Nonresponse in the 2020 Post-Enumeration Survey (PES)[1]

Mark Jost*, Shadie Khubba*, Courtney Hill*
*Decennial Statistical Studies Division, U.S. Census Bureau, Washington, DC, 20233

Abstract
The U.S. Census Bureau conducted the Post-Enumeration Survey (PES) to estimate the coverage of the 2020 Census. Due in part to the COVID-19 pandemic, the PES initially faced higher than expected nonresponse rates from housing units as well as a higher-than-expected rate of item nonresponse. To increase the overall response rate of the survey, a second wave of interviewing referred to as the PES Reopen was performed. In addition, the PES included a nonresponse adjustment to mitigate potential bias in survey estimates resulting from differing characteristics between responding and nonresponding housing units and applied imputation methods to address item nonresponse. This paper provides some background on the 2020 PES relevant to item and unit nonresponse, presents a breakdown of 2020 PES response rates, and examines some of the evidence for nonresponse bias in survey estimates.

Key Words: nonresponse bias, unit nonresponse, 2020 Post-Enumeration Survey, Census Bureau

1. Introduction

The U.S. Census Bureau conducted the 2020 Post-Enumeration Survey (PES) to estimate the coverage[2] of the 2020 Census. This survey involved numerous data collection efforts that were independent of the Census as well as several operations to follow up on census enumerations. Although great efforts were taken to get complete responses during all the data collection and follow-up efforts, including an additional interview stage referred to here as the PES Reopen, the 2020 PES had a higher degree of both unit and item nonresponse compared to the 2010 post-enumeration survey. Unit nonresponse for the purposes of this paper is defined as a case when a sampled housing unit did not adequately respond to the survey[3]. Item nonresponse occurred when certain questions were left unanswered. The increased levels of unit nonresponse were presumed to be related to the

[1] This paper is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors 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-FY23-008, CBDRB-FY22-347, CBDRB-FY22-216, and CBDRB-FY22-244)
[2] See Marra and Kennel (2022) for details on the target population of the PES. Components of coverage are out-of-scope for this paper.
[3] Housing units that were unintentionally left out of the frame were thought of as coverage error, not nonresponse.

COVID-19 pandemic and/or the recent trend of declining survey response (Marra and Khubba, 2022).

Nonresponse bias is often defined as

$$B(\bar{y}_r) = \bar{y}_r - \bar{y}_t = \left(\frac{n_{nr}}{n}\right)(\bar{y}_r - \bar{y}_{nr}) \tag{1}$$

where:
$\bar{y}_t$ = the mean based on all sample cases;
$\bar{y}_r$ = the mean based on only respondent cases;
$\bar{y}_{nr}$ = the mean based on only nonrespondent cases;
n = the number of cases in the sample; and
n_{nr} = the number of nonrespondent cases.

The Office of Management and Budget guidelines suggest a nonresponse bias analysis for any survey with a unit response rate of 80 percent or less (Office of Management and Budget, 2006). The 2020 PES had a higher unit response rate than this threshold. However, unlike in other surveys, PES item completeness plays a substantial role in generating key survey estimates. For the PES estimation process, matching people from the PES to the Census is needed. Item nonresponse would affect the ability of the PES to link to the Census, impacting the PES estimates. Thus, given the increased level of both item and unit data missingness in the survey, a nonresponse bias analysis was merited. Another reason such an analysis was necessary for the 2020 PES is the degree of the nonresponse compared to previous post-enumeration surveys. The equivalent survey in 2010, the 2010 Census Coverage Measurement (CCM), had a response rate of 96.6 percent for the U.S. The 2020 PES had a response rate of 83.2 percent for the U.S. Item nonresponse was also substantially higher compared to 2010.

Nonresponse bias can be thought of as a function of three factors: the response rate, differences between respondents and nonrespondents, and the relationship between these differences and the survey outcome of interest (Andridge and Little, 2011). The purpose of any nonresponse bias analysis is to determine if bias occurred and if so, to understand the impact of nonresponse on a survey's estimates. Typical nonresponse bias analyses use a combination of qualitative methods such as benchmarking, evaluating frame variables of respondents and nonrespondents, weighted response rates for groups, and analyzing characteristics of late responders (Mattingly, 2013). We implement a number of these methods, with some exclusions driven by PES design. For example, calibration to an external source is not feasible since the purpose of the PES is to be a benchmark. Some methods that directly estimate nonresponse bias require a rich set of auxiliary variables.

The scope of this paper is the 2020 PES in the U.S. which refers to all 50 states and D.C. The 2020 PES in Puerto Rico is not discussed in this paper.

This paper will lay out some of the theory underlying the PES estimation process in the context of nonresponse, present tables on unit and item nonresponse in the PES and examine some of the evidence for P-sample unit nonresponse bias. Item nonresponse bias is not within the scope of this paper. Section 2 gives an overview of the PES design. Section 3 describes the overall response rates. Section 4 explores the differences between respondents and nonrespondents. Section 5 provides a discussion on the presence of bias and Section 6 contains our conclusions.

