Usage of Administrative Record Modeling in the United States 2020 U.S. Census

Mary H. Mulry, Thomas Mule, Andrew Keller and Scott M. Konicki
U.S. Census Bureau

Abstract. A major innovation for the 2020 U.S. Census was the use administrative records (ARs) to enumerate some households. Methodology developed at the U.S. Census Bureau used ARs to classify some addresses as Occupied, Vacant, or Nonresidential and to create a roster of the residents with their characteristics for some of the addresses classified as Occupied. The Census Bureau began a research program in 2012 to pursue its goal of identifying a methodology for using administrative records to reduce the cost of the 2020 Census Nonresponse Follow-up (NRFU) operation while preserving data quality. NRFU is the census field operation where enumerators visit addresses that did not submit a self-response and attempt to obtain an interview. This document provides a high-level discussion that includes more details regarding the evolution of the methodology for AR enumeration and its implementation in the 2020 Census. By focusing on data quality during each phase of the research, development, and implementation, the chosen methodology enabled using ARs to enumerate 4.59 percent of the addresses. The investment in AR enumeration proved to be essential to the success of the 2020 Census.

1.Introduction

The 2020 Census is the first U.S. census to use administrative records (ARs) to enumerate some households. The innovation used ARs to classify some addresses as Occupied, Vacant, or Nonresidential and to create a roster of the residents with their characteristics for some of the addresses classified as Occupied. In addition, the AR methodology included assigning each AR roster a quality score that reflected how likely the AR household composition was to agree with what the household would have reported in a census response. Addresses that could not be classified received the designation ‘No Determination” and were retained for further processing in case additional information became available later in the processing operation. The more recent U.S. censuses also have used ARs in some aspects of census-taking, but not for the enumeration of households. Other statistical programs at the U.S. Census Bureau have used and continue to use ARs to support their data collection operations. This paper provides a high-level discussion of the research and methodology for using ARs in the enumeration of households at some addresses in the 2020 Census in the U.S. and some results from the implementation. The term “administrative records” refers to data collected by governmental or nongovernmental agencies while administering a program or service.

The motivation for the development of the methodology for using ARs in the 2020 Census enumeration was to reduce the increasingly high cost of Nonresponse Follow-up (NRFU), which is the census operation where census enumerators contact households that did not submit a census self-response. The NRFU operations in the 2000 Census and the 2010 Census illustrate the rising costs although there were minor differences in the two implementations. For the 2000 Census, the number of addresses in Nonresponse Follow-up was 53.8 million. The total cost of the 2000 Nonresponse Follow-up was $1.374 billion, which included $1.123 billion for the NRFU operation, and the remaining $251 million for a Reinterview operation, a Coverage Improvement operation, and a residual NRFU operation (Walker et al. 2012).

The Nonresponse Follow-up for 2010 Census included 47.2 million addresses that needed to be contacted. The total cost of NRFU was $1.589 billion of which $1.117 billion was for the field data collection and the remaining $454 million covered the costs of training, salaries, mileage, and other miscellaneous expenses (Walker et al. 2012).

Since the number of addresses in the 2000 NRFU differs from the number of addresses in the 2010 NRFU, the best strategy for comparing the cost of the two implementations of NRFU relies on identifying the cost per address in the main field data collection. Both implementations had additional reinterviews and checks, but the differences in the implementations make comparing the cost per case in these operations not possible in a reliable way. For the 2010 NRFU, the cost per address was $33.60 (Walker et al. 2012, p. 219). The cost per address for the 2000 NRFU was $26.09 (Walker et al. 2012, p. 2). However, an adjustment of the 2000 NRFU cost to 2010 dollars is needed for the comparison of the cost per case in the two implementations of NRFU. The Bureau of Labor Statistics Consumer Price Index inflation calculator, which may be found at https://www.bls.gov/data/inflation_calculator.htm, shows that the 2000 NRFU cost per address adjusted into 2010 dollars was $33.04. Taking the difference in the cost per address in the 2000 and 2010 implementations of NRFU shows that the 2010 cost per address was $0.56 higher than the cost per address in the 2000 NRFU when both costs are in 2010 dollars. The next step is to multiply the number of addresses in the 2010 NRFU, which was 47,235,198, by the increase in the cost per case of $0.56, which shows the total increase is the cost of the 2010 NRFU over the cost of the 2000 NRFU was $26,451,711 in 2010 dollars. The large increase in the cost of the

2010 NRFU motivated the Census Bureau to explore whether using administrative records to enumerate some addresses that did not self-respond could be used to avoid or reduce an increase in the cost of NRFU in the 2020 Census.

Advances in computer technology and software have enabled the expanded use of ARs. In addition, the Office of Management and Budget (OMB) has encouraged federal agencies to leverage the ARs collected by operations and services in their statistical programs (U.S. Office of Management and Budget 2014). The advances and support of OMB also have led to the development of agreements between the Census Bureau and other government agencies that permit the use of their ARs for enumeration of households under specified conditions. In the nongovernmental sector, new technology and software have enabled the construction of commercially available databases that contain information that may or may not overlap with data available to the Census Bureau in ARs housed in government agencies. The Census Bureau began a research program in 2012 to pursue its goal of identifying a methodology for using administrative records to reduce the cost of the 2020 Census NRFU operation while preserving data quality. Since NRFU was one of the most expensive operations in the 2010 Census, improving the efficiency of the NRFU operation was viewed as critical to the goal of a cost-effective census in 2020 (U.S. Census Bureau 2019a, 2019b).

The U.S. Census Bureau also attempted to reduce the workload and thereby the cost of the 2020 Census NRFU in several ways. The first strategy was to optimize self-response including via the internet. The 2020 Census was the first U.S. Census that actively encouraged response via the internet. Prior to 2020, only a small unannounced experiment in the 2000 Census offered Internet response where 65,000 households submitted a response (Whitworth 2001). For the 2020 Census, the Census Bureau promoted internet response by mailing a letter to each address with instructions on how to submit a census response using the internet. If a household did not submit a response by internet within a couple of weeks, the Census Bureau mailed a paper version of the census form to the household. In addition, the mailing included a Census form that a household member could complete and return by mail. Also, there was a telephone number that a household member could call and provide a census response. If self-response was not received for an address, then an enumerator visited the address during NRFU and attempted to get an interview.

This document discusses another reduction to the NRFU workload by using ARs in the enumeration of households at some addresses. A brief overview of AR enumeration may be found in the blog post by Mule (2021). In contrast, this document provides a high-level discussion that includes more details regarding the evolution of the methodology for AR enumeration and its implementation in the 2020 Census. Not covered is the use of records maintained by group quarters administrators who use their records to enumerate the people residing in their facilities. For information about the enumeration of group quarters facilities, see Stempowski and Christy (2021) and U.S. Census Bureau (2017). In addition, New Zealand applied a similar distance approach for unresolved addresses in its 2018 Census (Bycroft and Matheson-Dunning, 2020).

The research and development that the Census Bureau conducted from 2012 through 2018 to develop the methodology for using ARs to enumerate the households at some addresses in NRFU is discussed in this document. Also included is a short description of the implementation in the 2020 Census along with some results. A brief overview of previous uses of administrative records at the Census Bureau focuses on historical and recent uses, data quality, and data protection. In addition, there is a description of the statistical models used in the administrative records enumeration process. The original plan for AR enumeration in the 2020 Census required modifications and adaptations to cope with the unforeseen disruptions due to the COVID-19 pandemic causing an unexpected delay in the 2020 Census data collection and processing. A description of these modifications and adaptations is included.

Throughout this document, the descriptions of the research and development include the rationale behind the resulting decisions. By focusing on data quality during each phase of the research, development, and implementation, the chosen methodology was able to employ ARs to improve the quality of the census enumeration. The investment proved fortuitous when the COVID-19 pandemic and a series of natural disasters disrupted traditional NRFU operations. Without the strategic use of administrative records, more addresses would have received a final status of Unresolved which would have led to imputing the status of the address, the household size, and characteristics of the household members.

