Research on Reducing Correlation Bias in Dual-System Estimates[1]

Krista Heim*, Courtney Hill*
*Decennial Statistical Studies Division, U.S. Census Bureau, Washington, DC, 20233

Abstract
The concept of independence is central to dual-system estimation. One way this independence assumption can fail is if being included in the census makes someone more or less likely to be included in the subsequent post-enumeration survey. The resulting correlation bias could reflect a systematic overestimation or underestimation of the population by dual-system estimation. If violations to independence are present, adjustments can be made to improve estimates. Previous post-enumeration surveys conducted by the U.S. Census Bureau have relied on sex ratios from Demographic Analysis to increase adult male estimates that display signs of correlation bias. Different adjustments could be made to Demographic Analysis data to improve estimates, such as downweighting female estimates or proportionally adjusting both sex groups. This paper will survey some of the methods that have been used in the past and explore new ways to reduce correlation bias.

Key Words: post-enumeration survey; dual system estimation; correlation bias; census; coverage; demographic analysis

1. Introduction

The Post-Enumeration Survey (PES) is a survey conducted each decade to estimate the coverage of the decennial census and help inform future census processes. Measures of undercoverage and overcoverage are produced for various demographic groups and key census operations by comparing census counts to the dual system estimate (DSE) produced by the PES.

The concept of independence is central to dual-system estimation. Correlation bias occurs when the independence assumption underlying the DSEs fails due to either:

1) **causal dependence –** the act of being included in the Census makes someone more likely or less likely to be included PES, or
2) **heterogeneity** – the Census and PES inclusion probabilities vary for people within the same estimation group (Ahlo et al, 1993; Bell, 2001).

Causal dependence can lead to underestimation (overestimation), that is, a negative (positive) bias in the DSE, if being missed in the Census increases (decreases) one's chance of being missed in PES. At the same time, heterogeneity of the inclusion probabilities of people in an estimation group also leads to underestimation—a negative bias in the DSE.

---

[1] This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the author and not those of the U.S. Census Bureau. The U.S. Census Bureau reviewed this data product for unauthorized disclosure of confidential information and approved the disclosure avoidance practices applied to this release. (CBDRB-FY22-417, CBDRB-FY22-396, CBDRB-FY22-319, CBDRB-FY22-136, CBDRB-FY22-DSEP-001)

If violations to independence are present, adjustments can be made to improve PES estimates. Historical evidence of correlation bias in DSEs comes from comparisons against Demographic Analysis (DA) estimates for age-race-sex groups (Shores, 2002; Bell, 1993). In past post-enumeration surveys, the DA sex ratios (males to females) suggested the presence of correlation bias in the adult male DSEs, and thus adult male estimates were adjusted to align with the DA sex ratio.

There are other adjustments that could be made using DA data to improve estimates of net coverage. This paper reviews adjustments and results from previous post-enumeration surveys, and describes some methods researched for the 2020 PES. Section 2 gives an overview of the dual-system estimation used in the 2020 PES. Section 3 describes DA and PES populations and how they compare to each other. Section 4 explains past research and methods. Section 5 discusses the research done for 2020 PES. Section 6 compares results using different adjustment options. Section 7 describes potential future work.

2. Overview of Dual-System Estimation

To calculate the DSE, PES conducts an independent area-based sample of the population (known as the P sample) to compare to census housing unit and person enumerations within the same sample areas (known as the E sample). The history of the DSE in census coverage evaluation is well documented (Wolter 1986; Hogan 1993; U.S. Census Bureau 2004; Mule 2008). To formulate the DSE, the P- and E- sample records are classified into cells as shown in Table 1.

Table 1: Classification for Dual-System Estimation

		P Sample: Correctly Enumerated in the PES?		
		YES	NO	Total
E Sample: Correctly Enumerated in the Census?	YES	N_{11}	N_{12}	N_{1+}
	NO	N_{21}	N_{22}	N_{2+}
	Total	N_{+1}	N_{+2}	N_{++}

In theory, all cells are observable except those highlighted in gray (N_{22}, N_{2+}, N_{+2}, and N_{++}). N_{++} represents the true population total. Under the assumption of statistical independence between the PES and the census, we can estimate the total population, N_{++}, using the classic formulation of the DSE:

$$\hat{N}_{++} = N_{1+}\left({}^{N_{+1}}/_{N_{11}}\right) \quad (1)$$

where,

- $\hat{N}_{++}$ is an estimate of the total number of people.
- N_{1+} is the number of people correctly counted in the census.
- N_{+1} is the number of people correctly counted in the post-enumeration survey.
- N_{11} is the number of people counted in both the census and the post-enumeration survey (i.e., the number of matches).

The inverse of the term in parentheses is called the *match* rate. The term N_{1+} requires introduction of two terms into the DSE formula.