2. Overview of PES Design and Framework for Nonresponse Bias Analysis

The PES used the method of dual-system estimation to measure the coverage of the 2020 Census (Marra and Kennel, 2022). The first system in this context is the set of correct census enumerations. Correctness is estimated using the E (enumeration) sample, a sample of census enumerations from the same geographic areas as the PES sample. The second is an independent sample derived from PES operations called the P (population) sample. By determining which of the census enumerations were correct, and by accurately matching people from the P sample to the census, an estimate of the total population was produced. The classic dual-system estimator takes the form $N_{+1}N_{1+}/N_{11}$ and produces the estimate of the population total $\widehat{N}_{++}$, where N_{11} is the count of observations enumerated or captured in both systems and N_{+1} and N_{+1} are the counts of the observations captured in each of the systems (Hogan, 1992).

This estimator relies on an assumption of independence between the census and the PES error mechanisms for unbiasedness[4]. The assumption can fail in the cases of causal dependence and heterogeneous dependence. Causal dependence is the condition where inclusion in one system is not independent from inclusion in the other. This can be expressed as when $p_{i,ab} = p_{i,a+}p_{i,+b}$ does not hold for every $i = 1, \dots, N$, where $a, b \in (1,2)$, a or b equaling 1 represents an enumeration by the given system, a or b equaling 2 represents a miss by the given system, $p_{i,a+}$ and $p_{i,+b}$ are the marginal probabilities for either system[5], and N is the total number of individuals in the population (Wolter, 1986). One example of causal dependence is the case where individuals failed to respond to the PES, thinking they "helped enough" after responding to the census (Griffin, 2000). Heterogeneous dependence occurs when the marginal system enumeration probabilities are correlated, i.e., when $\sigma(p_{1+}, p_{+1})$ is nonzero[6] (Wolter, 1986). A sufficient condition for heterogeneous independence is homogeneity of capture probabilities within either the PES or census (Mulry et al., 1991).

The PES emphasized operational independence from the census. This involved constructing an independent list of addresses in sample areas and taking measures such as restricting census data from PES staff and conducting PES interviews after census operations in the area were complete (Zamora, 2022). The aim was that independence between census and PES operations would promote causal independence (Bell, 2008). The problem of heterogeneity in capture probabilities has been handled in prior post-enumeration surveys by using post-stratification, i.e., constructing cells of observations that are homogeneous with respect to measured characteristics such as race, age, sex, tenure, etc., calculating the dual-system estimate $\widehat{N}_{++}$ within cell, and aggregating for specific domains (Hogan, 1992).

Nonresponse bias stemming from P-sample unit nonresponse within the context of the post-stratification case can be thought of as a failure of the independence assumption. One potential driver of this bias is causal dependence. If, for example, there are nonrespondents to the PES that were also likely to not respond to the census within a given post-stratum,

---

[4] Assuming no ratio bias exists. See Marra and Kennel (2022) for a brief discussion of ratio bias.

[5] E.g., p_{i1+} and p_{i+1} are the marginal enumeration probabilities for each system, the P-sample and the census correct enumerations.

[6] $\sigma(p_{1+}, p_{+1}) = \frac{1}{N}\sum(p_{i,1+} - \overline{p_{1+}})(p_{i,+1} - \overline{p_{+1}})$.

the dual-system estimate will likely be biased in that stratum. Heterogeneous dependence can also drive bias in the context of nonresponse. For example, if there is a subpopulation b within post-stratum j that has a lower expected probability of being captured by the PES and the census compared to the rest of the stratum, heterogeneous independence is violated within j. If the nonrespondents are clustered within subpopulation b, heterogeneity bias can be exacerbated by nonresponse, leading to nonresponse bias.

Depending on the underlying census capture rates of the subpopulations that nonrespondents feature most heavily in, heterogeneity bias can be increased or reduced. In general, greater disparities between respondents and nonrespondents with respect to their underlying census capture rates and related characteristics increase the impact of nonresponse, raising the potential for additional bias. Further, violations of independence will have a smaller bias on the dual-system estimates if the two systems have high capture probabilities than if the two systems have low capture probabilities.

The 2020 PES used a modified form of the classic dual-system estimator

$$\sum_{i \in C} \hat{\pi}_{dd} \frac{\hat{\pi}_{ce}}{\hat{\pi}_m} \tag{2}$$

where $\hat{\pi}_m$ is the estimated "match probability", or the likelihood of a record matching from the P sample to the census, $\hat{\pi}_{ce}$ is the estimated correct enumeration probability, which is the likelihood of a census record being a real person and not a fictitious, duplicate, or out of scope record, $\hat{\pi}_{dd}$ is the estimated probability of being data-defined, and C is the universe of Census enumerations. The match probability is estimated by a logistic regression model using P sample records, and the correct enumeration probability by a logistic regression model using E sample records. See Zamora (2022) for more information on this estimator. The same general assumptions about independence should hold for this estimator to be unbiased.

The heterogeneous independence assumption takes a slightly different form when a logistic regression estimator is used. Since census inclusion probabilities, as measured in this case by the estimated match rate, are modeled at the individual level, the wording of the assumption shifts to one centered on model correctness. In particular, the specification of the model should be correct, and the model should adequately explain heterogeneity of census inclusion probabilities as measured by the match rate. Any unobserved variables excluded from the match model that are correlated to the match rate will lead to bias in the dual-system estimates produced (Alho, 1990; Chen and Tang, 2011). Causal dependence can also impact the estimates in the same way as in the post-stratification case.

