
     RANGE="dc_acs_2009_5yr_g00__data8$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

PROC IMPORT OUT= WORK.data9
            DATAFILE=
"H:\Thesis\Data\dc_acs_2009_5yr_g00\dc_acs_2009_5yr_g00__data9.xls"
            DBMS=EXCEL REPLACE;
     RANGE="dc_acs_2009_5yr_g00__data9$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
```

```
     USEDATE=YES;
     SCANTIME=YES;
RUN;

PROC IMPORT OUT= WORK.data10
            DATAFILE=
"H:\Thesis\Data\dc_acs_2009_5yr_g00\dc_acs_2009_5yr_g00__data10.xls"
            DBMS=EXCEL REPLACE;
     RANGE="dc_acs_2009_5yr_g00__data10$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

* Create the permanent FIPS data file. Separate the County Name and
State Abbreviation so that
files can be merged using FIPS numbers. Import the lists of states and
their codes into the dataset.;
PROC IMPORT OUT= WORK.States
            DATAFILE= "H:\Thesis\Data\States.xls"
            DBMS=EXCEL REPLACE;
     RANGE="Sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

* Merge the fips file with the states file to get the state
abbreviation in the file. */
* Parse the name of the county out of the county/state string.;
proc sort data=states;
      by state;
proc sort data=fips;
      by state;
proc sort data=regions;
      by state;
data a.fips;
      merge fips states regions;
      by state;
      * Correct Dona Ana County to remove the tilde;
      if geo_id2=35013 then geoname='Dona Ana County, New Mexico';
      countyname=scan(geoname, 1, ',');
      countycaps=upcase(countyname); *Create a fully capitalized county
name for use in merging.;
      where state ne 72; * Drop Puerto Rico;
run;

/* Verify that every county was assigned a state
proc print data=a.fips;
      var geo_id2;
      where state = .;
run; */
```

```
* Create a permanent SAS datafile that contains all of the MSM, poverty
and education data;
proc sort data=a.fips;
      by geo_id2;
data a.demomsmpoveduc;
      merge a.fips data1 data2 data3 data4 data5 data6 data7 data8
data9 data10;
      by geo_id2; *by FIPS code;
      where geo_id2 < 72000; *Exclude Puerto Rico;
      pov_tot= B17001_2_EST/B17001_1_EST;
run;

* Create another data set with only the data that I need;
data a.dmpefinal;
      set a.demomsmpoveduc;
      * Calculate the percent of unmarried male-male households;
      PerMMH = B11009_3_EST/B11009_1_EST;

      * Calculate the over-25 population for each race.
      These will be used as denominators for poverty and education;
      whitenhpop25=
B01001H_9_EST+B01001H_10_EST+B01001H_11_EST+B01001H_12_EST+

      B01001H_13_EST+B01001H_14_EST+B01001H_15_EST+B01001H_16_EST+B0100
1H_24_EST+B01001H_25_EST+B01001H_26_EST+B01001H_27_EST+

      B01001H_28_EST+B01001H_29_EST+B01001H_30_EST+B01001H_31_EST;;
      blackpop25=
      B01001B_9_EST+B01001B_10_EST+B01001B_11_EST+B01001B_12_EST+

      B01001B_13_EST+B01001B_14_EST+B01001B_15_EST+B01001B_16_EST+B0100
1B_24_EST+B01001B_25_EST+B01001B_26_EST+B01001B_27_EST+

      B01001B_28_EST+B01001B_29_EST+B01001B_30_EST+B01001B_31_EST;
      hisppop25=
B01001I_9_EST+B01001I_10_EST+B01001I_11_EST+B01001I_12_EST+

      B01001I_13_EST+B01001I_14_EST+B01001I_15_EST+B01001I_16_EST+B0100
1I_24_EST+B01001I_25_EST+B01001I_26_EST+B01001I_27_EST+

      B01001I_28_EST+B01001I_29_EST+B01001I_30_EST+B01001I_31_EST;
      hawpop25=
B01001E_9_EST+B01001E_10_EST+B01001E_11_EST+B01001E_12_EST+

      B01001E_13_EST+B01001E_14_EST+B01001E_15_EST+B01001E_16_EST+B0100
1E_24_EST+B01001E_25_EST+B01001E_26_EST+B01001E_27_EST+

      B01001E_28_EST+B01001E_29_EST+B01001E_30_EST+B01001E_31_EST;
      asianpop25=
B01001D_9_EST+B01001D_10_EST+B01001D_11_EST+B01001D_12_EST+

      B01001D_13_EST+B01001D_14_EST+B01001D_15_EST+B01001D_16_EST+B0100
1D_24_EST+B01001D_25_EST+B01001D_26_EST+B01001D_27_EST+

      B01001D_28_EST+B01001D_29_EST+B01001D_30_EST+B01001D_31_EST;
      aipop25=
B01001C_9_EST+B01001C_10_EST+B01001C_11_EST+B01001C_12_EST+
```

```
      B01001C_13_EST+B01001C_14_EST+B01001C_15_EST+B01001C_16_EST+
B01001C_24_EST+B01001C_25_EST+B01001C_26_EST+B01001C_27_EST+

      B01001C_28_EST+B01001C_29_EST+B01001C_30_EST+B01001C_31_EST;

      *Calculate race-specific rates of education;
      ed_hswnhtot=
(C15002H_4_EST+C15002H_5_EST+C15002H_6_EST+C15002H_9_EST+C15002H_10_EST
+C15002H_11_EST)/whitenhpop25;
      if hisppop25 = 0 then ed_hshtot=-9;
            else ed_hshtot=
(C15002I_4_EST+C15002I_5_EST+C15002I_6_EST+C15002I_9_EST+C15002I_10_EST
+C15002I_11_EST)/hisppop25;
      if blackpop25 = 0 then ed_hsbltot=-9;
            else
ed_hsbltot=(C15002B_4_EST+C15002B_5_EST+C15002B_6_EST+C15002B_9_EST+C15
002B_10_EST+C15002B_11_EST)/blackpop25;
      if aipop25 = 0 then ed_hsaitot=-9;
            else ed_hsaitot=
(C15002C_4_EST+C15002C_5_EST+C15002C_6_EST+C15002C_9_EST+C15002C_10_EST
+C15002C_11_EST)/aipop25;
      if asianpop25 = 0 then ed_hsastot=-9;
            else ed_hsastot=
(C15002D_4_EST+C15002D_5_EST+C15002D_6_EST+C15002D_9_EST+C15002D_10_EST
+C15002D_11_EST)/asianpop25;
      if hawpop25 = 0 then ed_hsnhtot=-9;
            else ed_hsnhtot=
(C15002E_4_EST+C15002E_5_EST+C15002E_6_EST+C15002E_9_EST+C15002E_10_EST
+C15002E_11_EST)/hawpop25;

