

Missing Black Men?

The Impact of Under-Reporting on Estimates of Black Male Labor Market Outcomes

Ariella Kahn-Lang *

November 10, 2018

Abstract

It is well documented by demographers that there is a long and persistent under-coverage of adult black males in the US Census. Despite this, the census undercount has been largely ignored in research using census data. Similar omission patterns also exist in other household based surveys, such as the Current Population Survey and the Survey of Income and Program Participation. I demonstrate that estimates of the undercount that rely on counts from vital statistics data are understated and provide estimates of the undercount of prime age black men in household-based survey data that are robust to under-coverage in vital statistics data. Because the incarcerated are automatically included in the Census, the population at risk of becoming incarcerated provides a unique opportunity for the identification of non-reporter characteristics. I use variation in incarceration by state and year to estimate the impact of increases in incarceration on non-reporting. I find that 90 percent of the incarcerated population would have been non-reporting had they not been incarcerated. Applying reasonable estimates of the size of the population at risk of incarceration, I conclude that non-reporting is almost entirely driven by the population at risk of incarceration. I then use data from the Survey of Inmates on labor market outcomes of inmates prior to incarceration to impute outcomes for the non-reporting population. Accounting for non-reporting meaningfully increases estimated gaps in black-white educational attainment, unemployment rates, and annual earnings.

*Kahn-Lang: Harvard Kennedy School, Harvard University. ariellakahnlang@fas.harvard.edu. I am indebted to my advisors, Christopher Avery, Edward Glaeser and Lawrence Katz for their guidance and encouragement. I thank Stephen Coussens, Roland Fryer, Claudia Goldin, Colin Gray, Ryan Hill, Helen Ho, Jennifer Kao, Jeff Liebman, Derek Neal, Jonathan Roth, Adrienne Sabety, Dev Patel, Sam Preston, Bruce Western, and participants in the Inequality and Social Policy seminar for their helpful comments. This paper also benefited from comments from participants at numerous seminars within Harvard. I particularly thank my family for their enduring support. I gratefully acknowledge funding from the Bradley Foundation.

1 Introduction

Household-based surveys are widely used by researchers and policy-makers to inform estimates of population counts, demographics, and labor market outcomes. In general, it is taken for granted that these data include a representative sample of the target population. In reality, there are often systematic coverage issues in these data that, when ignored, can lead to biased estimates. In this paper, I focus on the under-coverage of prime-age black men in household based survey data. I analyze data from the Census, the Current Population Survey, and the Survey of Income and Program Participation and find that adult black men are systematically missing from all three widely used datasets¹. The existence of an undercount alone does not prove that there is bias in statistics based on these surveys. If omitted individuals are randomly selected from the population, then analyses of labor market outcomes will be unbiased. I demonstrate that, in contrast, omitted individuals have lower levels of education, wages, and employment. Therefore, estimates for black men which ignore the undercount will be overstated.

I propose a novel method for quantifying the bias in estimates of labor market outcomes for black men. Quantifying the bias in estimates using household based survey data requires two key inputs. First, I provide updated estimates of the undercount of prime-age black men that are robust to incomplete vital statistics data. Second, I demonstrate the undercount is primarily driven by non-reporting of the population at risk of incarceration. I then adjust estimates of black male labor market outcomes to account for non-reporting by imputing outcomes for non-reporters using the population at risk of incarceration. I find that estimates of the black-white achievement gap are meaningfully underestimated. I show that adjusting for the undercount increases estimates of the black-white gap in male wages, unemployment, and education. Further, although unadjusted estimates suggest that the black-white male earnings gap has been rising since 1980, I find that after adjusting the undercount, this gap has been stable.

I first extend the prior demographic literature on the black male undercount by demonstrating that previous estimates have been understated. The primary focus of demographic research on the topic has been on estimating of the size of the undercount. Researchers have documented a large and persistent undercount of adult black males in the Census, but no similar undercount for black females or non-Hispanic whites. Specifically, past researchers have combined administrative records of all births and deaths with data on migration to estimate the full size of the population and of different population subgroups. They then

¹While I have not analyzed any other surveys that sample adult black men by households, I conjecture that the undercount exists across all datasets that rely on this sampling method.

compare this estimate to the observed population counts in the Census to estimate the undercount (see, for example, Preston et al, 2003, Robinson et al, 2002).

I argue that demographer estimates are subject to inaccuracies in the underlying data, as vital statistics data are known to be incomplete and estimates of immigration are noisy and subject to non-reporting bias. Further, I find that demographer estimates of non-reporting rates imply demonstrably inaccurate population gender ratios. I propose a new method of estimating the undercount based on true and observed gender ratios, which I refer to as the “Ratio Based Method.” The Ratio Based Method uses estimates of birth gender ratios, rather than absolute birth counts, to estimate the magnitude of the undercount. It therefore requires only the assumption that male and female births are reported at equal rates, rather than the stronger assumption that all births are reported. I further demonstrate that the Ratio Based Method is less sensitive to incomplete data on death and immigration counts. Using this approach, I estimate that the undercount for native born prime age black men ranges from 19.1 percent in 1970 to 8.4 percent in 2010.

Second, I show that the population at risk of incarceration primarily drives non-reporting. Ideally, I would have detailed data on the non-reporting population including demographics, economics outcomes, etc. However, almost by definition, there is very limited data on the non-reporting population. The incarcerated population provides a unique opportunity to study non-reporters. Given that the incarcerated population is reported to the Census by institutional employees, not individuals, the incarcerated population is automatically included in the Census. Therefore, men who would otherwise have been non-reporters are included in the Census if they become incarcerated.

To assess this possible connection, I propose a model of non-reporting where an individual's likelihood of not reporting to the Census is a function of latent risk characteristics and individual-specific risk of incarceration. I test this using variation in under-reporting and incarceration by state and year. I show that as the incarceration rate rises in a state and year, there is a corresponding risk in the number of black men counted. Controlling for observable factors, I estimate that a one percentage point increase in incarceration leads to a .90 percentage point decrease in under-reporting, suggesting that 90 percent of the incarcerated population did not report to the Census prior to incarceration. Given reasonable estimates of the share of prime age black men at risk of incarceration, these results suggest the population of non-reporters is nearly entirely made up of those at risk of incarceration.

These findings have natural implications for studies of relative education, earnings and employment rates of black and white men. I provide updated estimates for the educational attainment, employment and wages for black men that are adjusted for non-reporting. I compare these to my own calculations based on unadjusted census and CPS data using

methods standard in the literature (e.g. Bayer and Charles, 2018, Chandra, 2000). Based on my conclusion that non-reporters are primarily drawn from the population of men at risk of incarceration, I use data from the Survey of Inmates, including labor market outcomes prior to incarceration, to estimate labor market outcomes for the non-reporting population. Correcting for under-reporting meaningfully lowers estimates of educational attainment and wages and raises unemployment rates for black men ages 20-49. In 2010, adjusting for under-reporting raises the estimated black-white high school completion gap by 40 percent and the college attendance gap by 22 percent. Between 1980 and 2010, adjusting for under-reporting raises the estimated black-white wage gap by an average of 19 percent. Adjusting unemployment rates for non-reporting raises the black male unemployment rate by an average of 5 percent over my study period. I further find that, after adjusting for the undercount, the black-white earnings gap has been relatively stable since 1980. This is in contrast to unadjusted estimates which suggest that the black-white earnings gap grew by 18 percent between 1980 and 2010. I find no meaningful impact on the black-white employment gap. I also do not find a meaningful change in trends in any education or unemployment. This is driven by the fact that while omission rates are decreasing over the sample period, the discrepancy between the incarcerated population and the not incarcerated population is growing.

The rest of the paper proceeds as follows. Section two describes the undercount of black men. Section three describes the data used in this study. Section four argues that the undercount is driven by the population at risk of incarceration. Section five illustrates how the undercount biases estimates of the education attainment and labor market outcomes for black men. Section six concludes.

2 The Undercount

The left panel of Figure 1 shows the male to female ratio of non-institutionalized blacks by age for the years 1970-2010 based on census data. Taken at face value, this figure suggests that there is an approximately twenty percent drop in the ratio of males to females among blacks between ages 18 to 20 in all census years. This figure also suggests that the male to female ratio among blacks at each age has been very consistent over the past four decades. In fact, this pattern persists as far back as we have census data available, although prior to 1960 the male to female ratio among blacks also began to increase again around age 40. The right panel of Figure 1 shows that the sharp drop in males per female at age 18-20 is not caused by mass incarceration. This chart shows that the pattern generally persists even including the institutionalized population. There is also no evidence that there is a discrete

increase in mortality of black males between ages 18 and 20². Lastly, this is not driven by military enlistment³. This suggests that the drop in the male to female ratio seen in census data is not driven by a true change in the population but rather by a change in the reporting rates of male and females.

For as long as the Census has existed, there has been knowledge of its imperfection (Anderson and Fienberg, 1999). However, for most of the early census years, little was known about the size or nature of the undercount. Although the systematic undercounting of black males in the United States Census has likely existed for at least a century, it was only noticed in 1940. When in October of 1940 the government required that all men age 18 to 65 register for the selective service, government officials noticed that 2.8 percent of men registering for the selective service had not been counted by the 1940 Census. This number was 13 percent for black men (Anderson and Fienberg, 1999). This natural experiment highlighted issues of undercounting in the Census and led the Census Bureau to undertake efforts to evaluate the accuracy of census counts (Anderson and Fienberg, 1999). While there have been significant improvements made to increase the coverage of the Census, the black male undercount has continued to compromise census accuracy (Preston, 1998). For example, Robinson et al. found a .12 percent undercount overall in the 2000 Census but a 2.78 percent undercount for blacks. This number was 8.4 percent for black males ages 20 to 64.

Very little research has been done on the omission rates for household based surveys other than the Census. Given that many of these widely used surveys rely on sampling procedures based on the Census (e.g. the Survey of Income and Program Participation, the General Social Survey), it follows that these surveys also suffer from issues of undercounting prime age black males. Many surveys also provide weights for each respondent, to adjust samples to represent the population, which are based on census counts and therefore also under-count prime age black men (e.g. the Survey of Income and Program Participation, the Current Population Survey). Figure 2 illustrates the undercount of black males in two other widely used survey datasets, the Current Population Survey and the Survey of Income and Program Participation. The top left chart in Figure 2 shows the ratio of black males to females by age in the Census calculated as a five year moving average. Institutionalized individuals are excluded from this analysis to match the comparison datasets. The top middle and top right charts in Figure 2 show the male to female ratio for blacks by age in select years from the Current Population Survey and the Survey of Income and Program Participation,

²Based on author's analysis of Vital Statistics data on deaths available at:
<https://www.cdc.gov/nchs/nvss/deaths.htm>

³I note that those enlisted in the military are included in the Census. Regardless, the size of the military population is not large enough to explain the undercount

respectively. These charts show the same precipitous drop in the male to female ratio as in the Census. Given that this pattern in the Census reflects an undercount, the same pattern in other surveys suggests a similar undercount. Therefore, Figure 2 illustrates that the under-coverage is also present in these two surveys. The bottom row of Figure 2 shows the equivalent charts for whites in these datasets, illustrating that the same under-coverage does not exist among whites⁴.

2.1 Demographer Estimates of the Undercount

Today, demographers rely on two primary methods for estimating the undercount in the Census. Since 1980, the Census Bureau has conducted post-enumeration surveys to evaluate the coverage of each decennial Census. This approach involves matching results from independent in-depth surveys of small samples of the census coverage area with census results to identify omitted individuals and households. The Census then uses a logistic regression analysis based on a large set of covariates to estimate the extent of the full undercount. The second approach demographers use to estimate the undercount relies on vital statistics data. For decades, the United States has kept thorough records births and deaths. Demographers use vital statistics data to estimate the full size of the population and then compare counts to the census data to estimate the undercount. For this paper, I focus on the second method of estimation which uses vital statistics data for two reasons. The first is that post-enumeration surveys have only existed since the 1980 Census. The second is that the methods used in Census post-enumeration surveys changed between censuses. Therefore, using vital statistics data allows for more comparable estimates of the undercount between years.

I consider estimates of the undercount from 1970-2000 from two papers: Preston (1998) and Robinson et al. (2002). These two papers take very similar approaches to estimate the census undercount. To my knowledge, this analysis has not been done using the 2010 census data. Therefore, I replicate their methods for the 2010 Census. To estimate the true size of the population, Preston and Robinson et al. estimate the following equation:

$$Population = Births - Deaths + Immigrants - Emmigrants \quad (1)$$

To estimate total births and deaths, both papers rely on Vital Statistics data. To replicate methods for 2010, I also use vital statistics data. Preston and Robinson rely on estimates of legal migration based on Immigration and Naturalization Services data on legal immigration. Because this data does not include race, race is imputed using the racial distribution reported

⁴The exception to this occurs for the cohort of white med ages 18-30 in 1970. This is believed to represent non-reporting among draft-eligible men during the Vietnam war.

by immigrants from each country to the Census. Illegal immigration is estimated based on self reports to surveys. Emigration data is based on estimates of foreign-born emigration in the Census and US-born migration to Canada. For more information on estimates of migration, see Himes and Clogg (1992). I rely only on census data for immigration using the reported number of foreign born respondents. This is subject to less estimation error but excludes immigrants from the estimation of the undercount. For more details on the methods used to estimate the undercount in Preston, Robinson and my 2010 estimates, see Appendix B.

Table 1 presents estimates of the omission rate of the 1970-2010 US Censuses by age group for black men ages 5-49 based on the demography method. Consistent with observed male-female ratios, omission rates are low for black males ages 5-19. These rates are comparable to omission rates found for the equivalent age groups in black women (Preston, 1998, Robinson et al., 2002) and whites (Robinson et al., 2002). As Table 1 shows, from 1970 to 2010 omission rates for black men ages 20-49 ranged from five to fifteen percent with a median omission rate for prime age black men of eleven percent.

Although the demographer estimates suggest a substantially higher rate of non-reporting among adult black men, there is reason to believe they may not have captured the full undercount. Table 2 shows the implied true male-female ratios among blacks based on demographer estimates. These estimates suggest that the true ratio of black men to women increased by five percentage points between 1970 and 2010. For this to be the case, there would have to have been either a decrease in black male deaths relative to females or an increase in black female net migration relative to male. In contrast, the difference in the rates of black male and female deaths varied by less than half of a percentage point between 1970 and 2000. Similarly, the female and male immigrant shares of the black population moved roughly in parallel during this period. Appendix Table A.1 shows the true population gender ratios implied by death and immigration rates between 1970 and 2010. The estimated true gender ratios differ by less than .7 percentage points and were not increasing between 1970 and 2010.

One reason that the methods used by demographers may not capture the full population of black men is that estimates of total births may be incomplete. Vital statistics data on births is available starting in 1931. However, especially in early years, birth registration may be incomplete. Preston et al (2003) points out that comparisons of birth registrations to the 1940 and 1950 Censuses showed a substantial amount of underregistration, especially among blacks. It is also believed that, especially in earlier years, infant deaths may not have been registered as births (Preston et al, 2003). To estimate births prior to 1940, Preston uses a structural model which imposes that under-reporting rates for an age group

in a given year are the product of a constant year effect and a constant age group effect. This is a restrictive assumption which is inconsistent with patterns observed in the data for more recent years. Incomplete or inaccurate birth registration data has the potential to substantially bias results. As Robinson et al (1990) point out, “births are by far the largest component of population change involved in the demographic analysis system; thus even relatively small errors in the estimates of births can have significant effects on the demographic estimates of coverage.”

There is also reason to believe there may be measurement error in estimates of deaths and immigration. Estimates of immigration are based on estimates of legal immigration with imputed race based on country of origin and self reported survey data. Both sources are subject to risks of under-reporting and measurement error. Similarly, there is likely both under-coverage and measurement error in death data. For example, Robinson et al (1990) point out that prior to 1960 there was incomplete documentation of infant mortality. Robinson et al (1990) discuss the implications of measurement error for estimates of the undercount. They provide 95 percent confidence intervals for estimates of the undercount which demonstrate that measurement error may have a significant impact on estimates. For example, they estimate an undercount of 11.9% for black males ages 45-64 with a 95 percent confidence interval of 9.2% to 17.7%.

2.2 Ratio-Based Estimates of the Undercount

To address issues caused by incomplete vital statistics data on births and deaths, especially in early years of the available data, I propose an alternative method for estimating the undercount. Instead of relying on absolute counts in the birth data, this method uses ratios of male to female births. After adjusting these ratios for deaths and immigration, I compare predicted male to female ratios to those observed in the data to obtain estimates of the undercount. I can therefore estimate $U_{M,ay}$ as a function of the birth gender ratio, deaths, immigrants, emigrants, and the female undercount.

Assumption 1. *Following demographer methods, I assume that the total population for any subset of the population is*

$$P = B - D + I - E \tag{2}$$

where P is the total population, B is total births, D is total deaths, I is total immigrants, and E is total emigrants.

I denote the male (female) undercount as U_M (U_F) and the observed census population

as C , such that

$$P = \frac{C}{1 - U} \quad (3)$$

Proposition 1. *For a given age, a , and census year, y ,*

$$U_{M,ay} = 1 - \frac{C_{M,ay}}{\frac{B_{M,ay}}{B_{F,ay}} \left(\frac{C_{F,ay}}{(1 - U_{F,ay})} + D_{F,ay} - I_{F,ay} + E_{F,ay} \right) - D_{M,ay} + I_{M,ay} - E_{M,ay}} \quad (4)$$

Proof. Combining equations 2 and 3 for gender, G , gives

$$B_{G,ay} = \frac{C_{G,ay}}{1 - U_{G,ay}} + D_{G,ay} - I_{G,ay} + E_{G,ay} \quad (5)$$

Therefore

$$\frac{B_{M,ay}}{B_{F,ay}} = \frac{C_{M,ay}/(1 - U_{M,ay}) + D_{M,ay} - I_{M,ay} + E_{M,ay}}{C_{M,ay}/(1 - U_{F,ay}) + D_{F,ay} - I_{F,ay} + E_{F,ay}} \quad (6)$$

Solving for $U_{M,ay}$ gives proposition 1. □

Comparison of Ratio Method and Demographer Method

For the purposes of this paper, I use estimates of the undercount based on the ratio method as my primary estimates, although I also provide alternative estimates based on demographer methods. I argue that these estimates are superior to demographer estimates. The implied true gender ratios in the population are stable over time, consistent with what we would expect from birth and death rates. This is likely because these estimate are not impacted by incomplete birth data and are substantially less sensitive to issues of incomplete death and immigration data. Here, I consider the sensitivity of each method to incomplete data.

Assumption 2. *The share of uncounted births is the same for males and females.*

Proposition 2. *Let b be the share of births which are not included in the vital statistics birth records. Comparing the male undercount based on complete birth data, U^T , to the estimates of the undercount using the demographer and the ratio based method with complete death and immigration data gives*

$$\textbf{Demographer: } U^T - \hat{U}_D = \frac{C}{P^T} \times \frac{Bb}{P^T - Bb} \quad (7)$$

$$\textbf{Ratio based estimate: } U^T - \hat{U}_R = 0 \quad (8)$$

Proof. Recall that the true undercount is

$$U^T = 1 - \frac{C}{B - D + I - E} \quad (9)$$

If b is the share of births which are not included in vital statistics birth records, the observed number of births will be $B(1 - b)$. Assuming complete death and immigration data, demographers will estimate

$$\hat{U}_D = 1 - \frac{C}{B(1 - b) - D + I - E} = 1 - \frac{C}{PT - Bb} \quad (10)$$

Subtracting this from equation 9 gives Proposition 2.

In comparison, using the ratio method requires only the birth ratio as an input. It follows that because $\frac{B_M}{B_F} = \frac{B_M(1-b)}{B_F(1-b)}$, $U_T = \hat{U}_R$. □

Proposition 2 states that because the ratio based method uses only the birth gender ratio, not the total birth count, the estimate is unaffected by incomplete birth data. In contrast, the demographer method is highly sensitive to incomplete birth data. To illustrate the magnitude of this, take the example of estimates from vital statistics data from 1970. I will assume, just for the purposes of this example, that death and migration data for 1970 was complete. If 10 percent of births were missing from the vital statistics data, the demographer estimate of the undercount would be overestimated by 10 percentage points.

Although the ratio method for estimating the undercount still depends on the estimates of total deaths from vital statistics data, it is less sensitive to incomplete death and migration data than the demographer method is.

Proposition 3. *Let d be the share of deaths which are not included in the vital statistics death records. Comparing the male undercount based on complete death data, U^T , to the estimates of the undercount using the demographer and the ratio based method gives*

$$\textbf{Demographer: } U_M^T - \hat{U}_M = -\frac{C_M}{P_M^T} \times \frac{D_M d}{P_M^T + D_M d} \quad (11)$$

$$\textbf{Ratio based estimate: } U_M^T - \hat{U}_M = -\frac{C_M}{P_M^T} \times \frac{D_M d - \frac{B_M}{B_F} D_F d}{P_M^T + D_M d - \frac{B_M}{B_F} D_F d} \quad (12)$$

Proof. The proof for the impact of incomplete death data on demographer estimates in Proposition 3 is analogous to the proof for births in Proposition 2.

For the ratio based estimate, adding a undercount factor of d to the ratio based estimate of the male undercount gives

$$\hat{U}_M = 1 - \frac{C_M}{\frac{B_M}{B_F} \left(\frac{C_F}{(1-U_F)} + D_F(1-d) - I_F + E_F \right) - D_M(1-d) + I_M - E_M} = 1 - \frac{C_M}{P_M^T + D_M d - \frac{B_M}{B_F} D_F d} \quad (13)$$

Subtracting this from the true male undercount, U^T , gives the ratio based estimates from Proposition 3. □

These propositions demonstrate that, because it adjusts the undercount for female deaths as well as males, the ratio based estimate is less sensitive to undercounting deaths. To illustrate this, take the example of estimates from vital statistics data from 1970. I will assume, just for the purposes of this example, that birth data for 1970 was complete. If 25 percent of deaths were missing from the vital statistics data, the demographer estimate of the undercount would be overestimated by 6.1 percentage points. In contrast, the ratio based estimate would only be overestimated by 1.1 percentage point.

To calculate the ratio based estimates of the undercount, I estimate the number of total immigrants as the number of immigrants reporting to the Census. This is in contrast to Preston (1998) and Robinson et al (2002) who use legal immigration records supplemented with estimates of illegal immigration from survey datasets. Using census immigrant counts mechanically excludes estimates of non-reporting by the immigrant population from the estimate of the undercount, or implicitly assumes there is no undercount of immigrants. I choose this method because the size of immigrant population, especially the subset without legal status, is hard to measure accurately. Immigrants move frequently and are likely to be particularly at risk of non-reporting. Therefore, by excluding immigrant undercounts from the analysis, I am likely able to get a more accurate estimate of the native undercount. Therefore, for this paper I focus primarily on estimates of the undercount for the native born population.

One limitation of the ratio based method is that the estimation of the male undercount relies on having an estimate of the female undercount.

Proposition 4. *Let \hat{U}_F be the estimated underreporting rate for females. Comparing the male undercount based on the true female underreporting rate, U_F^T , to the estimates of the undercount using the ratio based methods based on \hat{U}_F gives*

$$U_M^T - \hat{U}_M = C_M \left(\frac{1}{P_M^T + \frac{B_M}{B_F} C_M \left(\frac{1}{1-U_F^T} - \frac{1}{1-\hat{U}_F} \right)} - \frac{1}{P_M^T} \right) \quad (14)$$

Proof. Recall that with complete vital statistics data,

$$\hat{U}_M = 1 - \frac{C_M}{\frac{B_M}{B_F} \left(\frac{C_F}{(1-\hat{U}_F)} + D_F - I_F + E_F \right) - D_M + I_M - E_M}$$

The estimate of U_M^T is the previous equation, substituting U_F^T for \hat{U}_F . Therefore, subtracting \hat{U}_M from U_M^F gives Proposition 4. □

Proposition 4 states that if I calculate the ratio based estimate of the male undercount based on an underestimate of the female undercount, I will obtain an underestimate. To illustrate this, I again take the example of 1970 data. I will assume, just for the purposes of this example, that vital statistics data on deaths and immigration are complete. If the true female undercount were the demographer estimate, 2.5 percent, but I incorrectly assumed that there was no female undercount, I would underestimate the male undercount by 2.2 percentage points.

Estimation of Undercount Based on Ratio Method

I estimate the birth gender ratio among blacks as the ratio of males to females born to black mothers in vital statistics data. I do this to avoid any discrepancy in fathers' reporting on birth certificates if, for example, men are more likely to be listed on the birth certificates of male children. I estimate total deaths using vital statistics mortality data. For each year and five year age group, I adjust total deaths by the share of black census respondents that were born outside the United States⁵. In non-census years, I adjust death rates by a weighted average of this share in the two closest census years. For 1921-1967, death data were only available by age range categorized as "white" and "non-whites." For these years, I adjust total deaths by the black share of non-whites from census data by age group. In non-census years, I adjust death rates by a weighted average of the black share of non-whites in the two closest census years. For these years, total deaths in an age group were also adjusted by the share of ages in a death age range falling into the age group.

I use census data on immigrants to measure the number of black respondents who were born outside of the United States in each year. I assume that the number of native born emigrants is negligible. There are no reliable figures on the number of native born emigrants or estimates of emigration by age and race (Jensen, 2013, Bhaskar et al, 2013). The Bureau of Consular Affairs estimates that there are nine million US citizens living overseas (Department of State, 2016). This represents 2.8 percent of the US population. However, many of these

⁵Relative rates of immigrant non-reporting and native will bias this slightly- expand

citizens were foreign born US citizens or returning naturalized US citizens (Bhaskar et al, 2013). Therefore, the true share of native born citizens who emigrate is likely to be under one percent. Bhaskar et al (2013) estimates net emigration of 18,000 native born US citizens between 2000 and 2010. This represents .006% of the population. There is also evidence that blacks are not more likely to emigrate than other races. For example, according to the 2008 CPS migration supplement, fewer all-black households had a member move abroad than all non-black households.

I propose two alternative estimates of the black female undercount. Table 3 shows estimates of the male undercount using the Ratio Based Method for each of these assumptions. The first set of estimates of the male undercount assumes that there is no female undercount. Estimates of the undercount using this method are decreasing from 16.6 percent in 1970 to 7.4 percent in 2010. The second set of estimates gives estimates of the male undercount using demographer estimates of the black female undercount. Here, estimates of the undercount start slightly higher, decreasing from 18.8 percent in 1970 to 7.4 percent in 2010. I note that both of these estimates are likely underestimates of the female undercount. Given my finding that demographers' estimates of the black male undercount are underestimated, it is likely that the same holds for black females. If true, this will cause my estimates of the undercount to be slightly understated.