Unit nonresponse in the PES P sample was addressed through a nonresponse weighting adjustment. This propensity stratification adjustment was used to transfer weights from nonresponding housing units to responding housing units with similar characteristics to minimize the potential for nonresponse bias. The response propensities were modeled using logistic regression on frame variables, which are auxiliary variables available for both respondents and nonrespondents. For more information on the nonresponse adjustment, see Beaghen et al., (2022). P-sample and E-sample item nonresponse could also lead to bias if the imputation models used to handle these cases were misspecified. While we list item response rates to give an idea of the levels of missingness which can end up impacting the matching operations and thus the main survey estimates, we focus on unit nonresponse in the P sample for this paper.

The concept of unit nonresponse did not apply to the E sample, which is a sample of census enumerations. Unlike in the P sample, insufficient information cases are kept in sample and treated as erroneous enumerations. Whole person imputations, which are similar to insufficient information cases, but have even less information recorded, are excluded from the E sample. There are noninterviews for E-sample cases, since information on E-sample nonmatch cases was collected in Person Interview and follow-up operations. However, these cases were not considered as a type of unit nonresponse, since a noninterview for an E-sample observation only results in a missing enumeration status. Thus, the E sample was only subject to item nonresponse (Beaghen et al., 2022).

3. Response Rates

3.1 Unit Response Rates

A unit response rate is the number of households that responded to a survey divided by the number of eligible households in the sample. The American Association for Public Opinion Research (2016) defines different ways to calculate unit response rates. The unit response rate for households we used from the American Association for Public Opinion Research (2016) is

$$RR6 = \frac{I + P}{(I + P) + (R + NC + O)} \tag{3}$$

where,
I is a complete interview.
P is a partial interview.
R is a refusal.
NC is a non-contact.
O is other.

For the PES, a complete interview is when the interview is finished. Every relevant question was asked during the interview, though not every question was necessarily answered. A partial interview is when only a portion of the questions were asked during the interview. For PES estimation, there was a stricter definition for a response than just the definitions for complete and partial interviews. For a response in PES estimation, at least one person in the interviewed household had to have sufficient information for matching between the PES and Census and the interviewed household could not have been full of fictitious people or duplicates. We call this stricter definition the final response status in this paper and use it for all P sample unit response rates. We considered unweighted unit response rates in this paper.

The PES had a lower response rate than past decades. The interview period started in September 2020, when the COVID-19 pandemic had an impact on the public's safety and health. The pandemic seemed to have an impact on peoples' comfort to conduct in-person interviews. Some areas were on lockdown throughout all of the planned PES interview schedule because of health and safety concerns. Because of the lower than expected response rate, the PES conducted an additional interview stage called the PES Reopen. PES interviewers were sent back into the field to conduct interviews for nonresponding units, including areas that were previously locked down. The PES Reopen improved the unit response rate for the PES (increased the response rate from 71.9 percent to 83.2 percent). All following tables include the PES Reopen.

Table 1 shows the unweighted unit response rates for the final response status of occupied households. The unit response rate as mentioned previously is 83.2 percent. It also shows the different types of nonresponses: when an interview was not conducted and when an insufficient interview was converted to a nonresponse. The interview was not conducted if a respondent could not be contacted, the interview was refused, or there was a language barrier. This resulted in a nonresponse rate of 3.6 percent. If an interview was conducted but was converted to a noninterview because of the stricter definition for estimation, the interview was deemed not sufficient and resulted in 13.5 percent of nonresponse. These breakdowns of nonresponse show the impact of item nonresponse on the PES unit response rate. If not for the item nonresponse that occurred for the interviews that were not sufficient, the nonresponse would have just been when an interview was not conducted (3.6 percent) and when a household was full of fictitious and duplicates (1.2 percent). The response rate would have included the interview and whole-household insufficient households and been calculated to be 95.2 percent.

Appendix Table 1 has the Person Interview outcome from field by the final response status from PES estimation for the U.S. for occupied households. The field interview status is broken down by type of respondent (household member or proxy) and completion status (complete or partial). It shows cases that were interviews in the field and were reclassified in PES estimation as nonresponses because of estimation's stricter definition of response. If all field interviews had sufficient information and it was determined that the household was not full of fictitious people or duplicates, then they would have all resulted in PES final responses. The occurrence of item nonresponse was the primary reason why some field interviews were defined as nonresponses in PES estimation.

Table 1: Person Interview Response Rates of Occupied Housing Units (Unweighted)

Final Response Status	Frequency	Percent
Total	137,000	100.0
Interview	114,000	83.2
Complete	101,000	73.7
Partial	12,500	9.1
Noninterview	23,000	16.8
Interview not conducted	4,900	3.6
Interview not sufficient	18,500	13.5
Whole-household insufficient	16,500	12.0
Other[1]	1,700	1.2

1 Other is households with a combination of fictitious, duplicate, or insufficient information.

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey.

3.2 Item Response Rates

In addition to unit nonresponse, item nonresponse impacts dual-system estimation. Equation (2) shows the dual-system estimate relies on the match rate of the P sample and correct enumeration rate of the E sample. Before these rates are calculated, P sample people need to be matched to Census people. Items such as age, sex, and other characteristics are instrumental in being able to accurately match the people in the P and E sample. If item nonresponse rates for these matching characteristics are large, we cannot accurately match as many people in the P sample to the E sample. This impacts the match status. The match status defines whether the P sample person matched to a correct person in the Census in the appropriate search area. The match status is used in the calculation of the match rate in the dual-system estimate. P sample people that we could not determine if they matched to an E sample person have an unresolved (missing) match status. This is another form of item nonresponse. See Beaghen (2022) for more detailed analysis of the missing data.