- A person record is a *correct enumeration* (CE) if it is accurately included in the census and at the location in which the person should have been counted[2].
- A person record is considered *data-defined* (DD) if it has sufficient information to be accepted for census processing. A person record must be data-defined to be considered a correct enumeration.

For each person enumeration in the census, we estimated the probability that the record was a match, a CE, and was DD using logistic regression modeling (Heim, 2022).

Assuming that the P and E samples are *independent* and each unit in the same pseudo post-stratum has the same chance of being in the E sample or the P sample, we can estimate our "true" population size for specific domain *C* using dual-system estimation as follows:

$$DSE_C = \sum_{j \in C} \frac{\pi_{dd,j} \times \pi_{ce,j}}{\pi_{m,j}} \qquad (2)$$

where,

- j is a census person enumeration.
- $\pi_{dd,j}$ is the probability that the census person enumeration is DD.
- $\pi_{ce,j}$ is the probability that the census person enumeration is a CE.
- $\pi_{m,j}$ is the probability that the census person enumeration is a match.

Past post-enumeration surveys (1980, 1990, and 2000) used post-stratification in part to minimize the impact of the failure of independence assumptions. Post-stratification, or grouping individuals likely to have similar inclusion probabilities, and calculating DSEs within post-strata was done to decrease heterogeneity and thus decrease correlation bias (Griffin, 2000). However, the post-stratification approach had limitations on the number of variables that could be included. Each variable and cross-classification increased the number of strata and thus reduced the sample size within strata. Additionally, only categorical variables or groupings of continuous variables could be used in the post-stratification.

Logistic regression is more flexible and has been shown to further reduce potential bias (Olson and Sands, 2012). Logistic regression uses a statistical modeling approach to estimate the relationship of a binary outcome to independent variables like race, age, and sex. Logistic regression incorporates additional variables that would have been infeasible in post-stratification, as some post-strata would contain too few sample cases. Logistic regression also allows for the inclusion of continuous variables and interaction terms. More details on post-stratification and logistic regression can be found in Zamora (2022). Logistic regression modeling was used in 2010 and 2020 post-enumeration surveys.

Correlation bias can remain after reducing the synthetic error using logistic regression modeling. In the original estimates used in the 2000 and 2010 post-enumeration surveys, adult males—but not adult females— in some age groups appeared to be underestimated, perhaps due to correlation bias. This was demonstrated by comparisons of PES estimates to DA estimates for those decades. To address this, the DSEs of males in those age groups were increased so that the sex ratio of the PES estimates matched that from DA. The DSEs for adult females in those age groups were not changed.

---

[2] Refer to Zamora (2022) for the criteria used in determining if a person record is a correct enumeration.

3. Demographic Analysis and Post-Enumeration Survey Populations

As stated previously, historical evidence of correlation bias in DSEs comes from comparisons against DA estimates for age, sex, and race groups. DA is another program from the U.S. Census Bureau that provides independent estimates of the population size. The DA program uses current and historical vital records, data on international migration, and Medicare records to produce national estimates of the population by age, sex, race, and Hispanic origin (Jensen et al., 2020). DA has the advantage that its estimates are constructed from administrative data sources, some of which are believed to be quite accurate.

Table 2 shows the DA estimates that were available for the 2020 population. Hispanic origin was not reported in vital records by all states until 1990, so DA does not provide estimates of this population beyond that time.

Table 2: Description of the three sets of 2020 DA Estimates

Populations	Characteristics	Age Groups
Black alone/non-Black alone	Age, sex, race	0-85+
Black alone or in combination/ non-Black alone or in combination	Age, sex, race	0-85+
Hispanic/non-Hispanic	Age, sex, Hispanic origin	0-29

Source: U.S. Census Bureau, Population Division, 2020 Demographic Analysis (December 2020 release)

Uncertainty in DA estimates is accounted for by publishing a low, middle, and high series for each set of official estimates. We historically compared PES estimates to **Middle Series** DA estimates.

Adjusting PES estimates to DA results requires PES creating consistently defined population groups. Since DA produces population estimates by sex and race for every individual year of age up to 85, PES needs to duplicate those groups as closely as possible. The following steps were taken to make PES estimates consistent with DA estimates:

1. The PES universe did not evaluate the coverage of deployed people and people in group quarters, remote Alaska, and in transitory locations (e.g., recreational vehicles in recreational vehicle parks, tents on campgrounds, etc.). These populations were removed from the DA estimates by subtracting the census counts of these populations from the DA estimates.

2. PES used DA estimates of Black alone or in combination/non-Black alone or in combination and adjusted PES race definitions. DA uses birth records and other sources to gather race, while PES had self-reported race with multiple categorical options. Race categories and definitions have also changed over time as the population has evolved. Efforts were made to align race classifications, but some differences still exist.