3 Data

To estimate the causes and impacts of the under-coverage of black men in household based survey datasets, I rely primarily on three data sources. The first two are based on household based surveys, and are therefore subject to the undercount. The third is not based on a household based survey, and therefore is not subject to the undercount. I describe each data source below.

3.1 Household Based Survey Datasets

Census Data and the American Community Survey

I rely heavily on the Integrated Public Use Microdata Series (IPUMS) which include data from the all census years since 1850 and the American Community Survey (ACS) since 2000. For most analyses, I rely on the 1 percent IPUMS sample from the 1970 Census and the 5 percent sample from the 1980, 1990 and 2000 Censuses. These datasets provide a 1 in 100 and 1 in 20 sample of all households covered in the Censuses, respectively. As there are no microdata available from the 2010 Census, I also rely on the 2010 ACS sample. The

ACS provides an annual sample of approximately 1 percent of the United States Population. The available data from the Censuses and ACS include information on all members of a household and allow for linking among household members. The data include basic demographic information, educational attainment, employment, and wages. To calculate the 2010 undercount, I rely on tabulations of the full 2010 Census⁶.

The Census and ACS both survey the full population, including those living in institutions such as prisons. Unfortunately, more recent years of the data do not include a indication of whether the respondent is incarcerated. Instead, there is an indicator for whether a respondent is institutionalized. Following common practice, I use institutionalization as a proxy for incarceration. This is justified in that, among prime age males, the incarcerated make up the overwhelming majority of the institutionalized population (Charles and Luoh, 2010).

The Current Population Survey

For labor market outcomes in non-census years or those not available in the Census, I rely on data from the Current Population Survey (CPS). The CPS is a household based survey collected monthly and is the primary source for many labor market statistics. The CPS covers only standard households and does not sample people in institutions, long term care hospitals, nursing homes, or the military. Therefore, the incarcerated population is not included in the CPS. For this study, I use data from the Annual Social and Economic Supplemental survey (ASEC) which is included in March surveys and includes more detailed information on earnings. I accessed the data through the IPUMS data collection website.

3.2 Non-Household Based Survey Dataset

The Survey of Inmates in Correctional Facilities

The Survey of Inmates in Correction Facilities (SOI) is a periodically collected nationally representative random survey of inmates conducted by the Bureau of Justice Statistics. Available surveys were conducted in 1974, 1979, 1986, 1991, 1997, and 2004. Data for 1974, 1979, and 1986 cover only inmates in state correctional facilities. The 1991, 1997, and 2004 surveys also include data on inmates in federal correctional facilities. The survey provides information on the current sentence, previous sentences, offense information, demographic information, and information about the inmates educational and employment status prior to incarceration.

⁶Available at: <http://factfinder.census.gov/faces/nav/jsf/pages/index.xhtml>

4 Who are the missing black men?

4.1 Prior Research

Most theories regarding why the undercount of black males exists are either speculative or based on ethnographic research with small sample sizes. This is likely because there is very little available data on the non-reporting population. In 1969, *Ebony* published an article suggesting that the black undercount was driven by the fact that enumerators were white and did not want to go into overcrowded black neighborhoods (Anderson and Fienberg, 1999). There was some hope that the transition to collecting census information by mail would improve coverage, but the undercount remained in census years relying on mail surveys (Anderson and Fienberg, 1999).

Today, there are two leading theories on what is driving under-reporting. The first, and probably the most widely-believed theory, is that many prime age black men are inadvertently missed by surveyors because of the circumstances they live in (Pettit, 2012). This theory suggests that many black men have tenuous attachment to households and are therefore neither considered a part of households nor homeless (Anderson and Fienberg, 1999, Pettit, 2012). By this theory, although these men are living in households, the heads of the households view this only as a temporary living arrangement and view these men as guests rather than members of the household. Therefore, when responding to surveys, the heads of households do not include these men as residents. However, given that they have a home to sleep in each night, they are also not estimated as part of the homeless population. Further, according to this theory, black men are also more likely to be homeless or live in densely populated urban areas, and therefore are more susceptible to omission.

There is significant ethnographic research supporting this theory. Many ethnographic studies have confirmed that low-income prime age black males often bounce among households, usually living with parents or girlfriends, although most do not directly address the topic of under-enumeration (e.g. Edin and Nelson, 2013, Goffman, 2009). Most of the direct ethnographic evidence on under-enumeration comes from Census Bureau initiatives to better understand the nature of the problem. In 1988, the Census Bureau launched a research and evaluation project called the “Ethnographic Evaluation of Behavioral Causes of Undercount.” Through this initiative, the Census funded over fifty research projects aimed at identifying threats to census coverage in areas thought to be particularly susceptible to under-coverage (Brownrigg and Puente, 1992). One consistent theme throughout these reports was the prevalence of unstable household arrangements. For example, in a study of a low income and predominantly black neighborhood in Flint, Michigan, Darden et al. (1991) describe “the presence of temporary residents in households” as one of the primary reasons

for the under-coverage. They also suggest that under-coverage was driven by households being misidentified as single units and units being misidentified as vacant. However, of the 25 people identified by Darden et al. as being omitted from the Census, 13 were female, suggesting that omissions patterns in their sample were not representative of overall trends. Similarly, Hamid (1991) studies a predominantly low-income black block in Central Harlem. He also cites missed housing units, unusual household, and enumerator fear as the sources of the undercount. Also, like Darden et al., he does not find that the under-coverage is driven by adult black men.

The second primary theory of non-reporting among prime-age black men is that men are actively not responding to surveys. According to this theory, men are worried about the consequences of giving information about themselves to government agencies (Pettit, 2012). Although the Census Bureau claims that census responses “can not be used against you by any government agency or court”⁷, it is probable that some black men do not believe this claim. This likely stems from a deep historical distrust of government officials (Pettit, 2012).

There is a significant literature on the historical distrust of government and other institutions in black communities. Gamble (1993) explains that African American distrust of medical professionals dates back to the times of slavery, when medical theories on blacks being inferior were used to justify their enslavement. More recently, when knowledge of the abuses of black men in the Tuskegee Syphilis study became widespread, there was a resurgence in the perception that the medical community should not be trusted by black men (Gamble, 1993). Similarly, the history of African American distrust of policing dates back to slavery when slave “patrols” were organized to enforce discipline on slaves (France, 2014). Today, there is a widespread belief that policing disproportionately targets black men through police prejudice (Fryer, 2016) and more systematically through policies such as the “War on Drugs” and “Stop and Frisk” (France, 2014). In an ethnographic study of Spanish Harlem, New York, Bourgois (1990) describes the more general distrust of “big government” in black communities. He describes a general disbelief that the government is working in the interest of minorities, especially in low-income communities. Bernstein (1994) argues that this belief is likely connected to the long deep history of government economic oppression of African Americans, including slavery and the Jim Crow laws.

Many studies have connected the distrust of government to non-reporting to the Census. Bourgois (1990) describes “widespread distrust” as the primary reason for non-reporting to the Census. According to Bourgois, residents in Spanish Harlem (nearly all black or Hispanic) overwhelmingly do not believe that “any greater benefits will accrue to them or to the community if a more accurate census is achieved.” Most of the active avoidance of the

⁷United States Census Bureau, available at: <http://www.census.gov/2010census/about/protect.php>

Census described in ethnographic studies is by people in fear of legal ramifications. Bourgois connects this to a general disbelief that information is confidential and that government agencies do not share this information. Ethnographers have observed a range of legal reasons individuals have for non-reporting. Darden et al. (1991) describe households concealing residents to protect public assistance. Bourgois observes non-reporting among individuals “intensely involved in the underground economy.” Goffman (2009) describes how a group of wanted poor men in Philadelphia avoid all formal identification systems. These ethnographic studies suggest that the population that is most likely to be worried about being tracked by the government is individuals who are involved in illegal activity and worried about being caught.

To my knowledge, no studies have attempted to systematically distinguish between these two theories. Therefore, although we have anecdotal evidence of what the characteristics of non-reporters are likely to be, this is not adequate to impute characteristics or outcomes for the non-reporting population. In this study I take a different approach by attempting to provide insight into what we know about the population of non-reporters. I argue that understanding who the non-reporters are, regardless of whether they are actively not reporting or being circumstantially omitted, is what is necessary to impute outcomes for this population, and therefore to understand the impact of non-reporting on measured outcomes.

4.2 The Population at Risk of Incarceration

The ideal data for this study would be a survey or administrative dataset which includes data on the characteristics of non-reporters. Unfortunately, this requires both having data on non-reporters and being able to identify non-reporters directly in a dataset. To my knowledge, there is no dataset up to the task. Given the lack of data on the characteristics of non-reporters, I instead look to the incarcerated population.

The incarcerated population provides a unique opportunity because, once incarcerated, even those who would otherwise be non-reporters are automatically included in the Census. Therefore, if a non-reporter is incarcerated, there will be an additional person counted in the Census compared to the counterfactual of that person not having been incarcerated. As a result, even if under-reporting were unrelated to incarceration, a rise in incarceration would lead to a small drop in under-reporting. However, if non-reporters are over-represented in the incarcerated population, then a rise in incarceration would lead to a larger drop in under-reporting. Figure 3 shows the relation between omission rates in the Census and institutionalization rates between 1970 and 2010. Each dot represents a cohort comprised of a given five year age group (ex: ages 20-24) in a given year. This graph shows a clear

strong negative relation between institutionalization rates and census omission rates. Given that institutionalization is a proxy for incarceration, this suggests that non-reporters are over-represented in the incarcerated population.

There are many reasons to think that the population at risk of incarceration might be over-represented among non-reporters. First, the population at risk of incarceration might be more likely to live in unstable living conditions that would lead to circumstantial omission. This is supported by many of the ethnographic studies. Second, ethnographic studies suggest that the population that is most likely to be worried about being tracked by the government is individuals who are involved in illegal activity and worried about being caught.

4.3 Model of Non-Reporting

To estimate the share of the black male undercount that is driven by the non-reporting of the population at risk of incarceration, I propose the following simple model of incarceration risk and non-reporting. I note that the values of each parameter in this model vary by race, gender and age. I estimate this model on the population of black men ages 20-49 in my data. It is possible to estimate this model for any age, gender and race group available in my data.

To set up this model, I assume that every individual, i , in the population in year y has some underlying risk level, a_{iy} . Factors impacting an individual's level of a include personal characteristics, location, social networks, etc. For simplicity, and without loss of generality, I assume that a_{iy} is uniformly distributed between 0 and 1, with 1 being the highest risk.

Incarceration Risk

Define C_i as an indicator for whether individual i takes an action that puts him at risk of incarceration. In many cases, C_i may be actual participation in a crime punishable by incarceration. In other cases, C_i may be a probation violation, failure to pay a fine, or simply being in the wrong place at the wrong time. I refer to these as “risky actions.” For a given value of a_{iy} , the probability of individual i taking a risky action is

$$c(a_{iy}) = P(C_{iy} = 1)$$

Define $f(a_{iy}, s, y)$ as the probability that individual i , who live in state s , is incarcerated in year y . It follows that

$$f(a_{iy}, s, y) = c(a_{iy}, s, y)g(a_{iy}, s, y)$$

where $g(a_{iy}, s, y)$ is the probability of being incarcerated conditional on taking a risky action. I note that the relationship between $c(a_{iy}, s, y)$ and $g(a_{iy}, s, y)$ is ambiguous. It is possible that as the risk of incarceration conditional on risky behavior goes up, individuals become less likely to take risky actions. Conversely, states may respond to high rates of crime by increasing the severity of punishments, including the likelihood of incarceration.

In order to estimate this model, I make the following two additional assumptions.

1. $g(a_{iy}, s, y) = \omega_{sy}$. This means that the likelihood of incarceration conditional on taking a risky action is a constant which may vary by state and year but is not impacted by a . This assumption may be overly restrictive if, for example, individuals with higher levels of a tend to participate in riskier behavior within the broad categories of risky actions. I discuss the alternative assumption that $g(a_{iy}, s, y) = \omega_{sy}g(a_{iy})$ at the end of this section. This is a weaker assumption which says that the risk of incarceration conditional on taking a risky action is a function of a which increases/decreases proportionally throughout the distribution of a .
2. $c(a_{iy}, s, y) = \delta_{sy}c(a_{iy})$. This means that changes in the share of men in a given state and year who take a risky action increases/decreases proportionally throughout the distribution. For example, if state 2 has twice the population taking risky actions as state 1, then each individual in state 2 also has twice the chance of taking a risky action as their counterpart in state 1.

Without loss of generality, I let $\int_0^1 c(a_{iy})da_{iy} = 1$ and therefore δ_{sy} be the share of the population in state s in year y who takes a risky action. Then, combining assumptions 1 and 2, I get that the probability of incarceration is

$$f(a_{iy}, s, y) = \omega_{sy}\delta_{sy}c(a_{iy}) \tag{15}$$

The Undercount

I define $r(a_{iy}, C_{iy}, s, y)$ as likelihood of person i being a non-reporter in state s and year y . Note that a non-reporter is defined as someone who does not report to household based surveys if not incarcerated. Non-reporters are reported automatically to surveys if they are incarcerated.

To estimate the model, I assume that non-reporting varies by state and year only through variations in C_{iy} . This means that two individuals in different states and years with the same underlying risk and involvement in risky behavior have the same likelihood of being a non-reporter. Formally, I assume that $r(a_{iy}, C_{iy}, s, y) = r(a_{iy}, C_{iy})$. This assumption could be

violated if there are within state-year factors that impact propensity to report conditional on involvement in risky activity. I discuss the robustness of this assumption and the sensitivity of my results to violations of this assumption in section 4.6.

Combining the estimates of incarceration and non-reporting, I get that the size of the undercount is equal to

$$U_{sy} = \int_0^1 (\delta_{sy}c(a_{iy})) (1 - \omega_{sy})r(a_{iy}, 1)da_{iy} + \int_0^1 (1 - \delta_{sy}c(a_{iy}))r(a_{iy}, 0)da_{iy} \quad (16)$$

Let $\phi = \int_0^1 c(a_{iy})r(a_{iy}, 1)da_{iy}$. ϕ can be interpreted as the weighted average non-reporting rate conditional on taking a risky action, weighted by the distribution of a in the population who takes risky actions. From above, I have assumed that there is a constant rate of incarceration for men who take in risky actions. Therefore, ϕ is also the weighted average non-reporting rate for the incarcerated population. Placing ϕ into equation 16 gives

$$U_{sy} = \delta_{sy}(1 - \omega_{sy})\phi + \int_0^1 (1 - \delta_{sy}c(a_{iy}))r(a_{iy}, 0)da_{iy} \quad (17)$$

To estimate equation 17 using a regression, I impose the additional assumption that $\delta_{sy} = \gamma * X_{sy} + \mu_{sy}$ where X is a set of controls including reported crime rates, state dummies, year dummies, and state/time specific controls such as unemployment rates. μ is an error term. Practically, this means that the share of the population who takes a risky action is a function of observable covariates. Note that these covariates include reported crime rates, which are likely to be highly correlated with the share of the population taking risky actions. By definition, because ϕ is a constant, μ is orthogonal to ϕ ⁸.

I can therefore estimate ϕ by regressing the omission rate on the incarcerated share of the population using the following regression equation

$$U_{sy} = \beta_0 + \beta_1 * X_{sy} + \delta_{sy}\omega_{sy}\phi + \epsilon_{sy} \quad (18)$$

where $\delta_{sy}\omega_{sy}$ is the share of the adult black male population that is incarcerated and ϵ_{sy} captures both μ_{sy} and any measurement error⁹.

I note that if, instead, I assume that the probability of incarceration conditional on taking

⁸If instead we assume that $\phi_{sy} = \phi + \epsilon_{sy}$ then this requires the assumption that that ϵ_{sy} is orthogonal to μ_{sy} .

⁹Here, $\beta_0 = \alpha r_1 + \int_0^1 r(a_{iy}, 0)da_{iy}$ and $\beta_1 = \gamma_{sy}r_1 + \gamma_{sy} \int_0^1 c(a_{iy})r(a_{iy}, 0)da_{iy}$

a risky action is $\omega_{sy}g(a_{iy})$, then equation 16 becomes

$$U_{sy} = \int_0^1 (\delta_{sy}c(a_{iy}))(1 - \omega_{sy}g(a_{iy}))r(a_{iy}, 1)da_{iy} + \int_0^1 (1 - \delta_{sy}c(a_{iy}))r(a_{iy}, 0)da_{iy} \quad (19)$$

Here, I instead define $\phi = \int_0^1 c(a_{iy})g(a_{iy})r(a_{iy}, 1)da_{iy}$. ϕ can now be interpreted as the average non-reporting rate conditional on taking a risky action, weighted by the distribution of a in the incarcerated population¹⁰.

4.4 Estimation

I estimate this model using variation by state and year. Although I can't directly see the omission rate by state given that there are between state movers, I use the male to female ratio by state of birth as a proxy for under-reporting in that state. Because the true under-reporting rate by state is not known, the size of the male population is also not known. Therefore, I estimate the share of the male population that is incarcerated as the total number of institutionalized men divided by the total number of females. Figure 4 shows the relationship between the male to female ratio and the institutionalization rate by state of birth. To adjust for the growth in incarceration rates over the study period, this plot compares the residuals from a regression on year fixed effects. Consistent with the evidence using the omission rate binned by year and age group, this chart shows a clear positive relationship between the male to female ratio by state of birth and the institutionalization rate.

Although this section describes the analysis in the context of prime-age black men, this method can be applied to any other race and age group with available data. For example, Figure 5 shows the relationship between the male to female ratio and the institutionalization rate for white men by state of birth. Like the previous figure, this also plots residuals from regressions on year fixed effects. Unlike the case of prime age black men, there is no visible relationship between institutionalization and male to female ratios¹¹. This suggests that even in comparing groups with similar risk of incarcerations, the same mechanism which leads to non-reporting in the black male population may not exist for white men. This method could also be applied to estimate non-reporting in other surveys and demographic groups where

¹⁰This interpretation requires the additional assumption that $\int_0^1 c(a_{iy})g(a_{iy}) = 1$ which I make without loss of generality. Therefore, $\omega_{sy}\delta_{sy}$ is the incarcerated share of the population and equation 17 becomes

$$U_{sy} = -\delta_{sy}\omega_{sy}\phi + \int_0^1 c(a_{iy})g(a_{iy})r(a_{iy}, 1)da_{iy} + \int_0^1 (1 - \delta_{sy}c(a_{iy}))r(a_{iy}, 0)da_{iy}$$

¹¹The slope of the best fit line is .02

we might expect differential reporting patterns between two groups.

To estimate the under-reporting rate for the population at risk of incarceration using between state and year variation, I adjust the above model to use the male to female ratio. I also adjust to account for the census duplication rate for the incarcerated population, d .¹² Incorporating the estimate of undercount from equation 17, I get the following equation for the observed male to female ratio by state.

$$\frac{M_{o,sy}}{F_{o,sy}} = \frac{M_{t,sy} * (1 - \delta_{sy}(1 - \omega_{sy})\phi - \int_0^1 (1 - \delta_{sy}c(a_{iy}))r(a_{iy}, 0)da_{iy} + \delta_{sy}\omega_{sy}d)}{F_{o,sy}} \quad (20)$$

where M_o (F_o) and M_t (F_t) represent the observed and true number of males(females), respectively.

Note that duplication is not a problem in the estimation of the institutionalized share of the population because I am using the size of the female population as a proxy for the true size of the male population in this measure.

Rearranging terms from equation 20 gives

$$\frac{M_{o,sy}}{F_{o,sy}} = \frac{M_{t,sy}}{F_{o,sy}} K + \frac{M_{t,sy}}{F_{o,sy}} \delta_{sy}\omega_{sy}(\phi + d) \quad (21)$$

where $K = 1 - \delta_{sy}\phi - \int_0^1 (1 - \delta_{sy}c(a_{iy}))r(a_{iy}, 0)da_{iy}$.

To estimate equation 21, I impose the following additional assumptions.

1. $F_{t,sy} \approx F_{o,sy}$: there is negligible non-reporting among women. This assumption is largely consistent with evidence from demographers who found under 3 percent non-reporting for black females ages 20-49 between 1970 and 2010. If there is positive non-reporting among women, this would cause me to slightly under-estimate the non-reporting for men. If non-reporting for women is correlated with non-reporting for men by state and year, this will lead me to understate the variation in non-reporting by state and therefore my estimates would be lower bounds.
2. $\frac{M_t}{F_t} \approx 1$. This means that the true gender ratio is approximately equal to one. This

¹²It is a known problem that individuals can be counted more than once in the Census (Heimel and King, 2012). Although the rates of double counting are negligible in the overall population, duplication is a particular problem for individuals living in group quarters, such as college dorms or prisons. It is speculated that this is because the head of household views the individual as having permanent residence in the household and temporary residence in the group quarters. For example, if a spouse or family member is incarcerated for a month, the head of household is likely to include them on the census form as a resident. However, if the same individual is incarcerated at the time of the census count, that individual will be counted by the Census twice. To adjust my analyses for duplicates, I use an estimate of duplicate reporting for the institutionalized population from Heimel and King (2012).

assumption is strongly supported by the data. As I showed in section 2.2, I estimate that the true male to female ratio for blacks ages 20-49 is between .97 and 1. If the true ratio of males to females is marginally one, this would also suggest that my estimates are slightly understated.

Imposing the above assumptions, and adding the model assumption that $\delta_{sy} = \gamma * X_{sy} + \mu_{sy}$, equation 21 simplifies to

$$\frac{M_{o,sy}}{F_{o,sy}} = 1 - \int_0^1 r(a_{iy}, 0) da_{iy} + (\gamma * X_{sy} + \mu_{sy})(\phi - \int_0^1 c(a_{iy}r(a_{iy}, 0) da_{iy}) + \delta_{sy}\omega_{sy}(\phi + d)) \quad (22)$$

I can therefore estimate ϕ by regressing the omission rate on the institutionalized share of the population using the following regression equation.

$$\frac{M_{o,sy}}{F_{o,sy}} = \beta_0 + \beta_1 * X_{sy} + \delta_{sy}\omega_{sy}(\phi + d) + \epsilon_{sy} \quad (23)$$

where $\delta_{sy}\omega_{sy}$ is the share of the adult black male population that is incarcerated and ϵ_{sy} captures both μ_{sy} and any measurement error in $\frac{M_{o,sy}}{F_{o,sy}}$ and $\delta_{sy}\omega_{sy}$.

4.5 Results

To estimate the above model of non-reporting, I regress the male to female ratio among blacks ages 20-49 by state of birth on the institutionalization rate by state of birth and a set of controls for each census year between 1970 and 2010¹³. By analyzing data at the state of birth level, I take advantage of between state variation but eliminate issues of differential mobility between states. In my preferred specification, I include a set of variables to control for economic factors including the black poverty rate, the white poverty rate and the unemployment rate. I also control for the violent crime rate and the black share of the population. My preferred specification also includes year dummies, state dummies, and state specific time trends. I limit my analysis to state-years with at least 25,000 black men ages 20-49 in the population to limit the noise in my data due to measurement error.

Table 4 shows the results of the regression analyses. Column 1 shows that, controlling for state, year, state time trends, and relevant controls, a 1 percentage point increase in the institutionalization rate in a state is associated with a .994 percentage point increase in the observed male to female ratio. Adjusting for potential duplicates, this suggests that 90.3 percent of black men who become institutionalized would otherwise have been non-

¹³For 2010, I use ACS data from 2009-2011 to limit the noise due to measurement error.

reporters. In other words, this suggests that there is an 90.3 percent non-reporting rate among the population at risk of incarceration. Columns 2 to 5 show the results of the regression analysis with more limited sets of controls. These results show that removing controls increases the coefficient on the institutionalization rate.

To estimate what share of total non-reporting is accounted for by the population at risk of incarceration, I return to the model. In the context of the model, the coefficient on institutionalization rate is interpreted as $\phi = .903$. From the model, we know that the non-reporting share of the population can be written as $U_{sy} = \delta_{sy}(1 - \omega_{sy})\phi + \int_0^1 (1 - \delta_{sy}c(a_{iy}))r(a_{iy}, 0)da_{iy}$. Therefore, the share of non-reporters who are at risk of incarceration is

$$\frac{\delta_{sy}(1 - \omega_{sy})\phi}{\delta_{sy}(1 - \omega_{sy})\phi + \int_0^1 (1 - \delta_{sy}c(a_{iy}))r(a_{iy}, 0)da_{iy}} \quad (24)$$

I estimate ϕ , the weighted average non-reporting rate for the population at risk of incarceration, by regressing the male to female ratio on the institutionalized share of the population. However, the key statistic needed for imputing outcomes for non-reporters is what share of non-reporters are at risk of incarceration. To estimate this, I also need an estimate of $\int_0^1 (1 - \delta_{sy}c(a_{iy}))r(a_{iy}, 0)da_{iy}$. I think about this by considering values for δ_{sy} and $r(a_{iy}, 0)$.

I start by considering the implied size of the population at risk of incarceration if $r(a_{iy}, 0)$, the non-reporting rate for the population not at risk of incarceration, is 0 for all values of a . I calculate that assuming $r(a_{iy}, 0) = 0$, meaning that all of the non-reporting is driven by the population at risk of incarceration, implies that 21.6 percent of black men ages 20-49 are at risk of incarceration over the sample period¹⁴. Recall that for each state and year, I defined $\delta_{sy} = \gamma * X_{sy} + \mu_{sy}$, so this is a weighted average over states and years.

To assess the implications of this estimate, I consider how 21.6 percent compares to plausible estimates of the size of the population at risk of incarceration. I consider two estimate of the size of this population. First, I consider the share of the black male population that will go to prison at some point in their lifetime. I use the calculation from a 2003 study

¹⁴Setting $r(a, 0) = 0$ for all a and replacing $c_{sy} = \gamma * X_{sy} + \mu_{sy}$, equation 22 simplifies to:

$$\frac{M_{o,sy}}{F_{o,sy}} = 1 - \phi * \delta_{sy} + \delta_{sy} * \omega_{sy} * (\phi + d) \quad (25)$$

Therefore, using the estimates of $r + d = .994$ and $r = .903$ from Table 4, I estimate:

$$\delta_{sy} = \frac{1 + \delta_{sy} * \omega_{sy} * .994 - \frac{M_{o,sy}}{F_{o,sy}}}{.903}$$

by the Bureau of Justice Statistics which found that if incarceration rates remained at 2001 levels, 32.2 percent of black men would be incarcerated at some point in their lifetime (Bonczar, 2003)¹⁵. However, even if there is zero non-reporting among the population not at risk of incarceration, a 32.2 percent rate of the population at risk of incarceration would predict more than the full amount of under-reporting. This suggests that even among the population who will be incarcerated at some point in their lives, they are not at risk of incarceration for their full adult lives.