      * Calculate race-specific poverty rates. ;
      pov_wnh = B17001H_2_EST/(B17001H_2_EST+B17001H_31_EST);
      if B17001I_2_EST = 0 and B17001I_31_EST = 0 then pov_hisp=-9;
            else pov_hisp =
B17001I_2_EST/(B17001I_2_EST+B17001I_31_EST);
      if B17001B_2_EST = 0 and B17001B_31_EST = 0 then pov_bl=-9;
            else pov_bl = B17001B_2_EST/(B17001B_2_EST+B17001B_31_EST);
      if B17001D_2_EST = 0 and B17001D_31_EST = 0 then pov_as=-9;
            else pov_as = B17001D_2_EST/(B17001D_2_EST+B17001D_31_EST);
      if B17001C_2_EST = 0 and B17001C_31_EST = 0 then pov_ai=-9;
            else pov_ai = B17001C_2_EST/(B17001C_2_EST+B17001C_31_EST);
      if B17001E_2_EST = 0 and B17001E_31_EST = 0 then pov_nh=-9;
            else pov_nh = B17001E_2_EST/(B17001E_2_EST+B17001E_31_EST);

      keep geo_id geo_id2 geoname state county name abbrev countyname
countycaps permmh ed_hswnhtot ed_hshtot
            ed_hsbltot ed_hsaitot ed_hsastot ed_hsnhtot pov_wnh pov_bl
pov_as pov_ai pov_nh pov_hisp;

run;

/* Print variable to check the math and look for errors;
*ods html file="h:\temp.xls";
proc print data=a.dmpefinal;
      var geo_id geo_id2 geoname state county name abbrev countyname
countycaps PerMMH B11009_3_EST B11009_1_EST
```

```
      whitenhpop25 B01001H_9_EST B01001H_10_EST B01001H_11_EST
B01001H_12_EST B01001H_13_EST B01001H_14_EST B01001H_15_EST
            B01001H_16_EST B01001H_24_EST B01001H_25_EST B01001H_26_EST
B01001H_27_EST B01001H_28_EST B01001H_29_EST B01001H_30_EST
B01001H_31_EST
      blackpop25 B01001B_9_EST B01001B_10_EST B01001B_11_EST
B01001B_12_EST
            B01001B_13_EST B01001B_14_EST B01001B_15_EST B01001B_16_EST
B01001B_24_EST B01001B_25_EST B01001B_26_EST B01001B_27_EST
            B01001B_28_EST B01001B_29_EST B01001B_30_EST B01001B_31_EST
      hisppop25 B01001I_9_EST B01001I_10_EST B01001I_11_EST
B01001I_12_EST
            B01001I_13_EST B01001I_14_EST B01001I_15_EST B01001I_16_EST
B01001I_24_EST B01001I_25_EST B01001I_26_EST B01001I_27_EST
            B01001I_28_EST B01001I_29_EST B01001I_30_EST B01001I_31_EST
      hawpop25 B01001E_9_EST B01001E_10_EST B01001E_11_EST
B01001E_12_EST
            B01001E_13_EST B01001E_14_EST B01001E_15_EST B01001E_16_EST
B01001E_24_EST B01001E_25_EST B01001E_26_EST B01001E_27_EST
            B01001E_28_EST B01001E_29_EST B01001E_30_EST B01001E_31_EST
      asianpop25 B01001D_9_EST B01001D_10_EST B01001D_11_EST
B01001D_12_EST
            B01001D_13_EST B01001D_14_EST B01001D_15_EST B01001D_16_EST
B01001D_24_EST B01001D_25_EST B01001D_26_EST B01001D_27_EST
            B01001D_28_EST B01001D_29_EST B01001D_30_EST B01001D_31_EST
      aipop25 B01001C_9_EST B01001C_10_EST B01001C_11_EST
B01001C_12_EST
            B01001C_13_EST B01001C_14_EST B01001C_15_EST B01001C_16_EST
B01001C_24_EST B01001C_25_EST B01001C_26_EST B01001C_27_EST
            B01001C_28_EST B01001C_29_EST B01001C_30_EST B01001C_31_EST
      ed_hswnhtot C15002H_4_EST C15002H_5_EST C15002H_6_EST
C15002H_9_EST C15002H_10_EST C15002H_11_EST
      ed_hshtot C15002I_4_EST C15002I_5_EST C15002I_6_EST C15002I_9_EST
C15002I_10_EST C15002I_11_EST
      ed_hsbltot C15002B_4_EST C15002B_5_EST C15002B_6_EST
C15002B_9_EST C15002B_10_EST C15002B_11_EST
      ed_hsaitot C15002C_4_EST C15002C_5_EST C15002C_6_EST
C15002C_9_EST C15002C_10_EST C15002C_11_EST
      ed_hsastot C15002D_4_EST C15002D_5_EST C15002D_6_EST
C15002D_9_EST C15002D_10_EST C15002D_11_EST
      ed_hsnhtot C15002E_4_EST C15002E_5_EST C15002E_6_EST
C15002E_9_EST C15002E_10_EST C15002E_11_EST
      pov_f  B17001_17_EST B17001_46_EST
      pov_wnh B17001H_2_EST B17001H_31_EST
      pov_hisp B17001I_2_EST B17001I_31_EST
      pov_bl B17001B_2_EST B17001B_31_EST
      pov_as B17001D_2_EST B17001D_31_EST
      pov_ai B17001C_2_EST B17001C_31_EST
      pov_nh B17001E_1_EST  B17001E_2_EST B17001E_31_EST;
run;
ods html close;   */

/**********************************************************************
*
* Drug use data;
***********************************************************************
/
```

```
* Import Sheet 1;
PROC IMPORT OUT= WORK.Drugdata
            DATAFILE= "H:\Thesis\Data\DrugUse.xls"
            DBMS=EXCEL REPLACE;
     RANGE="Sheet1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

* Import Sheet 2;
PROC IMPORT OUT= WORK.County
            DATAFILE= "H:\Thesis\Data\DrugUse.xls"
            DBMS=EXCEL REPLACE;
     RANGE="Sheet2$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

* Change specific data in the files so that the regions match;
* The new variable reg2 indicates the groupings of regions that are
used in reporting drug data;
data work.county;
      set work.county;
      length reg2 $100;
      if region in ('Area 1', 'Area 2') and state='AR' then
            reg2='Areas 1 and 2';
      else if state='AR' and (region='Area 3' OR region='Area 4') then
            reg2='Areas 3 and 4';
      else if state='AR' and (region='Area 5' OR region='Area 6') then
            reg2='Areas 5 and 6';
      else if state='AR' and (region='Area 7' OR region='Area 12' OR
region='Area 13') then
            reg2="Areas 7, 12, and 13";
      else if state='AR' and (region='Area 8' OR region='Area 10' OR
region='Area 11') then
            reg2='Areas 8, 10, and 11';
      else if state='CO' and (region='Region 2' OR region='Region 7')
then
            reg2='Regions 2 and 7';
      else if state='CO' and (region='Region 5' OR region='Region 6')
then
            reg2='Regions 5 and 6';
      else if state='FL' and (region='Circuit 3' OR region='Circuit 8')
then
            reg2='Circuits 3 and 8';
      else if state='FL' and (region='Circuit 11' OR region='Circuit
16') then
            reg2='Circuits 11 and 16';
      else if state='HI' and (region='Kauai' or region='Maui') then
            reg2='Kauai and Maui';
```