2. Background

2.1 History of ARs and the U.S. Census

The Census Bureau has a long history with ARs. The first study of consequence occurred during the early 1940s and used the demographic method of comparing aggregated totals. The study focused on a comparison of the number of males of military age in the 1940 Census to the number found in draft registration records using a clerical operation. The findings dispelled the prevailing assumption that the census had better coverage than records systems. The study estimated there were 14.9 percent more Black males of 21 to 35 years of age registered for the draft than were counted in the census and 2.8 percent more non-Black males in the same age category (Price 1947).

This result led to the development of census coverage evaluation methodologies. The first such evaluation method was Demographic Analysis. The estimates produced by Demographic Analysis are a sum of totals for subpopulations based on aggregating ARs from different record sources, such as birth and death records, to form an estimate of the total population that can be compared to the total from a census. The 1950 Census was the first census to have its coverage evaluated using Demographic Analysis (Coale 1955).

Demographic Analysis has been used to evaluate the coverage of every U.S. census at the national level since 1950 and is still used today although the methodology and data sources have improved over the years. The results of the Demographic Analysis estimates developed for evaluating the coverage of the 2020 Census may be found in Jensen et al. (2020). The need for estimates of census coverage for geographic and demographic subgroups led to the development of two other methods, the Reverse Records Check used by Canada (Statistics Canada 2007) and the Post-Enumeration Survey used by the U.S. and several other countries (Mulry 2014).

Over the years, the Census Bureau has conducted matches between different administrative lists and censuses to evaluate coverage and completeness. A few of the earlier studies are listed below:

- An administrative records match (ARM) evaluated the coverage of the 1960 Census for Social Security recipients (Marks & Waksberg 1996). The estimate of the number of Social Security recipients missed was 5.1 to 5.7 percent of those enumerated.
- Another ARM conducted in conjunction with the 1980 Census assessed the feasibility of using the 1979 Internal Revenue Service (IRS) file as a sampling frame for evaluating census coverage (Childers & Hogan, 1983). A sample from the IRS file was matched to the 1980 Census at their address in the IRS file. When a person was not found at the IRS address, the study attempted to trace them through the mail. However, 22 percent of the sample could not be traced, and the study did not make estimates of census undercount.
- An ARM conducted in conjunction with the 1996 Community Census Test focused on determining whether there were people in administrative records who were not listed in a census or a post-enumeration survey. The percentage of people from administrative records who were residents but not enumerated and not on the rosters in the post-enumeration interview ranged from 2.0 to 2.5 across the three sites in the test.

The reader needs to keep in mind that the U.S. does not have a single source of administrative records with high coverage of its entire population. After the 2000 Census, there was an attempt to create a census-like file by merging and unduplicating five federal sources of administrative records, called the Statistical Administrative Records System (StARS) (Leggieri, Pistiner, and Farber 2002). A comparison between the StARS and Census 2000 found that StARS covered 95 percent of the population in Census 2000 (Judson 2000).

The 2010 Census presented an opportunity for further research by creating a census-like administrative records file that merged both federal and commercial data sources and then compared the unduplicated administrative records file to census records. The results of the research, known as the 2010 Census Match Study, showed that 88.6 percent of the 308.7 million 2010 Census enumerations could be matched to an administrative record. The main reason for the low match rate appeared to be the inability to assign unique identification numbers to 9.6 percent of the census person records. The census-like administrative records file had 312.2 million records for unique people, but the study was not able to link 10.7 million to an address on the census file (Rastogi and O'Hara 2012).

2.2. Recent Uses of ARs

The 2010 Census used ARs in several operations. In one, the Census Bureau's Geography Division updated its Master Address File (MAF), which is the list of addresses with living quarters, using a database maintained by the U.S. Postal Service called the Delivery Sequence File (DSF) and five commercially available databases (U.S. Census Bureau 2016). The DSF is the list of residential addresses where the U.S. Postal Service delivers mail and was the primary source used in the update. A second 2010 operation that used ARs was Coverage Follow-up, which contacted households whose census questionnaire had ambiguities to clarify the information (Govern, Coombs and Glorioso 2012). Although the census questionnaires requested a telephone number, some households did not provide a number, and for others, the responses were not valid numbers. The Telephone Lookup operation used two commercially available databases, InfoUSA and QAS (formerly QuickAddress), to find telephone numbers for the households.

Other Census Bureau programs have expanded their use of ARs as computer technology and data processing improved to accommodate storing and manipulating larger data files. Some examples of current programs that use ARs are the Business Register (DeSalvo, Limehouse, Klimek 2016), Intercensal Population Estimates (U.S. Census Bureau 2012), Local Employment Dynamics (Coyle, E. 2019), Small Area Income and Poverty Estimates (U.S. Census Bureau 2019c), and Small Area Health Insurance Estimates (U.S. Census Bureau 2019d).

2.3 AR data quality and protection

2.3.1 AR data quality for AR enumeration

For the 2020 Census, the AR roster was used for enumeration only if its quality score was high and if a self-response was not received for the address or its status was not resolved during one visit of fieldwork enumeration. For the people whose enumerations were based on administrative records, the demographic characteristics of age, sex, race, Hispanic origin, and relationship were obtained from AR sources such as past census responses, the Social Security Numerical Identification (Numident) file, or datafiles from other Census Bureau programs such as the American Community Survey (ACS). For any of the characteristics that could not be assigned directly based on administrative record information, imputation procedures were used (U.S. Census Bureau Administrative Records Modeling Team 2017).

The Census Bureau's policies and procedures emphasize coverage of the population being studied and the accuracy of the data it collects (U.S. Census Bureau 2021a). The entire 2020 Census program was geared toward ensuring that the census numbers provide coverage of the

U.S. population as a whole and within each state. The AR enumeration was designed to ensure that the designation of addresses as Occupied, Vacant, or Nonresidential had a high probability of being accurate, and in doing so, AR enumeration contributed to the accuracy of coverage of the population. In addition, when a household was enumerated using ARs, the operation was required to ensure that there was a high probability that the AR records reflect the number of household members and their characteristics.

In this document, the term “high quality” ARs is used to mean that there is a high probability that the AR status assigned to an address is accurate. The assignment of the Occupied status means that there is a high probability that the address is occupied, and a high probability that the household size, composition, and characteristics are accurate. The assignment of the Vacant status means that there is a high probability that the address has living quarters, but no one resides there, while the Nonresidential status means that there is a high probability that address does not have living quarters.

2.3.2 Data protection

The Census Bureau is required by law to protect the personally identifiable information for a person or business that it acquires for use in the census and other statistical purposes. The Census Bureau is authorized to acquire or purchase records from states and third-party entities under Title 13, United States Code (U.S.C.). Title 13 also sets limitations on the type of data that the Census Bureau is permitted to share and the circumstances under which data can be shared. In particular, the Census Bureau is prohibited from sharing data that would enable identifying individuals, households, or businesses with law enforcement agencies. Other laws that apply to the protection of personal data are the Privacy Act of 1974 and the Confidential Information Protection and Statistical Efficiency Act (CIPSEA). In addition, other federal confidentially statutes permit agencies to share data with the U.S Census Bureau under strict, secure conditions (U.S. Census Bureau 2020a). When an agency agrees to share its data with the Census Bureau, the two parties enter into a binding agreement that specifies the data being shared and how the data may be used by the Census Bureau.

The Census Bureau grants access to personally identifiable information to employees when required for their work and for no other purposes. Every person who works with confidential information collected or acquired by the U.S. Census Bureau is sworn for life to uphold the law and not to illegally disclose any personally identifiable information. Violating these laws is a federal crime with serious penalties, including a federal prison sentence of up to five years, a fine of up to $250,000, or both (U.S Census Bureau 2020a).

Merging AR data from several sources requires linking records across datafiles to avoid having two or more records for the same person in the final file. A concern for this process is that a datafile may have records that need to be removed because they are invalid or out of date. Therefore, the Census Bureau developed the Personal Validation System (PVS) that processes incoming AR datafiles from other government agencies and commercial sources. The PVS has the dual purpose of attempting to validate the name and address combination on each record in incoming datafiles and to replace each person’s address and name with unique anonymized numbers (Wagner and Layne 2014).