We also need to consider item nonresponse for the E sample. An enumeration status defines whether the E sample person was enumerated correctly. The enumeration status is used in the calculation of the correct enumeration rate in the dual-system estimate. E sample people that we could not determine if they were correctly or erroneously enumerated have an unresolved (missing) enumeration status.

Table 2 provides the item nonresponse rates for some characteristics used in matching for the 2020 PES and 2010 CCM. Item nonresponse rates were calculated for responding people in households. For the P sample, the universe omits person records with insufficient information for survey processing (Phan and Lawrence 2022). For the E sample, the universe omits person records with all characteristics imputed. Item nonresponse rates are unweighted. There are higher item nonresponse rates for matching characteristics in 2020 than 2010.

Table 2: Item Nonresponse Rates for Matching Characteristics

Matching Characteristics	2020 PES P-Sample	2010 CCM P-Sample	2020 PES E-Sample	2010 CCM E-Sample
Sex missing	1.7	0.9	3.0	1.5
Age missing	11.4	4.5	8.3	4.9

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (March 2022 release) and 2010 Census Coverage Measurement Survey.

Appendix Table 2 shows the item nonresponse rates for the match status of P sample people in the U.S. for the 2020 PES and 2010 CCM. Appendix Table 3 shows the E-sample enumeration status item nonresponse rates for people in the U.S. for the 2020 PES and 2010 CCM. Both tables show higher levels of unresolved statuses in the 2020 PES than the 2010 CCM. This item nonresponse will be imputed before the statuses are used in the calculation of the dual-system estimate. Imputing a large number of people puts more emphasis on the models used in the imputation.

4. Differences Between Respondents and Nonrespondents

4.1 Response Rates by Variable Class

Differential response by variable class typically suggests that there could be nonresponse bias if those variables are correlated to survey estimates. No differential response is not necessarily indicative of no nonresponse bias. Response rates are calculated for multiple variable classes and are compared. If large differences occur between variable classes, it’s typically suggestive of nonresponse bias although dual-system estimation is resistant to nonresponse bias if the assumption of independence between PES and the Census holds.

Unweighted unit response rates were calculated for the P sample for households. The P-sample unit response rate for households was calculated as the unweighted number of responding households divided by the unweighted number of eligible households.

Table 3 shows the unweighted P-sample unit response rates for households by multiple variable classes for the U.S. The variable classes shown were used as covariates in the noninterview adjustment. Differential response is seen in the first phase sampling strata, housing unit type, and census region for the P sample. For first phase sampling strata, housing units in large non-owner blocks[7] had the lowest response rate. The P sample response rates are lower for the non-owner strata than the owner strata. For housing unit type, multi-unit address dwellings and other dwellings (i.e., tents, boats, etc.) had the lowest response rates. For census region, housing units in the West region also had lower response rates. Appendix Figure 1 shows the unweighted P sample unit response rates for households by state.

Table 3: 2020 PES P Sample Unit Response Rates by Multiple Variable Classes

Variable Class	P sample Unit Response Rate
First Phase Sampling Stratum	
Small Blocks	84.7
Medium Owner Blocks	86.5
Medium Non-Owner Blocks	82.5
Large Owner Blocks	84.4
Large Non-Owner Blocks	78.5
AIR[1] Blocks	80.7
Housing Unit Type	
Single Family	85.2
Multi-Unit Address	77.9
Trailers	83.8
Others	73.7
Census Region	
Northeast	83.2
Midwest	88.3
South	82.0
West	80.6

1 AIR is American Indians living on Reservations.
Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey.

4.2 Late Responder Characteristics

The PES Reopen was an unplanned event that created a new opportunity for analysis. It effectively created two classes of respondents: early and late respondents, i.e., those that were interviewed in the earlier person interview operation, and those that were interviewed in the PES Reopen operation. Because late respondents might be thought to be like nonrespondents, studying the differences between the early and late respondents can yield

[7] To be more precise, the first phase sampling strata was not comprised of 'blocks' but of 'basic collection units.' A basic collection unit was the smallest geographic level for 2020 Census data collection and roughly corresponded to a block.

interesting insights about survey results. If this similarity holds, any differences between the two can be thought of as another indicator of nonresponse bias. An advantage of this analysis is that differences in person-level characteristics can be used, which may give a more in-depth understanding of the nonrespondent set, since information on nonresponding housing units is limited to information about the housing unit and PES operational information. The match rate is of particular importance, as any differences in that rate between respondents and nonrespondents indicates a possible failure of the independence assumption, provided that these differences aren't accounted for in the match model. Of course, there is no guarantee that nonrespondents and late respondents are similar.

Appendix Table 6 shows the estimated results of the analysis. We estimated that late responders tended to belong to the categories of non-Hispanic Black and renter and tended to be younger than the full respondent set. Furthermore, we estimated late respondents also had a substantially lower match rate compared to the full respondent set. If late respondents were similar to nonrespondents, the inference could be made that nonrespondents also tended to have a lower theoretical match rate, which is a notable finding.