```
      else if state='KY' and (region='Adanta' or region='Cumberland 
River' or region='Lifeskills') then
            reg2='Adanta, Cumberland River, and Lifeskills';
      else if state='KY' and (region='Bluegrass' or region='Comprehend' 
or region='North Key') then
            reg2='Bluegrass, Comprehend, and North Key';
      else if state='KY' and (region='Communicare' or region='River 
Valley') then
            reg2='Communicare and River Valley';
      else if state='KY' and (region='Four Rivers' or 
region='Pennyroyal') then
            reg2='Four Rivers and Pennyroyal';
      else if state='KY' and (region='Kentucky River' or 
region='Mountain' or region='Pathways') then
            reg2='Kentucky River, Mountain, and Pathways';
      else if state='MN' and (region='Region 1' OR region='Region 2') 
then
            reg2='Regions 1 and 2';
      else if state='MN' and (region='Region 3' OR region='Region 4') 
then
            reg2='Regions 3 and 4';
      else if state='MN' and (region='Region 5' OR region='Region 6') 
then
            reg2='Regions 5 and 6';

      else if state='OH' and (region='Board 2' OR region='Board 46' OR 
region='Board 55' OR region='Board 68') then
            reg2='Boards 2, 46, 55, and 68';
      else if region in ('Board 3', 'Board 52', 'Board 85') and 
state='OH' then
            reg2='Boards 3, 52, and 85';
      else if region in ('Board 4', 'Board 78') and state='OH' then
            reg2='Boards 4 and 78';
      else if region in ('Board 5', 'Board 60') and state='OH' then
            reg2='Boards 5 and 60';
      else if region in ('Board 7', 'Board 15', 'Board 41', 'Board 79', 
'Board 84') and state='OH' then
            reg2='Boards 7, 15, 41, 79, and 84';
      else if region in ('Board 8', 'Board 13', 'Board 83') and 
state='OH' then
            reg2='Boards 8, 13, and 83';
      else if state='OH' and (region='Board 18' OR region='Board 47') 
then
            reg2='Boards 18 and 47';
      else if state='OH' and (region='Board 20' OR region='Board 32' OR 
region='Board 54' OR region='Board 69') then
            reg2='Boards 20, 32, 54, and 69';
      else if state='OH' and (region='Board 21' OR region='Board 39' OR 
region='Board 51' OR region='Board 70' OR region='Board 80') then
            reg2='Boards 21, 39, 51, 70, and 80';
      else if state='OH' and (region='Board 22' OR region='Board 74' OR 
region='Board 87') then
            reg2='Boards 22, 74, and 87';
      else if state='OH' and (region='Board 23' OR region='Board 45') 
then
            reg2='Boards 23 and 45';
```

```
      else if state='OH' and (region='Board 27' OR region='Board 71' OR
region='Board 73') then
            reg2='Boards 27, 71, and 73';
      else if state='OH' and (region='Board 28' OR region='Board 43' OR
region='Board 67') then
            reg2='Boards 28, 43, and 67';
      else if state='OH' and (region='Board 50' OR region='Board 76')
then
            reg2='Boards 50 and 76';

      else if region in ('Region 2', 'Region 3', 'Region 8', 'Region
9') and state='PA' then
            reg2='Regions 2, 3, 8, and 9';
      else if region in ('Region 4', 'Region 11', 'Region 37', 'Region
49') and state='PA' then
            reg2='Regions 4, 11, 37, and 49';
      else if region in ('Region 5', 'Region 18', 'Region 23', 'Region
24', 'Region 46') and state='PA' then
            reg2='Regions 5, 18, 23, 24, and 46';
      else if region in ('Region 6', 'Region 12', 'Region 16', 'Region
31', 'Region 35', 'Region 39', 'Region 45', 'Region 47') and state='PA'
then
            reg2='Regions 6, 12, 16, 31, 35, 39, 45, and 47';
      else if region in ('Region 7', 'Region 13', 'Region 20', 'Region
33') and state='PA' then
            reg2='Regions 7, 13, 20, and 33';
      else if region in ('Region 10', 'Region 14', 'Region 15', 'Region
27', 'Region 32', 'Region 43', 'Region 44') and state='PA' then
            reg2='Regions 10, 14, 15, 27, 32, 43, and 44';
      else if state='PA' and (region='Region 17' OR region='Region 21')
then
            reg2='Regions 17 and 21';
      else if region in ('Region 19', 'Region 26', 'Region 28', 'Region
42') and state='PA' then
            reg2='Regions 19, 26, 28, and 42';
      else if region in ('Region 22', 'Region 38', 'Region 40', 'Region
41', 'Region 48') and state='PA' then
            reg2='Regions 22, 38, 40, 41, and 48';
      else if state='PA' and (region='Region 25' OR region='Region 30')
then
            reg2='Regions 25 and 30';
      else if state='PA' and (region='Region 29' OR region='Region 34')
then
            reg2='Regions 29 and 34';

      else if region in ('Region 1', 'Region 3') and state='LA' then
            reg2='Regions 1 and 3';
      else if region in ('Region 2', 'Region 9') and state='LA' then
            reg2='Regions 2 and 9';
      else if region in ('Region 4', 'Region 5', 'Region 6') and
state='LA' then
            reg2='Regions 4, 5, and 6';
      else if region in ('Region 7', 'Region 8') and state='LA' then
            reg2='Regions 7 and 8';
      else if region in ('Western', 'Pathways') and state='MI' then
            reg2='Pathways and Western';
```

```
      else if region in ('West Central', 'Badlands') and state='ND'
then
            reg2='Badlands and West Central';
      else if region in ('South Central', 'Lake') and state='ND' then
            reg2='Lake Region and South Central';
      else if region in ('North Central', 'Northwest') and state='ND'
then
            reg2='North Central and Northwest';
      else if region in ('Northwest', 'Southwest') and state='OK' then
            reg2='Northwest and Southwest';
      else if region in ('Bear River', 'Northeastern', 'Summit',
'Tooele', 'Wasatch') and state='UT' then
            reg2='Bear River, Northeastern, Summit, Tooele, and
Wasatch';
      else if region in ('Central', 'Four Corners', 'San Juan',
'Southwest') and state='UT' then
            reg2='Central, Four Corners, San Juan, and Southwest';
      else if region in ('Northern A', 'Northern B') and state='WV'
then
            reg2='Northern A and B';
      else if region in ('Northern C', 'Northern D') and state='WV'
then
            reg2='Northern C and D';
      else if region in ('Southern I', 'Southern III') and state='WV'
then
            reg2='Southern I and III';
      else reg2=region;