When the PVS can identify the geographic location of the address on the internal Master Address File (MAF), the address receives the corresponding MAF identification number (MAFID), which is a unique anonymized address number. When the AR file includes Social Security Numbers (SSNs), the PVS checks the Census Bureau's version of the Social Security Administration Numident file to assure that each SSN is valid. When records in an AR file do not include SSNs, the PVS processing attempts to link the name to other sources and find an SSN. When an SSN can be assigned to the name, the PVS assigns a Protected Identification Key (PIK), which may be viewed as anonymized SSN (Wagner and Layne 2014).

Being able to assign a MAFID to an address, which led to placing it on the MAF, was viewed as confirming the existence of the address and made it eligible to be assigned an AR roster if needed. Similarly, a person who was assigned a PIK became eligible for inclusion in an AR roster at the address where they were found if the address was assigned a MAFID. In addition, MAFIDs and PIKs also were used to aid in identifying duplicate records on data files that were used in AR enumeration.

After the PVS processing, the records in the datafiles produced for statistical analyses contain each person's corresponding MAFID and PIK but *not* the person's name. The correspondence between the address and MAFID and the correspondence between person's name and PIK are retained in a separate file that is stored in a restricted area that is available only to the staff that assigns the MAFIDs and PIKs. In particular, the information is not available to employees assigned to perform the data analyses. However, if a project has a specific need for addresses or names or both, staff must submit a request with a justification to receive approval for use that is restricted to only the specified project and staff.

3. Constructing AR rosters

The first step in AR enumeration was to create AR rosters that could be used to enumerate addresses that were determined to be AR Occupied by using statistical models in a process that is discussed in Section 4. The construction of the AR rosters used the high-quality AR sources listed below in a process that included identifying and removing duplicate records at an address:

- IRS 1040 forms filed for Tax Year 2019
- IRS 1099 forms filed for Tax Year 2019
- Medicare Enrollment Database
- Indian Health Service Patient Database
- Household Composition Key File at the U.S. Census Bureau.

The Household Composition Key File (Deaver 2020) on the list above is a database created and maintained by Census Bureau staff using applications for Social Security Numbers (SSNs) from the Social Security Numerical Identification (Numident) File. The Census Bureau maintains a Census Numident file derived from the Social Security Numident File that contains names and SSNs that are used in assigning PIKs to records to enable child-to-parent linking for children under 18 years of age. A record for a child in the Household Composition Key File includes the names of the child’s mother and father, if available, but not the SSNs of the parents.

Some AR rosters were constructed using only one source while others had AR records that came from multiple sources. A concern about the reliability of using only one AR source in the construction of an AR roster arose during the research to develop the AR enumeration methodology. In response to the concern, the AR rosters received a new requirement, which was that at least one person on the AR roster must be corroborated at the address by two sources. Therefore, a second set of sources was identified to aid in satisfying the corroboration requirement. Uncertainty about the recency of some of the addresses in the second set of sources led to them being considered as good quality for use as corroboration but not of high enough quality for AR roster creation. The sources used in creating the AR rosters could be used to corroborate each other with the exception that records from the same agency could not corroborate each other. For example, IRS 1099s and IRS 1040s could not corroborate each other since they were from the same agency. The sources used only in corroboration are listed below:

- U.S. Postal Service National Change of Address File
- 2000 and/or 2010 Census Unedited File (CUF)
- Medicare Enrollment Database
- U.S. Department of Housing and Urban Development Public and Indian Housing
- U.S. Department of Housing and Urban Development Tenant Rental Assistance Certification System
- U.S. Department of Housing and Urban Development Computerized Home Underwriting Management System
- Indian Health Service Patient Database
- U. S. Department of Housing & Urban Development Public & Indian Tenant Rental Certification System

- Selective Service System Registration file
- Targus Wireless
- Targus Federal Consumer Targus Address File
- Veteran Service Group of Illinois Name and Address Resource file
- Veteran Service Group of Illinois Tracker Plus
- DAR (Data. Analytics. Research) Partners
- U.S. Census Bureau American Community Survey (ACS)

4. Use of statistical models in AR enumeration

The major result from the research program that ran from 2012 to 2018 was a multistep process that relied on statistical models to identify addresses that ARs indicated were occupied and had high-quality ARs to build rosters suitable for AR enumeration (Morris et al. 2016, Keller et al. 2018). This was the approach to NRFU that was viewed as the one most likely to save money and maintain quality when identifying the NRFU addresses with high quality ARs that could be used for enumeration when one contact attempt by a NRFU enumerator did not result in an interview. The identification of addresses that were vacant and nonresidential also used statistical models. The dataset used in fitting all the models covered the U.S.

The methodology chosen for implementation used a sequence of models to assign a status of Vacant, Nonresidential, or Occupied to each address and then constructed an AR roster for the addresses designated as Occupied. The models used data from administrative records and commercial data regarding the address and the household members, if there were any.

The assignment of statuses to addresses involved a sequence of models. The models are listed below in the order in which they were applied.

1) The AR Vacant-AR Nonresidential Model, the first model in the sequence, is a multi-level logistic regression model that that has three levels, which are Vacant, Nonresidential, and Occupied. The model produced an address-level estimate of the probability for each of the three statuses on Census Day.
2) The probabilities produced by the AR Vacant-AR Nonresidential Model were used in two Euclidean distance measures; one was used to assign the status of AR Vacant and the other was used to assign the status of AR Nonresidential.
3) The Person-Place model is a logistic regression model that produced an estimate of the probability that each person on the AR roster was correctly enumerated for each of addresses under consideration.
4) The Household Composition Model is multinomial logistic regression model with eight levels that reflect possible household compositions of adults and children for the address. The model uses variables based on 2010 Census data and administrative records data to produce an estimate of the probability that each household composition agreed with what a Census response would report for the address.
5) The assignment of AR Occupied to an address relied on a Euclidean distance measure that used estimates from the Person-Place Model and the Household Composition Model to produce a single score that indicated whether to assign AR Occupied to the address. Although the Person-Place Model produced a probability for each person, it was summarized to the address-level by taking the lowest estimated person probability for all AR persons on the roster.

The application of the model resulted in the assignment of a probability for each of the three statuses of Vacant, Nonresidential and Occupied. In most cases, the address received the status that received the highest probability of being correct.

4.1 Development of methodology

As part of the preparations for the 2020 Census, Census Bureau researchers pursued developing a method for identifying addresses with high quality AR rosters suitable for enumeration. If a self-response was not submitted for an address, but a high-quality AR roster that listed the residents was available, the address was removed from NRFU fieldwork, and the AR roster was inserted for the enumeration of that address. The initial work on identifying high quality rosters occurred during the 2015 census test and used the predictive modeling in a linear optimization approach to identify high quality AR rosters. The goal was to assign a quality score to each AR roster for use in ranking the AR rosters. The rankings would permit developing a cutoff that identified AR rosters that were suitable for AR enumeration (Morris et al. 2016).

The researchers also used a linear optimization approach to develop models for use in assigning the status of AR Occupied, AR Vacant, or AR Nonresidential to addresses. Later, the researchers were encouraged by outside reviewers to investigate a distance function approach to assigning an AR status to addresses and AR rosters.

Subsequently, the Census Bureau researchers developed a distance function approach and used 2010 Census data to compare the results from the distance function approach to the results from a linear optimization approach. The study found a high level of agreement between the quality assessments of the AR rosters assigned by the two approaches. The results showed that 91 percent of the addresses assigned the status of AR Vacant by the distance function approach also were assigned the AR Vacant status by the linear optimization approach. A similar study that compared the assignments of AR Occupied to the 2010 NRFU status found an agreement rate of 94.4 percent among the 500,000 addresses with the lowest scores, which indicated the highest probability of being occupied. The high agreement rates on the assignments of AR Vacant and AR Occupied provided a measure of validity to the overall approach. The distance function approach also had the advantage of being easier to implement and more flexible if adaptations were needed during the Census. Therefore, the distance function approach was selected for implementation in the Census.