DA sex ratios are overall thought to be accurate and invariant to assumptions in DA estimates. DA estimates of totals can be affected by uncertainty about the level of emigration or undocumented immigration, among other limitations of administrative data. Assumptions about these populations are thought to impact males and females similarly. For this reason, DA sex ratios are thought to be more accurate than DA population estimates (Bell, 2001).

4. Past Research and Methods

4.1 2000 Research and Methods

Early research into adjusting the DSEs from post-enumeration surveys to account for correlation bias is documented in Bell (1993). In this research, four alternative DSEs were considered using DA results to adjust PES estimates. Each had different underlying assumptions about what parameters were held constant over male age-race groupings.

As part of the research for the 2000 post-enumeration survey, known as 2000 Accuracy and Coverage Evaluation (A.C.E.), Bell (1999) documents using the “two group” method to adjust logistic regression DSEs from 1990 PES data. The two-group model assumes heterogeneity within any post-stratum arises from there being two groups of people in the post-stratum, with these groups having different census inclusion probabilities. (These groups could be thought of as “easy to count" and "hard to count.") The two-group model has two advantages relative to other models for combining of DA estimates with DSEs. First, it is a simple approach to rescale the male DSEs that is constant across poststrata within age-race groups. Second, it is expected to have low variance relative to other combining models.

2000 A.C.E. used three age categories: 18-29, 30-49, and over 50. It adjusted for correlation bias in the DSEs for adult males using the 2000 DA sex ratios, assuming no correlation bias for females or children. Table 3 shows the sex ratios (Male/Female) from DA and 2000 A.C.E for Blacks and non-Blacks. DA sex ratios for adult Blacks exceeded those for A.C.E., suggesting correlation bias in DSEs for adult Black males. DA sex ratios for non-Blacks 30-49 and 50 and over also slightly exceed those from A.C.E., suggesting at most small amounts of correlation bias.

Table 3: Sex Ratios (Male/Female) from DA and 2000 A.C.E. (non-adjusted DSE)

Age Group	Black		Non-Black	
	A.C.E.	DA	A.C.E.	DA
18-29	0.83	0.90	1.05	1.04
30-49	0.81	0.89	0.99	1.01
50+	0.72	0.76	0.85	0.86

Source: Shores (2002).
DA sex ratios were calculated using modified DA estimates.

Within each age-race group, the DA sex ratio was multiplied by the female DSE to produce a “control total.” The aggregated male DSE times a constant was equated with this control total, and solving for the constant gives the estimated adjustment needed to correct for correlation bias within the age-race group i:

$$CBA_i = \frac{DA_{male,i}}{DA_{female,i}} \times \frac{DSE_{female,i}}{DSE_{male,i}} \tag{3}$$

where,

- $\frac{DA_{male,i}}{DA_{female,i}}$ is the DA sex ratio for a particular age-race group i.
- CBA_i provides an estimate of a multiplicative correction factor to correct each male DSE in the age-race group for correlation bias i (Shores, 2002).

Large adjustments were made for correlation bias for Blacks. For non-Blacks, a relatively small adjustment for correlation bias was made in the two older age categories (30-49 and 50 and over). Because of inconsistencies between DA and the estimates for non-Blacks, it was not possible to reliably adjust for correlation bias for non-Blacks in the 18-29 age category.

Some additional work was needed to better align race categories in 2000. A.C.E. subtracted estimates of Black Hispanics from the DA totals for Blacks and added these estimates to the DA totals for non-Blacks (Shores, 2002). We needed this adjustment because A.C.E. assigned Black Hispanics to its Hispanic race domain, not its Black race domain. The DA estimates of Black Hispanics for 2000 were obtained by inflating the Census counts of Black Hispanics by adjustment factors corresponding to the DA estimates of Black undercount, since separate DA estimates of Hispanic undercount were not available.

The adjustment for correlation bias decreased the A.C.E. net overcount estimate of the total population from 3.0 million (1.10%) to 1.3 million (0.48%) (Shores, 2002).

4.2 2010 Research and Methods

As part of the research for the 2010 post-enumeration survey, known as Census Coverage Measurement (CCM), we tested adjusting by individual years of age for the adult Black and adult non-Black male populations (Mule, 2008). This differs from the previous adjustment that used only three age categories. This was motivated by the assumption that CCM would create population estimates by individual years of age. Also, there were detectable differences in coverage within some population groups, most noticeably between the younger and older members of the 18-29 group. Use of post-stratified groups also created artificial jumps in coverage adjustment at the borderline ages.

Concerns about heaping for adult ages divisible by five, as well as irregularity caused by minor errors in age reporting, motivated the desire to smooth factors using a moving average (Olson, 2008). All population counts and estimates for individual age k that went into the adjustment used a 5-year moving average. For example,

$$DSE_k = \sum_{k-2}^{k+2} DSE_i \tag{4}$$

However, the decision to create population estimates by individual years of age was reversed, and age groups were used instead. Thus, sex ratio adjustment factors were calculated within the same age groups defined previously.