Second, I consider the share of black men ages 20-49 who are either arrested or incarcerated in a given year. This is also a rough estimate. There are likely men who are arrested for petty crimes or no crime at all who were not at risk of incarceration. Similarly, there are also likely men who were at risk of incarceration but managed to avoid arrest. I estimate the share of black men ages 20-49 who are either incarcerated or arrested in a given year to be approximately 21.8 percent¹⁶.

Taken together, these two estimates suggest that 21.6 percent is a reasonable, if not somewhat low, estimate of the share of the population at risk of incarceration. I therefore conclude that $r(a, 0)$ is either equal to 0 or very close to 0 for all values of a . This suggests that the non-reporting population is drawn completely or almost completely from the same population that ends up incarcerated. Therefore, I can use the characteristics of the incarcerated population to estimate the characteristics of the non-reporters who are not incarcerated. This analysis is done in section 5.

4.6 Sensitivity and Robustness

In this section I address the assumption that non-reporting varies by state and year only through variations in C_{iy} . This means that there is the same reporting rate of all men who participate in a risky activity and for all men who do not, across states and years. Formally, I write this assumption as $r(a_{iy}, C_{iy}, s, y) = r(a_{iy}, C_{iy})$. One concern with this assumption is that there may be state and year specific factors that impact the propensity for black men

¹⁵Bonczar calculates the share of black men who would be incarcerated in their lifetime at 2001 incarceration levels by comparing total prison admissions to the population. To account for issues of under-reporting in estimating the population size, Bonczar adjusts population counts using estimates from the 1990 Post Enumeration Survey (Bonczar, 2003). These estimates are similar in magnitude to the estimates that I use in this study (1990 Post Enumeration Survey).

¹⁶To estimate the share of black men who are either incarcerated or arrested in a given year I add the share of black men who are arrested in that year to the share who were incarcerated in the beginning of the year. I calculated the share of black men who were arrested as the total number of arrests of black males times the share of total arrests to men ages 20-49. I divide this by the average number of arrests per year, conditional on being arrested, from the NLSY. To avoid duplication, I subtract an estimate of the share of black men who were incarcerated at the beginning of the year and arrested later in the year.

to report. For example, in areas with higher mistrust of government officials, reporting may be lower.

To address concerns with the assumption of constant non-reporting rates within group, I consider the possibility that $r(a_{iy}, C_{iy}, s, y) = r(a_{iy}, C_{iy}) + \eta_{sy}$ where η_{sy} is a state-year specific error. In this case, the assumption necessary for identification of ϕ , the non-reporting rate for the population at risk of incarceration, is that η_{sy} is orthogonal to ϕ . In the context of the model, this means that η_{sy} is also orthogonal to ϵ_{sy} .

First, assume instead that in states where there is a higher black incarceration rate, there is also greater mistrust of government. This greater mistrust leads to higher non-reporting. This suggests that there is a positive correlation between incarceration risk and non-reporting rates, and therefore a positive correlation between η_{sy} and ϵ_{sy} . In this case, I would under-estimate ϕ . Given that the current estimate of ϕ is over 90%, and the maximum possible value of ϕ is 1 (full non-reporting), it seems implausible for ϕ to be meaningfully underestimated.

I now consider the arguably less intuitive scenario in which there is a negative correlation between η_{sy} and ϵ_{sy} . This may be either driven by a negative correlation between non-reporting and the size of the population at risk of incarceration, or by a negative correlation between non-reporting and the likelihood of incarceration conditional on being at risk. If this correlation can be captured by the included covariates, the model will still give an unbiased estimate of ϕ . For example, if there are state specific errors in non-reporting which correlate with incarceration rates, these will be picked by state fixed effects. However, if there are unobservable state-year specific factors which are not captured by observables, ϕ will be overestimated.

I consider the possible role of unobservables using methodology developed by Oster (2017), building on Altonji et al (2005). Altonji et al demonstrate that one can evaluate the robustness of results by estimating how important unobservables would have to be to eliminate the treatment effect. Oster expands on this by providing a method for estimating bounds on the treatment effect based on the assumption that the relationship between treatment and unobservables is similar to that between treatment and observables. I use this to estimate a lower bound on the non-reporting rate.

The Oster bounding estimates requires two inputs. The first, the “proportional selection assumption,” dictates how strongly the unobservable characteristics can correlate with the treatment relative to the observables. I follow Altonji et al, who default this to 1, suggesting that the observables and unobservables have the same correlation with treatment. The second input is the maximum possible R-squared of the model. Altonji et al assume a value of 1, suggesting that the true model would be perfectly predictive of outcomes in the data.

Oster points out that in the real world, we are limited by things like measurement error. To estimate a maximum R-squared, I run a simulations by adding measurement error to a regression of the gender ratio on state and year fixed effects, year time trends and the true residual.

To estimate Oster bounds, I use a residual regression of the gender ratio on state and year fixed effects and state time trends. Because state and year effects are so strong, it is unreasonable to assume that unobservables within state and year would behave similarly. I first estimate a maximum R-squared by estimating the residual regression as the true ratio with simulated measurement error on the true residual. I run 1,000 simulations and take the average of .62 as the maximum plausible R-squared. Based on this, I estimate a lower bound on the non-reporting rate for the population at risk of incarceration of 52%. Although the lower bound is meaningfully lower than the model estimate of 90%, this still implies that non-reporting is primarily driven by the population at risk of incarceration¹⁷.

Finally, I consider the possibility that both black and white incarceration rates are impacted by the strictness of the legal system, but only black incarceration rates are also impacted by the status race relations. If non-reporting rates are also impacted by race relations, this could cause a correlation between the black incarceration rate and non-reporting. To address this possibility, I instrument for the black incarceration rate with the white incarceration rate. I find no meaningful impact of this on my results, although the specification including state time trends is underpowered. The full results of this analysis are in Appendix Tables A.2 and A.3.

5 Updating Measures of Black-White Relative Outcomes

A large and growing literature has investigated the relative education, earnings and employments of blacks and whites. This literature has generally focused on black males given the contemporaneous changing status of women in the work force (e.g. Chandra, 2000, Bayer and Charles, 2018). The general consensus among labor economists has been that the earnings gap between black and white workers has been decreasing over the past forty years (see, for example, Margo, 2016). Trends in educational achievement suggest that the gap in college attendance between blacks and whites has been decreasing since 1970 but that relative high

¹⁷Assuming a 52% non-reporting rate for the population at risk of incarceration and no non-reporting for the population no at risk of incarceration implies that 32% are at risk of incarceration. This is consistent with the Bonczar (2003) estimate of the share of the black male population who would be incarcerated in their lifetime at 2001 incarceration rates.

school dropout rates have been more steady (Lee, 2002). However, some researchers have argued that this does not represent true progress in racial equality, pointing to the decreasing employment rates among black males due to incarceration and discouraged labor market drop outs (Chandra, 2000). A growing literature challenges labor market statistics based on the Current Population Survey which does not sample from the prison population. In a series of papers, Western and Pettit provide updated statistics on relative earnings and wages corrected to include the prison population (Western and Pettit, 2010, Western and Pettit, 2000). Consistent with this, in a comprehensive analysis of relative earnings and earnings ranks of blacks and whites, Bayer and Charles (2018) analyze earnings for black and white men. They find that the gap in median earnings between blacks and whites has been growing since the 1970s. Corrections have also been made to educational attainment for blacks and whites and voter participation (Pettit, 2012). However, to my knowledge, no one has attempted to address issues of the under-coverage of black men in household based-survey data. Although Pettit (2012) acknowledges the undercount in her book, “Invisible Men,” her updated statistics only adjust for the currently incarcerated population. Therefore, her calculations, like most other existing calculations, rely on the unlikely assumption that the non-reporters are representative of the population.

There is a significant body of research that has emerged addressing how we should think about black-white disparities. These studies generally take the educational, earnings and wage gaps as known and attempt to explain them with differences in educational attainment (e.g. Maxwell, 1994), school quality (e.g. Card and Krueger, 1991), human capital (e.g. Neal and Johnson, 1995), discrimination (e.g. Pager et al., 2009), criminal records (e.g. Pager, 2003), etc. However, as my research shows, researchers should be careful to think about how survey coverage may impact the accuracy of estimates.

I correct statistics on education, earnings and employment gaps to adjust for the under-reporting of prime age black men. As I showed in section 4, my research suggests that the non-reporters are primarily taken from the population of black men at risk of incarceration. I therefore use data from the Survey of Inmates (SOI) to impute education, wages and employment for the non-reporters. These calculations are based on the assumption that labor market outcomes before prison admission are representative of labor market outcomes for the population at risk of incarceration. This assumption seems reasonable given that I am estimating the size of the population at risk of incarceration as the share of the population directly at risk. As I discussed in section 2, my primary estimates are based on estimates of the native undercount using the ratio based method with demographer estimates of the female undercount. As a result, I limit my primary analysis to the native born population. I present all results based on alternative estimates of the undercount in Appendix D. Also

consistent with estimates of the undercount, I limit my analyses to respondents listed as “black only” or “white only.”

My imputation strategy also assumes that under-reporters are randomly distributed within the population of prisoners. In reality, it is likely that under-reporters are negatively selected among the prison population. If this is the case, my adjusted estimates are likely to over-estimate achievement for black men and under-estimate achievement differentials.

5.1 Educational Attainment

Figure 6 shows the unadjusted distribution of educational attainment for native-born blacks and whites ages 25-29 between 1970 and 2010 based on census data. Not taking into account under-reporting, this figure shows that in 1970 blacks were 29 percent less likely to finish high school than whites. By 2010 that number was 13 percent. Similarly, in 1970 blacks were 56 percent less to go to college than whites but by 2010 that number was down to 26 percent. However, adjusting for under-reporting shows that educational attainment of blacks was even lower relative to whites.

To adjust rates of educational attainment to account for under-reporting, I use the educational attainment distribution for prisoners prior to admission from the Survey of inmates. Appendix Table C.1 shows the educational distribution from each year of the SOI. For 1970 and 2010, I impute the educational distribution from the closest SOI - 1974 and 2003, respectively. For 1980-2000, I estimate the educational distribution of non-reporters as the weighted average educational distribution of inmates from the SOI immediately prior to the census year. Appendix Table C.2 shows these values. I note that incarceration rates increased dramatically over the time period of this study. Therefore it is likely that the incarceration population was selected differently across years. There is evidence of this in the data. For example, controlling for age at admission, the education of inmates actually decreases over time despite the education of the general population increasing. If the incarcerated population in 2003 is more representative of the full population at risk, I am slightly over-estimating the educational attainment for non-reporters in earlier years. I limit my population to inmates admitted in the survey year in order to avoid over-weighting inmates who are incarcerated for longer periods of time. I also limit my sample to inmates admitted at age 20 or older. For additional details on the estimation of non-reporter education levels, see Appendix C.

Table 5 shows the adjusted rates of high school completion and college attendance for black men ages 25-29 incorporating under-reporting. As this table shows, black men are now an average of 4.5 percentage points less likely to complete high school and 3.2 percentage

points less likely to attend college than unadjusted statistics show. Figure 7 shows the adjusted gap in high school completion rates and college attendance for blacks to whites. As Figure 7 shows, after adjusting to incorporate under-reporting, trends still show improvements in black educational attainment relative to whites, however the relative levels are now lower. Even by 2010, blacks were still 19 percent less likely to complete high school and 31 percent less likely to attend college. Adjusting for non-reporting, in 2010 the black-white gap in high school completion is 40 percent larger accounting for under-reporting and the gap in college attendance is 22 percent larger.

5.2 Employment

The employment gap between blacks and whites is an important piece in the story of relative achievement of blacks and whites. I therefore provide updated estimates of both the employment rate for black men and the unemployment rate for black men.

Employment Rates

Figure 8 shows the unadjusted employment rates for black and white men ages 20-49 between 1970 and 2010. This figure, consistent with Chandra (2000), shows that black male employment is decreasing relative to white men. Much of this decrease is driven by high levels of incarceration of black men. However, as I showed in section 3, these figures are missing a substantial share of the population of black men who were at risk of incarcerated but not actually incarcerated. Therefore, upon incarceration, men are being added to the count of total men who were not previously included. In order to understand how this impacts employment rates, I impute the employment rates of the population at risk of incarceration.

To calculate the employment rates for the non-reporting population, I use the employment rates for prisoners prior to incarceration based on the SOI. One key challenge with this imputation strategy is that SOI data is not available in census years. Therefore, I use ratios of employment rates for the incarcerated population to employment rates for the non-incarcerated population in SOI years to impute employment in non-SOI years. I estimate by age group, and then take the average ratio to get a single employment ratio by education level. These calculations are shown in Appendix Table C.3. I then estimate employment for the non-responders as the estimated non-reporter wages by education group, weighted by the share of non-reporters in a given education group. Because the Census is not available in SOI years, I use the CPS as the source for employment data. I supplement this with institutionalization rates from the Census to get total employment. For additional details on the estimation on non-reporter employment levels, see Appendix C.

Table 6 shows the adjusted and non-adjusted employment levels for black men. As Table 6 shows, adjusting for non-reporting actually increases employment rate for black men. This is largely because much of the non-employment for black men is due to incarceration and non-reporters are by definition not incarcerated. Further, adjusting for age and education, employment rates for prisoners prior to incarceration are comparable to the overall population. Figure 9 shows the adjusted and unadjusted gap in employment for black and whites. As Figure 9 shows, the black-white employment gap is increasingly slightly more than unadjusted estimates suggest. By 2010, the adjusted and unadjusted black-white employment gap differ by less than 1 percent.

Unemployment Rates

Figure 10 shows the unadjusted unemployment rates for black and white men ages 20-49 between 1970 and 2010 based on the BLS definition of unemployment rates¹⁸. The unemployment rates over this time period vary significantly. In 1970 the white unemployment rate was only 3.3 percent and the black unadjusted unemployment rate was only 6.0 percent. In contrast, in 2010, during the Great Recession, the white unemployment rate was at 10.8 percent and the black unadjusted unemployment rate was as high as 23.1 percent. Over this period the black unadjusted unemployment rate was consistently significantly higher than the white unemployment rate, averaging over two times as high.

I use the same method that I used to impute employment rates for non-reporters to impute unemployment rates. I calculate the share of inmates reporting to have been without employment and looking for work prior to incarceration. I use this to predict the relationship between the unemployment share of the population between the SOI and the CPS by age group. I then calculate the unemployed share of the population for non-reporters by multiplying the unemployment rate for each age and education group in a year by this ratio. I calculate the total labor force share in a given age, education, year cohort as the sum of my imputed employed and unemployed percentage. For additional details on the estimation of non-reporter unemployment rates, see Appendix C.

Table 7 shows the adjusted and unadjusted unemployment rates for black men. Adjusting the unemployment rate for black men to account for under-reporting changes the estimated unemployment rate by an average of .6 percentage points. These are meaningful differences for unemployment rates. These changes in the unemployment rate are driven by the fact that there are much higher unemployment rates for high school drop outs than high school

¹⁸The Bureau of Labor Statistics defines the unemployment rate as the total number of employed persons divided by the total number either employed or looking for work. For more information see “How the Government Measures Unemployment” https://www.bls.gov/cps/cps_htgm.htm

graduates, and that high school drop outs are significantly over-represented in the population of non-reporters. Figure 11 shows the adjusted and unadjusted ratio of black to white unemployment rates. Adjusting the unemployment rate to account for under-reporting shows that the ratio of black to white unemployment is an average of 5 percent higher than the previously calculated.

5.3 Earnings

Relative earnings of blacks and whites is a topic of great interest among labor economists. Many studies have attempted to explain why black men make significantly less than white men. However, these studies generally ignore any under-coverage issues of surveys. I therefore adjust these estimates to account for under-reporting. Figure 12 shows the unadjusted average annual earnings, shown as real earnings in 2010 dollars, for working black and white men ages 20-49. On average, over the time period 1970-2010 blacks made 28 percent less than whites based on the unadjusted figures. Although the earnings of black men increased between 1980 and 2010, the earnings of white men increased at a faster pace. Therefore, after a drop between 1970 and 1980, the unadjusted estimate of the black-white earnings gap increased steadily between 1980 and 2010. This is consistent with the findings of Bayer and Charles (2018), who found that the median wage gap was increasing over this period.

As with employment, to calculate earnings for the non-reporting population, I compare reported earnings prior to incarceration by inmates in the SOI. Appendix Table C.7 shows the relative earnings reported by inmates compared to average earnings in the general population from the CPS by education level weighted by age to match the age distribution in the SOI. Like employment, to calculate earnings for non-reporters I take the average ratio of inmate earnings to the general population by education level and apply this ratio to earnings in the CPS by age group, education level and year cohorts. For additional details on the estimation of non-reporter earnings levels, see Appendix C.

Table 8 shows the adjusted and unadjusted annual earnings for working black men ages 20-49. Although inmates report having similar employment rates to the general population, their reported annual earnings are less than half of the general population. Therefore, adjusting for under-reporting significantly lowers the estimated average annual earnings for black men. On average, adjusting for non-reporting lowers average annual earnings for black men by \$2,270 or seven percent. Figure 13 shows the adjusted and unadjusted ratio of black to white earnings. Adjusting for under-reporting, I find that the black-white earnings gap is significantly larger than previous estimates. Further, I find after adjusting for non-reporting, there is no longer evidence that the earnings gap has been meaningfully increasing since 1980.

In fact, between 1980 and 2010, the earnings gap varied by less than two percentage points. This suggests that smaller earnings

6 Conclusion

In this paper, I demonstrate the importance of considering the impact of undercoverage of surveys in analyses of household based survey datasets. I demonstrate that trends in the undercount suggest that the non-reporting population is primarily drawn from the population at risk of incarceration. I show that adjusting statistics to account for the under-reporting of prime age black men to survey datasets can lead to highly biased estimates of outcomes for this population. I show disparities between blacks and whites in educational attainment, unemployment rates and annual earnings are understated due to omitting the non-reporting population. I do not find a significant bias in the calculation of employment rates.

It is likely that I have understated the impact of under-reporting on calculated outcomes for black men. My analyses are based on the assumption that non-reporters are evenly distributed among the population at risk of incarceration. If non-reporters are actually negatively selected among this population, the true impact of non-reporting is actually even greater. Also, the true impact of non-reporting on measured outcomes is likely greater if labor market outcomes for the incarcerated population are worse after incarceration.

In this paper, I have focused on a small number of outcomes for a single population - prime age black men. Although I have only demonstrated the impact of under-reporting on estimates of educational attainment, employment and earnings, it is likely that many other relevant statistics calculated from household based datasets are severely biased by under-reporting. Further, there are other populations and surveys which are compromised by undercoverage. I hope this paper highlights the importance of considering survey coverage when estimating statistics based of survey datasets. This paper also highlights the importance of collecting administrative datasets, and pursuing other alternative measures of labor market statistics.

References

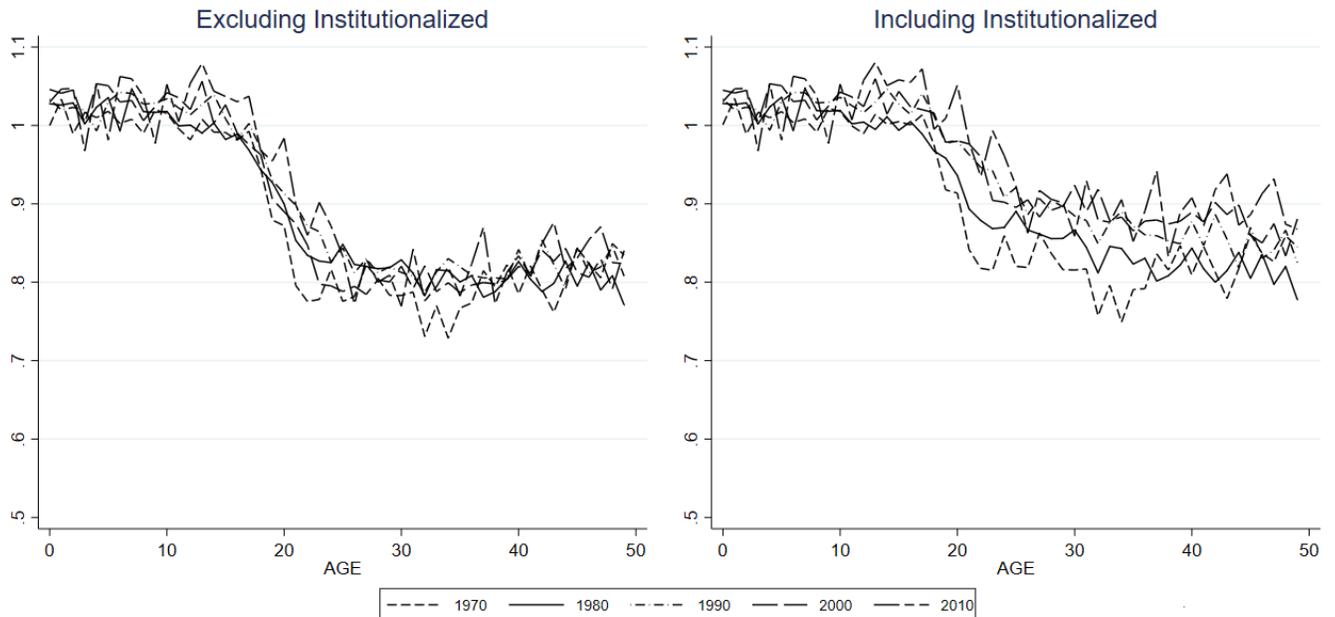
1. Anderson, M., and Fienberg, S. E. (1999). *Who counts?: The politics of census-taking in contemporary America*. Russell Sage Foundation.
2. Bernstein, D. E. (1994). *Licensing Laws: A Historical Example of the Use of Government Regulatory Power Against African Americans*. *San Diego L. Rev.*, 31, 89.

3. Bonczar, T. P. (2003). Prevalence of Imprisonment in the US Population, 1974-2001. US Department of Justice, Office of Justice Programs. NCJ 197976.
4. Bourgois, P. (1990) Hypotheses and Ethnographic Analysis of Concealment in the Underground Economy: The Economic and Ideological Dynamics of the Census Undercount. Final Report for Joint Statistical Analysis 88-24.
5. Brownrigg, L. and de la Puente, M. (1990). Alternative Enumeration Methods and Results: Resolution and Resolved Population by Site.
6. Bureau of Labor Statistics (2016) How the Government Measures Unemployment. Available at: http://www.bls.gov/cps/cps_htgm.htm#unemployed
7. Card, D., and Krueger, A. B. (1991). School quality and black-white relative earnings: A direct assessment (No. w3713). National Bureau of Economic Research.
8. Chandra, A. (2000). Labor-market dropouts and the racial wage gap: 1940-1990. *The American Economic Review*, 90(2), 333-338.
9. Charles, K. K., and Luoh, M. C. (2010). Male incarceration, the marriage market, and female outcomes. *The Review of Economics and Statistics*, 92(3), 614-627.
10. Cortes, A., Henry, M., de la Cruz, R. and S. Brown (2012) The 2012 Point in Time Estimates of Homelessness. 2012 Annual Homeless Assessment Report, Vol 1.
11. Darden, J., Jones, L., and Price, J. (1991). Ethnographic Evaluation of the Behavioral Causes of Undercount in a Black Ghetto of Flint, Michigan. Prepared under Joint Statistical Agreement, 89, 43.
12. Edin, K., and Nelson, T. J. (2013). *Doing the best I can: Fatherhood in the inner city*. Univ of California Press.
13. France, K. (2014) *Blacks and Law Enforcement: A History*. Killing the Breeze.
14. Fryer Jr, R. G. (2016). An empirical analysis of racial differences in police use of force (No. w22399). National Bureau of Economic Research.
15. Gamble, V. N. (1993). A legacy of distrust: African Americans and medical research. *Am J Prev Med*, 9(6 Suppl), 35-38.
16. Goffman, A. (2009). On the run: Wanted men in a Philadelphia ghetto. *American Sociological Review*, 74(3), 339-357.

17. Hamid, A. (1991). Ethnographic Follow-up of a Predominantly African American Population in a Sample Area in Central Harlem, New York City: Behavioral Causes of the Undercount of the 1990 Census: Final Report for Joint Statistical Agreement 89-28 (No. 11). Center for Survey Methods Research, Bureau of the Census.
18. Juhn, C. (2003). Labor market dropouts and trends in the wages of black and white men. *Industrial and Labor Relations Review*, 56(4), 643-662.
19. Lee, J. (2002). Racial and ethnic achievement gap trends: Reversing the progress toward equity?. *Educational researcher*, 31(1), 3-12.
20. Maxwell, N. L. (1994). The effect on black-white wage differences of differences in the quantity and quality of education. *Industrial and labor relations review*, 47(2), 249-264.
21. Moore, W., Pedlow, S., Krishnamurty, P., and Wolter, K. (2000) National Longitudinal Survey of Youth 1997 (NLSY97) Technical Sampling Report. National Opinions Research Center.
22. Neal, D. A., and Johnson, W. R. (1995). The role of pre-market factors in black-white wage differences (No. w5124). National Bureau of Economic Research.
23. Pager, D. (2003). The mark of a criminal record1. *American journal of sociology*, 108(5), 937-975.
24. Pager, D., Western, B., and Bonikowski, B. (2009). Discrimination in a low-wage labor market a field experiment. *American sociological review*, 74(5), 777-799.
25. Passel, J. 1990. Demographic Analysis: A Report on Its Utility for Adjusting the 1990 Census. The Urban Institute.
26. Pettit, B. (2012). *Invisible men: Mass incarceration and the myth of black progress*. Russell Sage Foundation.
27. Preston, S. H., Elo, I. T., Foster, A., and Fu, H. (1998). Reconstructing the size of the African American population by age and sex, 1930-1990. *Demography*, 35(1), 1-21.
28. Preston, S. H., Elo, I. T., Hill, M. E., and Rosenwaike, I. (2013). *The Demography of African Americans 1930-1990*. Springer Science and Business Media.

29. Robinson, J. G., Admlakha, A. and West, K. (2002) Coverage of Population Census 2000: Results from Demographic Analysis. Report for the Annual Meeting of the Population Association of America.
30. Robinson, J. G., Ahmed, B., Gupta, P. D., and Woodrow, K. A. (1993). Estimation of population coverage in the 1990 United States Census based on demographic analysis. *Journal of the American Statistical Association*, 88(423), 1061-1071.
31. Wagner, P. (2014) Tracking State Prison Growth in 50 States. Prison Policy Initiative. Available at: <https://www.prisonpolicy.org/reports/overtime.html>
32. Western, B., and Pettit, B. (2000). Incarceration and racial inequality in men's employment. *Industrial and Labor Relations Review*, 54(1), 3-16.
33. Western, B., and Pettit, B. (2005). Black-White Wage Inequality, Employment Rates, and Incarceration. *American Journal of Sociology*, 111(2), 553-578.
34. Western, B., and Pettit, B. (2010). Incarceration and social inequality. *Daedalus*, 139(3), 8-19.