The development of the methodology for AR enumeration produced a screening tool that was flexible enough to permit adjustments of the threshold for identifying AR rosters suitable for enumeration if required by circumstances such as the cost or time constraints to complete NRFU fieldwork. Being able to raise or lower the quality standard for the AR rosters was an advantage of using a distance function. The approach involved creating a housing unit level data set for the addresses that contained variables obtained from administrative records and commercial data sources. The variables included characteristics of each address including whether it was occupied or vacant, characteristics of the household such as the number and age of household members, and whether one or more of the household members was found at the address in administrative records in the current and previous year. Then the model was applied to the data for each address using characteristics of each address to produce a quality score for the AR roster for the address.

These models are described in the next section. A more extensive discussion of the research concerning the models and their variables may be found in Keller, Mule, Morris, and Konicki (2018) and Mulry, Mule, Keller, and Konicki (2021).

4.2 Applying the models

This section provides more information about the application of the various models and measures in this analysis. The variables used in Keller et al. (2018) were also used in the modeling for the AR enumeration discussed in this section.

4.2.1 AR Vacant and AR Nonresidential Model

The identification of AR Vacant and AR Nonresidential addresses was made in a process that considered all NRFU addresses. The first step was to fit a multinomial logistic regression model using data from the 2010 Census as training data to aid in identifying addresses that were AR Vacant and AR Nonresidential. The model produced estimates of the probabilities of an address having each of the three statuses of Occupied, Vacant, or Nonresidential on Census Day. Prior to the 2020 Census, the regression models used for AR Vacant and AR Nonresidential identification were tested and refined by running simulations with the 2010 Census data. The probability of occupied was used in the two distance measures, one that supported the identification of addresses qualifying for the AR Vacant status and the other supporting the identification of addresses qualifying for the AR Nonresidential status. Section 4.2.2 contains a discussion of these two distance measures.

This model, called the Vacant-Nonresidential model, relied heavily on UAA codes assigned by mail carriers for the U.S. Postal Service when they could not deliver mail to an address. Examples of UAA codes are Vacant, No Such Number, and Refused. Additional independent variables included characteristics of the address and its neighborhood from other data sources. One NRFU field visit was required for all addresses classified as AR Vacant or AR

Nonresidential to avoid misclassifying occupied addresses. The field visit was done even when the U.S. Postal Service returned an Undeliverable-As-Addressed (UAA) status for an initial mailing.

When a status could not be assigned, the address received the status of No Determination and was retained in case more data about the address became available for an additional attempt to use the models to assign a status.

4.2.2 Distance Measures for AR Vacant and AR Nonresidential Identifications

The identification of AR Vacant for an address relied on a distance measure that used the vacant probability, $\hat{p}_{h,vac}$, and occupied probability, $\hat{p}_{h,occ}$, estimated via the Vacant-Nonresidential model discussed in Section 4.2.1. These predicted probabilities lie in the interior of the two-dimension simplex $\{(x, y): x \geq 0, y \geq 0, x + y \leq 1\}$. Based on the two probabilities, each address would have a point in this two-dimensional space. The addresses most likely to be vacant would be those that have shortest distance to the point where the occupied probability equals 0 and the vacant probability equals 1 (i.e., the (0,1) point). As a result, the definition of the Euclidean vacant distance, $\delta_{h,vac}$, for each unit *h*, was defined as

$$\delta_{h,vac} = \sqrt{\left(1 - \hat{p}_{h,vac}\right)^2 + \left(\hat{p}_{h,occ}\right)^2} \qquad (1)$$

The formulation of the vacant distance using equal weighting of the vacant and occupied probabilities was adopted after using unequally weighted combinations did not improve model predictions. For the nonresidential identification, the vacancy probability is replaced by the nonresidential probability from the model in Section 2.1. The Euclidean nonresidential distance $\delta_{h,nr}$, for each unit h, is

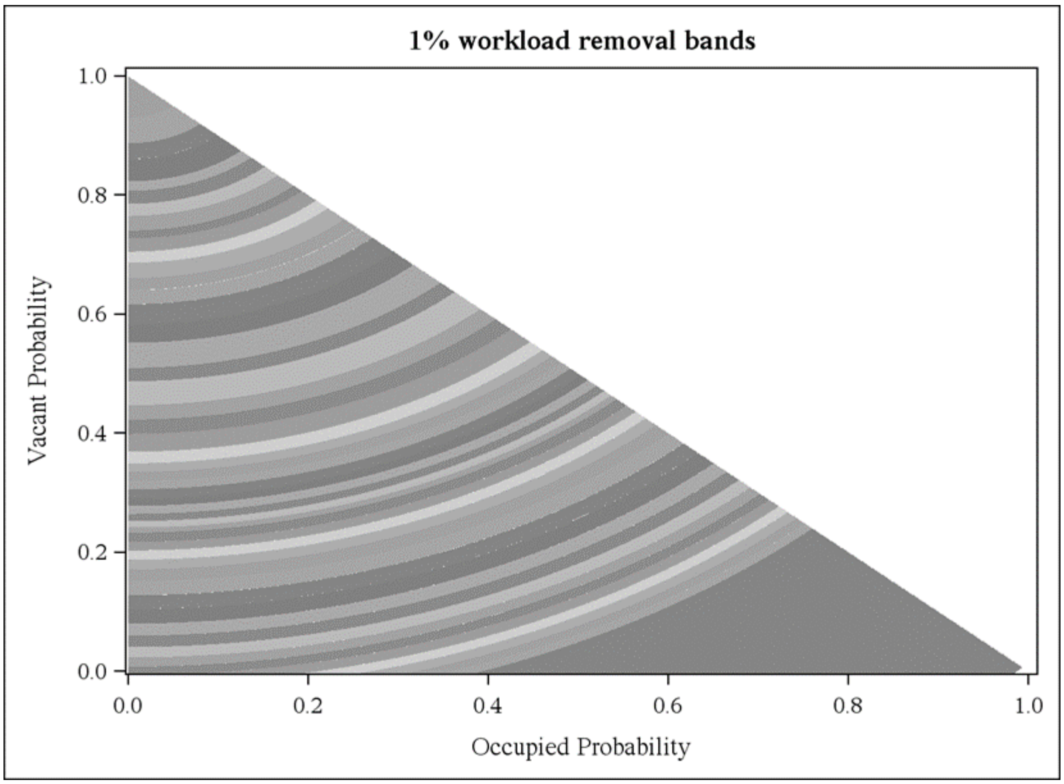
$$\delta_{h,nr} = \sqrt{\left(1 - \hat{p}_{h,nr}\right)^2 + \left(\hat{p}_{h,occ}\right)^2} \qquad (2)$$

Figure 1 demonstrates how the distance function approach was used to determine if an address could be classified as vacant. The **vertical** axis is the probability – estimated from the multinomial model – that an address is vacant. The (1,0) point on the upper left represents the probability of being vacant as 1 and the probability of being occupied as 0. The **horizontal** axis is the estimated probability that an address is occupied, also from the multinomial model. The (0,1) point on the lower right represents the probability of being Occupied as 1 and the probability of being Vacant as 0. The darker bands in the upper left quadrant of the figure indicate the address is more likely to be vacant.

The goal was to identify some of the addresses as vacant if appropriate. The requirement for a vacant identification was that the probability of being vacant was high and the probability of being occupied was low. The bands in Figure 1 show the workload for different cutoffs of the distance from points on the plot to the point on the upper left where the probability of being vacant is 1. Research and testing throughout the decade helped identify the threshold to delineate the distance cutoffs used for AR Vacant for the 2020 Census.

The designation of nonresidential applied the same distance function approach used to identify the vacant addresses with the vacant probability replaced by the nonresidential probability.

Figure 1. Visualization of the distance measure used to identify AR Vacant addresses



Source: Keller et al. (2018).

4.2.3 AR Occupied Identification

This section focuses on the identification of AR Occupied addresses, which included forming an AR roster for each address that did not submit a self-response and one NRFU contact attempt did not resolve the status. In addition, the address was not identified as AR Vacant or AR Nonresidential. The methodology used one binary logistic regression model and one multinomial logistic regression model. Both were fit using 2010 Census data. The logistic regression model was called the "Person-place model" and was used to assign each person in the AR household a probability of being enumerated in the correct place. The minimum of these person probabilities among all the AR people at an address was assigned to the address.