4.3 Comparing Frame Variable Distributions

Comparing the distributions of frame variables is one way to check for nonresponse bias (White, 2003). Two stages of weights are mentioned for the comparisons in this section. Base weight refers to the sampling weight before any adjustment from nonresponse or imputation is done. The adjusted weights refer to the sampling weights after the nonresponse adjustment procedure. As previously mentioned, the nonresponse adjustment transferred weight from nonrespondents to similar respondents. The significance tests use a 90 percent confidence interval constructed using 80 replicate samples. Three comparisons of frame variable distributions are performed and analyzed. All comparisons and tables (Appendix Tables 4 and 5[8]) in this section are performed at the household level:

- Respondents using base weights versus the full sample using base weights (outcome of significance test in column "Are the respondents and full sample significantly different?").
- Respondents using base weights vs late respondents using base weights (outcome of significance test in column "Are the respondents and late responders significantly different?").
- Full sample using base weights vs respondents post adjustment using noninterview adjusted weights.

The first two comparisons see if the nonrespondents have different frame variable distributions compared to the full sample (respondents and nonrespondents). If there is a difference, then this is typically considered to be evidence of nonresponse bias. The second comparison requires the assumption that late respondents and nonrespondents have a similar frame variable distribution. This assumption is used to test if nonrespondents have a different distribution compared to the respondents. The third comparison sees if the full sample is represented by the respondents post nonresponse adjustment. Similar frame variable distributions between these groups shows evidence of the differences between the two groups being reduced by the noninterview adjustment. As seen in the analysis below,

---

[8] All comparisons in this section are for areas that did not include American Indian reservations and trust lands. This is due to the fact that during the PI, some reservations areas were not accessible (health and safety concerns being of the COVID-19 pandemic) during the original interview period. Late response in this case was exogenously imposed. Thus, late respondent comparisons are not appropriate for areas with reservations.

these comparisons and formal tests give mixed results. These comparisons are seen as qualitative as opposed to quantitative and are intended to be used with other qualitative results.

Table 4 in the Appendix show the comparisons of categorical frame variables. Housing unit type, sampling strata, and region are the frame variables shown. The point estimate and standard error (SE) are included. The first test compares if the respondents and full sample are statistically significantly different. There is a mix of statistically significant and not statistically significant differences for the levels of the frame variables. There are estimated distributional differences in the proportion of single and multi-family housing units, the housing units in the Midwest, and the medium owner and non-large owner strata. The other variable classes showed no statistically significant differences. Therefore, some frame variable classes have some evidence of nonresponse bias. The second test compares the respondents and the late respondents. There are estimated distributional differences between single family and multi-unit households and the regions besides the Northeast. The final comparison in Appendix Table 4 is between the full sample and the respondents using the adjusted weights (this is not a formal test). The estimated distributional differences between the respondents post adjustment and the full sample are very minor[9]. This suggests the nonresponse adjustment is making the respondents representative of the full sample (including nonrespondents).

A similar analysis was performed on binary and continuous operational and geographic variables; Appendix Table 5 shows the results. The variables listed include the initial match status of the housing unit before the case was sent to the Initial Housing Unit Followup operation, the final initial housing unit match status, and demographic geography-based variables from the ACS Planning Database. The largest estimated differences between the respondent and full sample were in the Percent Black Alone, Percent Hispanic, and Percent Renter Occupied variables. In all cases, these estimated differences were reduced by the adjustment.

5. Discussion

Our analysis did show some of the classic indicators of nonresponse bias: a higher-than-expected level of nonresponse, differential response by variable class, distributional differences between respondents and the full sample, and differences found between late respondents and the full sample. The higher response rates in housing units in the Midwest and the lower response rates among housing units in multi-unit dwellings and large blocks point to a geographic dimension and an urban dwelling dimension. A similar trend was found when looking at distributional differences in continuous frame variables. Nonrespondents tended to come from geographic areas with higher proportions of minorities and renters living in them compared to the full sample, another indicator of an urban dimension. Minority renters have been shown in prior research to be undercounted in the census compared to other groups, raising the question of whether causal dependence is an issue (Hogan, 1992).

The general pattern of nonresponse held when examining the person-level late respondent characteristics. Late respondents' higher proportion of minority and renter status again

[9] The distributions of region between the full sample and the respondents post adjustment are identical. This is due to the fact that state (which is used to create region) is included in the post stratification for areas with no American Indian reservations.

point to an urban dimension. In the group of late respondents, we get direct confirmation of a lower census capture rate compared to the full set of respondents. This is significant because differential census capture probabilities for nonrespondents increase potential for nonresponse bias if heterogeneous independence does not hold. Although late respondents are in general thought to be similar to nonrespondents, this assumption is hard to verify since person-level data on nonrespondents is not available. The mechanisms of nonresponse for the two could have been impacted by, for example, the time-varying factors of the pandemic in different ways, or they could simply be comprised of distinct subpopulations. Frame variable distributions, though limited in scope, seem to suggest nonrespondents and late respondents do have a degree of similarity.

Despite the differences being observed between respondents and nonrespondents using the base weight, we are unable to draw any strong conclusions about the existence of nonresponse bias. First, the differences observed between respondents and nonrespondents were relatively small in magnitude and tended to be reduced after the nonresponse adjustment was applied. This was expected, since the propensity model in the adjustment explicitly included most of the frame variables used in our analysis in addition to modeling local geographic effects. The hope with this adjustment was that by bringing the respondent sample to be more in line with the full sample of respondents and nonrespondents in terms of their frame characteristics, this will also adjust for differences in observed and unobserved person-level characteristics, including the match rate. This would be expected to reduce the potential for bias resulting from nonresponse.