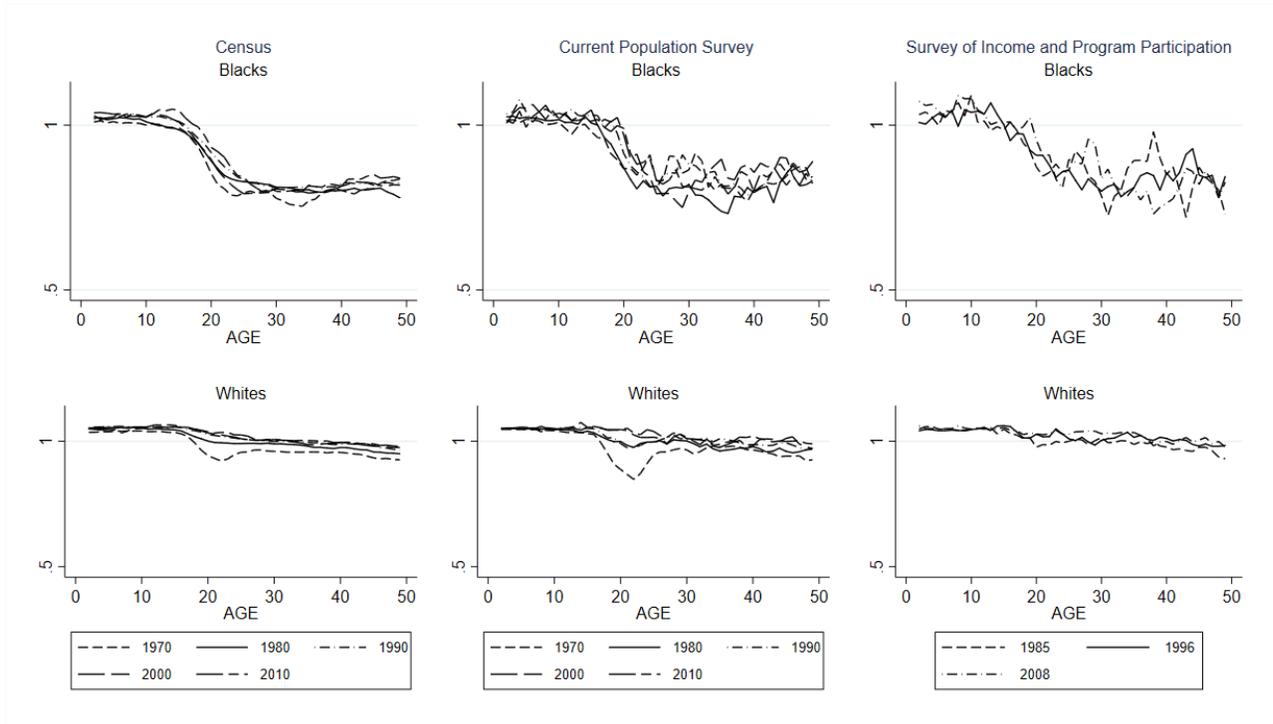
Figure 1
Male to Female Ratio by Age
Black Respondents in Census Data



Notes: Respondents were considered to be black if black was the only race listed on the Census form.

Source: IPUMS Census data (1970-2000) and IPUMS ACS data (2010)

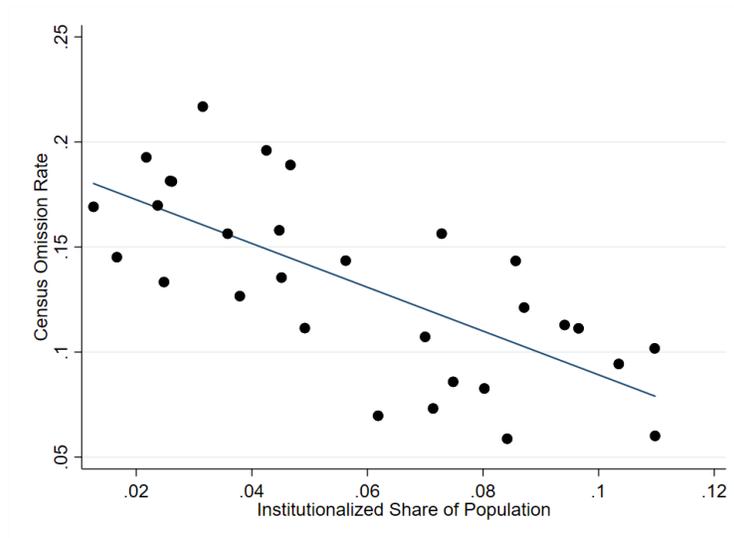
Figure 2
Male to Female Ratio by Age
Five Year Moving Average



Notes: Respondents were considered to be black if black was the only race listed. Respondents listed as living in group quarters defined as institutions were excluded. Estimates are weighted using survey provided person weights. 1985 estimates from the SIPP include data from the 1984, 1985 and 1986 panels.

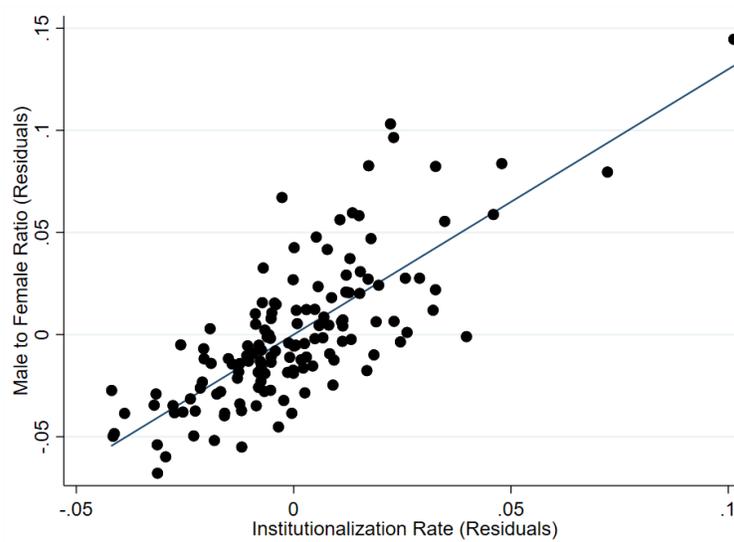
Sources: Census: IPUMS Census data (1970-2000) and IPUMS ACS data (2010). Survey of Income and Program Participation (SIPP): 1984, 1985, 1986m 1996 and 2008 SIPP panels. Current Population Survey (CPS): IPUMS Current Population Survey data.

Figure 3
 Institutionalization Rate and Census Omission Rate
 Black Men Ages 20-49
 1970-2010



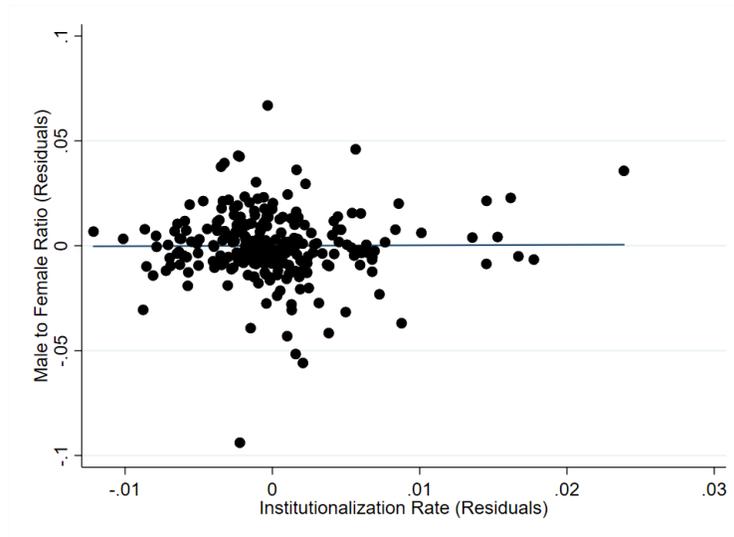
Notes: Observation is a five year age group-year. Institutionalization rate is measured as the total number institutionalized men divided by the total number of females. Data for 2010 are based on averages from 2009-2011.
 Source: Institutionalization rate: IPUMS Census data (1970-2000) and IPUMS ACS data (2010). Omission rate: See section 2.2

Figure 4
 Institutionalization Rate and Male to Female Ratio
 Black Men Ages 20-49



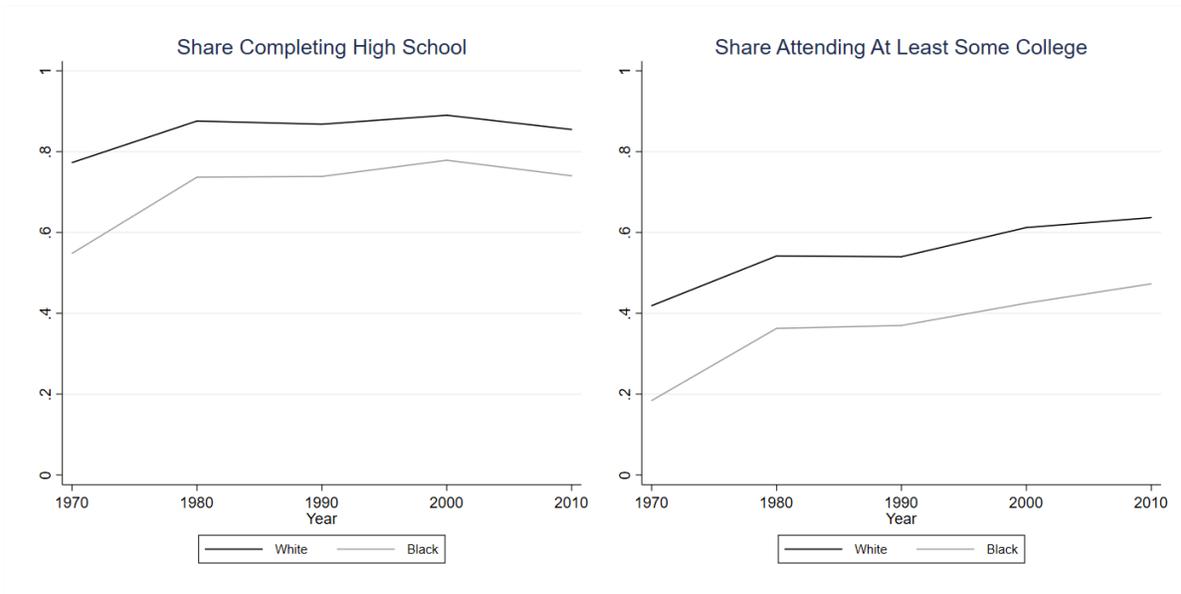
Notes: X axis shows the residuals of a regression of institutionalization rate on year fixed effects. Y axis shows the residuals of a regression of male-female ratio on year fixed effects. Each observation is a state of birth-year. Analysis includes all state of birth-years in which there were at least 25,000 black men. Incarceration rate is measured as the total number institutionalized men divided by the total number of females. Data for 2010 are based on averages from 2009-2011.
 Source: IPUMS Census data (1970-2000) and IPUMS ACS data (2010)

Figure 5
Institutionalization Rate and Male to Female Ratio
White Men Ages 20-49



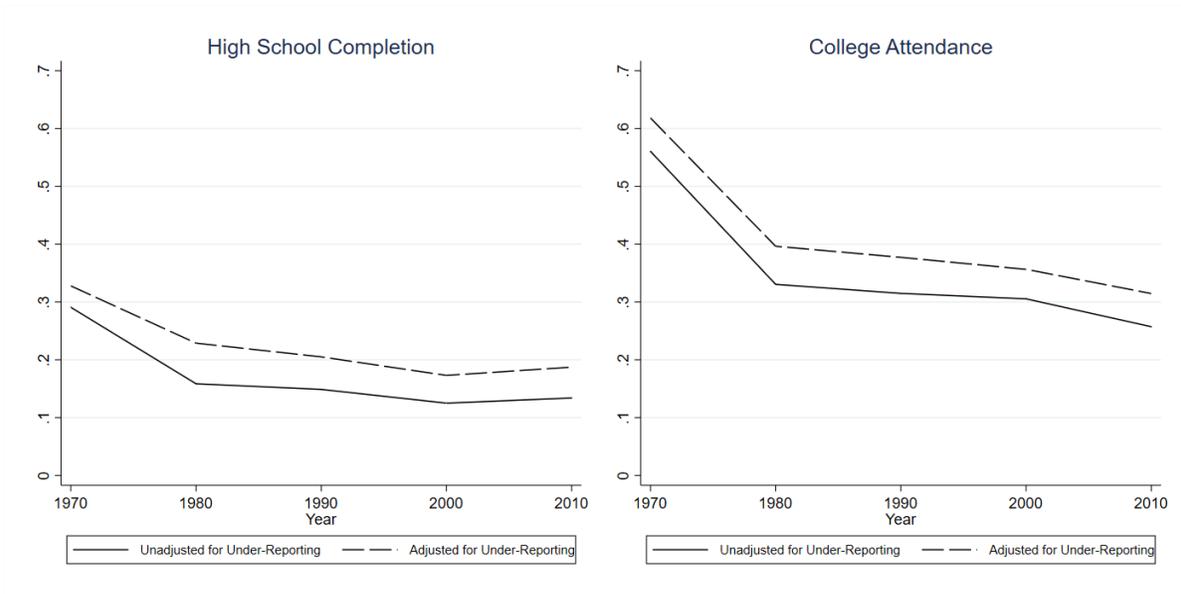
Notes: X axis shows the residuals of a regression of institutionalization rate on year fixed effects. Y axis shows the residuals of a regression of male-female ratio on year fixed effects. Each observation is a state of birth-year. Analysis includes all state of birth-years in which there were at least 25,000 white men. Incarceration rate is measured as the total number institutionalized men divided by the total number of females. Data for 2010 are based on averages from 2009-2011.
Source: IPUMS Census data (1970-2000) and IPUMS ACS data (2010)

Figure 6
 Unadjusted Educational Attainment
 Native-born Men Ages 25-29



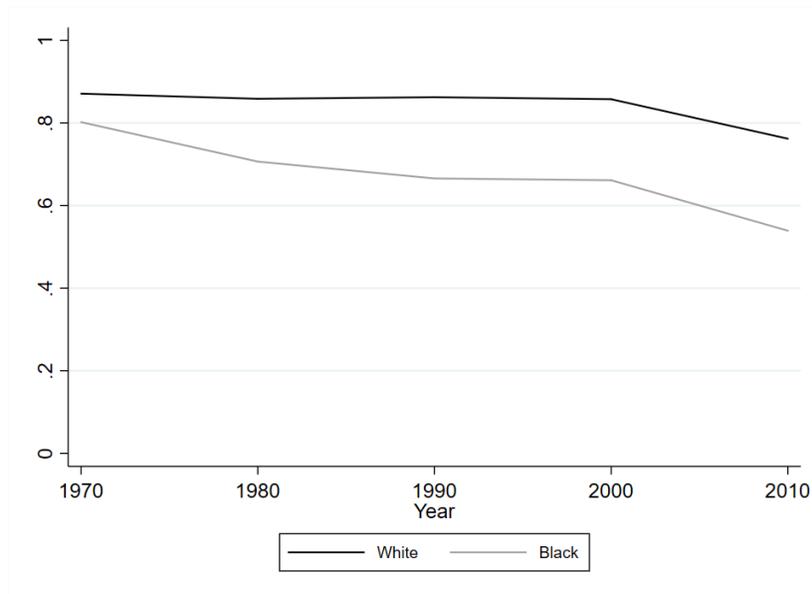
Notes: Respondents are considered black (white) if black (white) is the only race identified.
 Source: IPUMS Census data (1970-2000) and IPUMS ACS data (2010)

Figure 7
 Black-White Educational Attainment Gap
 Native-born Men Ages 25-29
 Ratio Based Method - Demographer Female Undercount



Notes: See Appendix C for estimation methods and sources.

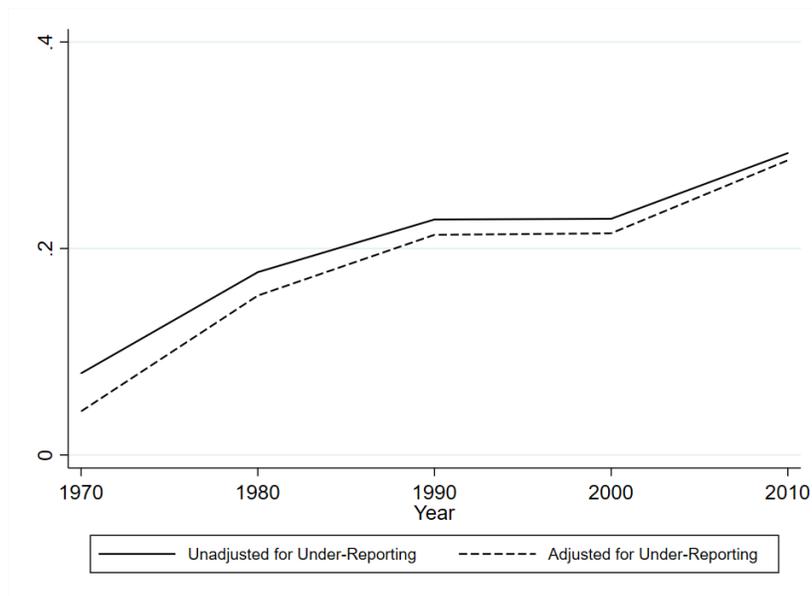
Figure 8
 Unadjusted Employment Rates
 Native-born Men Ages 20-49



Notes: Respondents are considered black (white) if black (white) is the only race identified. Employment rates are calculated as the share of the non-institutionalized population based on the CPS, adjusted for the size of institutionalized based on the Census.

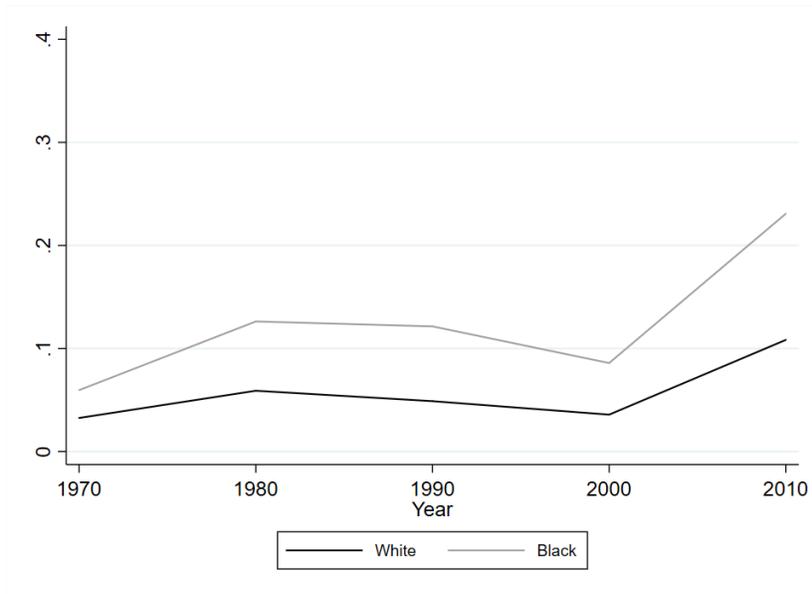
Source: IPUMS Census data (1970-2000) and IPUMS ACS data (2010), IPUMS Current Population Survey data

Figure 9
 Black-White Employment Gap
 Native-born Men Ages 20-49
 Ratio Based Method - Demographer Female Undercount



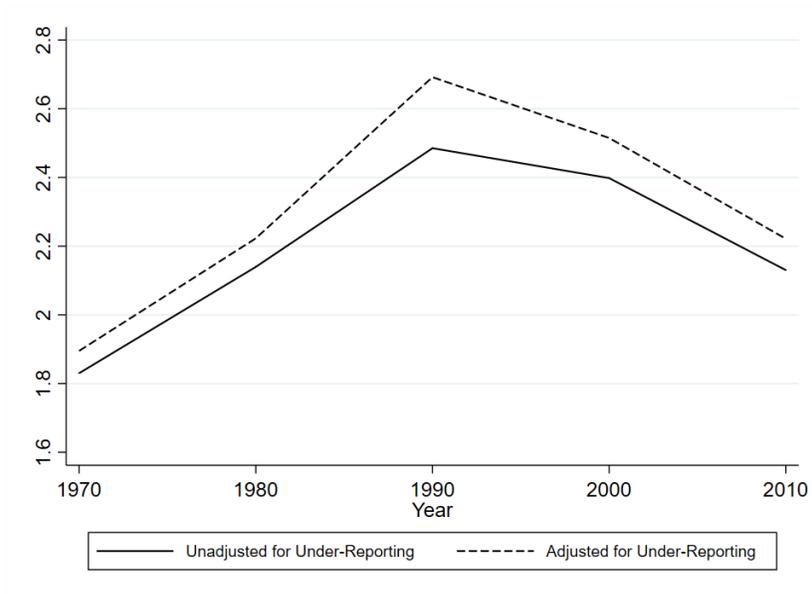
Notes: See Appendix C for estimation methods and sources.

Figure 10
 Unadjusted Unemployment Rates
 Native-born Men Ages 20-49



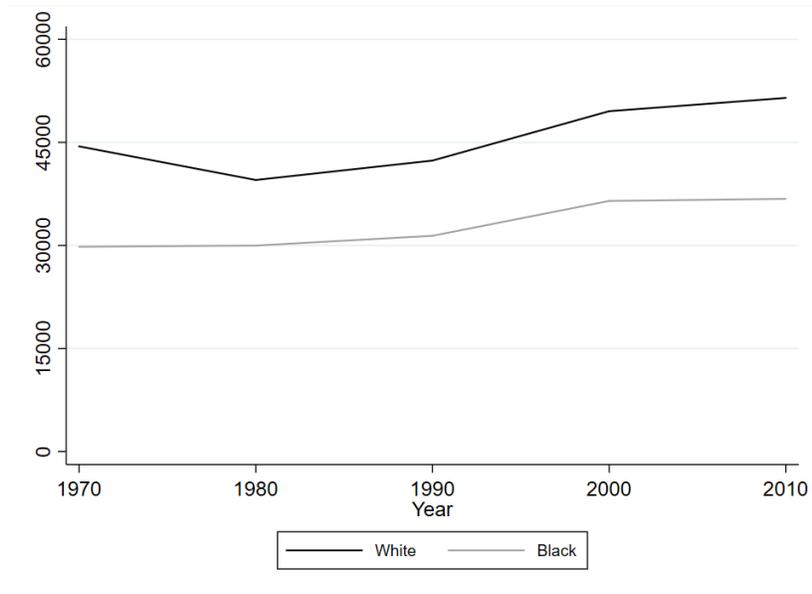
Notes: Respondents are considered black (white) if black (white) is the only race identified.
 Source: IPUMS Current Population Survey data

Figure 11
 Ratio of Black to White Unemployment
 Native-born Men Ages 20-49
 Ratio Based Method - Demographer Female Undercount



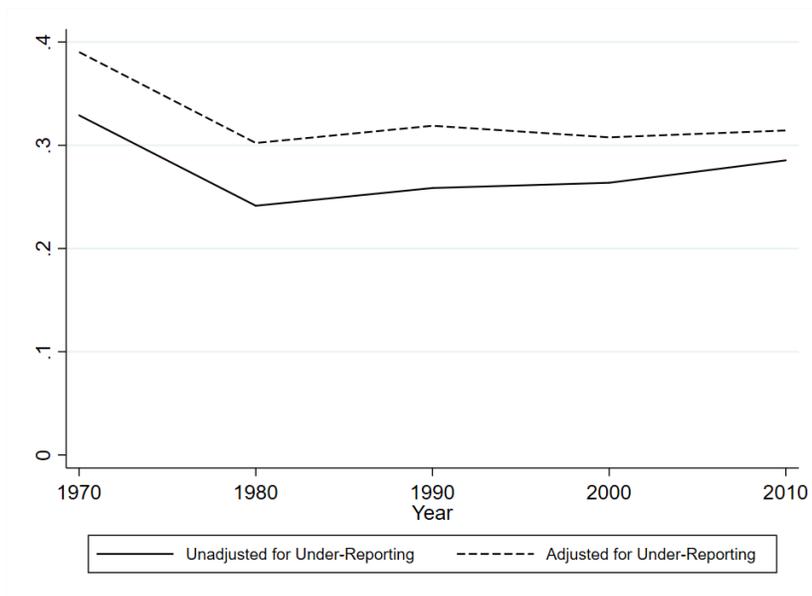
Notes: See Appendix C for estimation methods and sources.

Figure 12
 Unadjusted Earnings Among Working
 Native-born Men Ages 20-49



Notes: Respondents are considered black (white) if black (white) is the only race identified. Earnings are average earnings for respondents reporting to be employed. Real earnings are reported in 2010 dollars.
 Source: IPUMS Current Population Survey data

Figure 13
 Black-White Earnings Gap
 Native-born Men Ages 20-49
 Ratio Based Method - Demographer Female Undercount



Notes: See Appendix C for estimation methods and sources.

Table 1
Omission Rates for Black Men from US Census
Demographer Method

Age	1970	1980	1990	2000	2010
5 to 9	.069	.057	.077	.014	-
10 to 14	.036	.013	.042	-.015	-.013
15 to 19	.043	-.002	-.001	-.017	-.062
20 to 24	.115	.086	.055	.053	.061
25 to 29	.153	.123	.120	.088	.079
30 to 34	.153	.120	.133	.092	.090
35 to 39	.134	.132	.115	.103	.065
40 to 44	.119	.125	.101	.108	.067
45 to 49	.127	.121	.114	.091	.063

Sources: 1970-1990: Preston et al 2003. 2000: Robinson et al 2002. 2010: Author's calculation (see "Appendix A" for details on calculation).

Table 2
Estimates of True Gender Ratio for Blacks Ages 20-49
Demographer Method

	Male	Female	Census gender ratio	Implied true gender ratio
1970	.133	.025	.825	.929
1980	.114	.014	.850	.945
1990	.106	.027	.885	.964
2000	.090	.008	.894	.974
2010	.071	.000	.910	.979

1. Sources: 1970-1990: Preston et al 2003. 2000: Robinson et al 2002. 2010: Author's calculation (see "Appendix A" for details on calculation)
2. Source: IPUMS Census data
3. Implied true gender ratio estimated as $\frac{C_M/(1-U_M)}{C_F/(1-U_F)}$ where C_M (C_F) is the number of male (female) census respondents and U_M (U_F) is the male (female) undercount.