      * Since the fips file includes the word "County","Parish", etc.
as part of the countyname, change county in the
      drugdata file to reflect this.;
      if state='LA' then
            county2=upcase(catx(' ', county, 'Parish'));
      else if county in ('Skagway', 'Anchorage') and state='AK' then
            county2=upcase(catx(' ', county, 'Municipality'));
      else if county in ('Aleutians West', 'Southeast Fairbanks',
'Dillingham', 'Hoonah-Angoon', 'Wade Hampton', 'Nome', 'Prince of
Wales-Hyder',
                  'Petersburg', 'Yukon-Koyukuk', 'Valdez-Cordova',
'Bethel') and state='AK' then
            county2=upcase(catx(' ', county, 'Census Area'));
      else if county in ('Aleutians East', 'Bristol Bay', 'Denali',
'Fairbanks North Star', 'Haines', 'Kenai Peninsula', 'Matanuska-
Susitna',
                  'Lake and Peninsula', 'North Slope', 'Kodiak Island',
'Northwest Arctic', 'Ketchikan Gateway') and state='AK'  then
            county2=upcase(catx(' ', county, 'Borough'));
      else if state='AK'
            then county2=upcase(catx(' ', county, 'City and Borough'));
      else if (state='VA' and indexw(county, 'City')>0 and county ne
'James City' and county ne 'Charles City') or state='DC' or
                  (state='MD' and indexw(county, 'City')>0)
                  or (state='MO' and indexw(county, 'City')>0) or
(state='NV' and indexw(county, 'City')>0)
            then county2=upcase(county);
      else county2=upcase(catx(' ', county, 'County'));
      rename region=druguseregion;
```

```
run;

* Sort the data sets. ;
* Merge the two datasets into a single file.;
proc sort data=drugdata;
      by state reg2;
proc sort data=county;
      by state reg2;
data drugdata2;
      merge drugdata county;
      by state reg2;
      rename county2=countycaps
                  state=abbrev;
      drop county;
run;

/*Print a list of all counties missing data;
proc print data=a.drugdata;
      where du=. or region='';
run;*/

proc sort data=drugdata2;
      by abbrev countycaps;
proc sort data=a.fips;
      by abbrev countycaps;
data a.drugdatafinal;
      merge drugdata2 a.fips;
      by abbrev countycaps;
      where countycaps ne "OUTER KETCHIKAN CITY AND BOROUGH" and
countycaps ne "CLIFTON FORGE CITY";
      * Drop these two counties because they don't have FIPS codes;
run;

/*Check which counties are missing drug data;
proc print data=a.drugdatafinal;
      where du=. and abbrev ne 'PR';
run;
*/

/**********************************************************************
****************************
*
*     This section creates a file for the HIV rate data.
*
***********************************************************************
***************************/
PROC IMPORT OUT= WORK.hivTEMP
            DATAFILE= "H:\Thesis\Data\County rate living with HIV
2008_update_8 2011.csv"
            DBMS=CSV REPLACE;
      GUESSINGROWS=3000;
    GETNAMES=YES;
    DATAROW=2;
RUN;

data hiv;
      set hivtemp;
```

```
      if county="District of Columbia"
            then state="District of Columbia";
      if (state='New Mexico' and indexw(county, 'Ana')>0)
            then county="Dona Ana County";
      countycaps=upcase(county);
      name=upcase(state);
      rename      state=statename
                  county=countyname;
      where state ne "";
run;

proc sort data=hiv;
      by name countycaps;
proc sort data=a.fips;
      by name countycaps;
data a.hivdata;
      merge hiv a.fips;
      by name countycaps;
      where name ne 'PUERTO RICO';
run;

proc print data=a.hivdata;
      where countypop=.;
run;


/*********************************************************************
******************
* This section creates SAS data files for Health Care Data;
*********************************************************************
*****************/
* Data for percent uninsured.;
* Updated 1/28/12 to include 2009 data instead of 2007 data;
PROC IMPORT OUT= WORK.INS
            DATAFILE= "H:\Thesis\Data\SAHIE2009.txt"
            DBMS=TAB REPLACE;
     GETNAMES=YES;
     DATAROW=2;
RUN;

/*ods html file="h:\temp.xls";
proc freq data=ins;
      tables agecat*racecat*sexcat*iprcat/list;
run;
*ods html close;*/

*Update FIPS codes for 3 counties with incorrect codes;
data work.ins2;
      set ins;
      if statefips=2 and countyfips=201
            then countyfips=198; *Prince of Wales County;
      *Skagway-Hoonah-Angoon was FIPS 232, but was split into FIPS 230
and 105.;
      if statefips=2 and countyfips=232
            then countyfips=105; *Skagway-Hoonah-Angoon;
      *Wrangell-Petersburg was FIPS 280, but was split into FIPS 195
and 275.;
```

```
      if statefips=2 and countyfips=280
            then countyfips=195; *Petersburg;
run;

proc sort data=work.ins2;
      by statefips countyfips agecat racecat sexcat iprcat;
data work.insurance;
      set work.ins2;
      *Include only county level data. Remove all stratficiation. and
only include all incomes. Don't stratify by poverty.;
      where geocat=50 and iprcat=0 and racecat=0 and agecat=0 and
sexcat=0;
      * Create FIPS codes in the dataset.;
      geo_id2 = (statefips*1000) + (countyfips);
      keep geo_id2 pctui;
run;

* Create the final insurance data set;
data a.insurance;
      set work.insurance; * work.alaskains;
run;

/***********************************************************************
*****************
* This section creates SAS data files for Urban Rural Data;
************************************************************************
*****************/

* Rename the FIPS code in the Urban-Rural data set so that it can be
merged;
data work.urban;
      set a.nchsurbrural_v2006;
      attrib geo_id2 length=8;
      geo_id2 = input(fipscode,8.);
run;

*Update FIPS codes for 3 counties with incorrect codes;
data work.urban2;
      set work.urban;
      if geo_id2=2201
            then geo_id2=2198; *Prince of Wales County;
      *Skagway-Hoonah-Angoon was FIPS 232, but was split into FIPS 230
and 105.;
      if geo_id2=2232
            then geo_id2=2105; *Skagway-Hoonah-Angoon;
      *Wrangell-Petersburg was FIPS 280, but was split into FIPS 195
and 275.;
      if geo_id2=2280
            then geo_id2=2195; *Petersburg;
      keep geo_id2 urbruralcode;
run;

* Add the two Alaska counties who were split. Same as Insurance
coverage;
data alaskaurb;
      input geo_id2 urbruralcode;
      cards;
```

```
      2230 6
      2275 6
      ;
run;

data a.urbanfinal;
      set work.urban2 work.alaskaurb;
run;

/*********************************************************************
*****************
* Correctional data;
**********************************************************************
*****************/
*Import the Census 2010 correctional data;
PROC IMPORT OUT= WORK.censuscorr
            DATAFILE= "H:\Thesis\Data\2010CensusCorrectional.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=6;
     GUESSINGROWS=1000;
RUN;