Table 4 shows the sex ratios (Male/Female) from DA and 2010 CCM for Blacks and non-Blacks. The DA sex ratios for Black adults exceeded those sex ratios for CCM. This strongly suggested the presence of correlation bias in the adult Black male DSEs (Konicki, 2012). Additionally, the DA sex ratios for Non-Blacks aged 30 or older also exceeded those sex ratios for CCM, but to a lesser degree than for Black adults. This suggests a relatively small amount of correlation bias at the national level in the Non-Black male DSEs for these age groups. For Non-Blacks aged 18 to 29, the DA sex ratio was slightly less than the CCM sex ratio. Because of this, we did not make an adjustment for correlation bias for non-Black males aged 18 to 29.

Table 4: Sex Ratios (Male/Female) from DA and 2010 CCM (non-adjusted DSE)

Age Group	Black		Non-Black	
	CCM	DA	CCM	DA
18-29	0.89	0.94	1.03	1.03
30-49	0.81	0.91	0.99	1.02
50+	0.76	0.80	0.88	0.89

Source: Konicki (2012).
DA sex ratios were calculated using modified DA estimates.

The correlation bias adjustment for 2010 CCM was calculated as follows. If the ratio of males to females in CCM was smaller than ratio of males to females in DA, then correlation bias adjustment inflated the male DSE until the DA sex ratio was obtained. Within each age-race group *i*, the adjustment was defined as:

$$CBA_i = max\left(\frac{DA_{male,i}/DA_{female,i}}{DSE_{male,i}/DSE_{female,i}}, 1\right). \quad (5)$$

Then, the adjusted male DSE was:

$$Adjusted\ DSE_C = \sum\nolimits_{j \in C} \pi_{dd,j} \frac{\pi_{ce,j}}{\pi_{m,j}} * CBA_j \quad (6)$$

for adult male age-race group *j*. Once again, the children and adult female estimates were not adjusted.

The definition of Black that we used for the CCM correlation bias adjustment differed from the definition used for A.C.E. (Konicki, 2012). For 2010 CCM, we defined Black as Black alone-or-in-combination with other races. This is a broader definition of Black than what was used for A.C.E., and thus the adjustment factors for Blacks were applied to a greater portion of the population.

The adjustment for correlation bias decreased the CCM estimate of nation-level net undercount, which changed from 0.87% without the correlation bias adjustment to a net undercount of 0.01% with the adjustment (Konicki, 2012).

5. 2020 PES Research and Methods

Table 5 displays the sex ratios from DA and 2020 PES for Black and non-Black age groups. Like the previous post-enumeration survey, the Black population was identified by any census enumeration who reported their race as Black alone or in combination with other races, while the non-Black population was any census enumeration that did not report a race of Black. We also continued to not adjust for age groups less than 18, as the sex ratios for these groups were the same between 2020 PES and DA.

Historically, three age categories were formed for post-stratification as well as correlation bias adjustment: ages 18 to 29, ages 30 to 49, and ages 50 and over. In 2020 PES, we analyzed correlation bias in five age groups. The 18 to 29 age group was separated into 18 to 24 and 25 to 29, as we could see large differences between these age categories for both PES and DA sex ratios, particularly for the Black population. This would also allow us to analyze college-aged adults separately. Since there are more females in college than males, sex ratios do differ for this age group (since PES excludes group quarters, which includes colleges). There were also challenges with counting the college-aged population in the census during COVID-19 while many students were being sent home. We also saw differences in sex ratios with the 50 and over group. We decided to divide this group into adults aged 50-64 and 65 and over.

Table 5: 2020 PES and Demographic Analysis Sex Ratios (Males to Females) by Age and Race

Age	Black		Non-Black	
Group	PES	DA	PES	DA
0 to 4	1.02	1.03	1.05	1.05
5 to 9	1.03	1.03	1.05	1.05
10 to 17	1.03	1.03	1.05	1.05
18 to 24	0.98	0.99	1.02	1.05
25 to 29	0.89	0.93	1.00	1.02
30 to 49	0.85	0.92	0.99	1.00
50 to 64	0.83	0.90	0.95	0.97
65+	0.70	0.71	0.84	0.84

Sources: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey. DA sex ratios were calculated using modified DA estimates based on estimates from U.S. Census Bureau, Population Division, 2020 Demographic Analysis (December 2020 release).

Compared to previous decades, we saw similar trends in 2020 for the PES and DA sex ratios as we did in the previous decades for most of the age categories. However, as the data was being collected and processed for 2020 PES, there were differences in the patterns we were seeing for the overall male and females estimates compared to previous post-enumeration surveys. Table 6 gives an example of trends we were seeing with fictitious data, with female DSEs larger than DA estimates.