Table 3
Omission Rates for Black Men from US Census
Ratio Based Method

Year	Census Gender Ratio	<u>No Female Undercount</u>			<u>Demographer Female Undercount</u>		
		True Ratio	Total Undercount	Native Undercount	True Ratio	Total Undercount	Native Undercount
1970	.825	.990	.166	.170	.991	.188	.191
1980	.850	.981	.134	.140	.982	.147	.153
1990	.885	.982	.098	.106	.983	.124	.132
2000	.894	.976	.083	.092	.976	.091	.099
2010	.910	.983	.074	.084	.983	.074	.084

Notes: Total undercount is estimated as

$$U_{M,as} = 1 - \frac{C_{M,ay}}{\frac{B_{M,ay}}{B_{F,ay}} \left(\frac{C_{F,ay}}{(1-U_{F,ay})} + D_{F,ay} - I_{F,ay} \right) - D_{M,ay} + I_{M,ay}}$$

Native only undercount estimated as

$$U_{M,as} = 1 - \frac{C_{M,ay} - I_{M,ay}}{\frac{B_{M,ay}}{B_{F,ay}} \left(\frac{C_{F,ay} - I_{F,ay}}{(1-U_{F,ay})} + D_{F,ay} \right) - D_{M,ay}}$$

See section 2.2 for additional details of calculation.

Table 4
Regression of Male to Female Ratio on Male Incarceration Rates
Black Respondents Ages 20-49
By State and Year, 1970-2010

	[1]	[2]	[3]	[4]	[5]
Institutionalization Rate	0.994 (0.227)	1.070 (0.159)	1.107 (0.247)	1.517 (0.174)	1.244 (0.073)
Institutionalization Rate (Adjusted for Duplicates)	0.903 (0.227)	0.980 (0.159)	1.016 (0.247)	1.426 (0.174)	1.153 (0.073)
Poverty Rate (Black, 100%)	0.003 (0.307)	0.076 (0.181)	-	-	-
Poverty Rate (Black, 200%)	0.994 (0.227)	1.070 (0.159)	-	-	-
Poverty Rate (White, 100%)	0.778 (0.459)	0.696 (0.293)	-	-	-
Unemployment Rate	-0.232 (0.389)	-0.072 (0.246)	-	-	-
Violent Crime Rate	-0.001 (0.239)	0.026 (0.164)	-	-	-
Black/White Population Ratio	-0.369 (0.197)	-0.021 (0.116)	-	-	-
State/Year Fixed Effects	Yes	Yes	Yes	Yes	No
State Time Trend	Yes	No	Yes	No	No
N	136	136	136	136	136

Notes: Analysis includes all state-years in which there were at least 25,000 black men. Poverty rate is calculated as the share of men ages 20-49 reporting to be under 100% (200%) of the poverty rate in the Census IPUMS sample by state of birth. Unemployment rate is calculated as the share of respondents in the labor market reporting to be unemployed by state of birth. Violent crime rates are calculated as number of violent crimes per 10,000 residents by state using data from the Uniform Crime Reporting Statistics. Black population share is calculated as the total number of black female respondents divided by the total number of black and white female respondents. Data for 2010 are based on averages from 2009-2011.

Source: 1970-2000: IPUMS Census data. 2010: IPUMS 2009-2011 ACS data. Uniform Crime Reporting Statistics.

Table 5
 Educational Attainment by Race
 Native-born Men Ages 25-29
 Ratio Based Method - Demographer Female Undercount

	High School Completion			College Attendance		
	White	Black		White	Black	
		Unadjusted	Adjusted		Unadjusted	Adjusted
1970	.773	.548	.520	.419	.184	.160
1980	.876	.737	.675	.542	.363	.327
1990	.868	.739	.690	.540	.370	.336
2000	.890	.779	.736	.612	.425	.394
2010	.855	.740	.695	.637	.473	.436

Notes: See Appendix C for estimation methods and sources.

Table 6
 Employment by Race
 Native-born Men Ages 20-49
 Ratio Based Method - Demographer Female Undercount

	White	Black	
		Unadjusted	Adjusted
1970	.871	.802	.834
1980	.859	.707	.726
1990	.863	.666	.679
2000	.858	.661	.673
2010	.762	.539	.545

Notes: See Appendix C for estimation methods and sources.

Table 7
 Unemployment by Race
 Native-born Men Ages 20-49
 Ratio Based Method - Demographer Female Undercount

	White	Black	
		Unadjusted	Adjusted
1970	.033	.060	.062
1980	.059	.126	.131
1990	.049	.121	.132
2000	.036	.086	.090
2010	.108	.231	.241

Notes: See Appendix C for estimation methods and sources.

Table 8
 Earnings by Race
 Native-born Men Ages 20-49
 Ratio Based Method - Demographer Female Undercount

	White	Black	
		Unadjusted	Adjusted
1970	44,423	29,807	27,085
1980	39,527	29,985	27,581
1990	42,353	31,400	28,846
2000	49,539	36,477	34,297
2010	51,472	36,781	35,292

Notes: See Appendix C for estimation methods and sources.

Appendix A: Additional Tables

Table A.1
Death and Immigration Rates

Year	Population - Adjusted by Demographer Undercount		Deaths				Immigration				Birth Gender Ratio	Implied True Gender Ratio
			Counts		Share		Counts		Share			
			Male ¹	Female ¹	Male ²	Female ²	Male	Female	Male ³	Female ³		
1970	4,077,240	4,390,482	803,882	652,828	18.9%	13.5%	84,950	103,600	2.1%	2.4%	102.6%	96.0%
1980	5,572,353	5,894,535	814,002	570,378	14.8%	9.3%	250,820	260,180	4.5%	4.4%	102.3%	96.2%
1990	7,094,163	7,360,300	864,628	550,363	13.0%	7.7%	542,585	522,138	7.6%	7.1%	102.3%	96.6%
2000	8,064,363	8,277,325	899,352	547,559	12.0%	6.8%	724,591	763,092	9.0%	9.2%	102.4%	96.0%
2010	8,551,157	8,734,578	731,575	447,977	9.5%	5.5%	998,367	1,083,494	11.7%	12.4%	102.7%	96.2%

1. Sources: Census: IPUMS Census data (1970-2000) and IPUMS ACS data (2010). Demographer undercount data: 1970-1990: Preston et al 2003. 2000: Robinson et al 2002. 2010: Author's calculation (see "Appendix B" for details on calculation).

2. Source: Vital statistics data. Deaths for all years were adjusted by the non-immigrant share of the black population from Census data by age group. In non-census years, death rates were adjusted by a weighted average of non-immigrant share of the black population in the two closest census years. For 1921-1967, death data were only available by age range for non-whites. For these years, total deaths were adjusted by the black share on non-whites from Census data by age group. In non-census years, death rates were adjusted by a weighted average of the black share of non-whites in the two closest census years. For these years, total deaths in an age group were also adjusted by the share of ages in a death age range falling into the age group.

3. Source: Census IPUMS data. Immigrants defined as state of birth outside of 50 US states and Washington DC (bpl 100-999). Population data is based on demographer estimates of the full population.

Table A.2
Instrumental Variables Approach - First Stage
Regression of Black Incarceration Rates on White Incarceration Rates (Instrument)
Men Ages 20-49
By State and Year, 1970-2010

	[1]	[2]	[3]	[4]	[5]
White Male Incarceration Rate	0.693 (0.60)	1.678 (0.51)	0.706 (0.59)	1.856 (0.47)	3.553 (0.42)
Poverty Rate (Black, 100%)	0.089 (0.16)	0.292 (0.12)	-	-	-
Poverty Rate (Black, 200%)	-0.070 (0.15)	-0.188 (0.11)	-	-	-
Poverty Rate (White, 100%)	-0.093 (0.28)	-0.185 (0.22)	-	-	-
Unemployment Rate	0.055 (0.20)	0.134 (0.18)	-	-	-
Violent Crime Rate	0.270 (0.13)	0.271 (0.12)	-	-	-
Black/White Population Ratio	-0.122 (0.17)	-0.011 (0.11)	-	-	-
State/Year Fixed Effects	Yes	Yes	Yes	Yes	No
State Time Trend	Yes	No	Yes	No	No
N	136	136	136	136	136

Notes: Analysis includes all state-years in which there were at least 25,000 black men. Poverty rate is calculated as the share of men ages 20-49 reporting to be under 100% (200%) of the poverty rate in the Census IPUMS sample by state of birth. Unemployment rate is calculated as the share of respondents in the labor market reporting to be unemployed by state of birth. Violent crime rates are calculated as number of violent crimes per 10,000 residents by state using data from the Uniform Crime Reporting Statistics. Black population share is calculated as the total number of black female respondents divided by the total number of black and white female respondents. Data for 2010 are based on averages from 2009-2011.
Source: 1970-2000: IPUMS Census data. 2010: IPUMS 2009-2011 ACS data. Uniform Crime Reporting Statistics.

Table A.3
Instrumental Variables Approach
Regression of Male to Female Ratio on Black Male Incarceration Rates
Instrument: White Incarceration Rates
Respondents Ages 20-49
By State and Year, 1970-2010

	[1]	[2]	[3]	[4]	[5]
Institutionalization Rate	1.121 <i>(1.038)</i>	1.092 <i>(0.365)</i>	1.423 <i>(1.070)</i>	1.388 <i>(0.322)</i>	1.157 <i>(0.116)</i>
Institutionalization Rate (Adjusted for Duplicates)	1.030 <i>(1.038)</i>	1.001 <i>(0.365)</i>	1.332 <i>(1.070)</i>	1.297 <i>(0.322)</i>	1.066 <i>(0.116)</i>
Poverty Rate (Black, 100%)	0.107 <i>(0.210)</i>	0.070 <i>(0.176)</i>	-	-	-
Poverty Rate (Black, 200%)	-0.249 <i>(0.184)</i>	-0.238 <i>(0.142)</i>	-	-	-
Poverty Rate (White, 100%)	0.973 <i>(0.353)</i>	0.809 <i>(0.269)</i>	-	-	-
Unemployment Rate	-0.352 <i>(0.246)</i>	-0.051 <i>(0.220)</i>	-	-	-
Violent Crime Rate	0.042 <i>(0.319)</i>	0.063 <i>(0.181)</i>	-	-	-
Black/White Population Ratio	-0.079 <i>-(0.253)</i>	0.076 <i>(0.130)</i>	-	-	-
State/Year Fixed Effects	Yes	Yes	Yes	Yes	No
State Time Trend	Yes	No	Yes	No	No
N	136	136	136	136	136

Notes: Analysis includes all state-years in which there were at least 25,000 black men. Poverty rate is calculated as the share of men ages 20-49 reporting to be under 100% (200%) of the poverty rate in the Census IPUMS sample by state of birth. Unemployment rate is calculated as the share of respondents in the labor market reporting to be unemployed by state of birth. Violent crime rates are calculated as number of violent crimes per 100,000 residents by state using data from the Uniform Crime Reporting Statistics. Black population share is calculated as the total number of black female respondents divided by the total number of black and white female respondents. Data for 2010 are based on averages from 2009-2011.
Source: 1970-2000: IPUMS Census data. 2010: IPUMS 2009-2011 ACS data. Uniform Crime Reporting Statistics.

Appendix B: Estimation of 2010 Undercount Using Demography Method

To my knowledge, no one has attempted to quantify the undercount by age, sex and race for 2010 using demographer methodology. Therefore, I provide my own estimates of the 2010 undercount for black males and females. Following demographer methods, I estimate the total 2010 population as

$$Population = Births - Deaths + Immigrants - Emmigrants$$

Table B.1 shows the estimated undercount by age group with each input. The following describes the methods for estimating each input.

Births

I estimate total births using birth records from the National Vital Statistics System. Although estimates of total births are thought to be nearly complete, there are significant differences across datasets in the classification of race. In the births data, there is no indicator for the race of the child, and race is imputed based on the race of the mother and father. However, in many cases, information for the father is not available in the births data. Preston and Robinson et al. use “the father rule” which assigns race to a child based on the race of their father, unless this information is missing, in which case the race of the mother is used. This was found to be the most accurate method for matching vital statistics data to Census trends through 1990 when respondents were only allowed to indicate one race on the Census form (Passel, 1990). Starting in 2000, respondents were allowed to “mark one or more” races on the Census form¹⁹. To account for this fact, I limit my analysis to only men in the Census reporting to be “black only.” To match counts from vital statistics data to Census counts, I calculate the following equation:

$$B = \sum_i (P_i + M_i \cdot S_m + F_i \cdot S_f) \quad (26)$$

Where B is the total number of black men born, P_i is a dummy equal to one if both the

¹⁹To account for this change, Robinson et al. calculate the Census population using two models. The first counts only those identifying as “black only” as black. The second counts anyone identifying as black, regardless of other racial identification. For the purposes of this paper, I use the average of these two models, which is presented in the paper. The authors note that the differences in the population estimates are the largest for the younger age groups suggesting that this will be a bigger threat to identification in subsequent Census years.

child's parents identified as black, M_i and F_i are dummies equal to one if person i had a mother (father) who identified as black and a father (mother) who identified as non-black, and S_m and S_f represent the share of children in the 2010 Census identifying as black only who had mother (father) who identified as black and a father (mother) who identified as non-black.

Deaths

I estimate deaths using mortality data from the National Vital Statistics System. To account for immigrant deaths, I adjust total deaths by the immigrant share of the black respondents in the Census by year and five year age group. In non-census years, I use a weighted average of the immigrant share of black respondents in the two closest census years. For 1921-1967, death data were only available by age range²⁰. I assume that deaths were evenly distributed across all ages within age group. For 1921-1967, deaths are only listed as "white" and "non-white." For these years, I adjust the number of non-white births by the black share of non-white respondents in the Census. In non-census years, I use a weighted average of the black share of non-whites in the two closest census years.

Immigrants

I estimate the number of immigrants using the share of respondents to the ACS. I define immigrants as anyone who reports having a birthplace outside of the United States. For the purposes of this calculation, to match birth data, I also define respondents born in Puerto Rico, or US territories to be immigrants. This method assumes that there is no non-reporting among immigrants. In reality, it is likely that there is substantial non-reporting among the immigrant population, especially among illegal immigrants. However, given the nature of illegal immigration, quantifying the amount of non-reporting among immigrants would be a large undertaking and would certainly qualify as a standalone paper. Because of this shortcoming, I consider the native born estimates of non-reporting to be the primary estimates in this paper.

Emmigrants

I assume that the number of native born emigrants is negligible and therefore I estimate $Emmigrants \approx 0$. There are no reliable figures on the number of native born emigrants or estimates of emigration by age and race (Jensen, 2013, Bhaskar et al, 2013). The Bureau of

²⁰Deaths are grouped into ten year age bins with the exception of "under 1 year" and "1-4 years."

Consular Affairs estimates that there are nine million US citizens living overseas (Department of State, 2016). This represents 2.8 percent of the US population. However, many of these citizens were foreign born US citizens or returning naturalized US citizens (Bhaskar et al, 2013). Therefore, the true share of native born citizens who emigrate is likely to be under one percent. For example, Bhaskar et al (2013) estimates net emigration of 18,000 native born US citizens between 2000 and 2010. This represents .006% of the population. There is evidence that blacks are not more likely to emigrate than other races. For example, according to the 2008 CPS migration supplement, fewer all black households had a member move abroad than all non-black households.

Figure B.1
Estimate of the 2010 Black Respondent Undercount
Demographer Method

Gender	Age	Birth ¹	Death ²	Immigrants ³	Population ⁴	Census ⁵	Undercount ⁶
Male	10-14	1,550,205	29,911	62,461	1,582,755	1,603,746	-1.3%
Male	15-19	1,659,283	42,327	93,515	1,710,472	1,816,259	-6.2%
Male	20-24	1,625,964	59,869	141,326	1,707,422	1,604,113	6.1%
Male	25-29	1,442,404	73,690	150,872	1,519,586	1,400,211	7.9%
Male	30-34	1,360,342	96,464	164,819	1,428,697	1,300,313	9.0%
Male	35-39	1,302,185	121,151	178,892	1,359,926	1,271,729	6.5%
Male	40-44	1,373,300	158,369	180,183	1,395,114	1,301,039	6.7%
Male	45-49	1,499,556	222,032	182,275	1,459,799	1,367,345	6.3%
<i>Male</i>	<i>20-49</i>			<i>998,367</i>	<i>8,870,543</i>	<i>8,244,750</i>	<i>7.1%</i>
Female	10-14	1,504,035	23,671	66,513	1,546,877	1,550,700	-0.2%
Female	15-19	1,610,381	31,679	94,952	1,673,655	1,756,659	-5.0%
Female	20-24	1,577,066	38,794	136,054	1,674,326	1,635,676	2.3%
Female	25-29	1,401,755	42,884	172,638	1,531,510	1,511,200	1.3%
Female	30-34	1,323,647	53,527	184,959	1,455,079	1,442,786	0.8%
Female	35-39	1,266,626	68,517	203,518	1,401,627	1,434,691	-2.4%
Female	40-44	1,340,619	99,018	196,639	1,438,239	1,450,544	-0.9%
Female	45-49	1,465,466	145,237	189,686	1,509,915	1,533,878	-1.6%
<i>Female</i>	<i>20-49</i>				<i>9,010,696</i>	<i>9,008,775</i>	<i>0.0%</i>

Notes: 1. Source: Vital statistics data. For 1961-1967, total births classified as "Negro" were included. For 1967, total births were only available for "non-whites." In this case, non-white births were adjusted by the share of the non-white population identifying as black from the decennial Census. For 1968-2000, black births were defined as $B + M * S_M + F * S_F$ where B is the total number of births to two black parents or one black parent with the other's race missing, M (F) is the total number of births to a mother (father) who identified as black and a father (mother) who identified as non-black, and S_M (S_F) is the share of children in the 2010 Census identifying as black only who had a mother (father) who identified as black and a father (mother) who identified as non-black.

2. 2. Source: Vital statistics data. Deaths for all years were adjusted by the non-immigrant share of the black population from Census data by age group. In non-census years, death rates were adjusted by a weighted average of non-immigrant share of the black population in the two closest census years. For 1921-1967, death data were only available by age range for non-whites. For these years, total deaths were adjusted by the black share on non-whites from Census data by age group. In non-census years, death rates were adjusted by a weighted average of the black share of non-whites in the two closest census years. For these years, I also assume that deaths are evenly distributed across ages in an age group.

3. Source: Census IPUMS data. Immigrants defined as state of birth outside of 50 US states and Washington DC.

4. Population = Births + Deaths - Immigrants.

5. Source: US Census Bureau

6. Undercount is estimated as $(Population - Census)/Population$.

Appendix C: Estimation Details of Education, Employment and Earnings Outcomes

Figure C.1
Education Distribution
Survey of Inmates
Native-born Black Men Ages 25-29

Survey Year	College	High School Graduate	High School Dropout
1974	6.1%	34.1%	59.8%
1979	14.2%	20.4%	65.4%
1986	10.8%	24.9%	64.4%
1991	14.2%	26.6%	59.2%
1997	9.2%	29.6%	61.2%
2003	14.5%	18.6%	66.9%

Notes: Sample is limited to black male inmates ages 25-29 who were admitted in the survey year. College is defined as having attended college for any amount of time.
Source: Survey of Inmates

Figure C.2
Estimation of Adjusted Education Distribution
Black Men Ages 25-29

Year	Unadjusted ¹		Survey of Inmates ³			Adjusted ⁵	
	College ²	High School Completion	College ²	High School Completion	Undercount ⁴	College ²	High School Completion
1970	.184	.548	.061	.402	.196	.160	.520
1980	.363	.737	.137	.347	.158	.327	.675
1990	.370	.739	.135	.398	.143	.336	.690
2000	.425	.779	.119	.359	.102	.394	.736
2010	.473	.740	.145	.331	.111	.436	.695

Notes:

1. Analysis is limited to the native born population. Respondents are considered black (white) if black (white) is the only race identified. Source: 1970-2000 IPUMS Census data, 2010: IPUMS ACS data.
2. College is defined as having attended college for any amount of time.
3. Education estimates for Survey of Inmates are a weighted average of the two adjacent survey years. For example, 1980 estimates are a weighted average of estimates from the 1979 and 1986 SOI, with linear weights based on distance from 1980.
4. Undercount estimates are based on the "Ratio Method" described in section 2, using only the native born population. Female undercount is estimated using demographer based estimates.
5. Adjusted educational outcomes are estimated as $Unadjusted * (1 - Undercount) + SOI * Undercount$.

Figure C.3
 Estimation of Non-Reporter Share Employed
 Native-born Black Men Ages 20-49

Year	Age Group	CPS ¹			Non-Respondents ³			SOI Share ⁴			Non-Respondent Total ⁵
		College ²	High School Grad	High School Dropout	College ⁴	High School Grad	High School Drop	College ²	High School Grad	High School Drop	
1970	20-24	.560	.818	.750	.554	.816	.934	.070	.228	.702	.881
1970	25-29	.839	.823	.844	.830	.821	1.000	.070	.228	.702	.947
1970	30-34	.934	.883	.904	.924	.881	1.000	.070	.228	.702	.968
1970	35-39	.967	.854	.891	.957	.852	1.000	.070	.228	.702	.963
1970	40-44	.885	.880	.867	.876	.878	1.000	.070	.228	.702	.963
1970	45-49	.872	.862	.838	.862	.860	1.000	.070	.228	.702	.959
1980	20-24	.553	.647	.521	.547	.645	.648	.111	.193	.695	.636
1980	25-29	.796	.761	.642	.787	.759	.799	.111	.193	.695	.790
1980	30-34	.811	.783	.715	.803	.781	.891	.111	.193	.695	.860
1980	35-39	.796	.870	.654	.787	.868	.814	.111	.193	.695	.822
1980	40-44	.920	.904	.773	.910	.902	.962	.111	.193	.695	.945
1980	45-49	.873	.899	.742	.864	.897	.924	.111	.193	.695	.912
1990	20-24	.580	.628	.424	.573	.626	.528	.154	.228	.618	.558
1990	25-29	.801	.784	.517	.792	.782	.644	.154	.228	.618	.699
1990	30-34	.815	.769	.499	.807	.767	.621	.154	.228	.618	.683
1990	35-39	.864	.754	.643	.855	.753	.800	.154	.228	.618	.798
1990	40-44	.856	.720	.602	.847	.718	.750	.154	.228	.618	.758
1990	45-49	.868	.818	.671	.859	.816	.836	.154	.228	.618	.835
2000	20-24	.672	.561	.370	.665	.560	.461	.143	.215	.642	.511
2000	25-29	.828	.737	.542	.819	.735	.675	.143	.215	.642	.709
2000	30-34	.889	.816	.470	.880	.814	.585	.143	.215	.642	.676
2000	35-39	.876	.694	.686	.867	.693	.855	.143	.215	.642	.822
2000	40-44	.816	.783	.516	.807	.782	.643	.143	.215	.642	.696
2000	45-49	.852	.795	.540	.843	.793	.673	.143	.215	.642	.723
2010	20-24	.483	.472	.216	.478	.471	.269	.147	.215	.637	.343
2010	25-29	.712	.580	.307	.704	.579	.383	.147	.215	.637	.472
2010	30-34	.758	.582	.377	.750	.581	.470	.147	.215	.637	.535
2010	35-39	.773	.610	.501	.764	.609	.624	.147	.215	.637	.641
2010	40-44	.732	.591	.283	.724	.590	.352	.147	.215	.637	.458
2010	45-49	.724	.628	.498	.716	.626	.620	.147	.215	.637	.635

Notes:

1. For 2000 and 2010, the native employment rate is calculated directly using the native born population. For 1970-1990, employment rates are calculated as the employment rate for the full population scaled by the average native-full population ratio from the Census. Native-Full Population employment ratios are .999 for 1970, .999 for 1980, and .990 for 1990. Respondents are considered black if black is the only race identified.
2. College is defined as having attended college for any amount of time.
3. Non-respondent employed share is estimated as the employed rate from the CPS times the average ratio of employment from the SOI to the CPS. This ratio is estimated by age and education level. Employment in the SOI is estimated as the share of respondents reporting to be employed in the month before arrest. Analysis is limited to black men over age 20 who were admitted in the survey year. Average ratio is weighted by the age distribution in the SOI. The average SOI-CPS employment ratios are .989 for college attenders, .998 for high school graduates, and 1.245 for high school dropouts.
4. The SOI education distribution is estimated based on black male respondents ages 20-49 who were admitted in the survey year.
5. Non-respondent total is estimated as the sum of imputed non-respondent employment weighted by the education distribution.

Figure C.4
 Estimation of Adjusted Employment Distribution
 Native-born Black Men Ages 20-49

Year	Unadjusted			Non-reporters ⁴	Undercount ⁵	Adjusted ⁶
	CPS ¹	Institutionalized ²	Total ³			
1970	.837	.042	.802	.970	.191	.834
1980	.737	.041	.707	.834	.153	.726
1990	.720	.075	.666	.763	.132	.679
2000	.749	.107	.661	.782	.099	.673
2010	.617	.098	.539	.602	.084	.545

Notes:

1. For 2000 and 2010, the native employment rate is calculated directly using the native born population. For 1970-1990, employment rates are calculated as the employment rate for the full population scaled by the average native-full population ratio from the Census. Native-Full Population employment ratios are .999 for 1970, .999 for 1980, and .990 for 1990. Respondents are considered black if black is the only race identified.
2. Institutionalize share of the population is estimated from the Census data for the native population only. Respondents are considered black (white) if black (white) is the only race identified.
3. Total unadjusted employment share is estimated as $CPS * (1 - Institutionalized)$.
4. Employment estimates for non-reporters is described in table C.3. Total annual employment rate is estimated as the average over all age groups, weighted by the total size of the non-reporting population in each age group.
5. Undercount estimates are based on the "Ratio Method" described in section 2, using only the native born population. Female undercount is estimated using demographer based estimates.
6. Adjusted employment shares outcomes are estimated as $Unadjusted * (1 - Undercount) + SOI * Undercount$.