* Import the Census 2010 Population Data;
PROC IMPORT OUT= WORK.censuspop
            DATAFILE= "H:\Thesis\Data\2010CensusPop.csv"
            DBMS=CSV REPLACE;
     GETNAMES=YES;
     DATAROW=6;
     GUESSINGROWS=1000;
RUN;

proc sort data=work.censuscorr;
      by geo_id2;
proc sort data=work.censuspop;
      by geo_id2;
data work.censusprison;
      merge work.censuscorr work.censuspop;
      * Calculate the percent of the adult pop by county in
correctional facilities;
      adultpop2010=sum(of s005-s019);
      corrpop_pct=100*d003/adultpop2010;
      rename d003=adultcorrpop;
      keep geo_id2 corrpop_pct d003 adultpop2010;
      if geo_id<72000;
run;

/*********************************************************************
*****************
* Create the final analysis file that merges everything together;
* Merge these one at a time to verify that merges occur correctly.
**********************************************************************
*****************/
proc sort data=a.fips;
      by geo_id2;
proc sort data=a.dmpefinal;
      by geo_id2;
```

```
data temp;
      merge a.fips a.dmpefinal;
      by geo_id2;
run;

proc sort data=a.drugdatafinal;
      by geo_id2;
data temp2;
      merge temp a.drugdatafinal;
      by geo_id2;
run;

proc sort data=a.insurance;
      by geo_id2;
data temp3;
      merge temp2 a.insurance;
      by geo_id2;
run;

/*proc print data=work.temp3;
      where pctui=. or geoname="";
      var geo_id2 countycaps abbrev;
run;*/

proc sort data=a.hivdata;
      by geo_id2;
data temp4;
      merge temp3 a.hivdata;
      by geo_id2;
run;

proc sort data=a.urbanfinal;
      by geo_id2;
data temp6;
      merge temp4 a.urbanfinal;
      by geo_id2;
run;

proc sort data=work.censusprison;
      by geo_id2;
data temp7;
      merge temp6 work.censusprison;
      by geo_id2;
run;

proc sort data=a.arf;
      by geo_id2;
data temp8;
      merge temp7 a.arf;
      by geo_id2;
run;

/*
proc print data=work.temp8;
      var geo_id2 geoname pov_tot exppercap f04437;
      where exppercap=. or pov_tot=.;
run;
```

```
*/

proc sort data=work.income;
      by geo_id2;
data temp9;
      merge temp8 work.income;
      by geo_id2;
      rename      B19013_1_EST=MedIncome
                  B19013A_1_EST=MedIncome_Wh
                  B19013B_1_EST=MedIncome_Bl
                  B19013C_1_EST=MedIncome_AIAN
                  B19013D_1_EST=MedIncome_Asian
                  B19013E_1_EST=MedIncome_PacIsl
                  B19013H_1_EST=MedIncome_WhNH
                  B19013I_1_EST=MedIncome_Hisp
                  B19301_1_EST=PerCapIncome
                  B19301A_1_EST=PerCapIncome_Wh
                  B19301B_1_EST=PerCapIncome_Bl
                  B19301C_1_EST=PerCapIncome_AIAN
                  B19301D_1_EST=PerCapIncome_Asian
                  B19301E_1_EST=PerCapIncome_PacIsl
                  B19301H_1_EST=PerCapIncome_WhNH
                  B19301I_1_EST=PerCapIncome_Hisp;
      drop B19013_1_MOE B19013A_1_MOE B19013B_1_MOE B19013C_1_MOE
B19013D_1_MOE B19013E_1_MOE B19013F_1_EST B19013F_1_MOE
             B19013G_1_EST B19013G_1_MOE B19013H_1_MOE B19013I_1_MOE
B19301_1_MOE B19301A_1_MOE B19301B_1_MOE B19301C_1_MOE
             B19301D_1_MOE B19301E_1_MOE B19301F_1_EST B19301F_1_MOE
B19301G_1_EST B19301G_1_MOE B19301H_1_MOE B19301I_1_MOE;
run;

proc sort data=work.gini;
      by geo_id2;
data temp10;
      merge temp9 work.gini;
      by geo_id2;
      drop ginimoe;
run;

data a.thesisrates;
      set temp10;
      povtotpct=pov_tot*100;
      pct_wnh=wnhpop/countypop;
      pct_bl=bnhpop/countypop;
      pct_asnh=asnhpop/countypop;
      pct_aian=aianpop/countypop;
      pct_nhpip=nhpipop/countypop;
      pct_hisp=hisppop/countypop;

      blackrr=bnhrate/wnhrate;
      logblackrr=log(blackrr);
      if pov_bl ~= -9 then do;
            blpovratio=pov_bl/pov_wnh;
            blpovdiff=pov_bl-pov_wnh;
      end;
      else do;
            blpovratio=.;
```

```
            blpovdiff=.;
      end;
      if ed_hsbltot ~= -9 then
            bleducratio=ed_hsbltot/ed_hswnhtot;
      else bleducratio=.;

      blincomeratio=percapincome_bl/percapincome_wh;

      * Remove Puerto Rico;
      if geo_id2 < 72000;
run;

proc means data=a.thesisrates;
      var pov_bl ed_hsbltot;
run;

/*proc print data=temp6;
      where urbruralcode=. or geo_id2=. or countycaps="";
      var geo_id2 countycaps abbrev;
run; */

*----------------------------------------------------------------------
----*;
* Now create a file from this data that contains one observation per
race;
* Split the original SAS data file into separate files for each race;
data white;
      set a.thesisrates;
      rename      ed_hswnhtot=ed_race
                  wnhrate=rate_race
                  wnhcase=case_race
                  wnhpop=pop_race
                  pct_wnh=pct_race;
      where wnhrate>0 and wnhpop>0; * Only include those counties
reporting this data with a county pop>0;
      pov_race=pov_wnh*100;
      povdiff= pov_race-povtotpct;
      keep geo_id2 ed_hswnhtot wnhrate wnhcase wnhpop pct_wnh pov_race
povdiff;
run;

data hisp;
      set a.thesisrates;
      rename      ed_hshtot=ed_race
                  hisprate=rate_race
                  hispcase=case_race
                  hisppop=pop_race
                  pct_hisp=pct_race;
      where hisprate>0 and hisppop>0; * Only include those counties
reporting this data;
      pov_race=pov_hisp*100;
      keep geo_id2 ed_hshtot hisprate hispcase hisppop pct_hisp
pov_race;
run;

data black;
      set a.thesisrates;
```

```
      rename      ed_hsbltot=ed_race
                  bnhrate=rate_race
                  bnhcase=case_race
                  bnhpop=pop_race
                  pct_bl=pct_race;
      where bnhrate>0 and bnhpop>0; * Only include those counties
reporting this data;
      pov_race=pov_bl*100;
      keep geo_id2 ed_hsbltot bnhrate bnhcase bnhpop pct_bl pov_race;
run;