Table 6: Example with 2020 Patterns (Fictitious Data)

Domain	Demographic Analysis	Pre-Adjusted DSE	Adjusted DSE
Females	100	110	110
Males	100	96	110
Total	200	206	220

Historically, since female DSEs were close to DA, the correlation bias adjustment increased the male DSEs to achieve the DA sex ratio. Since the 2020 female DSEs appeared higher than DA in 2020, this historic correlation bias adjustment could have resulted in male DSEs that were larger than DA totals. Refer to Figure 3 for comparisons of the 2020 DSE/DA ratio compared to 2010.

We tested several different adjustment methods to account for the patterns we were seeing and to be prepared to make a different adjustment than in the past. We included the following options in our research:

0) No Correlation Bias Adjustment: This assumes that no adjustment is needed and that there is independence of PES estimates for both males and females for age-race groups.
1) Male Adjustment: This is the adjustment done in the previous post-enumeration survey (see Equation (5)). If the PES male-to-female sex ratio is smaller than DA male-to-female sex ratio, we weighted up. If the PES male-to-female sex ratio was larger than DA male-to-female sex ratio, we did no adjustment.
2) Female Adjustment: Instead of inflating the male estimate, this adjustment would decrease the female estimate. If the PES female-to-male sex ratio was larger than DA female-to-male sex ratio, we weighted the female estimate down. If the PES female-to-male sex ratio was smaller than DA female-to-male sex ratio, we did not adjust. Thus, for females only,

$$CBA_i = min\left(\frac{DA_{female,i}/DA_{male,i}}{DSE_{female,i}/DSE_{male,i}}, 1\right). \quad (7)$$

3) Proportion Adjustment: This method would adjust the PES proportions of both adult males and adult females to match the DA sex ratio. This method would not change the total PES DSE for the age group, but it changes the PES DSE for males and females.

For males, this adjustment is calculated as:

$$CBA_i = \frac{DA_{male,i}/\left(DA_{male,i} + DA_{female,i}\right)}{DSE_{male,i}/\left(DSE_{male,i} + DSE_{female,i}\right)}. \quad (8)$$

For females, this adjustment is calculated as:

$$CBA_i = \frac{DA_{female,i}/\left(DA_{male,i} + DA_{female,i}\right)}{DSE_{female,i}/\left(DSE_{male,i} + DSE_{female,i}\right)}. \quad (9)$$

4) Distance Adjustment: If the male PES estimate in a specific age-race group is farther from the DA Estimate than the female PES estimate in the same age-race group, adjust males and not females. Likewise, if the female PES estimate in a specific age-race group is farther from the DA Estimate than the male PES estimate in the same age-race group, adjust females and not males.

$$if\ |DSE_{male,i} - DA_{male,i}| > |DSE_{female,i} - DA_{female,i}|$$
$$then\ CBA_{male,i} = \frac{DA_{male,i}/DA_{female,i}}{DSE_{male,i}/DSE_{female,i}}. \tag{10}$$

$$if\ |DSE_{male,i} - DA_{male,i}| < |DSE_{female,i} - DA_{female,i}|$$
$$then\ CBA_{female,i} = \frac{DA_{female,i}/DA_{male,i}}{DSE_{female,i}/DSE_{male,i}}. \tag{11}$$

5) Calibration Adjustment: This adjusts the DSE totals to match the DA totals within sex, age, and race groups. The adjustment factor for age-race group i is calculated as:

$$CBA_{male,i} = \frac{DA_{male,i}}{DSE_{male,i}}. \tag{12}$$

$$CBA_{female,i} = \frac{DA_{female,i}}{DSE_{female,i}}. \tag{13}$$

The final overall DSEs for age, race, sex, and total population would be equal to the DA estimates with this method. Note that DA totals are susceptible to errors due to completeness of birth records or undocumented immigration, among other limitations of administrative data. DA sex ratios are overall thought to be accurate and invariant to assumptions in DA estimates, unlike DA totals.

6. Results

Figures 1 and 2 illustrate the adjustment factors for Black and non-Black adults using the different methods discussed in Section 5. The initial DSEs are multiplied by the adjustment factors to produce the final adjusted DSEs. If the factor is greater than one, we are inflating the DSE for that group. If the factor is less than one, we are decreasing the DSE for that group. No adjustment falls along the line at one.

The 2010 male adjustment inflates the DSEs the most, especially for Black males 25-29, 30-49, and 50-64, as well as non-Black males 18-24. The female adjustment decreases the DSEs the most, particularly for those same age and race categories, although the adjustment is slightly smaller than the male adjustment. The proportion method splits the difference between the male and female adjustments. This adjustment does not impact the total DSE but adjusts males up and females down. The distance adjustment weights down female Black estimates, but weights up most male non-Black estimates. The calibration adjustment greatly decreases the DSE for the 18-24 age category for both males and female and for both Black and non-Black groups.