Figure C.5
 Estimation of Non-Reporter Share Unemployed
 Native-born Black Men Ages 20-49

Year	Age Group	CPS ¹			Non-Respondents ³			SOI Share ⁴			Non-Respondent Total ⁵
		College ²	High School Grad	High School Dropout	College ²	High School Grad	High School Dropout	College ²	High School Grad	High School Drop	
1970	20-24	.076	.052	.163	.098	.061	.159	.046	.214	.740	.135
1970	25-29	.013	.086	.086	.016	.101	.084	.062	.217	.720	.083
1970	30-34	.029	.049	.030	.038	.057	.029	.069	.221	.710	.036
1970	35-39	.000	.027	.057	.000	.032	.056	.071	.211	.718	.047
1970	40-44	.047	.065	.051	.061	.077	.049	.067	.224	.709	.056
1970	45-49	.058	.042	.023	.076	.050	.023	.067	.216	.716	.032
1980	20-24	.158	.229	.268	.205	.269	.261	.088	.196	.716	.257
1980	25-29	.099	.146	.218	.129	.171	.212	.104	.183	.713	.196
1980	30-34	.084	.139	.131	.109	.163	.128	.104	.186	.711	.133
1980	35-39	.096	.033	.156	.125	.039	.152	.106	.190	.703	.128
1980	40-44	.036	.053	.050	.047	.062	.048	.110	.192	.698	.051
1980	45-49	.041	.045	.077	.053	.053	.075	.106	.191	.703	.069
1990	20-24	.119	.218	.344	.154	.257	.335	.133	.241	.626	.292
1990	25-29	.064	.092	.286	.084	.108	.279	.153	.229	.618	.210
1990	30-34	.081	.085	.300	.105	.100	.292	.160	.225	.615	.219
1990	35-39	.046	.106	.133	.059	.124	.129	.158	.219	.623	.117
1990	40-44	.040	.119	.185	.052	.140	.180	.158	.214	.628	.151
1990	45-49	.065	.097	.094	.085	.114	.091	.159	.230	.610	.096
2000	20-24	.082	.230	.386	.106	.270	.376	.144	.216	.640	.314
2000	25-29	.080	.112	.285	.104	.132	.277	.144	.201	.655	.223
2000	30-34	.058	.062	.117	.076	.073	.113	.184	.256	.560	.096
2000	35-39	.033	.132	.078	.043	.155	.076	.148	.197	.655	.086
2000	40-44	.043	.042	.045	.056	.050	.044	.170	.247	.583	.047
2000	45-49	.011	.033	.062	.014	.039	.061	.165	.238	.596	.048
2010	20-24	.241	.372	.587	.312	.437	.571	.143	.247	.610	.501
2010	25-29	.156	.285	.495	.203	.335	.481	.162	.229	.609	.403
2010	30-34	.152	.255	.476	.197	.300	.463	.154	.198	.648	.390
2010	35-39	.135	.240	.302	.175	.282	.294	.154	.203	.643	.273
2010	40-44	.151	.228	.384	.196	.269	.374	.197	.257	.546	.312
2010	45-49	.156	.182	.211	.202	.214	.206	.162	.214	.623	.207

Notes:

1. For 2000 and 2010, the native unemployment rate is calculated directly using the native born population. For 1970-1990, unemployment rates are calculated as the unemployment rate for the full population scaled by the average native-full population ratio from the Census. Native-Full Population employment ratios are 1.008 for 1970, 1.012 for 1980, and 1.024 for 1990. Respondents are considered black if black is the only race identified.

2. College is defined as having attended college for any amount of time.

3. Non-respondent unemployed share is estimated as the unemployment rate from the CPS times the average ratio of unemployment from the SOI to the CPS. This ratio is estimated by age and education level. Unemployment in the SOI is estimated as the share of respondents reporting to be unemployed in the month before arrest divided by the share reporting to be in the labor force. Analysis is limited to black men over age 20 who were admitted in the survey year. Average ratio is weighted by the age distribution in the SOI. The average SOI-CPS employment ratios are 1.298 for college attenders, 1.176 for high school graduates, and .973 for high school dropouts.

4. The SOI share for each education level, j , is estimated as $\frac{s_j * lf_j}{\sum_{i=1}^3 s_i * lf_i}$ where s_j is the share of non-responders with education j and lf_j is the share of the non-responders with education j who are in the labor force. The SOI education distribution is estimated based on black male respondents ages 20-49 who were admitted in the survey year. These raw values can be found in Table C.3. The share of non-responders in the labor force is estimated using the same method as for unemployment, as described above.

5. Non-respondent total is estimated as the sum of imputed non-respondent unemployment weighted by the education distribution.

Figure C.6
 Estimation of Adjusted Unemployment Distribution
 Native-born Black Men Ages 20-49

Year	Unadjusted			Non-reporters			Adjusted ⁷
	Unemployment ¹	Labor Force ²	Institutionalized ³	Unemployment ⁴	Labor Force ⁵	Undercount ⁶	
1970	.060	.898	.042	.071	.975	.191	.062
1980	.126	.853	.041	.156	.948	.153	.131
1990	.121	.844	.075	.194	.901	.132	.132
2000	.086	.810	.107	.128	.802	.099	.090
2010	.231	.780	.098	.344	.781	.084	.240

Notes:

1. For 2000 and 2010, the native unemployment rate is calculated directly using the native born population in the CPS. For 1970-1990, unemployment rates are calculated as the labor force participation rate for the full population in the CPS scaled by the average native-full population ratio from the Census. Native-Full Population employment ratios are 1.008 for 1970, 1.012 for 1980, and 1.024 for 1990. Respondents are considered black if black is the only race identified.
2. For 2000 and 2010, the native labor force participation rate (excluding the institutionalized population) is calculated directly using the native born population in the CPS. For 1970-1990, labor force participation rates are calculated as the labor force participation rate for the full population in the CPS scaled by the average native-full population ratio from the Census. Native-Full Population labor force participation ratios are 1.008 for 1970, 1.012 for 1980, and 1.024 for 1990. Respondents are considered black if black is the only race identified.
3. Institutionalize share of the population is estimated from the Census data for the native population only. Respondents are considered black (white) if black (white) is the only race identified.
4. Unemployment estimates for non-reporters is described in table C.5. Total annual unemployment rate is estimated as the average over all age groups, weighted by the total size of the non-reporting members of the labor force in each age group.
5. Undercount estimates are based on the "Ratio Method" described in section 2, using only the native born population. Female undercount is estimated using demographer based estimates.
6. Adjusted unemployment rate is calculated as

$$\frac{Unemployment_{Unadjusted} * (1 - Undercount - Institutionalized) * lf_{Unadjusted} + Unemployment_{Nonreporters} * Undercount * lf_{Unadjusted}}{(1 - Undercount - Institutionalized) * lf_{Unadjusted} + Undercount * lf_{Unadjusted}}$$

Figure C.7
 Estimation of Non-Reporter Earnings Among Working
 Native-born Black Men Ages 20-49

Year	Age Group	CPS ¹			Non-Respondents ³			SOI Share ⁴			Non-Respondent Total ⁵
		College ²	High School Grad	High School Dropout	College ²	High School Grad	High School Dropout	College ²	High School Grad	High School Drop	
1970	20-24	4,143	4,464	3,560	2,334	2,764	2,331	.044	.212	.745	2,423
1970	25-29	6,828	5,508	4,411	3,846	3,410	2,888	.061	.198	.741	3,050
1970	30-34	7,442	6,096	5,292	4,192	3,774	3,465	.066	.208	.726	3,578
1970	35-39	8,250	6,411	4,654	4,648	3,970	3,047	.069	.202	.729	3,344
1970	40-44	9,351	6,518	5,233	5,268	4,036	3,427	.063	.208	.729	3,670
1970	45-49	7,197	6,269	4,648	4,054	3,882	3,043	.063	.205	.732	3,279
1980	20-24	7,070	6,823	5,783	3,983	4,225	3,787	.096	.196	.708	3,892
1980	25-29	12,405	10,916	7,514	6,988	6,759	4,921	.111	.186	.703	5,492
1980	30-34	15,917	11,865	8,967	8,966	7,347	5,872	.104	.176	.720	6,453
1980	35-39	18,105	14,425	8,602	10,199	8,932	5,633	.107	.204	.689	6,794
1980	40-44	17,371	13,563	10,196	9,786	8,398	6,677	.107	.185	.708	7,328
1980	45-49	17,353	13,235	10,733	9,775	8,195	7,029	.105	.190	.704	7,540
1990	20-24	13,378	10,814	6,652	7,536	6,696	4,356	.159	.256	.585	5,460
1990	25-29	21,276	15,091	10,336	11,985	9,344	6,769	.175	.255	.570	8,339
1990	30-34	22,934	16,722	11,431	12,919	10,354	7,485	.182	.256	.562	9,210
1990	35-39	24,444	19,424	13,239	13,770	12,027	8,670	.165	.215	.620	10,235
1990	40-44	28,158	20,786	16,716	15,862	12,871	10,947	.172	.216	.611	12,210
1990	45-49	30,575	22,387	14,941	17,224	13,862	9,784	.159	.223	.618	11,875
2000	20-24	13,987	14,942	7,414	7,879	9,252	4,855	.186	.235	.579	6,453
2000	25-29	28,172	22,145	12,010	15,870	13,712	7,865	.165	.223	.612	10,491
2000	30-34	36,689	32,149	18,455	20,668	19,906	12,086	.186	.259	.555	15,706
2000	35-39	40,539	25,408	21,132	22,837	15,732	13,838	.151	.181	.668	15,539
2000	40-44	39,205	26,749	19,927	22,085	16,562	13,049	.166	.241	.593	15,396
2000	45-49	38,576	28,451	17,241	21,731	17,616	11,291	.167	.236	.597	14,523
2010	20-24	16,999	15,402	9,002	9,576	9,537	5,895	.205	.295	.499	7,726
2010	25-29	33,305	31,425	19,847	18,762	19,458	12,997	.220	.264	.516	15,969
2010	30-34	47,248	25,680	18,662	26,616	15,901	12,221	.206	.234	.560	16,053
2010	35-39	49,607	36,497	18,299	27,945	22,599	11,983	.175	.205	.620	16,955
2010	40-44	59,836	32,552	25,577	33,707	20,156	16,749	.233	.277	.490	21,638
2010	45-49	61,542	31,430	20,068	34,668	19,461	13,142	.166	.212	.622	18,055

Notes:

1. Earnings are calculated as average earnings reported from wages and salary among the population reporting to be employed. For 2000 and 2010, the native earnings are calculated directly using the native born population. For 1970-1990, earnings are calculated as the average earnings for the full population scaled by the average native-full population ratio from the Census. Native-Full Population employment ratios are 1.001 for 1970, .997 for 1980, and 1.001 for 1990. Respondents are considered black if black is the only race identified.
2. College is defined as having attended college for any amount of time.
3. Non-respondent average earnings share is estimated as the average earnings from the CPS times the average ratio of earnings from the SOI to the CPS. This ratio is estimated by age and education level. Earnings in the SOI is estimated as the average earnings of respondents reporting to be employed in the month before arrest. Analysis is limited to black men over age 20 who were admitted in the survey year. Average ratio is weighted by the age distribution in the SOI. The average SOI-CPS employment ratios are .563 for college attenders, .619 for high school graduates, and .655 for high school dropouts.
4. The SOI share for each education level, j , is estimated as $\frac{s_j * e_j}{\sum_{i=1}^3 s_i * e_i}$ where s_j is the share of non-responders with education j and e_j is the share of the non-responders with education j who are employed. The SOI education distribution is estimated based on black male respondents ages 20-49 who were admitted in the survey year. These raw values can be found in Table C.3. The share of non-responders who are employed and the estimation methods are in table C.3.
5. Non-respondent total is estimated as the sum of imputed non-respondent average earnings weighted by the education distribution.

Figure C.8
 Estimation of Adjusted Earnings Among Working
 Native-born Black Men Ages 20-49

Year	Unadjusted		Non-reporters		Undercount ⁴	Adjusted	
	Earnings ¹	Employed ²	Earnings ³	Employed ²		Nominal ⁵	Real ⁶
1970	5,304	.802	3,121	.970	.191	4,819	27,085
1980	11,331	.707	6,165	.834	.153	10,423	27,581
1990	18,821	.666	8,476	.763	.132	17,290	28,846
2000	28,806	.661	13,899	.782	.099	27,084	34,297
2010	36,781	.539	20,803	.602	.084	35,292	35,292

Notes:

1. Earnings are calculated as average earnings reported from wages and salary among the population reporting to be employed. For 2000 and 2010, the native earnings are calculated directly using the native born population. For 1970-1990, earnings are calculated as the average earnings for the full population scaled by the average native-full population ratio from the Census. Native-Full Population employment ratios are 1.001 for 1970, .997 for 1980, and 1.001 for 1990. Respondents are considered black if black is the only race identified.
2. The unadjusted employment share and the estimation methods are in table C.4.
3. Average earnings estimates for non-reporters is described in table C.7. Total annual average earnings is estimated as the average over all age groups, weighted by the total size of the non-reporting employed men in each age group.
4. The non-reporter employment share and the estimation methods are in tables C.3 and C.4.
5. Adjusted average earnings is calculated as

$$\frac{Earnings_{Unadjusted} * (1 - Undercount) * Employed_{Unadjusted} + Earnings_{Nonreporters} * Undercount * Employed_{Unadjusted}}{(1 - Undercount) * Employed_{Unadjusted} + Undercount * Employed_{Unadjusted}}$$

6. Real earnings are reported in 2010 dollars.

Appendix D: Estimates of Education, Employment and Earnings Using Alternative Undercount Estimates

The following shows adjusted education, employment, unemployment and earnings outcomes using alternative estimates of the undercount. For descriptions and estimates of each undercount method, see section 2. For details on data sources and estimation details, see Appendix C.

D.1 Demographer Based Undercount

Figure D.1.1
Unadjusted Educational Attainment
Men Ages 25-29
Demographer Based Undercount

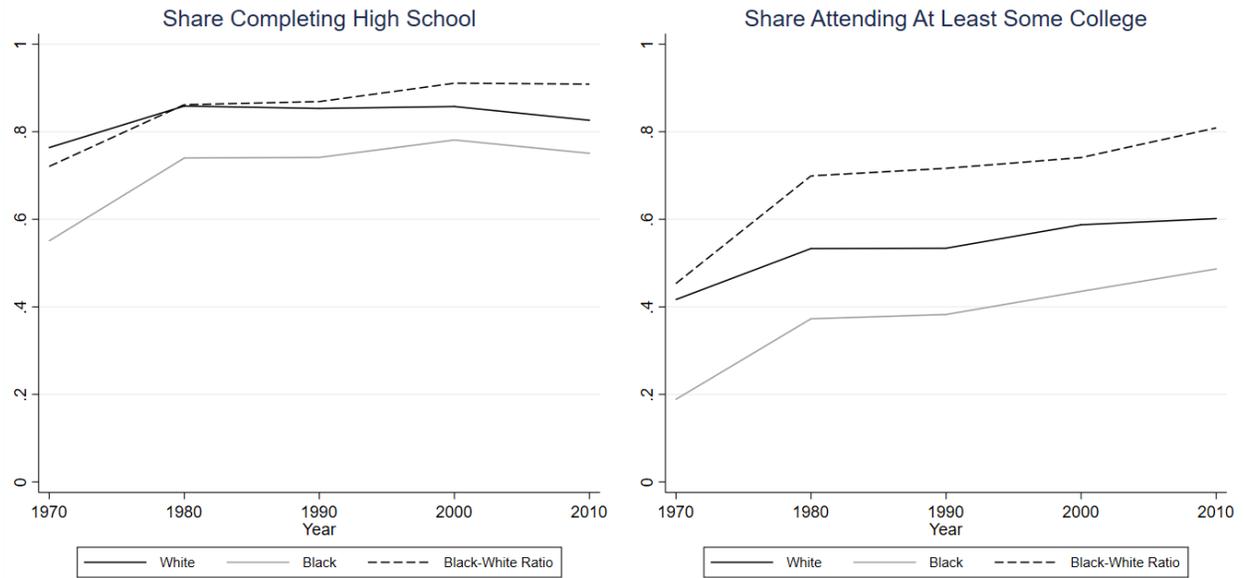


Figure D.1.2
 Ratio of Black to White Educational Attainment
 Men Ages 25-29
 Demographer Based Undercount

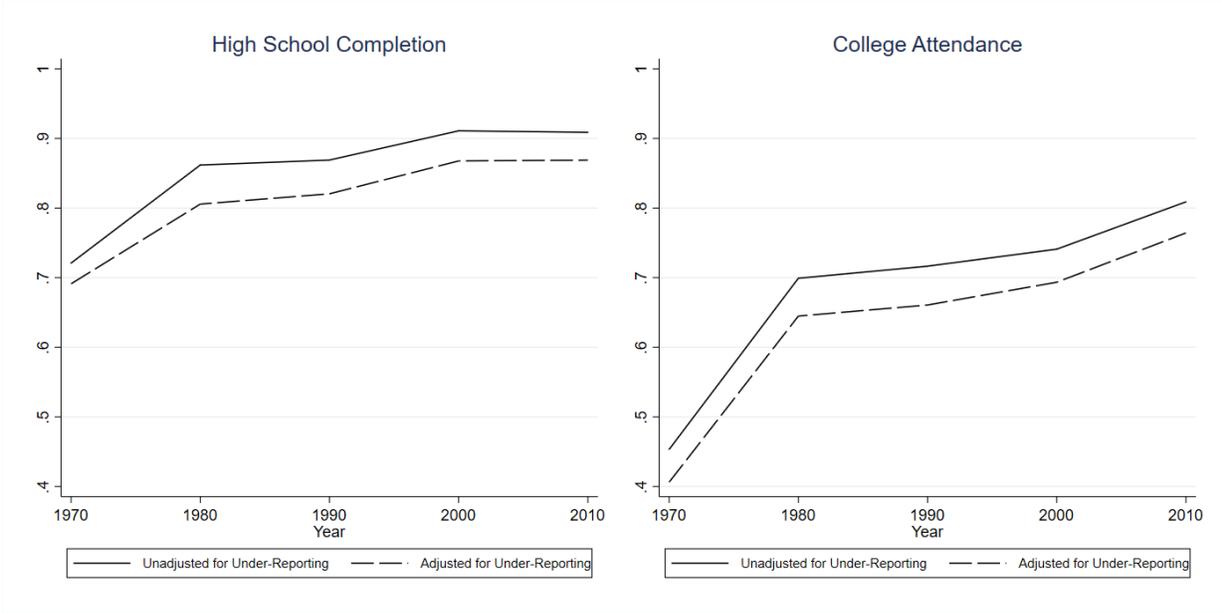


Figure D.1.3
 Unadjusted Employment Rates
 Men Ages 20-49
 Demographer Based Undercount

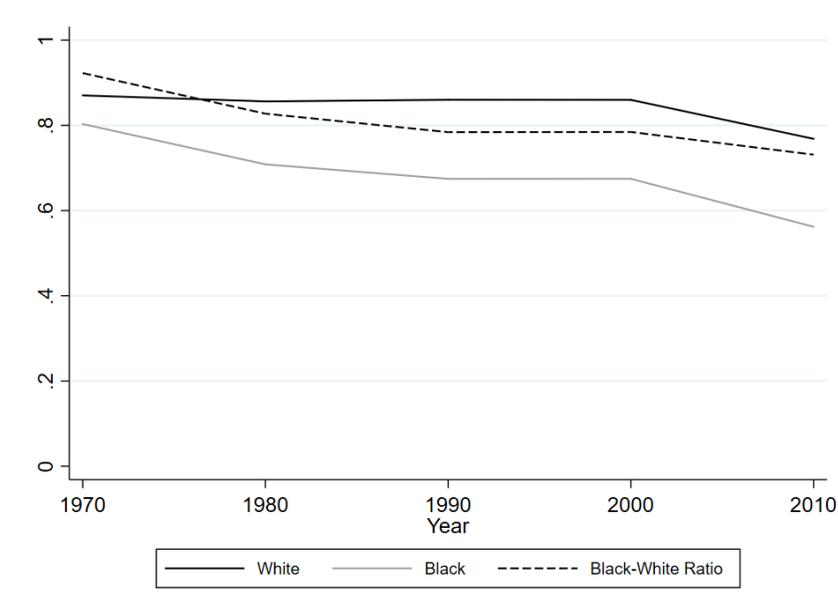


Figure D.1.4
 Ratio of Black to White Employment
 Men Ages 20-49
 Demographer Based Undercount

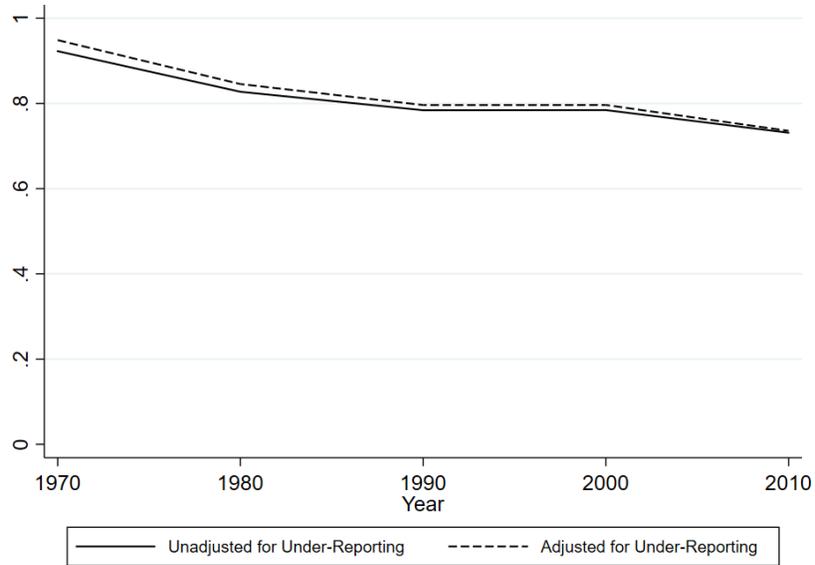


Figure D.1.5
 Unadjusted Unemployment Rates
 Men Ages 20-49
 Demographer Based Undercount

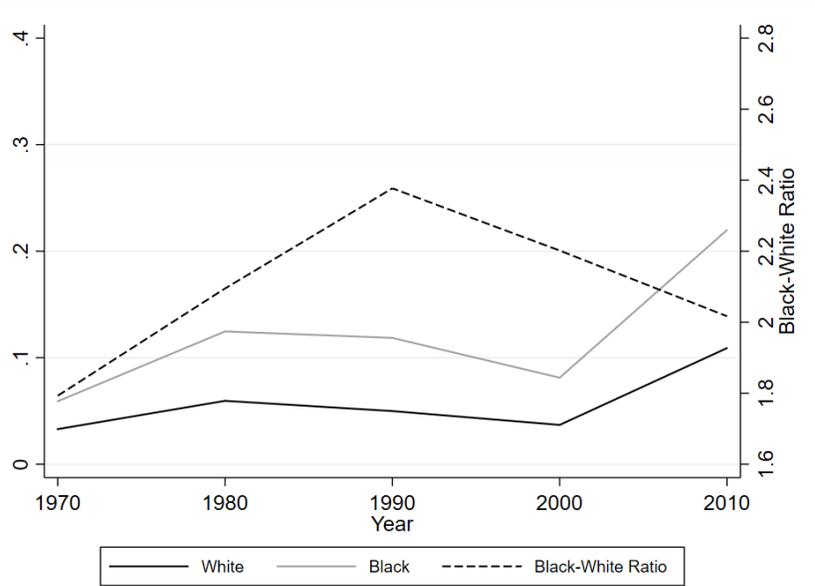


Figure D.1.6
 Ratio of Black to White Unemployment
 Men Ages 20-49
 Demographer Based Undercount

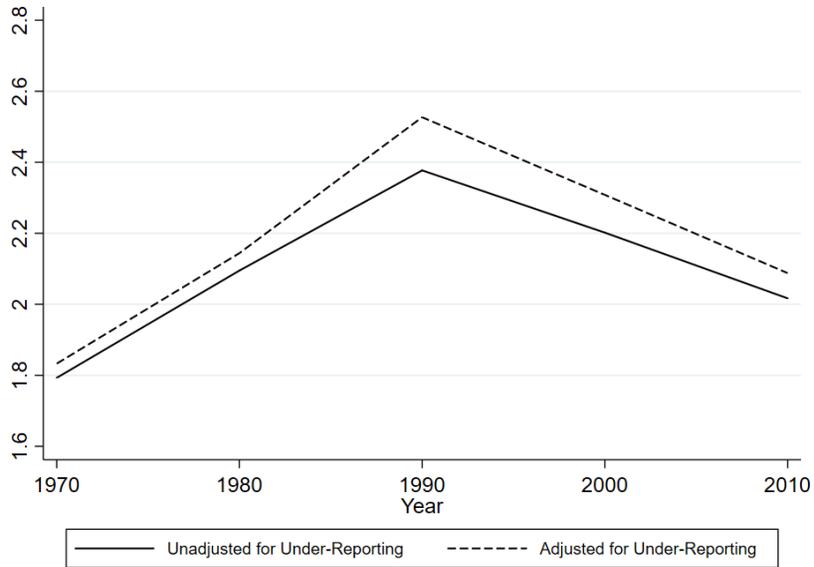


Figure D.1.6
 Unadjusted Earnings Among Working
 Men Ages 20-49
 Demographer Based Undercount

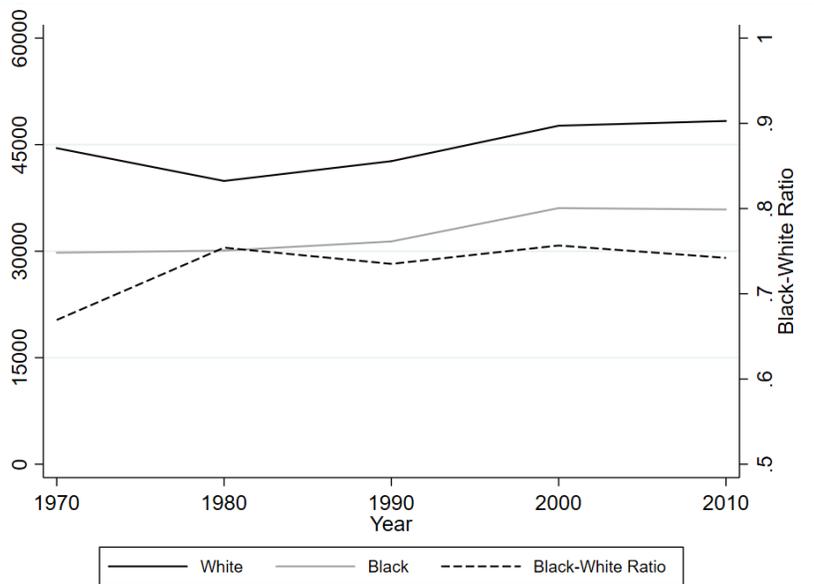


Figure D.1.7
 Ratio of Black to White Earnings
 Men Ages 20-49
 Demographer Based Undercount

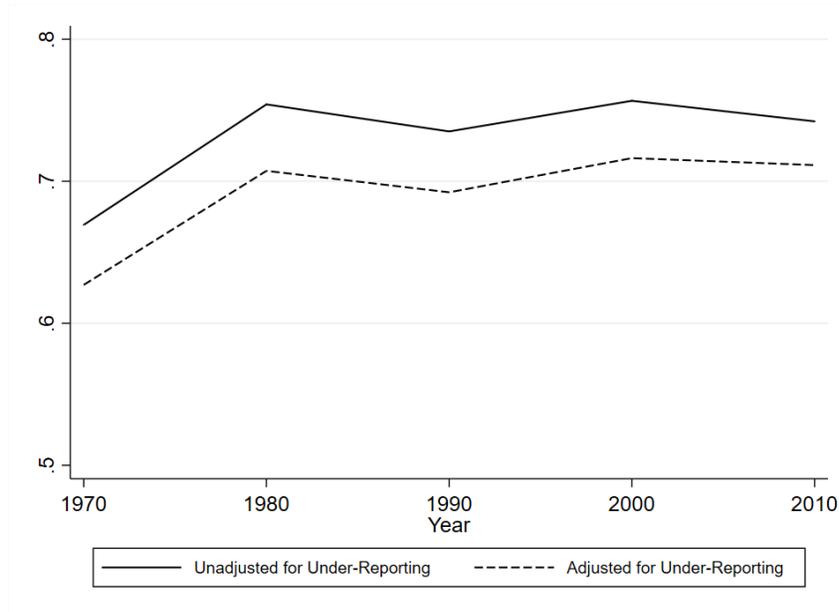


Table D.1.1
 Educational Attainment by Race
 Men Ages 25-29
 Demographer Based Undercount