data asian;
      set a.thesisrates;
      rename      ed_hsastot=ed_race
                  asnhrate=rate_race
                  asnhcase=case_race
                  asnhpop=pop_race
                  pct_asnh=pct_race;
      where asnhrate>0 and asnhpop>0; * Only include those counties
reporting this data;
      pov_race=pov_as*100;
      keep geo_id2 ed_hsastot asnhrate asnhcase asnhpop pct_asnh
pov_race ;
run;

data pacific;
      set a.thesisrates;
      rename      ed_hsnhtot=ed_race
                  nhpirate=rate_race
                  nhpicase=case_race
                  nhpipop=pop_race
                  pct_nhpip=pct_race;
      where nhpirate>0 and nhpipop>0; * Only include those counties
reporting this data;
      pov_race=pov_nh*100;
      keep geo_id2 ed_hsnhtot nhpirate nhpicase nhpipop pct_nhpip
pov_race ;
run;

data indian;
      set a.thesisrates;
      rename      ed_hsaitot=ed_race
                  aianrate=rate_race
                  aiancase=case_race
                  aianpop=pop_race
                  pct_aian=pct_race;
      where aianrate>0 and aianpop>0; * Only include those counties
reporting this data;
      pov_race=pov_ai*100;
      keep geo_id2 ed_hsaitot aianrate aiancase aianpop pct_aian
pov_race ;
run;

* Combine all the race specific files into one big file;
data allraces;
      set white (in=a) black (in=b) hisp (in=c) asian (in=d) pacific
(in=e) indian(in=f);
```

```
      * Remove all labels;
      attrib _all_ label=' ';
      if a then race=1;
      else if b then race=2;
      else if c then race=3;
      else if d then race=4;
      else if e then race=5;
      else if f then race=6;
run;

*Merge this to include county level data in the final data set;
* First create a smaller dataset from ThesisFinal;
data county;
      set a.thesisrates;
      keep geo_id2 geoname state abbrev subregion region edhstot_pct
permmh du pctui ctyrate ctycase countypop urbruralcode urb2
adultcorrpop
            corrpop_pct povtotpct exppercap gini;
run;

proc sort data=county;
      by geo_id2;
proc sort data=allraces;
      by geo_id2;
data a.racerates;
      merge allraces (in=a) county (in=b);
      by geo_id2;
      if a and b; * Only include observations in both datasets;
run;
```

```
/***********************************************************************
************************
* Create by Adam Vaughan, Fall 2011 and Spring 2012
* This SAS file contains the code needed for the final analysis of
thesis data;
* It is based on the analysis from 2/20/12;
* This analysis restricts included counties to those counties with
either (white and black)
* or (white and hispanic) or all three;
* There is a single model for all three races stratified on
urbabization;
************************************************************************
***********************/

libname a "h:\thesis\data";
%include "h:\thesis\data\collin_2011.sas";
options mprint symbolgen mlogic;


* Create a new dataset that only includes counties with (black or
Hispanic) and white;
* Exclude Broomfield County, CO because it doesn't have drug use data;
* This is the only county with incomplete data.;
data racerates;
      set a.thesisrates;
      where (wnhrate ~= . and (bnhrate ~= . or hisprate ~= .)) and
geo_id2 ~= 8014;
run;

proc print data=racerates;
      var wnhrate bnhrate hisprate;
run;

* Pull all white rates into a dataset;
data white;
      set racerates;
      rename      ed_hswnhtot=ed_race
                  wnhrate=rate_race
                  wnhcase=case_race
                  wnhpop=pop_race
                  pct_wnh=pct_race;
      pov_race=pov_wnh*100;
      povdiff= pov_race-povtotpct;
      keep geo_id2 ed_hswnhtot wnhrate wnhcase wnhpop pct_wnh pov_race
povdiff numrates;
      where wnhrate ~= .;
run;

* Pull all Hispanic rates into a datasetthe county;
data hisp;
      set racerates;
      rename      ed_hshtot=ed_race
                  hisprate=rate_race
                  hispcase=case_race
                  hisppop=pop_race
                  pct_hisp=pct_race;
      pov_race=pov_hisp*100;
```

```
      * povdiff is the difference between the race's poverty and the
county overall poverty;
      povdiff= pov_race-povtotpct;
      keep geo_id2 ed_hshtot hisprate hispcase hisppop pct_hisp
pov_race povdiff numrates;
      where hisprate ~= .;
run;

data black;
      set racerates;
      rename      ed_hsbltot=ed_race
                  bnhrate=rate_race
                  bnhcase=case_race
                  bnhpop=pop_race
                  pct_bl=pct_race;
      pov_race=pov_bl*100;
      povdiff= pov_race-povtotpct;
      keep geo_id2 ed_hsbltot bnhrate bnhcase bnhpop pct_bl pov_race
povdiff numrates;
      where bnhrate ~= .;
run;

* Combine all the race specific files into one file. Only include if
white rate plus at least one other (numrates);
data threeraces;
      set white (in=a) black (in=b) hisp(in=c);
      * Remove all labels;
      attrib _all_ label=' ';
      if a then race=1;
      else if b then race=2;
      else if c then race=3;
run;

*Merge this to include county level data in the final data set;
* First create a smaller dataset from ThesisFinal;
data county;
      set a.thesisrates;
      keep geo_id2 geoname state abbrev subregion region ed_hstot
permmh du pctui ctyrate ctycase countypop urbruralcode adultcorrpop
            corrpop_pct povtotpct exppercap gini;
run;

* This data set has three observations for each county, even though
many are missing;
proc sort data=county;
      by geo_id2;
proc sort data=threeraces;
      by geo_id2;
data work.pairedcounties;
      merge threeraces (in=a) county (in=b);
      by geo_id2;
      logracepop=log(pop_race);
      if a and b; * Only include observations in both datasets;
run;

proc means data=pairedcounties sum;
      var case_race;
```

```
      class urbruralcode race;
run;

* Check for collinearity;
proc sort data=pairedcounties;
      by urbruralcode;
run;

ods output genmod.parminfo=parms;
ods output genmod.covb=covdsn;
proc genmod data=pairedcounties;
      class race (ref="1") / param=ref;
      model case_race = pov_race race ed_race corrpop_pct permmh du
pctui exppercap pov_race*race /
            dist=negbin link=log offset=logracepop covb;
      by urbruralcode;
run;
ods output close;

%COLLIN(COVDSN=COVDSN, PROCDR=GENMOD, PARMINFO=Parms, OUTPUT=model1);

* Create a file to use in preparing a map of included counties;
data work.maprates;
      set work.racerates;
      if bnhrate ~= . and hisprate ~= . then mapnum=1;
      else if bnhrate ~= . then mapnum=2;
      else if hisprate ~= . then mapnum=3;
      keep geo_id2 geoname mapnum urbruralcode;
run;

PROC EXPORT DATA= WORK.MAPRATES
            OUTFILE= "H:\Thesis\Maps\CountiesForMapping.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=YES;
RUN;