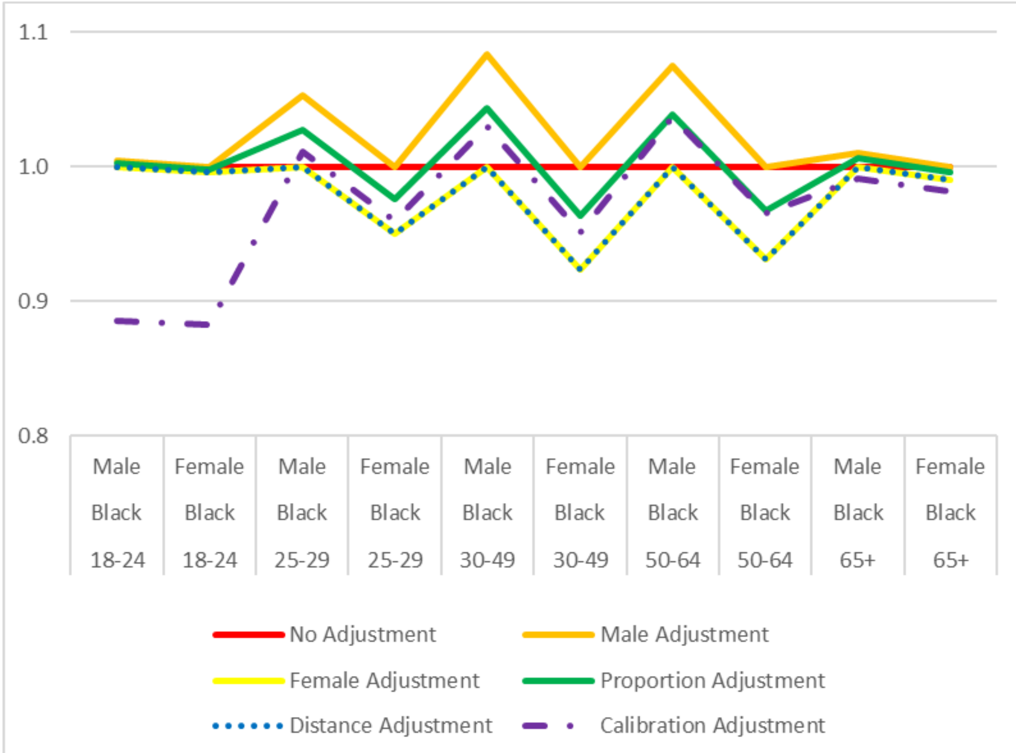


Figure 1: Adjustment Factor for Black Population

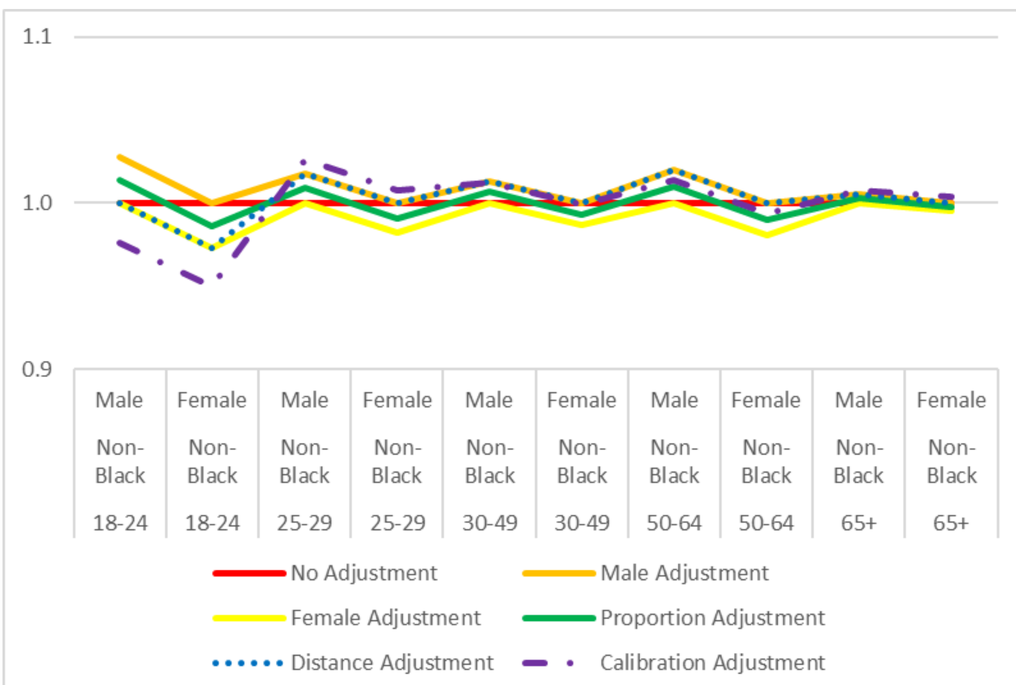
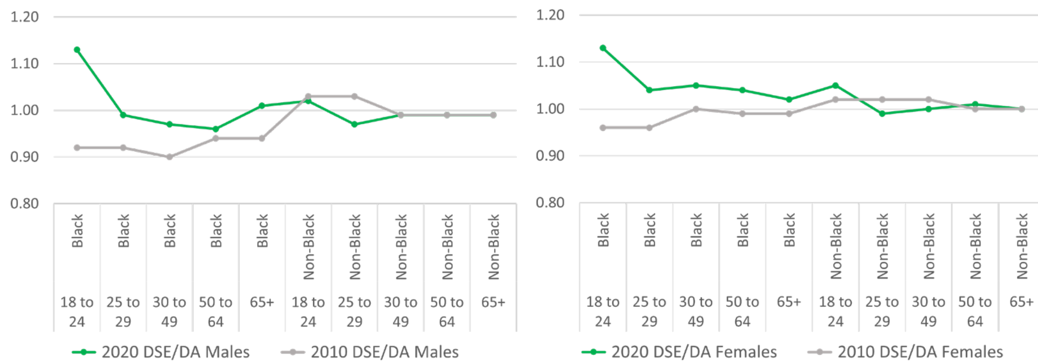