	High School Completion			College Attendance		
	White	Black		White	Black	
		Unadjusted	Adjusted		Unadjusted	Adjusted
1970	76.4%	55.1%	52.8%	41.7%	18.9%	16.9%
1980	85.9%	74.0%	69.2%	53.3%	37.3%	34.4%
1990	85.3%	74.1%	70.0%	53.4%	38.3%	35.3%
2000	85.8%	78.1%	74.4%	58.8%	43.5%	40.7%
2010	82.6%	75.1%	71.8%	60.2%	48.7%	46.0%

Table D.1.2
Employment by Race
Men Ages 25-29
Demographer Based Undercount

	White	Black	
		Unadjusted	Adjusted
1970	87.0%	80.3%	82.6%
1980	85.6%	70.9%	72.4%
1990	86.0%	67.4%	68.5%
2000	86.0%	67.5%	68.5%
2010	76.9%	56.2%	56.5%

Table D.1.3
Unemployment by Race
Men Ages 25-29
Demographer Based Undercount

	White	Black	
		Unadjusted	Adjusted
1970	\$44,510	\$29,789	\$27,910
1980	\$39,892	\$30,084	\$28,218
1990	\$42,678	\$31,372	\$29,543
2000	\$47,670	\$36,071	\$34,146
2010	\$48,337	\$35,875	\$34,386

Table D.1.4
 Earnings by Race
 Men Ages 25-29
 Demographer Based Undercount

	White	Black	
		Unadjusted	Adjusted
1970	\$44,510	\$29,789	\$27,910
1980	\$39,892	\$30,084	\$28,218
1990	\$42,678	\$31,372	\$29,543
2000	\$47,670	\$36,071	\$34,146
2010	\$48,337	\$35,875	\$34,386

D.2 Ratio Based Method - No Female Undercount

Figure D.2.1
 Unadjusted Educational Attainment
 Native-born Men Ages 25-29
 Ratio Based Method - No Female Undercount

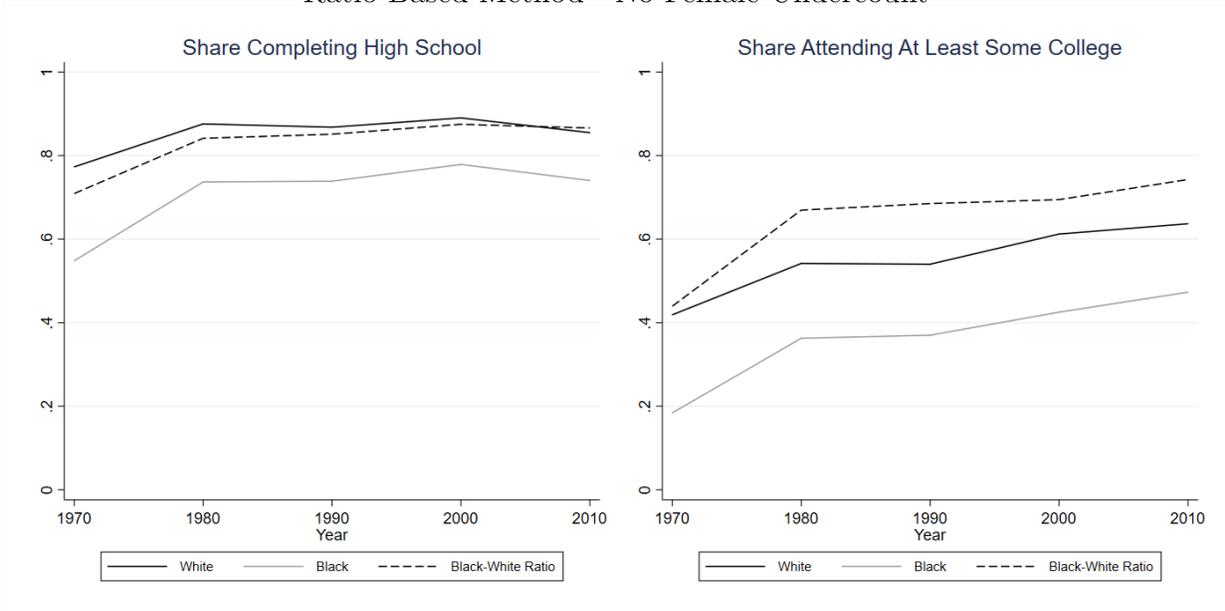


Figure D.2.2
 Ratio of Black to White Educational Attainment
 Native-born Men Ages 25-29
 Ratio Based Method - No Female Undercount

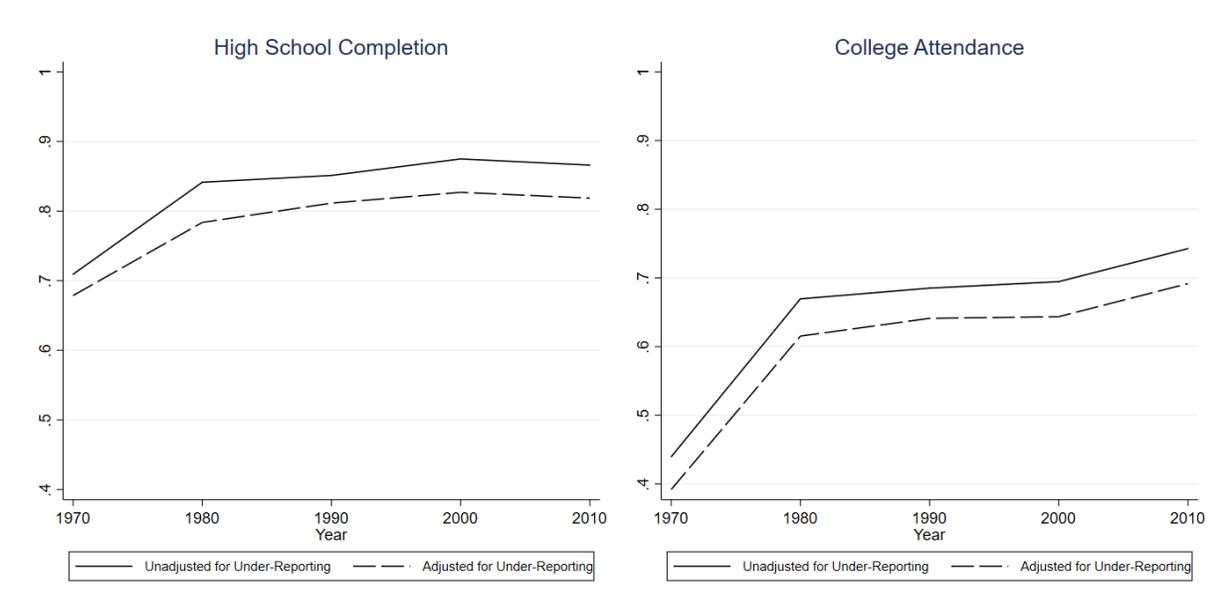


Figure D.2.3
 Unadjusted Employment Rates
 Native-born Men Ages 20-49
 Ratio Based Method - No Female Undercount

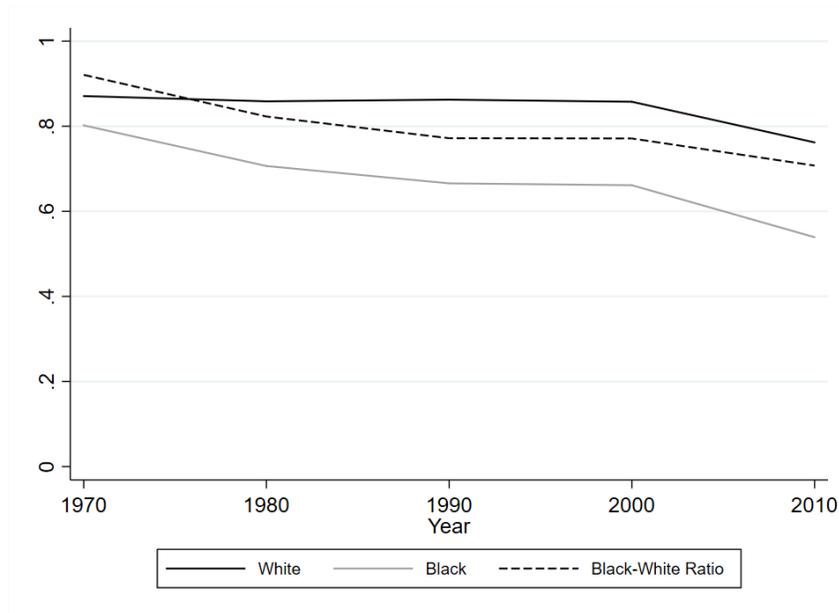


Figure D.2.4
 Ratio of Black to White Employment
 Native-born Men Ages 20-49
 Ratio Based Method - No Female Undercount

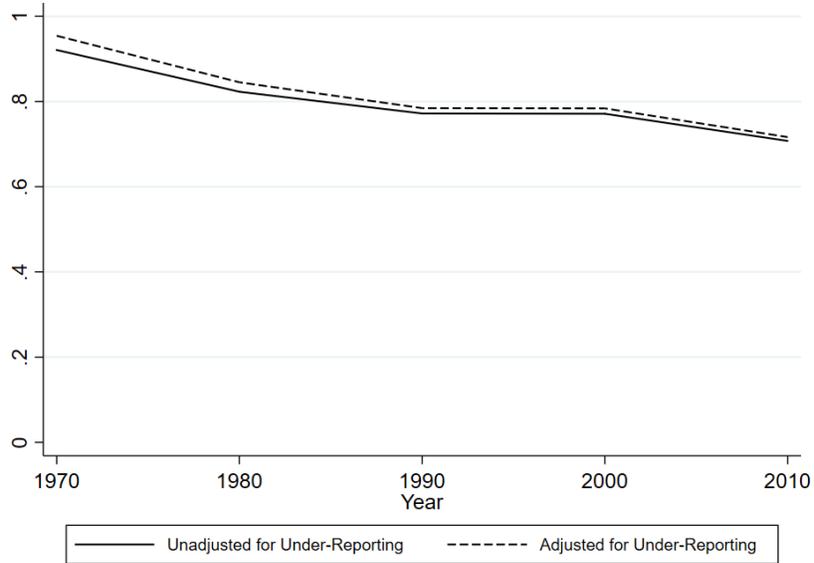


Figure D.2.5
 Unadjusted Unemployment Rates
 Native-born Men Ages 20-49
 Ratio Based Method - No Female Undercount

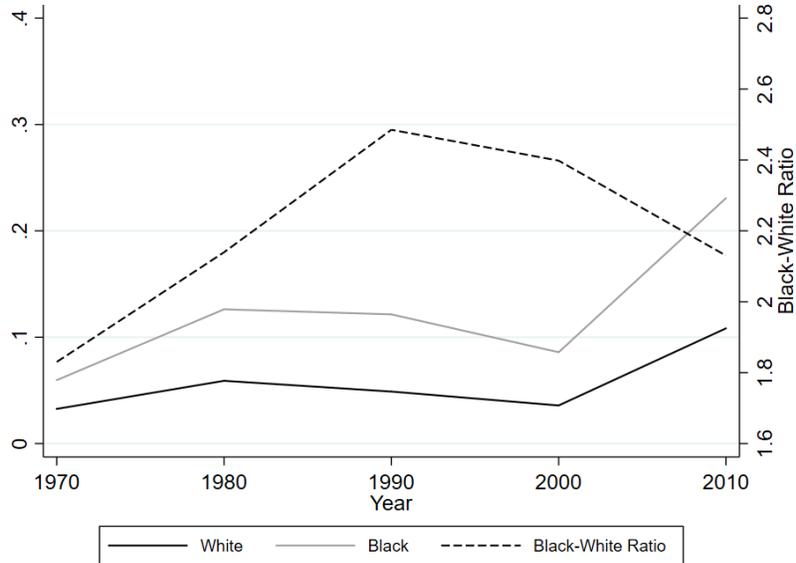


Figure D.2.6
 Ratio of Black to White Unemployment
 Native-born Men Ages 20-49
 Ratio Based Method - No Female Undercount

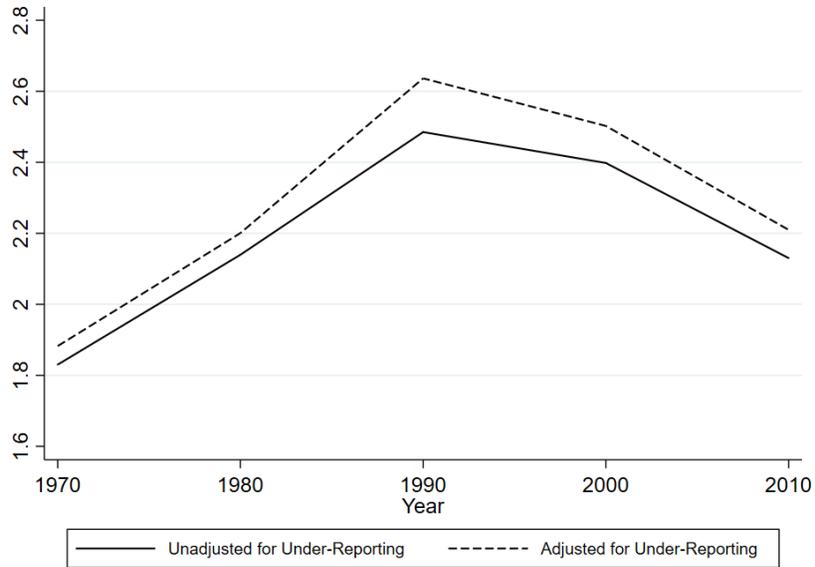


Figure D.2.6
 Unadjusted Earnings Among Working
 Native-born Men Ages 20-49
 Ratio Based Method - No Female Undercount

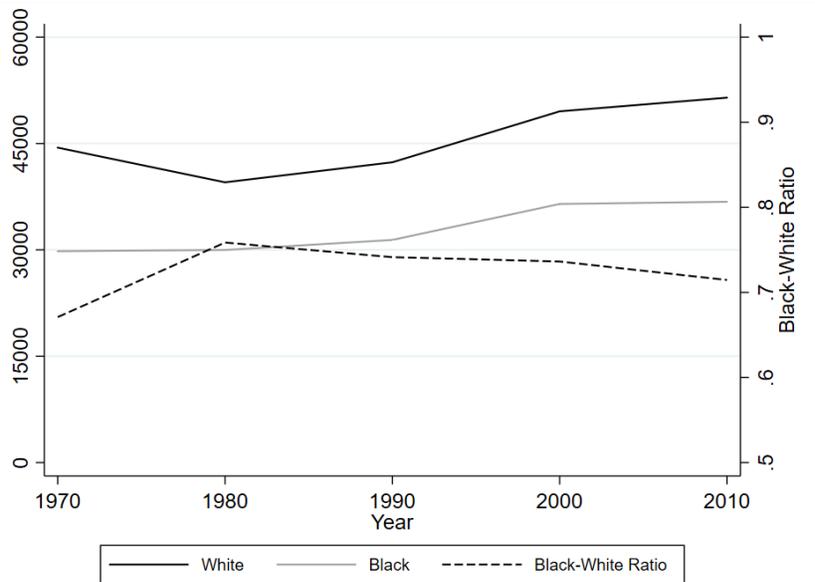


Figure D.2.7
 Ratio of Black to White Earnings
 Native-born Men Ages 20-49
 Ratio Based Method - No Female Undercount

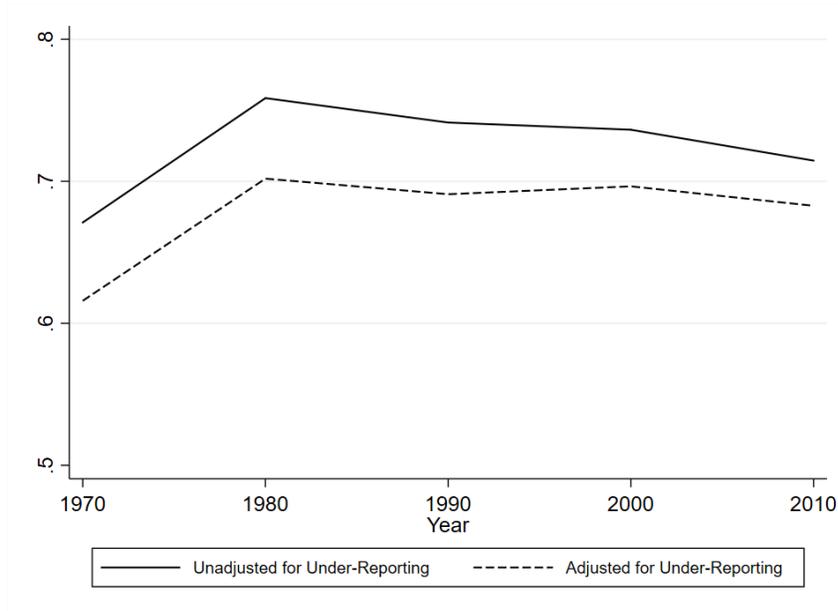


Table D.2.1
 Educational Attainment by Race
 Native-born Men Ages 25-29
 Ratio Based Method - No Female Undercount

	High School Completion			College Attendance		
	White	Black		White	Black	
		Unadjusted	Adjusted		Unadjusted	Adjusted
1970	77.3%	54.8%	52.5%	41.9%	18.4%	16.4%
1980	87.6%	73.7%	68.6%	54.2%	36.3%	33.3%
1990	86.8%	73.9%	70.4%	54.0%	37.0%	34.6%
2000	89.0%	77.9%	73.6%	61.2%	42.5%	39.4%
2010	85.5%	74.0%	70.0%	63.7%	47.3%	44.0%

Table D.2.2
Employment by Race
Native-born Men Ages 25-29
Ratio Based Method - No Female Undercount

	White	Black	
		Unadjusted	Adjusted
1970	87.1%	80.2%	83.1%
1980	85.9%	70.7%	72.6%
1990	86.3%	66.6%	67.7%
2000	85.8%	66.1%	67.2%
2010	76.2%	53.9%	54.6%

Table D.2.3
Unemployment by Race
Native-born Men Ages 25-29
Ratio Based Method - No Female Undercount

	White	Black	
		Unadjusted	Adjusted
1970	\$44,423	\$29,807	\$27,358
1980	\$39,527	\$29,985	\$27,743
1990	\$42,353	\$31,400	\$29,262
2000	\$49,539	\$36,477	\$34,501
2010	\$51,472	\$36,781	\$35,143

Table D.2.4
 Earnings by Race
 Native-born Men Ages 25-29
 Ratio Based Method - No Female Undercount

	White	Black	
		Unadjusted	Adjusted
1970	\$44,423	\$29,807	\$27,358
1980	\$39,527	\$29,985	\$27,743
1990	\$42,353	\$31,400	\$29,262
2000	\$49,539	\$36,477	\$34,501
2010	\$51,472	\$36,781	\$35,143

D.2 Ratio Based Method - No Female Undercount

Figure D.2.1
 Unadjusted Educational Attainment
 Men Ages 25-29
 Ratio Based Method - No Female Undercount

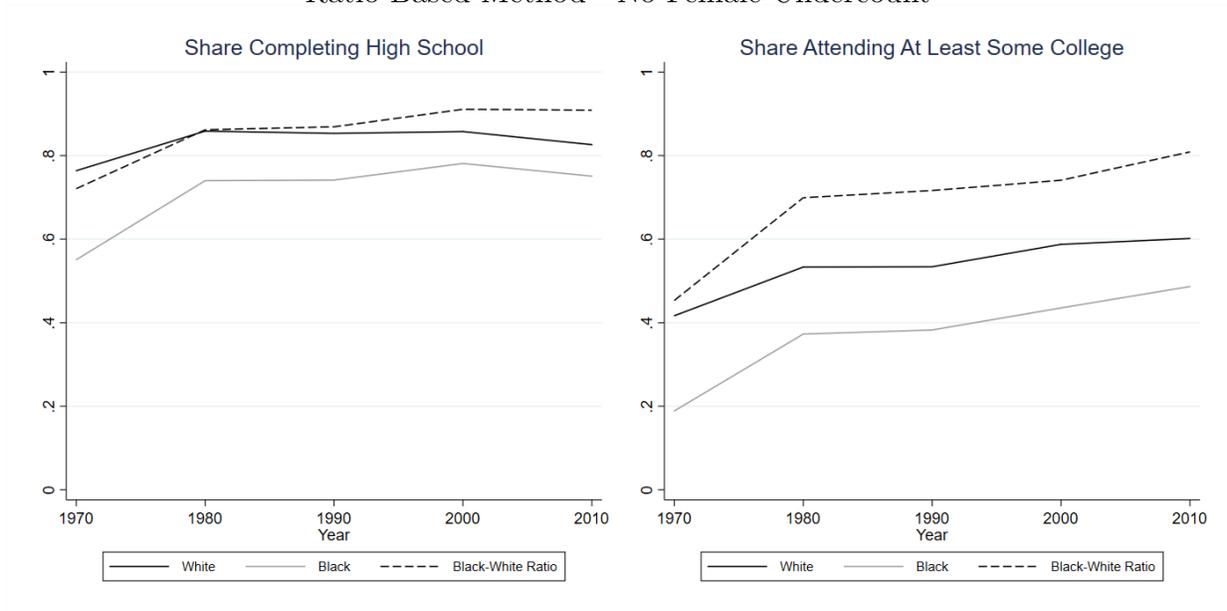


Figure D.2.2
 Ratio of Black to White Educational Attainment
 Men Ages 25-29
 Ratio Based Method - No Female Undercount

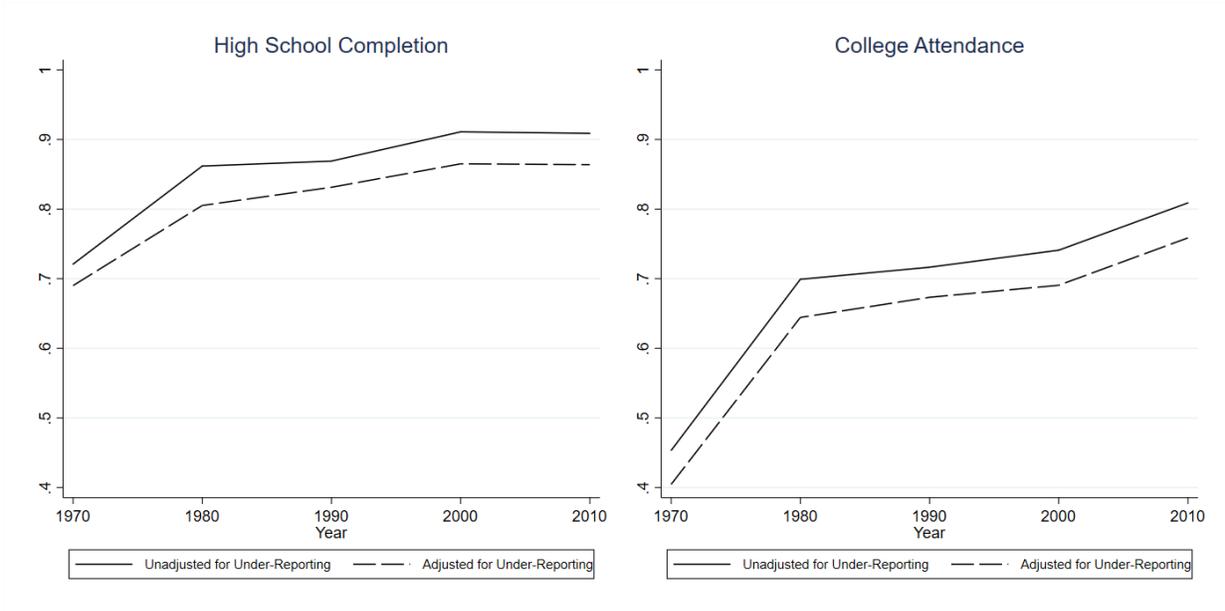


Figure D.2.3
 Unadjusted Employment Rates
 Men Ages 20-49
 Ratio Based Method - No Female Undercount

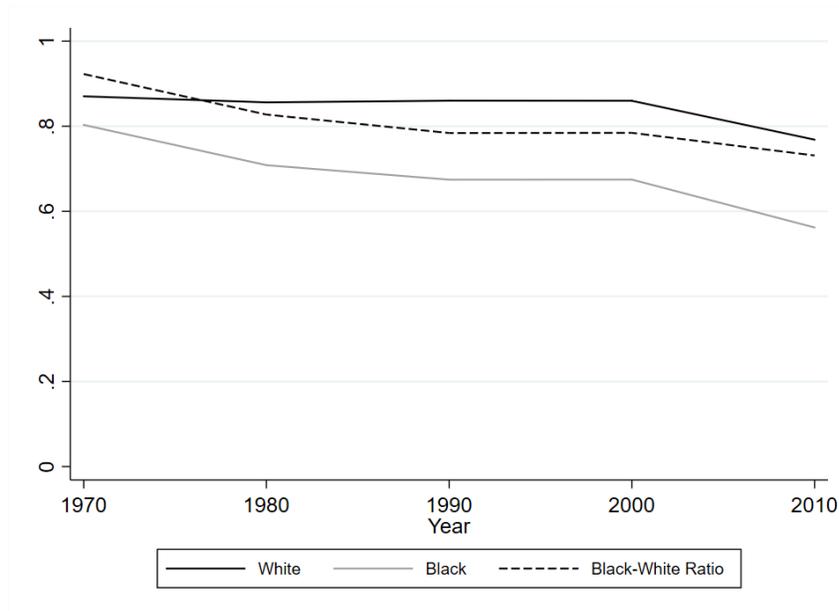


Figure D.2.4
 Ratio of Black to White Employment
 Men Ages 20-49
 Ratio Based Method - No Female Undercount