* Get the mean and variance of the outcome variable;
proc univariate data=pairedcounties;
      var case_race;
run;

*-----------------------------------------------------------------------
------------------------;
* Now run the analysis;
* Create a macro to run the analysis with different sets of variables;
* Poverty, race and their interaction term are always in the model.;
* Only the potential confounders change;
* The model in this macro is not a correlated analysis;
%macro ratemodel(modeloutput, covar);
ods output estimates=&modeloutput;
proc sort data=pairedcounties;
      by urbruralcode geo_id2 race;
proc genmod data=pairedcounties;
      class geo_id2 race (ref="1") / param=ref;
      model case_race = pov_race race pov_race*race &covar/
            dist=negbin link=log offset=logracepop ;
      by urbruralcode;
```

```
      estimate "Black vs. White, Pov=10" pov_race 0 race 1 0 
pov_race*race 10 0;
      estimate "Black vs. White, Pov=20" pov_race 0 race 1 0 
pov_race*race 20 0;
      estimate "Black vs. White, Pov=30" pov_race 0 race 1 0 
pov_race*race 30 0;
      estimate "Hisp vs. White, Pov=10" pov_race 0 race 0 1 
pov_race*race 0 10;
      estimate "Hisp vs. White, Pov=20" pov_race 0 race 0 1 
pov_race*race 0 20;
      estimate "Hisp vs. White, Pov=30" pov_race 0 race 0 1 
pov_race*race 0 30;
run;
%mend;


* Run the full model. This is the Gold Standard;
%ratemodel(model1, corrpop_pct permmh du pctui exppercap);

*************************************************************;
* Evaluate confounding;
* Model 2a: Drop corr;
%ratemodel(model2a, permmh du pctui exppercap);

* Model 2b: Drop mmh;
%ratemodel(model2b, corrpop_pct du pctui exppercap);

* Model 2c: Drop du;
%ratemodel(model2c, corrpop_pct permmh pctui exppercap);

* Model 2d: Drop ui;
%ratemodel(model2d, corrpop_pct permmh du exppercap);

* Model 2e: Drop exp;
%ratemodel(model2e, corrpop_pct permmh du pctui);

* Model 2f: Drop corr mmh;
%ratemodel(model2f, du pctui exppercap);

* Model 2g: Drop corr du;
%ratemodel(model2g, permmh pctui exppercap);

* Model 2h: Drop corr ui;
%ratemodel(model2h, permmh du exppercap);

* Model 2i: Drop corr exp;
%ratemodel(model2i, permmh du pctui);

* Model 2j: Drop mmh du;
%ratemodel(model2j, corrpop_pct pctui exppercap);

* Model 2k: Drop mmh ui;
%ratemodel(model2k, corrpop_pct du exppercap);

* Model 2l: Drop mmh exp;
%ratemodel(model2l, corrpop_pct du pctui);
```

```
* Model 2m: Drop du ui;
%ratemodel(model2m, corrpop_pct permmh exppercap);

* Model 2n: Drop du exp;
%ratemodel(model2n, corrpop_pct permmh pctui);

* Model 2o: Drop ui exp;
%ratemodel(model2o, corrpop_pct permmh du);

* Model 2p: Drop corr mmh du;
%ratemodel(model2p, pctui exppercap);

* I accidentally left q out of this list;

* Model 2r: Drop corr mmh exp;
%ratemodel(model2r, du pctui);

* Model 2s: Drop mmh du ui;
%ratemodel(model2s, corrpop_pct exppercap);

* Model 2t: Drop mmh du exp;
%ratemodel(model2t, corrpop_pct pctui);

* Model 2u: Drop du ui exp;
%ratemodel(model2u, corrpop_pct permmh);

* Model 2v: Drop corr du ui;
%ratemodel(model2v, permmh exppercap);

* Model 2w: Drop corr du exp;
%ratemodel(model2w, permmh pctui);

* Model 2x: Drop corr ui exp;
%ratemodel(model2x, permmh du);

* Model 2y: Drop mmh ui exp;
%ratemodel(model2y, corrpop_pct du);

* Model 2z: Drop corr mmh du ui;
%ratemodel(model2z, exppercap);

* Model 2aa: Drop corr mmh du exp;
%ratemodel(model2aa, pctui);

* Model 2bb: Drop corr du ui exp;
%ratemodel(model2bb, permmh);

* Model 2cc: Drop mmh du ui exp;
%ratemodel(model2cc, corrpop_pct);

* Model 2dd: Drop corr mmh ui exp;
%ratemodel(model2dd, du);

* Model 2ee: Drop corr mmh ui exp du;
%ratemodel(model2ee, );

* Now combine each model with the GS to assess confounding;
```

```
* Create a macro that combines the gold standard model with each of the
other models;
* The macro also evaluates +/-10% and precision;
%macro combinemodels(modelinput);
data &modelinput;
      merge model1 &modelinput;
      by urbruralcode label;
      pct_conf=100*((meanestimategs-meanestimate)/meanestimategs);
      /* Confounding toward the null */
      pot_conf_twd=(pct_conf <= -10);

      /* Confounding away from the null */
      pot_conf_away=(pct_conf >= 10);
      pot_conf=pot_conf_twd + pot_conf_away;

      precisiongs=meanupperclgs/meanlowerclgs;
      precision=meanuppercl/meanlowercl;
      better_precision=(precision <= precisiongs);
run;
%mend;

* Rename fields in the GS;
data model1;
      set model1;
      rename meanestimate=meanestimategs
                  meanlowercl=meanlowerclgs
                  meanuppercl=meanupperclgs;
      keep urbruralcode label meanestimate meanlowercl meanuppercl;
run;

* Combine each model's results with the GS and evaluate 10% and
precision;
%combinemodels(model2a);
%combinemodels(model2b);
%combinemodels(model2c);
%combinemodels(model2d);
%combinemodels(model2e);
%combinemodels(model2f);
%combinemodels(model2g);
%combinemodels(model2h);
%combinemodels(model2i);
%combinemodels(model2j);
%combinemodels(model2k);
%combinemodels(model2l);
%combinemodels(model2m);
%combinemodels(model2n);
%combinemodels(model2o);
%combinemodels(model2p);
%combinemodels(model2r);
%combinemodels(model2s);
%combinemodels(model2t);
%combinemodels(model2u);
%combinemodels(model2v);
%combinemodels(model2w);
%combinemodels(model2x);
%combinemodels(model2y);
%combinemodels(model2z);
```