The next step was to fit a multinomial logistic regression model was called the "Household composition model." This model produced probabilities for six types of adult-child household compositions being correct at an address. The addresses with household compositions that had between 1 and 6 persons were used in the modeling. Since there were not enough households with 7 or more persons in the 2010 Census for the modeling to produce reliable estimates, the addresses with 7 or more residents were not used in subsequent modeling. Each address with 6 or fewer household members was assigned a probability corresponding to their observed household composition. Then the two estimated probabilities for an address were used in a Euclidean distance metric developed to identify addresses with high-quality AR rosters that met the requirement for assigning the address an AR Occupied status. As with the AR Vacant and AR Nonresidential models, prior to the 2020 Census, the regression models used for AR Occupied identification were improved by incorporating the results of simulations with the 2010 Census data.

4.2.4 Person-place Model

The Person-place model provided an estimate of the probability that person i at the h^{th} address was enumerated at the correct location $\hat{p}_{hi}^{person}$ where $i = 1, \ldots n_h$ is an index for the people on the AR roster of size n_h, and $h = 1, \ldots N$, is an index of the addresses where N equals the number of addresses under consideration. The person-place probability $\hat{p}_h^{person}$ assigned to address h was the minimum value of the estimated probabilities $\hat{p}_{hi}^{person}$ over all the n_h people on the AR roster for the address. Using the minimum estimated person-place probability for an address in subsequent calculations aided in assuring the use of only AR rosters where each person had a high probability of residing at the address on Census Day.

4.2.5 Household Composition Model

The "Household composition model" was used to estimate the probability that the address had the same household composition (number of adults and children) as would have been determined by NRFU fieldwork. A multinomial logistic model was fit using the housing unit-level data from the 2010 Census (i.e., the training data) where the dependent variable was defined as follows:

$$y_h^{occ2} = \begin{cases} 0 & \text{if housing unit } h \text{ has 0 occupants in the 2010 Census} \\ 1 & \text{if housing unit } h \text{ has 1 adult, 0 children in the 2010 Census} \\ 2 & \text{if housing unit } h \text{ has 1 adult, } >0 \text{ children in the 2010 Census} \\ 3 & \text{if housing unit } h \text{ has 2 adults, 0 children in the 2010 Census} \\ 4 & \text{if housing unit } h \text{ has 2 adults, } >0 \text{ children in the 2010 Census} \\ 5 & \text{if housing unit } h \text{ has 3 adults, 0 children in the 2010 Census} \\ 6 & \text{if housing unit } h \text{ has 3 adults, } >0 \text{ children in the 2010 Census} \\ 10 & \text{otherwise} \end{cases} \qquad (3)$$

where the *occ2* superscript denotes the household composition model for determining occupied units and the h subscript indexes the housing unit. The multinomial model produced a probability for each household composition type in the variable y_h^{occ2}. Each address was assigned the estimated household composition probability $\hat{p}_h^{HHC}$ that corresponded to its AR household composition. Note that the construction of the dependent variable in the multinomial model assumed that age was available for all household members in all housing units. This assumption was satisfied because the dependent variable for the modeling used an edited file that included imputed age for any nonresponse.

4.2.6 Distance Measure for AR Occupied Identifications

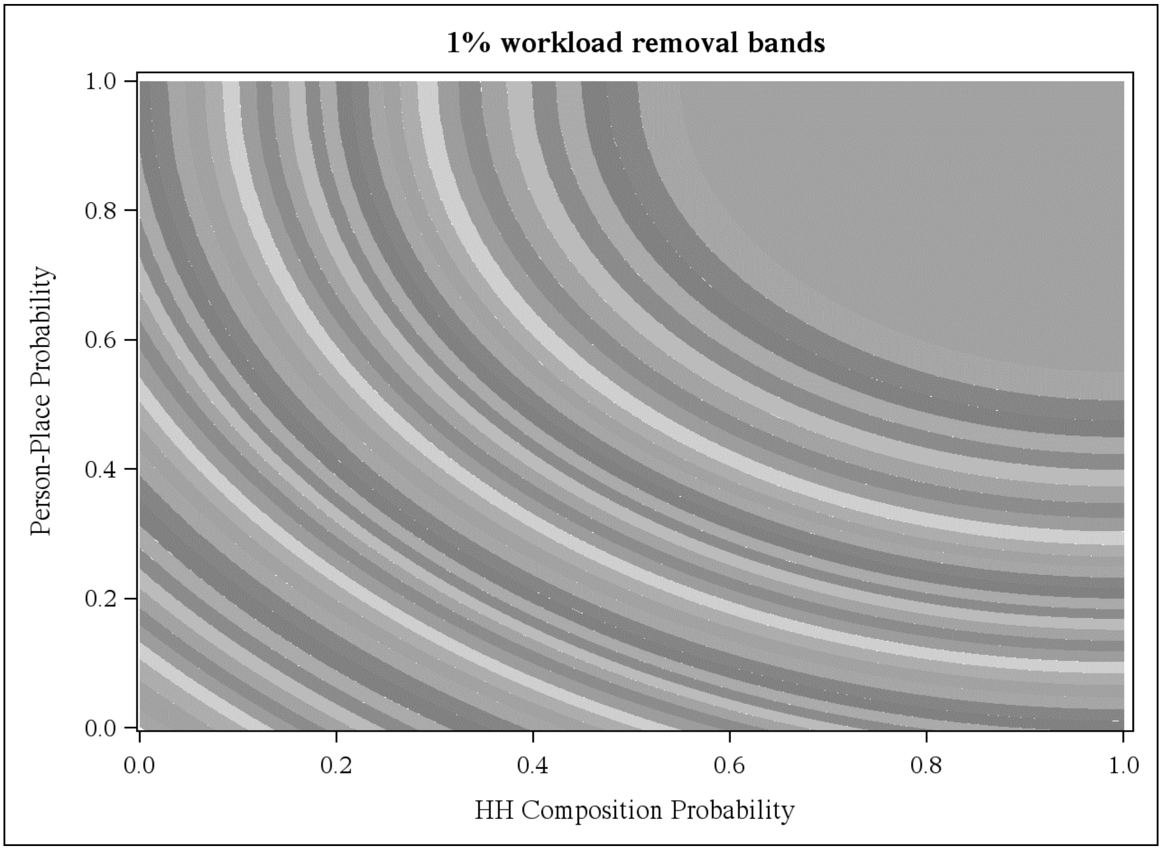
The distance measure used in identifying the AR Occupied status combined the estimated probabilities from the person-place model and the household composition model to create a single score that indicated the quality of the AR roster for an address. The score was a combination of the estimated household person-place probability $\hat{p}_h^{person}$ defined in Section 4.2.4 and the estimated household composition probability $\hat{p}_h^{HHC}$ defined in Section 4.2.5. These predicted probabilities combined into a single score, $\delta_{h,occ}$, using a distance function. The score for address h was defined using the distance function selected for this application as follows:

$$\delta_{h,occ} = \sqrt{\left(1 - \hat{p}_h^{person}\right)^2 + (1 - \hat{p}_h^{HHC})^2}. \qquad (4)$$

The addresses assigned a score $\delta_{h,occ}$ that was less than a specified threshold were considered to have a high-quality AR roster suitable for census enumeration. Note that lower scores of $\delta_{h,occ}$ indicated higher values of the estimated probabilities $\hat{p}_h^{person}$ and $\hat{p}_h^{HHC}$ and thereby, higher quality of the AR roster. Analyses with the 2010 Census data determined the threshold values. The AR rosters that had a score less than the threshold were called One-Visit AR Rosters because if one visit to the address by a NRFU interviewer did not produce an interview with a household member, the AR roster for the address was used for census enumeration.