Figure 2: Adjustment Factor for Non-Black Population

The choice of adjustment factor is motivated in part by comparing the unadjusted PES estimates to DA estimates by sex. Figure 3 shows the ratio of DSE and the DA estimate for both 2020 PES and 2010 CCM. If the ratio is greater than 1, the unadjusted DSE was larger than the DA estimate. If the ratio is less than 1, the DA estimate was larger than the DSE. In the previous decade, male DSEs were often lower than male DA estimates, particularly for Black males. In 2020, the DSEs were closer to the DA estimates than in 2010, and in some cases the 2020 DSEs even exceeded the DA estimates. For Black females, the 2020 DSEs exceeded the DA estimates whereas the 2010 DSEs were at or below the DA estimates. The relatively high male and female unadjusted DSE estimates compared to DA indicated maybe something else occurring apart from correlation bias in 2020.



2010 data source: Konicki (2012).

Figure 3: Male (left) and Female (right) Adult Unadjusted DSE/DA Ratio

Based on this information, we concluded that the proportion adjustment method was the most reasonable option. We did not see evidence of correlation bias for adult male estimates. Additionally, this option adjusted to the DA sex ratio without changing the overall DSE. With the challenges of COVID-19, there was more missing data making it more difficult to match, so retaining the national DSE total made sense.

Table 7 displays the percent net coverage error in 2020 for specific domains for the different adjustment options described in Section 5. The adjustment used in our published results (Khubba et al., 2022) is the proportion adjustment. If we had used the same adjustment that was used in 2010, PES would have estimated a statistically significant net undercount of 1 percent overall. The proportion adjustment gives the same percent net coverage error as having no adjustment. The proportion adjustment, the selected method, leads us to conclude that there is a significant undercount of males 18-29, and no significant undercount of females 30-49. These would not be the same conclusions we would draw if we had not made any adjustments.

Table 7. Percent Net Coverage Error for Specific Domains

Adjustment Option	0 None	1 Male	2 Female	3 Proportion	4 Distance	5 Calibration[a]
Domain						
U.S. Total	-0.2	*-1.0	*0.6	**-0.2**	-0.2	-0.3
Non-Black	0.3	-0.3	*0.9	**0.3**	0.0	0.8
Black alone or in combination	*-3.3	*-5.0	*-1.3	***-3.3**	-1.3	-6.7
18-29 Males	-1.1	*-3.4	-1.1	***-2.3**	*-1.7	0.1
18-29 Females	*-2.1	*-2.1	0.2	***-1.0**	-0.5	1.3
30-49 Males	*-1.9	*-4.1	*-1.9	***-3.1**	*-3.0	-3.2
30-49 Females	*-1.0	*-1.0	*1.3	**0.1**	0.2	-0.2
50+ Males	*1.5	-0.2	*1.5	***0.6**	0.3	0.2
50+ Females	*1.8	*1.8	*3.5	***2.6**	*2.4	2.2

** Denotes a (percent) net coverage error that is significantly different from zero.*

[a] *Calibration Adjustment estimates are the DA published estimates or are calculated from DA and Census P.L. 94-171 published estimates. Estimates from DA do not include significance tests.*

7. Future Work

This research found that the female DSEs were larger than they have been in the past relative to males, as well as larger than 2020 DA female estimates. This motivated the decision to move away from the previous correlation bias adjustment that solely inflated adult male estimates to achieve the DA sex ratio. This increase in female estimates could have been caused in part by complications with PES matching operations due to COVID-19. Further work is needed to explore why this population had a high estimate.