```
%combinemodels(model2aa);
%combinemodels(model2bb);
%combinemodels(model2cc);
%combinemodels(model2dd);
%combinemodels(model2ee);

data confound;
      set model2a (in=a) model2b (in=b) model2c (in=c) model2d (in=d)
            model2e (in=e) model2f (in=f) model2g (in=g) model2h (in=h)
            model2i (in=i) model2j (in=j) model2k (in=k) model2l (in=l)
            model2m (in=m) model2n (in=n) model2o (in=o) model2p (in=p)
            model2r (in=r) model2s (in=s) model2t (in=t)
            model2u (in=u) model2v (in=v) model2w (in=w) model2x (in=x)
            model2y (in=y) model2z (in=z) model2aa (in=aa) model2bb
(in=bb)
            model2cc (in=cc) model2dd (in=dd) model2ee (in=ee);
      format model $ 2.;
      if a then model="a";
      else if b then model="b";
      else if c then model="c";
      else if d then model="d";
      else if e then model="e";
      else if f then model="f";
      else if g then model="g";
      else if h then model="h";
      else if i then model="i";
      else if j then model="j";
      else if k then model="k";
      else if l then model="l";
      else if m then model="m";
      else if n then model="n";
      else if o then model="o";
      else if p then model="p";
      else if r then model="r";
      else if s then model="s";
      else if t then model="t";
      else if u then model="u";
      else if v then model="v";
      else if w then model="w";
      else if x then model="x";
      else if y then model="y";
      else if z then model="z";
      else if aa then model="aa";
      else if bb then model="bb";
      else if cc then model="cc";
      else if dd then model="dd";
      else if ee then model="ee";
run;

proc sort data=confound;
      by model;
proc means data=confound noprint;
      var pot_conf_twd pot_conf_away pot_conf better_precision
pct_conf;
      by model;
      output out=confound_summ
```

```
                  sum= pot_conf_twd pot_conf_away pot_conf
better_precis sum_pct_conf
                  mean= mean_pot_conf_twd  mean_pot_conf_away
mean_pot_conf mean_better_prec mean_pct_conf;
run;

proc print data=work.confound_summ;
run;

* Based on this, no model adequately controls for confounding. So, use
the gold standard.;
%ratemodel(finalmodelresults, corrpop_pct permmh du pctui exppercap);

*****************************************************
* The code below performs Wald tests on the difference between urban
groups at 10% poverty;
* This code give all possible group comparisons;

* This macro creates a dataset to use in Wald testing;
* It first splits the data into 6 different data sets - one for each
* level of urbanization. It them combines these into a single data set;
%macro resultsbyurb(modelinput, labeltext, outputset);
      %do i=1 %to 6;
            %let estname=%sysfunc(cat(logrr,&i));
            %let errname=%sysfunc(cat(stderr,&i));
            %let outputname=%sysfunc(cat(urb,&i));
            data &outputname;
                  set &modelinput;
                  rename lbetaestimate = &estname
                        stderr = &errname;
                  where urbruralcode=&i and label=&labeltext;
                  keep lbetaestimate stderr;
            run;
      %end;

      data &outputset;
            merge urb1 urb2 urb3 urb4 urb5 urb6;
      run;
%mend;

* This macro uses the results of the above dataset to generate Wald
statistics and pvalues;
%macro prrwaldtest(dataset, bonferroni);
      %let outputname=%sysfunc(cat(&dataset,results));
      data &outputname;
            set &dataset;
            %do i=1 %to 6;
                  %do j=&i+1 %to 6;
                        %let a=%sysfunc(cat(logrr,&i));
                        %let b=%sysfunc(cat(logrr,&j));
                        %let sea=%sysfunc(cat(stderr,&i));
                        %let seb=%sysfunc(cat(stderr,&j));
                        prr_a=&i;
                        prr_b=&j;
                        wald=(&a-&b)**2/(&sea**2+&seb**2);
                        pwald=1-probchi(wald, 1);
                        sig05 = (pwald <= 0.05);
```

```
                        sigbon = (pwald <= 0.05/&bonferroni);
                        output;
                  %end;
            %end;
      run;
%mend;

* Evaluate RRs for black;
%resultsbyurb(finalmodelresults, "Black vs. White, Pov=20", black20);

* Get Wald results for black 20%;
%prrwaldtest(black20, 30);

* Evaluate RRs for Hispanics;
* Create a dataset that includes a single observation with every Hisp
RR at pov=20;
%resultsbyurb(finalmodelresults, "Hisp vs. White, Pov=20", hisp20);

* Get Wald results for Hispanic 20%;
%prrwaldtest(hisp20, 30);

*********************************************************************;
* This section calculates results needed for the paper.
* Count the number of counties in the sample;
* geo_id2 counts the number of counties, abbrev counts the number of
states;
ods select nlevels;
proc freq data=work.pairedcounties nlevels;
      tables geo_id2 abbrev;
      where geo_id2 ~= 8014;
run;

* Code for Table 2 (JAMA): Counties by urbanization;
* Get the number of counties in each urbanization classification;
ods select nlevels;
proc freq data=work.pairedcounties nlevels;
      tables geo_id2;
      by urbruralcode;
      where geo_id2 ~= 8014; * Exclude Broomfield County, CO b/c this
model includes drugs;
run;

proc freq data=work.pairedcounties;
      tables race*urbruralcode;
      where geo_id2 ~= 8014; * Exclude Broomfield County, CO b/c this
model includes drugs;
run;

* Poverty by race;
proc sort data=pairedcounties;
      by race;
proc means data=pairedcounties;
      var pov_race;
      class race;
run;

proc boxplot data=pairedcounties;
```

```
   plot pov_race*race/boxwidth=20;
run;

*Total cases by urbanization;
proc sort data=pairedcounties;
      by urbruralcode;
proc means data=work.pairedcounties sum;
      var case_race;
      class urbruralcode;
      where geo_id2 ~= 8014;
run;

* Cases by race and urbanization;
proc sort data=pairedcounties;
      by race urbruralcode;
proc means data=work.pairedcounties sum;
      var case_race;
      class race urbruralcode;
      where geo_id2 ~= 8014;
run;

/*proc print data=work.racegini;
      var geoname case_race pov_race race permmh du pctui exppercap ;
      where urbruralcode=2 and geo_id2 ~= 8014;
run;

proc print data=work.pairedcounties;
      where urbruralcode=2;
run; */

* Calculate crude PRRs by urbanization;
proc sort data=pairedcounties;
      by race urbruralcode;
proc univariate data=pairedcounties;
      var case_race pop_race;
      class race urbruralcode;
      output out=sums sum=case_sum pop_sum;
run;

data sums;
      set sums;
      rate=case_sum/pop_sum;
run;

* Create a data set with one obs per urb level;
proc sort data=sums;
      by race urbruralcode;
run;

data white;
      set sums;
      where race=1;
      rename rate=whiterate;
run;

data black;
      set sums;
```

```
      where race=2;
      rename rate=blackrate;
run;

data hisp;
      set sums;
      where race=3;
      rename rate=hisprate;
run;

data sums2;
      merge white black hisp;
      by urbruralcode;
      blackprr=blackrate/whiterate;
      hispprr=hisprate/whiterate;
      drop race case_sum pop_sum;
run;

PROC EXPORT DATA= WORK.SUMS2
            OUTFILE= "H:\Thesis\CrudeRates.csv"
            DBMS=CSV REPLACE;
     PUTNAMES=YES;
RUN;
```