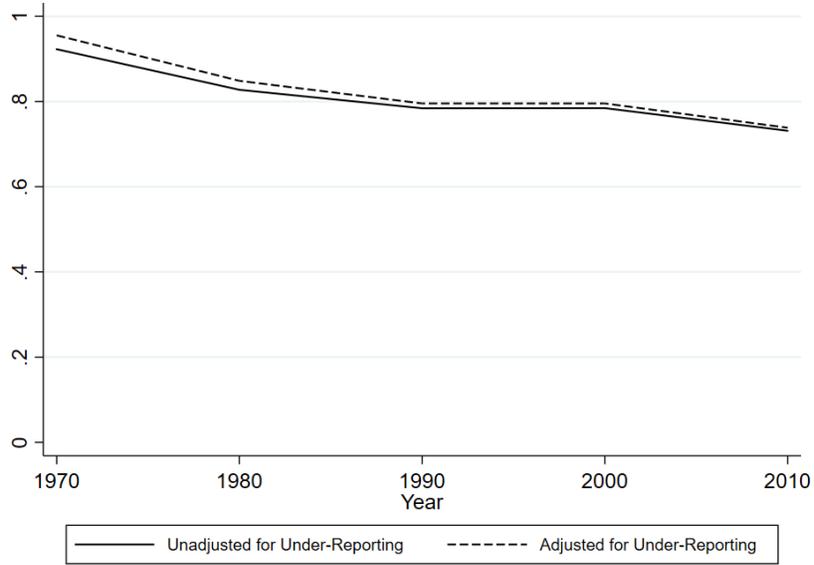


Figure D.2.5
 Unadjusted Unemployment Rates
 Men Ages 20-49
 Ratio Based Method - No Female Undercount

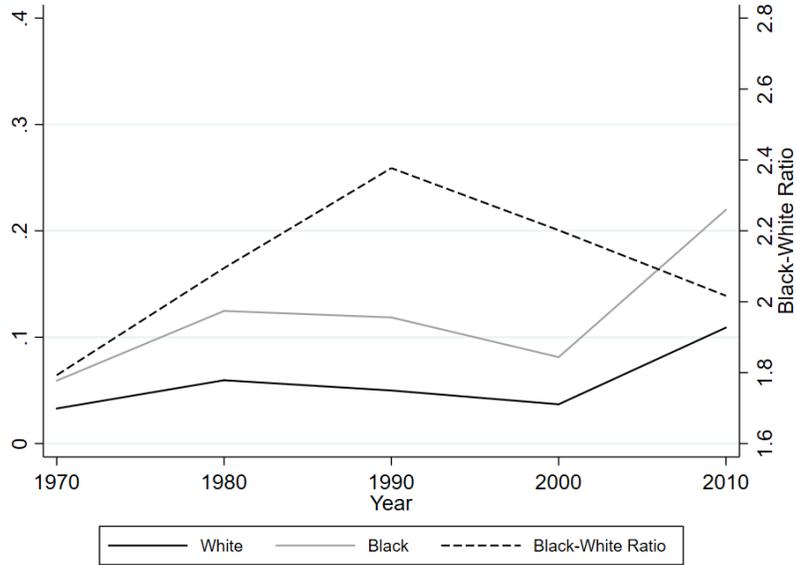


Figure D.2.6
 Ratio of Black to White Unemployment
 Men Ages 20-49
 Ratio Based Method - No Female Undercount

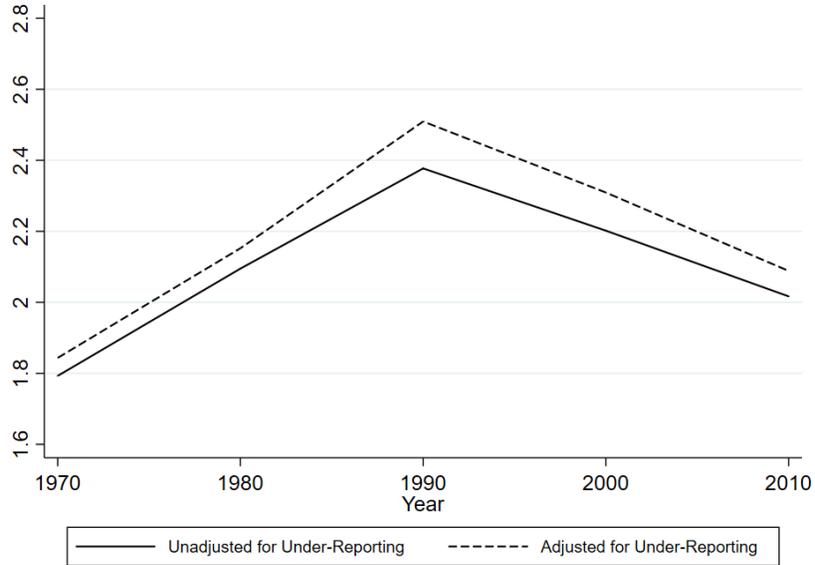


Figure D.2.6
 Unadjusted Earnings Among Working
 Men Ages 20-49
 Ratio Based Method - No Female Undercount

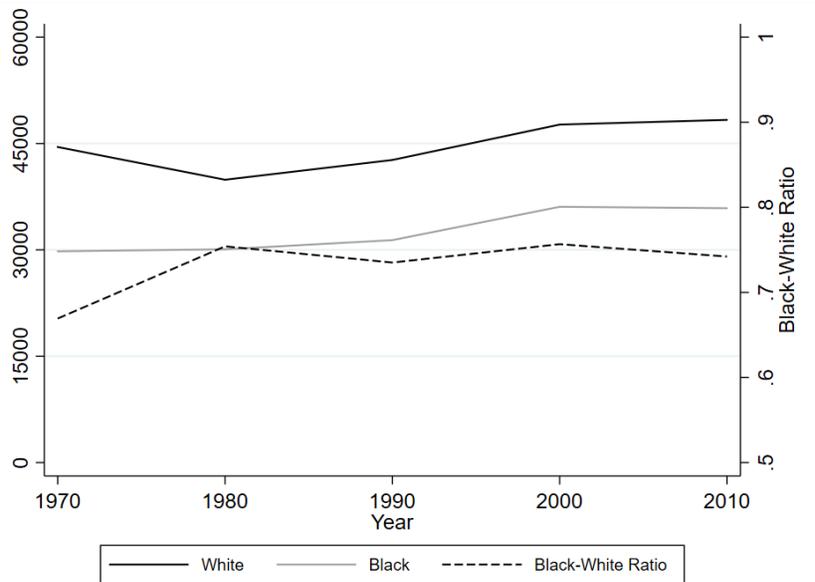


Figure D.2.7
 Ratio of Black to White Earnings
 Men Ages 20-49
 Ratio Based Method - No Female Undercount

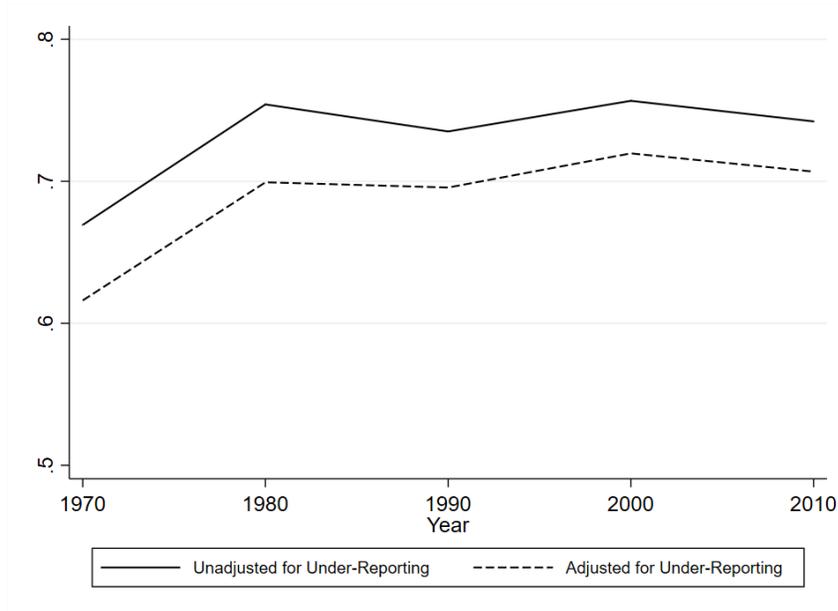


Table D.2.1
 Educational Attainment by Race
 Men Ages 25-29
 Ratio Based Method - No Female Undercount

	High School Completion			College Attendance		
	White	Black		White	Black	
		Unadjusted	Adjusted		Unadjusted	Adjusted
1970	76.4%	55.1%	52.7%	41.7%	18.9%	16.9%
1980	85.9%	74.0%	69.1%	53.3%	37.3%	34.4%
1990	85.3%	74.1%	70.9%	53.4%	38.3%	35.9%
2000	85.8%	78.1%	74.2%	58.8%	43.5%	40.6%
2010	82.6%	75.1%	71.4%	60.2%	48.7%	45.6%

Table D.2.2
Employment by Race
Men Ages 25-29
Ratio Based Method - No Female Undercount

	White	Black	
		Unadjusted	Adjusted
1970	87.0%	80.3%	83.1%
1980	85.6%	70.9%	72.7%
1990	86.0%	67.4%	68.4%
2000	86.0%	67.5%	68.4%
2010	76.9%	56.2%	56.8%

Table D.2.3
Unemployment by Race
Men Ages 25-29
Ratio Based Method - No Female Undercount

	White	Black	
		Unadjusted	Adjusted
1970	\$44,510	\$29,789	\$27,423
1980	\$39,892	\$30,084	\$27,897
1990	\$42,678	\$31,372	\$29,687
2000	\$47,670	\$36,071	\$34,307
2010	\$48,337	\$35,875	\$34,166

Table D.2.4
 Earnings by Race
 Men Ages 25-29
 Ratio Based Method - No Female Undercount

	White	Black	
		Unadjusted	Adjusted
1970	\$44,510	\$29,789	\$27,423
1980	\$39,892	\$30,084	\$27,897
1990	\$42,678	\$31,372	\$29,687
2000	\$47,670	\$36,071	\$34,307
2010	\$48,337	\$35,875	\$34,166

D.2 Ratio Based Method - Demographer Female Undercount

Figure D.2.1
 Unadjusted Educational Attainment
 Men Ages 25-29
 Ratio Based Method - Demographer Female Undercount

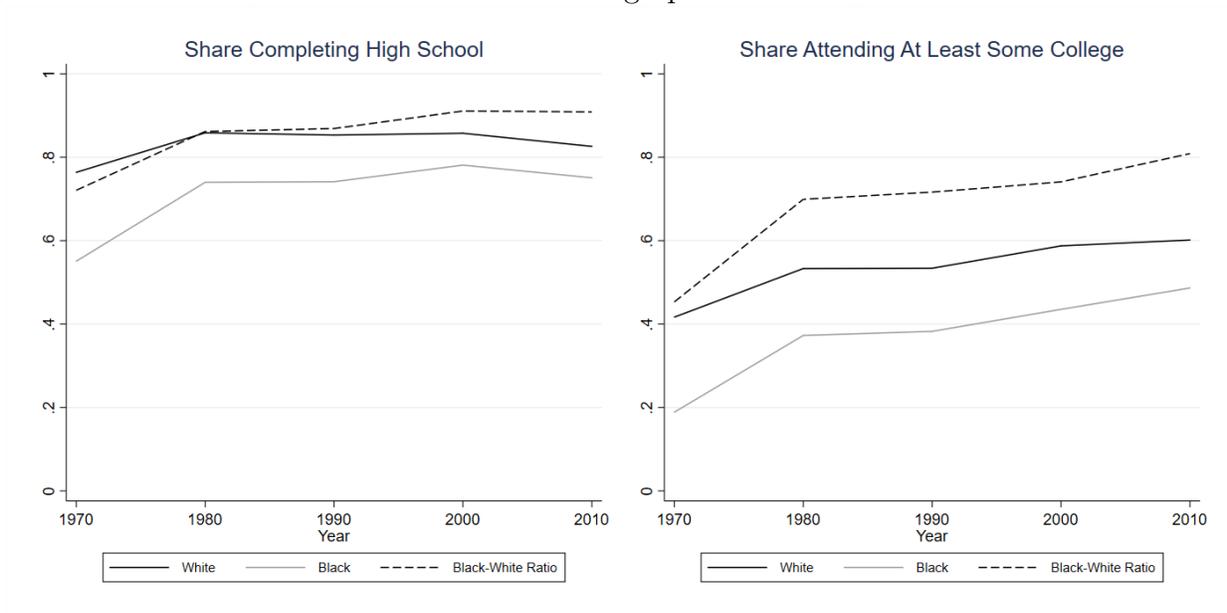


Figure D.2.2
 Ratio of Black to White Educational Attainment
 Men Ages 25-29
 Ratio Based Method - Demographer Female Undercount

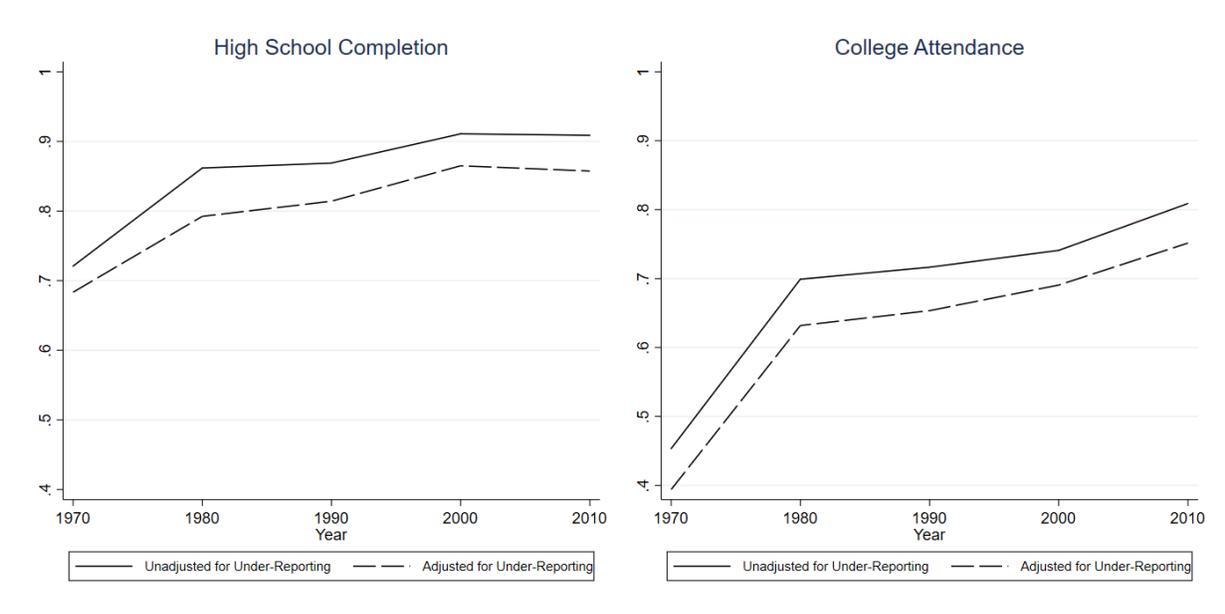


Figure D.2.3
 Unadjusted Employment Rates
 Men Ages 20-49
 Ratio Based Method - Demographer Female Undercount

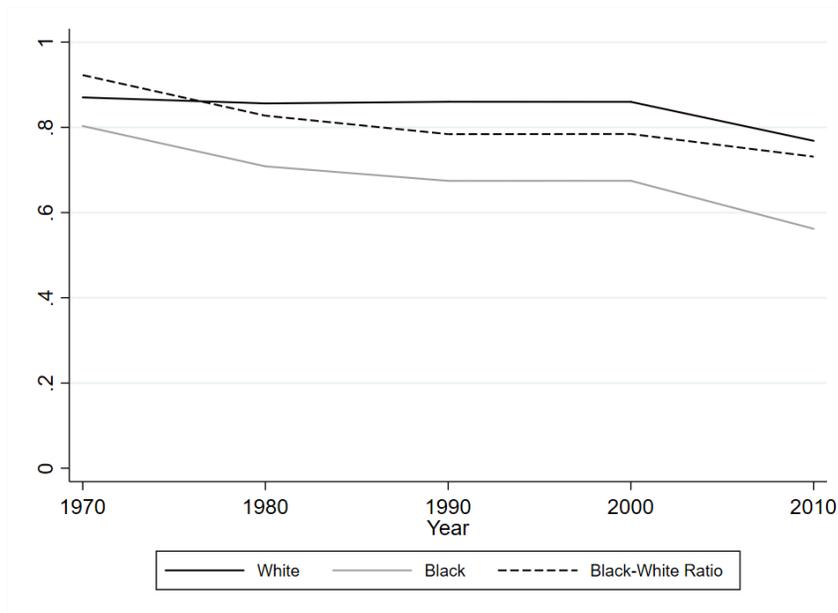


Figure D.2.4
 Ratio of Black to White Employment
 Men Ages 20-49
 Ratio Based Method - Demographer Female Undercount

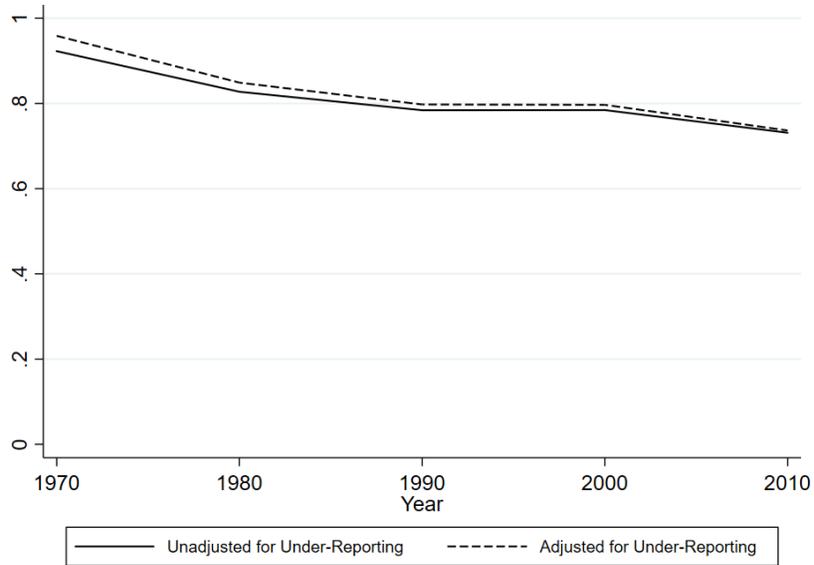


Figure D.2.5
 Unadjusted Unemployment Rates
 Men Ages 20-49
 Ratio Based Method - Demographer Female Undercount

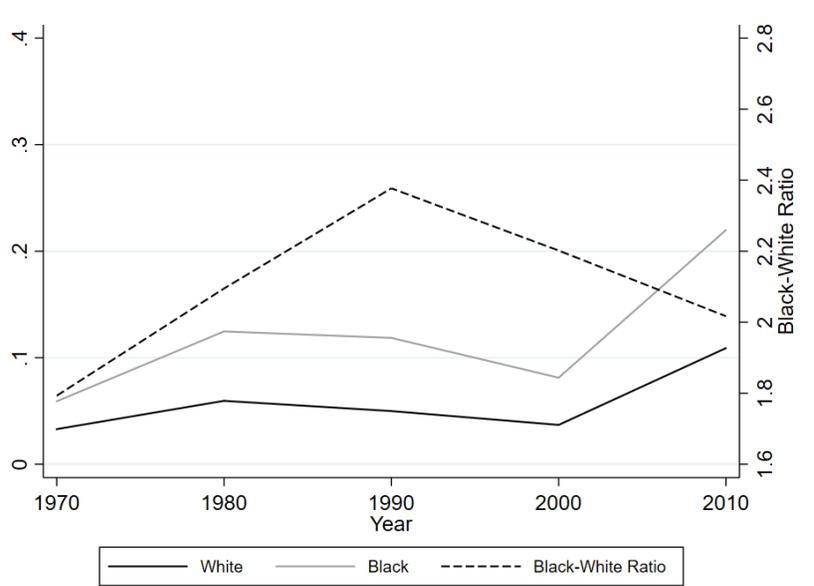


Figure D.2.6
 Ratio of Black to White Unemployment
 Men Ages 20-49
 Ratio Based Method - Demographer Female Undercount

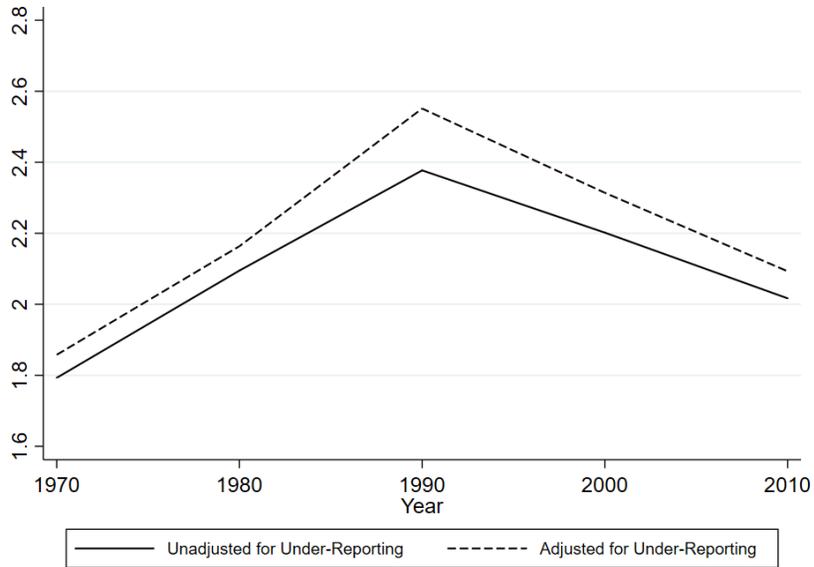


Figure D.2.6
 Unadjusted Earnings Among Working
 Men Ages 20-49
 Ratio Based Method - Demographer Female Undercount

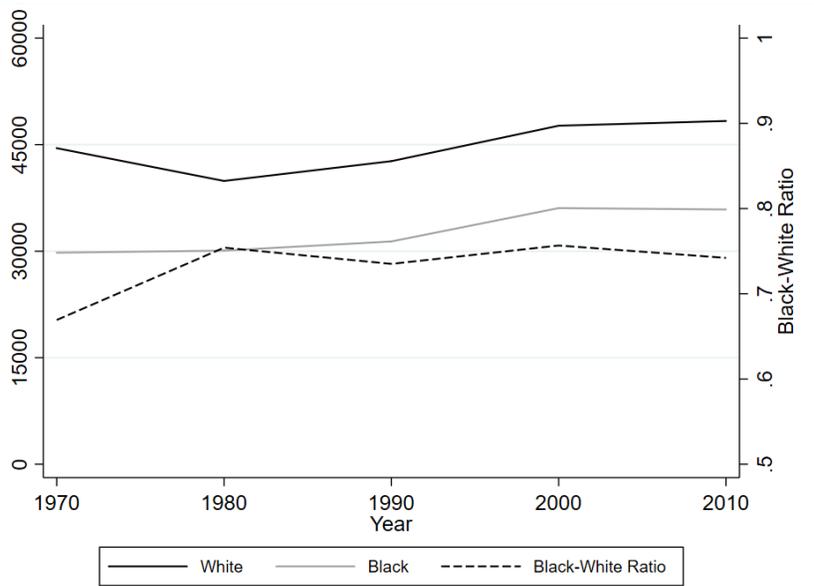


Figure D.2.7
 Ratio of Black to White Earnings
 Men Ages 20-49
 Ratio Based Method - Demographer Female Undercount

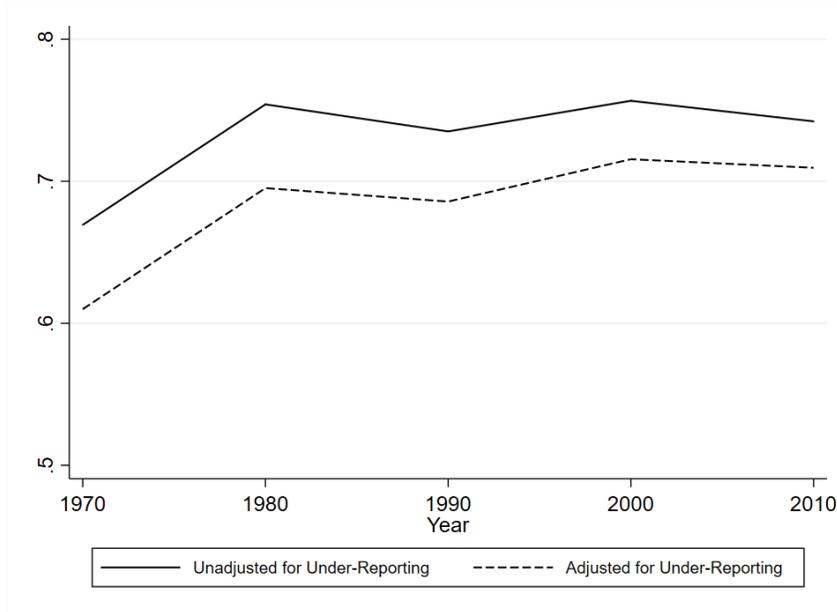


Table D.2.1
 Educational Attainment by Race
 Men Ages 25-29
 Ratio Based Method - Demographer Female Undercount

	High School Completion			College Attendance		
	White	Black		White	Black	
		Unadjusted	Adjusted		Unadjusted	Adjusted
1970	76.4%	55.1%	52.2%	41.7%	18.9%	16.4%
1980	85.9%	74.0%	68.0%	53.3%	37.3%	33.7%
1990	85.3%	74.1%	69.5%	53.4%	38.3%	34.9%
2000	85.8%	78.1%	74.2%	58.8%	43.5%	40.6%
2010	82.6%	75.1%	70.8%	60.2%	48.7%	45.2%

Table D.2.2
Employment by Race
Men Ages 25-29
Ratio Based Method - Demographer Female Undercount

	White	Black	
		Unadjusted	Adjusted
1970	87.0%	80.3%	83.5%
1980	85.6%	70.9%	72.7%
1990	86.0%	67.4%	68.6%
2000	86.0%	67.5%	68.5%
2010	76.9%	56.2%	56.6%

Table D.2.3
Unemployment by Race
Men Ages 25-29
Ratio Based Method - Demographer Female Undercount

	White	Black	
		Unadjusted	Adjusted
1970	\$44,510	\$29,789	\$27,149
1980	\$39,892	\$30,084	\$27,734
1990	\$42,678	\$31,372	\$29,265
2000	\$47,670	\$36,071	\$34,109
2010	\$48,337	\$35,875	\$34,296

Table D.2.4
 Earnings by Race
 Men Ages 25-29
 Ratio Based Method - Demographer Female Undercount

	White	Black	
		Unadjusted	Adjusted
1970	\$44,510	\$29,789	\$27,149
1980	\$39,892	\$30,084	\$27,734
1990	\$42,678	\$31,372	\$29,265
2000	\$47,670	\$36,071	\$34,109
2010	\$48,337	\$35,875	\$34,296