Our research only focused on correlation bias for Black and non-Black groups for the adult population 18 years or older. This research did not find correlation bias for young children based on DA sex ratios. However, this population is historically undercounted in both the census and the PES. Further research could be done to analyze the bias of the population under 18 years of age. Additionally, with the ever-increasing availability of DA data for the Hispanic population, it could be beneficial to analyze this group separately from the Black and non-Black populations.

We could also compare PES estimates to administrative records sources. DA produces estimates by a limited number of characteristics (i.e., age, race, sex). Adjusting DSEs using DA estimates could drive up errors for other groups not accounted for in DA. Using additional information could help detect and measure correlation bias better.

References

Alho, J., M. Mulry, K. Wurdeman, and J. Kim (1993), "Estimating Heterogeneity in the Probabilities of Enumeration for Dual-System Estimation," *Journal of the American Statistical Association*, Vol. 88, No. 423, 1130-1136.

Bell, W. (1993), "Using Information from Demographic Analysis in Post-Enumeration

Survey Estimation," *Journal of the American Statistical Association*, Vol. 88, No. 423, 1106-1118.
Bell, W. (1999), "Accuracy and Coverage Evaluation Survey: Combining Logistic Regression DSEs (Target Model) with 1990 Demographic Analysis Results," DSSD Census 2000 Procedures and Operations Memorandum Series Q-12.
Bell, W. (2001), "Accuracy and Coverage Evaluation: Correlation Bias," DSSD Census 2000 Procedures and Operations Memorandum Series B-12.
Griffin, R. (2000), "Accuracy and Coverage Evaluation Survey: Definition and Explanation of Correlation and Related Heterogeneity Bias," DSSD Census 2000 Procedures and Operations Memorandum Series #Q-35.
Heim, K. (2022), "2020 Post-Enumeration Survey Estimation Methods: Net Coverage Estimation," DSSD 2020 Post-Enumeration Survey Memorandum Series #2020-J-07.
Hogan, H. (1993), "1990 Post-Enumeration Survey: Operation and Results," *Journal of the American Statistical Association,* Vol. 88, No. 423, 1047–1060.
Jensen, E., Knapp, A., King, H., Armstrong, D., Johnson, S., Sink, L., and Miller, E. (2020), "Methodology of the 2020 Demographic Analysis Estimates." https://www2.census.gov/programs-surveys/popest/technical-documentation/methodology/2020da_methodology.pdf
Khubba, S., J. Hong, and K. Heim (2022), "2020 Post-Enumeration Survey Estimation Report: National Census Coverage Results for People in the United States by Demographic Characteristics," DSSD 2020 Post-Enumeration Survey Memorandum Series #2020-G-01, U.S. Census Bureau.
Konicki, S. (2012), "2010 Census Coverage Measurement Estimation Report: Adjustment for Correlation Bias," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-G-11. https://www2.census.gov/programs-surveys/decennial/2010/technical-documentation/methodology/g-series/g11.pdf
Mule, T. (2008), "2010 Census Coverage Measurement Estimation Methodology," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-E-18. https://www2.census.gov/programs-surveys/decennial/2010/technical-documentation/methodology/ccm-workshop/2010-e-18.pdf
Olson, D. (2008), "Correlation Bias Adjustment by Individual Year of Age," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-E-19. https://www2.census.gov/programs-surveys/decennial/2010/technical-documentation/methodology/ccm-workshop/2010-e-19.pdf
Olson, D. and R. Sands (2012), "2010 Census Coverage Measurement Estimation Report: Net Coverage Comparison with Post-stratification," DSSD 2010 Census Coverage Measurement Memorandum Series #2010-G-12, U.S. Census Bureau.
Shores, R. (2002), "A.C.E. Revision II: Adjustment for Correlation Bias," DSSD A.C.E. Revision II Memorandum Series #PP-53. https://www2.census.gov/programs-surveys/decennial/2000/program-management/5-review/ace/pp-53r.pdf
Wolter, K.M. (1986), "Some Coverage Error Models for Census Data," *Journal of the American Statistical Association*, Vol. 81, 1986, 338-353.
U.S. Census Bureau (2004), Accuracy and Coverage Evaluation of Census 2000: Design and Methodology, Washington, D.C.
Zamora, J. (2022), "2020 Post-Enumeration Survey: Estimation Design," DSSD 2020 Post-Enumeration Survey Memorandum Series #2020-J-03, U.S. Census Bureau.