Further, the match model included person-level covariates that seemed to adequately capture the geographic and urban dimension of the difference observed across respondent sets. The model included terms for age, sex, race/Hispanic origin, geographic area, family relationship, and tenure, among others, as well as various interaction terms. Variable selection was a combination of subjective assessment and model fit. There were some limitations. For example, variables that were included in the match model also had to be used in the correct enumeration model, ruling out variables that exist only in either the P sample or the E sample. For more information on the match model, please see the PES net coverage estimation and design memos (Heim, 2022; Zamora, 2022).

Most importantly, if independence between the PES and the census holds, nonresponse, which is a subset of those people missed by the P sample, will not lead to bias. This property implies that the PES estimation procedure used is relatively robust to nonresponse bias if independence generally holds. Even if independence is violated, the presence of nonresponse bias is not a foregone conclusion. For example, depending on the census capture rates of survey nonrespondents, heterogeneity bias can actually be reduced by nonresponse in certain cases. Thus, it is difficult to ascertain with the usual methods of nonresponse bias analysis whether our estimation procedure and adjustments were robust to nonresponse bias, or whether both the degree of nonresponse and the differences between respondents and nonrespondents put too much strain on our independence assumption, resulting in additional bias.

A number of possible sources of nonresponse bias based on unmeasured characteristics, which would be assumed to be less accounted for in our estimation procedures, are speculated. These characteristics would likely have to be uncorrelated to both the match model covariates and the frame variables used in the nonresponse adjustment as well as correlated to P-sample response propensities and census inclusion probabilities to cause appreciable bias. Possible targets are characteristics such as political orientation, income level, and time-varying factors correlated to census capture probabilities. Any data collected that shed further light onto both respondents and nonrespondents beyond the current set of collected data could be useful in helping to reduce nonresponse bias.

6. Conclusion

The 2020 PES had a higher level of unit and item nonresponse compared to previous post-enumeration surveys. This was in part due to the COVID-19 pandemic and the public's concern with having in-person interviews. The higher response rates among housing units in the Midwest and the lower response rates among housing units in multi-unit dwellings and large non-owner blocks point to a geographic dimension and an urban dwelling dimension of nonresponse. Late respondents were more often made up of renters and minority groups and had a lower match rate compared to early respondents. If assumptions of causal independence and heterogeneous independence between the PES and the census hold, then these observed differences will not contribute to nonresponse bias in our dual-system estimates. The PES tried to ensure operational independence between the census and the survey and properly account for heterogeneity in census capture probabilities in its estimation procedure. In addition, a nonresponse adjustment was applied to limit the potential for nonresponse bias. While we can measure differences between respondents and nonrespondents, it is difficult to speculate, given the methods used for this analysis, about the direction and magnitude of nonresponse bias. This is because its existence is directly linked to the validity of the independence assumption. However, we generally assume the potential for nonresponse bias to be low given the robust PES estimation procedure used.

A challenge to any nonresponse bias analysis is the lack of information on nonrespondents. Subsequent post-enumeration surveys should collect as much information on people and housing units as is feasible to limit the potential for nonresponse bias and allow for a richer post-survey nonresponse bias analysis. Future research on nonresponse bias in the 2020 PES should attempt to directly estimate nonresponse bias using frame variables in addition to these qualitative analyses presented. One such method involves the use of proxy-pattern mixture models (Andridge and Little, 2011). The PPM method would have to be adapted to take into account the particular estimator used in the PES as well as the survey design. Nonresponse bias analyses provide useful insights and will become more valuable over time with declining survey response.

7. Appendix

Appendix Table 1: Person Interview Outcome by Final Response Status (Unweighted)

Person Interview Outcome	Final Response Status: Response	Final Response Status: Nonresponse	Total
Total	114,000	23,000	137,000
Interview	114,000	18,500	132,000
Complete Interview with HH[1] Respondent	97,000	1,900	99,000
Partial interview with HH[1] Respondent	8,700	7,800	16,500
Complete Interview with Proxy[2] Respondent	4,000	450	4,400
Partial interview with Proxy[2] Respondent	4,000	8,200	12,000
Noninterview	0	4,900	4,900

1 HH is household member.

2 Proxy is a non-household member such as a neighbor or landlord.

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey.