Figure 2 shows how the distance function can be visualized as successive bands of values of the distance function emanating from the point (1,1) in the top right corner. The vertical axis is for the Person-Place probability assigned to an address, which is the minimum probability of counting everyone in the household at the right place. The horizontal axis is for the Household Composition probability for the address, which is the probability of ARs having the same household composition as the census. The different shaded bands on graph represent NRFU workloads. The (1,1) point in the upper right-hand corner is where both the Person-Place probability and the Household Composition probability equal 1. That is the ideal. The next step is to calculate the distance between the point determined by the two probabilities for the address and the ideal (1,1) point. Research and testing throughout the decade helped in identifying the threshold to delineate the distance cutoffs used for AR Occupied for the 2020 Census. The bands in the upper right quadrant of the figure indicate addresses that are more likely to be occupied.

Figure 2. Visualization of the distance measure used to identify AR Occupied addresses



Source: Keller et al. (2018).

5. NRFU Closeout AR statuses and corroboration

After the 2018 Census Test, which included a test of AR enumeration methodology, the researchers identified a need for creating a set of AR Closeout statuses that relaxed the threshold used for the One-Visit Multiple Source AR rosters to help finish the NRFU operation. This strategy permitted AR rosters of slightly less quality than the One-Visit Multiple Source AR rosters to be used for AR enumeration. Fortunately, the distance function approach discussed in Section 4.1 and selected for the AR modeling had the flexibility to create the required AR Closeout statuses in the timeframe available to complete the census enumeration. The development of the AR Closeout statuses was based on research with 2010 Census data and used an approach similar to the methodology discussed in Section 4 for the assignment of AR Occupied, AR Vacant, and AR Nonresidential (Keller 2019). The addresses assigned an AR Closeout status would not receive additional visits during NRFU Closeout that occurred at the end of NRFU when a specified percentage of the addresses in an area were resolved. However, when the NRFU operation entered its Closeout Phase, these addresses would be assigned their AR Closeout status and would not be reopened for additional contact attempts. The assignment of an AR Closeout roster to an address that received the AR Closeout Occupied status required corroboration of at least one person at the address by two sources.

In September 2020 as the AR processing entered its final stage, IRS approved the use of its data to determine the number of household members at an NRFU address when the household could not be corroborated with other data sources. Medicare and the Indian Health Service also approved the use of their data to determine only household size when corroboration was not possible. Using these sources to obtain a household size provided higher quality data than could possibly be obtained through NRFU Closeout fieldwork or imputation. The rosters created by this procedure were called Household Size Only AR rosters.

The addresses that met the criteria for AR Occupied but their AR rosters were formed using IRS data that could not be corroborated by another source received the designation One-visit Single Source AR Occupied. Similarly, the addresses that met the criteria for AR Closeout Occupied but had AR rosters that could not be corroborated by another source received the status AR Closeout Occupied Single Source.

Despite the use of Closeout statuses to complete NRFU, 91.95% of the AR rosters used for enumeration were in the category One-visit Multiple Sources, which contains the AR rosters of the highest quality. For the remaining 9.05% of the AR rosters, 3.83% were Closeout Multiple Source, 3.87% were Closeout Household Size Only, and 0.34% were classified as Other.

6. Results

This section summarizes the use of ARs in the 2020 Census enumeration. Table 1 shows the distribution of the final status by mode and respondent type for addresses in the 2010 and 2020 Censuses. Addresses considered resolved had a status of occupied, vacant, or nonresidential. Several different census operations could have produced the resolved status of an address. These operations were self-response, NRFU, and others such as Update Enumerate, Remote Alaska, Coverage Follow-up, and self-response quality assurance. The enumeration types were self-responses, household member interviews, proxy interviews, and AR enumerations. Addresses that provided a NRFU household member interview always received a status of occupied. AR enumeration was used only for the 2020 Census. For the 2010 Census, the respondent for the address sometimes was not recorded and therefore is unknown.

Table 1 shows the percentage of addresses that were occupied and had a self-response increased from 61.05 percent in 2010 to 64.28 percent in 2020. In NRFU and other operations, the percentage of interviews that had a household member as the respondent decreased from 18.79 percent in 2010 to 10.84 percent in 2020. The percentage of addresses that were assigned the status of vacant based on a NRFU proxy interview decreased from 10.92 percent in 2010 to 6.82

percent in 2020. The identification of nonresidential addresses during NRFU and other operations increased from 3.55 percent in 2010 to 6.86 percent in 2020.

Table 1: 2010 and 2020 Census Enumerations by Type

Total Addresses	**2010**	**2020**
	136,700,000	151,800,000
Self-Response	61.05%	65.28%
Self-Response Occupied	61.05%	64.28%
Self-Response Vacant/Nonresidential	n/a	1.00%
All NRFU and Others	38.41%	33.64%
Household Member Interview	18.79%	10.84%
Proxy Interview	19.51%	18.21%
Occupied	5.03%	4.53%
Vacant	10.92%	6.82%
Nonresidential	3.55%	6.86%
Unknown Respondent Type	0.10%	n/a
Administrative Records	n/a	4.59%
Occupied	n/a	3.20%
Vacant	n/a	1.15%
Nonresidential	n/a	0.24%
Unresolved Housing Units	0.38%	0.93%
Unresolved, data collection	0.38%	0.23%
Unresolved, person unduplication	n/a	0.71%

Note: Data collection operations for the 2010 and 2020 Censuses did not have the same design
Source: 2020 Census Quality Metrics Release 1 (U.S. Census Bureau 2021)

Table 1 also shows that the status of 4.59 percent (6.968 million) of the 151.8 million addresses in the 2020 Census was assigned by using administrative records. The distribution of the statuses of these addresses was that 3.20 percent (4.859 million) were occupied, 1.15 percent (1.746 million) were vacant, and 0.24 percent (0.364 million) were nonresidential. Among the 3.20 percent of addresses resolved as occupied using administrative records in the 2020 Census, further groupings delineate the quality of the roster.

Table 2 provides information about the AR roster categories and the number of addresses receiving AR enumeration in each category. Addresses enumerated with a One-Visit AR roster were considered to have the highest quality information. The NRFU operation was instructed to visit the address only once to attempt to obtain a response from a household member. If the single visit did not resolve the status of the address, and absent any other response, the address was enumerated using its One-Visit AR Roster. The justification for requiring only one visit was due to the AR modeling determining that the AR roster for the address was of the highest quality. Closeout AR rosters had lower quality than the One-visit AR rosters but were of sufficient quality that they could be used for AR enumeration for AR occupied addresses if the status of the address remained unresolved after six visits. See Section 4.3 for more information. Finally, other occupied addresses received AR enumerations using data received from college and universities through the 2020 Census Off-Campus submission program. In addition, some occupied administrative record addresses were completed to compensate for the truncated NRFU field operations in four southwest Louisiana parishes affected by Hurricane Laura in late August 2020 and Hurricane Delta in early October 2020. Table 2 shows the distribution of the administrative record occupied enumerations by the quality category of the AR rosters. One-Visit addresses accounted for 92% of the AR enumerations which indicates that the administrative records used for enumeration were predominantly of the highest quality.

Table 2: 2020 Administrative Record Occupied Enumerations by AR Roster Quality Category

Administrative Records Occupied Roster Quality Category	Count	Percent of 4.859 million Administrative Records Occupied Addresses	Percent of 151.8 million Total Addresses
One-Visit with Rosters	4,468,000	92.0	2.9
Close-out with Rosters	186,000	3.8	0.1
Close-out Household Size Only	188,000	3.9	0.1
Other	16,500	0.3	<0.1
Total	4,859,000	100.0	3.2

Note: Due to rounding percentages may not add to 100 percent. CBDRB-FY22-172.

Table 3 shows the results concerning the availability of characteristics for AR enumerations at AR Occupied addresses. When characteristics for household members on AR rosters were not available in the source where they were found, characteristics from other sources of administrative records were used when available as an alternative to imputation. One advantage concerning administrative records was the availability of characteristics. Table 3 shows the presence of characteristics for persons enumerated using administrative records. Note that the availability of age and sex is less than 100 percent because household size was the only information available for 3.9 percent of AR Occupied addresses.

Table 3: Availability of characteristics for assignment in 2020 AR Enumerations

Characteristic	Percentage of people that had characteristic assigned from administrative records
Age	96.4
Sex	96.5
Race or Hispanic Origin	83.6

Note: CBDRB-FY22-172

Table 4 shows that the national percentage of addresses that were determined to be AR Occupied was 3.20 percent. The percentage of addresses that are AR Occupied varies across the 50 states and the District of Columbia with the percentages range from 1.32 percent in Hawaii to 4.46 percent in Louisiana (U.S. Census Bureau 2021a). In addition, Table 4 shows the national percentage for AR Vacant was 1.15 percent. The percentage of addresses that are AR Vacant varies for the 50 states and the District of Columbia. These percentages range from 0.62 percent in California to 2.11 percent in Maine (U.S. Census Bureau 2021a).The next analysis focuses on the use of the results of AR modeling in the identification of occupied addresses in NRFU. Table 4 shows how the sources used to determine the status of addresses in the 2020 NRFU operation compared to those used in 2010. For the 2010 Census, 74.88 percent of the occupied addresses were identified by a NRFU household member interview, while proxy interviews identified 24.71 percent of the occupied addresses. For occupied addresses in the 2020 Census, NRFU household member interviews identified 55.48 percent, proxy interviews identified 26.07 percent, and administrative records identified 18.44 percent. While the 2020 Census introduced ARs as a mode of data collection, the 2020 Census was conducted during the COVID-19 pandemic which could have affected the ability to resolve the status of addresses through NRFU interviews with a household member.

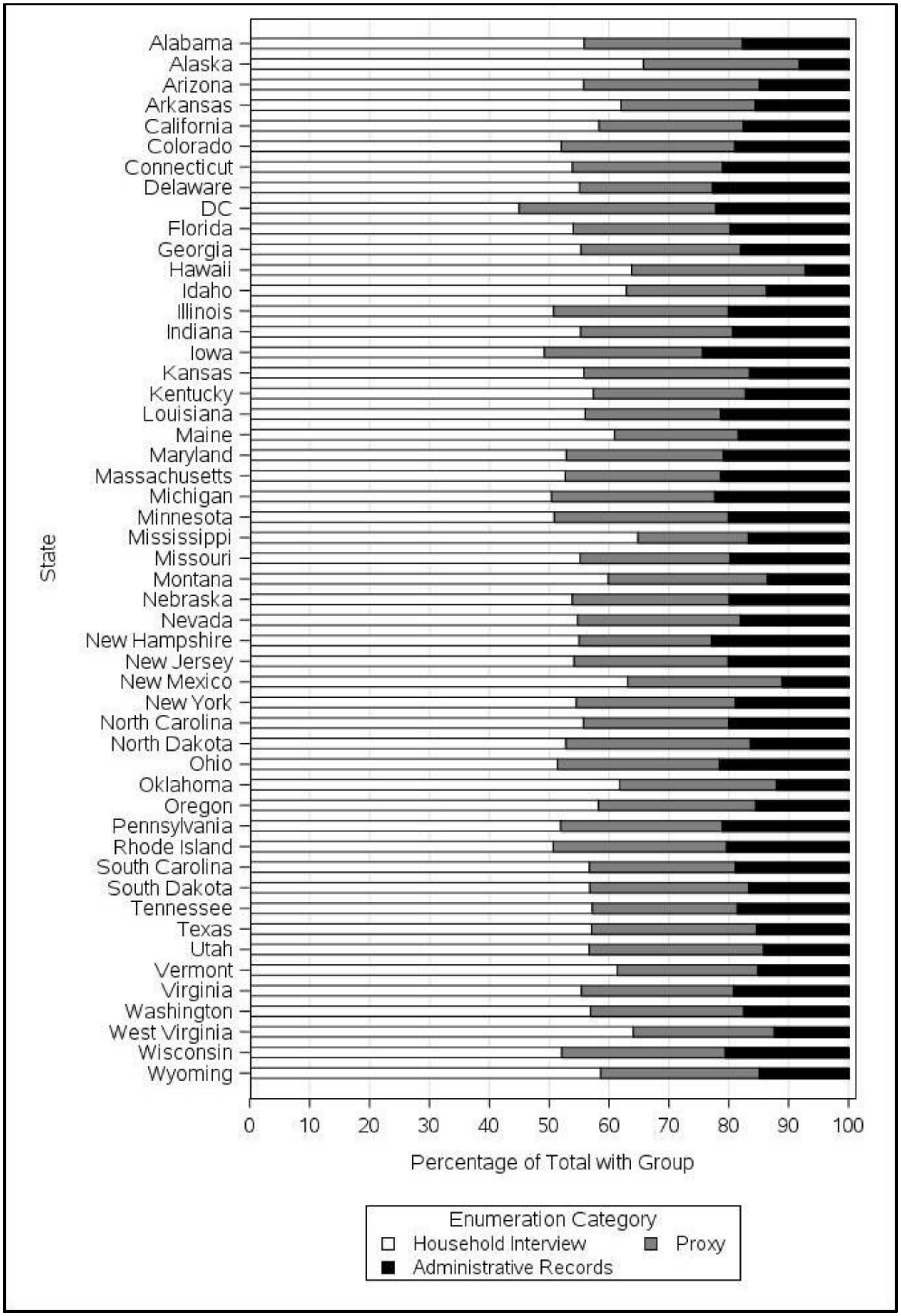
Table 4: Percentage of Nonresponse Follow-up Addresses Assigned Occupied in the 2010 and 2020 Censuses by Respondent Type

Respondent Type	2010	2020
Household Member Interview	74.88%	55.48%
Proxy Interview	24.71%	26.07%
Unknown Respondent	0.41%	n/a
Administrative Records Roster	n/a	18.44%
Total	100.00%	100.00%

Source: 2020 Census Quality Metrics Release 1 (U.S. Census Bureau 2021)

Figure 3 shows the distribution of the sources of information used in the 2020 Census for addresses identified as occupied during NRFU by state. The percentage of addresses in NRFU that were administrative record occupied ranged from 7.23 percent in Hawaii to 24.44 percent in Iowa.

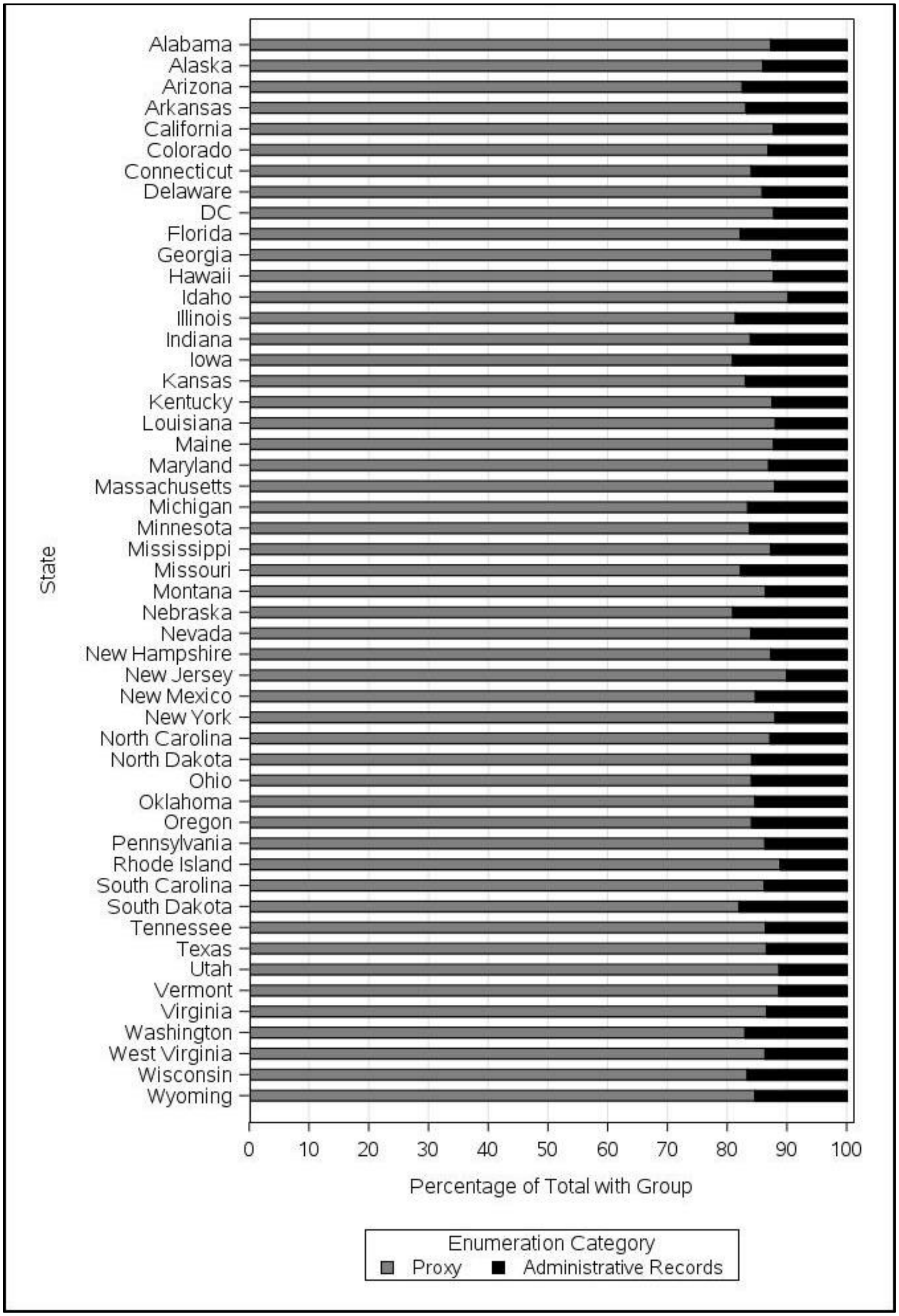
Figure 3. For addresses resolved as Occupied during NRFU, the percentages in each of three enumeration categories by state