Appendix Table 2: Item Nonresponse Rates for P-Sample Match Status

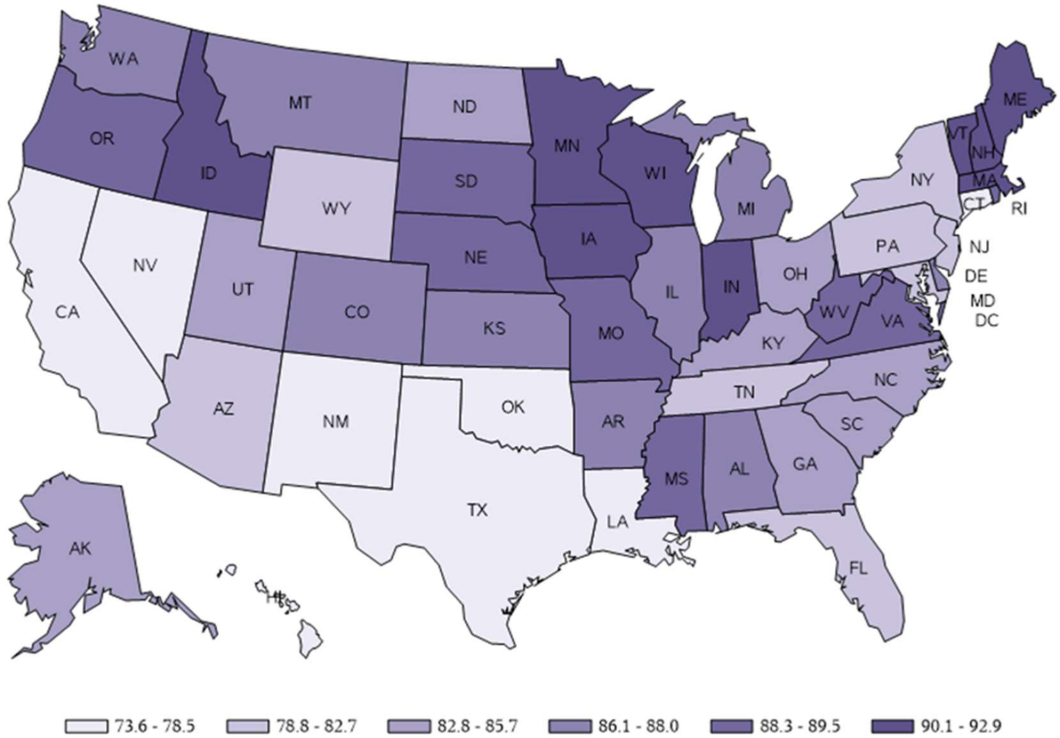
Unresolved Match Status	2020 PES	2010 CCM
Unresolved match status	5.0	1.9
Unresolved nonmover[1] match status	3.4	1.0
Unresolved inmover[2] match status	24.6	11.8

1 Nonmovers were people who lived at the same address on April 1 and months later on the day the Person Interview was conducted.
2 Inmovers were people who moved into the sample address after April 1.
Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release) and 2010 Census Coverage Measurement Survey.

Appendix Table 3: Item Nonresponse Rates for E-Sample Enumeration Status

Survey	Unresolved Enumeration Status
2020 PES	11.6
2010 CCM	4.8

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey (May 2022 release) and 2010 Census Coverage Measurement Survey.



Appendix Figure 1: 2020 PES P sample Unit Response Rate (unweighted)

Appendix Table 4: Comparing Frame Variable Distributions (US, Excluding American Indian reservations and trust lands)

	Respondents (Base Weight)		Respondents and Nonrespondents (Base Weight)		Late Respondents (Base Weight)		Respondents (Adjusted Weight)		Are the respondents and full sample significantly different?	Are the respondents and late respondents significantly different?
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE		
Housing unit type										
Single family	73.13	0.59	71.36	0.63	70.03	1.18	71.27	0.64	yes	yes
Multi-unit address	23.14	0.55	24.93	0.59	26.35	1.14	25.09	0.60	yes	yes
Other	3.72	0.22	3.71	0.21	3.63	0.37	3.64	0.21	no	no
First phase sampling										
Small	0.33	0.07	0.33	0.07	0.24	0.07	0.34	0.09	no	no
Medium Owner	41.41	0.36	40.21	0.37	40.29	1.09	40.00	0.38	yes	no
Medium Non-Owner	10.48	0.20	10.61	0.20	10.54	0.63	10.67	0.21	no	no
Large Owner	28.49	0.45	28.22	0.45	28.30	1.09	28.14	0.45	no	no
Large Non-Owner	19.28	0.35	20.64	0.39	20.63	0.91	20.84	0.40	yes	no
Region										
Northeast	17.55	0.32	17.60	0.35	16.36	0.73	17.60	0.35	no	no
Midwest	22.88	0.36	21.59	0.34	20.70	0.80	21.59	0.34	yes	yes
South	37.59	0.47	38.20	0.49	48.84	0.98	38.20	0.49	no	yes
West	21.98	0.38	22.61	0.38	14.09	0.85	22.61	0.38	no	yes

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey.

Appendix Table 5: Comparing Continuous and Binary Frame Variable Means (US, Excluding American Indian reservations and trust lands)

	Respondents (Base Weight)		Respondents and Nonrespondents (Base Weight)		Late Respondents (Base Weight)		Respondents (Adjusted Weight)		Are the respondents and full sample significantly different?	Are the respondents and late respondents significantly different?
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE		
IHU BFU Match	90.79	0.27	90.21	0.29	90.04	0.8	90.33	0.3	no	no
IHU Final Match	97.86	0.15	97.56	0.17	97.24	0.63	97.73	0.16	no	no
Percent Black Alone	11.12	0.25	11.69	0.25	13.85	0.43	11.71	0.26	no	yes
Percent Hispanic	15.1	0.25	15.75	0.26	16.01	0.55	15.74	0.26	yes	no
Percent Below Poverty	12.69	0.16	12.9	0.17	13.35	0.25	12.93	0.17	no	yes
Percent Renter Occ	33.85	0.33	34.9	0.35	34.91	0.58	34.92	0.35	yes	no
Percent Over 65	16.75	0.17	16.52	0.17	16.01	0.24	16.52	0.17	no	yes
Percent Mobile Homes	5.71	0.17	5.59	0.15	6.45	0.31	5.59	0.16	no	yes
Percent Census Return	68.35	0.17	68.03	0.17	67.27	0.28	68.03	0.17	no	yes

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey.

Appendix Table 6: Comparing Variable Distributions for Late Responders (US, Excluding American Indian reservations and trust lands)

	Respondents (Base Weight)		Late Respondents (Base Weight)		Respondents (Adjusted Weight)		Are the respondents and late respondents significantly different?
	Estimate	SE	Estimate	SE	Estimate	SE	
Male Sex	48.76	0.08	48.31	0.27	48.7	0.08	no
American Indian Nation	0.58	0.03	0.69	0.1	0.6	0.04	no
Hispanic	17.65	0.36	18.15	0.76	18.4	0.38	no
Non-Hispanic Black	11.6	0.28	15.55	0.53	12.16	0.31	yes
Native Hawaiian / Pacific Islander	0.31	0.03	0.21	0.05	0.34	0.03	yes
Asian	5.67	0.23	4.59	0.29	5.93	0.23	yes
White/Other	64.19	0.41	60.8	0.95	62.57	0.42	yes
0-4	5.72	0.08	5.76	0.17	5.69	0.08	no
5-9	6.28	0.08	6.24	0.14	6.26	0.08	no
10-17	10.54	0.09	10.97	0.24	10.49	0.1	no
18-29	14.75	0.16	14.76	0.32	14.81	0.17	no
30-49	25.98	0.13	27.49	0.32	26.22	0.13	yes
50+	36.73	0.29	34.78	0.54	36.53	0.3	yes
Renter	30.73	0.4	34.09	0.9	32.09	0.43	yes
Owner	69.27	0.4	65.91	0.9	67.91	0.43	yes
Match	85.19	0.15	81.88	0.41	85.04	0.16	yes

Source: U.S. Census Bureau, Decennial Statistical Studies Division, 2020 Post-Enumeration Survey.

References

Alho, J. 1990. Logistic Regression in Capture-Recapture Models. In *Biometrics*, Vol. 46, No. 3 (Sep., 1990), pp. 623-63.

Andridge, R., R. Little. 2011. Proxy Pattern Mixture Analysis for Survey Nonresponse. In *Journal of Official Statistics*, Vol. 27, No. 2, 2011, pp. 153–180.

Beaghen, M., R. Turner, M. Jost, and E. Marra, 2020 Post-Enumeration Survey Estimation Methods and Results: Missing Data for Person Estimates, *In 2020 Post-Enumeration Survey Memorandum Series* #2020-J-05, U.S. Census Bureau, 2022.

Bell, R., M. Cohen. 2008. Coverage Measurement in the 2010 Census, Chapter 2. The National Academies Press.

Chen, S., C. Tang. 2011. Properties of Census Dual System Population Size Estimators. In *International Statistical Review*, Volume 79, No. 3, pp. 336-361.

Griffin, R. 2000. Accuracy and Coverage Evaluation Survey: Definition and Explanation of Correlation and Related Heterogeneity Bias. In *DSSD Census 2000 Procedures and Operations Memorandum Series* #Q-35, U.S. Census Bureau.

Heim, K. 2022. 2020 Post-Enumeration Survey Estimation Methods: Net Coverage Estimation, In *DSSD 2020 Post-Enumeration Survey Memorandum Series* #2020-J-07, U.S. Census Bureau.

Hogan, H. 1992. The 1990 Post-Enumeration Survey: An Overview. In *The American Statistician*, Vol. 46, No. 4 (Nov., 1992), pp. 261-269.

Marra, E. and T. Kennel. 2022. Source and Accuracy of the 2020 Post-Enumeration Survey Person Estimates, In *2020 Post-Enumeration Survey Memorandum Series* #2020-J-01, U.S. Census Bureau.

Marra, E., S. Khubba. 2022. COVID-19 Impacts on the Post-Enumeration. In *JSM Proceedings*. Pages forthcoming.

Mattingly, T. 2013. Measuring Nonresponse Bias in the Demographic Surveys," U.S. Census Bureau.

Mulry, M. and B. Spencer. 1991. Total Error in PES Estimates of Population. In *Journal of the American Statistical Association*, Vol. 86, No. 416 (Dec., 1991), pp.839-855.

Office of Management and Budget. 2006. Standards and Guidelines for Statistical Surveys.

Phan, N. and J. Lawrence. 2022. 2020 Post-Enumeration Survey Estimation Methods: Characteristic Editing and Imputation, In *2020 Post-Enumeration Survey Memorandum Series* #2020-J-04, U.S. Census Bureau.

The American Association for Public Opinion Research. 2016. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 9th edition. AAPOR.

White, M. 2010. 2003 National Survey of College Graduates Nonresponse Bias Analysis. In *Proceedings from the Survey Research Methods Section of Joint Statistical Meetings* 4793-4807.

Wolter, K. 1986. Some Coverage Error Models for Census Data. In *Journal of the American Statistical Association*, Vol. 81, No. 394 (Jun., 1986), pp.338-346.

Zamora, J. 2022. 2020 Post-Enumeration Estimation Survey Estimation Methods: 2020 Post-Enumeration Survey Estimation Design, In *DSSD 2020 Post-Enumeration Survey Memorandum Series* #2020-J-03, U.S. Census Bureau.