Source: 2020 Census Quality Metrics Release 1 (U.S. Census Bureau 2021)

The last analysis focuses on the assignment of a vacant status to addresses in the NRFU operation through a comparison of results from the 2020 Census to the results from the 2010 Census shown in Figure 4. For the 2010 Census, the only way an address received a vacant status was through a proxy interview. For the 2020 Census, an address could be assigned a vacant status through a proxy interview or administrative records. As a result, administrative records were used to determine the status of 14.51 percent of the vacant addresses. Proxy interviews reported the vacant status of the remaining 85.50 percent of the vacant addresses. Within states, the percentages of vacant address statuses resolved using administrative records ranged from 9.77 percent in Idaho to 19.20 percent in Nebraska.

Figure 4. Percent of addresses assigned a Vacant Status based on a proxy interview and percent assigned using administrative records by state



Source: 2020 Census Quality Metrics Release 1 (U.S. Census Bureau 2021)

7. Summary

This paper discusses how the Census Bureau incorporated the use of administrative records for enumeration into the 2020 Census. The self-response rate for occupied housing units in the 2020 Census was 64.28% which was higher than the rate of 61.05% observed in the 2010 Census when the Internet mode of response was not available (Letourneau 2012). In a world of declining survey response rates, maintaining the census response rate and even increasing it a little is an accomplishment. The innovation of internet self-response likely is the reason for the 2020 self-response rate being comparable to the rate observed in 2010.

The COVID-19 pandemic caused the Census Bureau to delay NRFU interviewing originally planned for May 13 to July 24 of 2020 to be conducted from July 16 to October 15. The innovation that provided NRFU interviewers with cell phones for data collection, and immediate transmission of responses to the Census Bureau for processing enabled NRFU managers to know quickly which addresses needed to be contacted again. In previous censuses, the shipping of completed NRFU paper questionnaires to the Census Bureau and subsequent processing took much longer and field supervisors had to use information they collected from interviewers to decide how to allocate work assignments.

In addition, the COVID-19 pandemic caused IRS to delay its deadlines for companies to send W2 and 1099 forms to taxpayers and to delay the deadline for filing a tax return from April 15 to July 15. While the Census Bureau continued to receive monthly deliveries from IRS, the extended IRS filing deadline led to a delay in some of the IRS data being delivered to the Census

Bureau. The AR processing staff was able to modify some of their procedures and prepare AR rosters that were used for enumeration of some addresses. In the end, the Census Bureau was able to overcome the delays and meet the deadline for producing the 2020 Census data files for states to use in redistricting of their Congressional seats.

Fortunately, the research concerning the use of ARs for census enumeration enabled the Census Bureau to provide alternative paths to enumerate some households compared to traditional approaches to census enumeration. For example, four parishes in Louisiana were badly damaged by two hurricanes. The damaged homes and other structures in these areas caused many residents to leave to find housing elsewhere which made using a traditional approach to NRFU practically impossible. The Census Bureau’s AR modeling staff and subject matter experts adapted the planned procedures for AR enumeration to allow extended use of ARs for enumeration in ways that had not been planned. The innovation of using ARs to enumerate some households created the capability and the flexibility to implement AR processing for unforeseen scenarios, and this was an essential factor in the success of the 2020 Census.

References:

Childers, D. and H. Hogan. 1983. “Census experimental match studies.” In *Joint Statistical Meetings Proceedings, Survey Research Methods Section.* Alexandria, VA: American Statistical Association, 173–176. http://www.asasrms.org/Proceedings/papers/1983_032.pdf

Coale, A.J. 1955. The population of the United States in 1950 classified by age, sex, and color - a revision of census figures. *Journal of the American Statistical Association*, 50(1), 16–54.

Coyle, E. 2019. *Transcript: OnTheMap: The Road to Local Employment Dynamics*. Washington, DC: U.S. Census Bureau. https://www2.census.gov/about/training-workshops/2019/2019-05-15-led-transcript.pdf

Deaver, Karen D. 2021. “Administrative Data Used in the 2020 Census.” Decennial Census Programs Directorate. August 11, 2021. https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/administrative-data-used-in-the-2020-census.pdf

Department of Commerce vs United States House. 98-404. Supreme Court of the U.S. 1999. https://www.law.cornell.edu/supremecourt/text/525/326

Govern, K., J. Coombs, and R. Glorioso. 2012. “2010 Census Coverage Follow-up Assessment Report.” 2010 Census Planning Memoranda Series, No. 197. Washington, DC: U.S. Census Bureau. https://www2.census.gov/programs-surveys/decennial/2010/program-management/5-review/cpex/2010-memo-197.pdf

Keller, A. 2019. “Analyzing Tradeoff Between Administrative Records Enumeration and Count Imputation.” *Proceedings of the 2019 Joint Statistical Meetings, Government Statistics Section.* Alexandria, VA: American Statistical Association. 2166-2178.

Keller, A., V. T. Mule, D. S. Morris, and S. Konicki. 2018. "A Distance Metric for Modeling the Quality of Administrative Records for Use in the 2020 U.S. Census." *Journal of Official Statistics*, 34: 599-624. DOI: http://dx.doi.org/10.2478/JOS-2018-0029

Leggieri, C., A. Pistiner, and J. Farber. 2002. "Methods for conducting an administrative records experiment in Census 2000." In *Joint Statistical Meetings Proceedings, Survey Research Methods Section.* Alexandria, VA: American Statistical Association. 2709–2713. http://www.asasrms.org/Proceedings/y2002/Files/JSM2002-000858.pdf

Marks, E. and J. Waksberg. 1996. "Evaluation of coverage in the 1960 Census of Population through case-by-case checking." In *Joint Statistical Meetings Proceedings*. Alexandria, VA: American Statistical Association. 62-70.

Morris, D. S., A. Keller, and B. Clark. 2016. An Approach for Using Administrative Records To Reduce Contacts in the 2020 Decennial Census. *Statistical Journal of the IAOS*, 177-188. DOI 10.3233/SJI-161002.

Mule, T. 2021. *Administrative Records and the 2020 Census.* Random Samplings Blog dated April 1, 2021. Washington, DC: U.S. Census Bureau. https://www.census.gov/newsroom/blogs/random-samplings/2021/04/administrative_recor.html.

Mulry, M. H., Mule, T., Keller, A., and Konicki, S. 2021. *Administrative Records Modeling in the 2020 Census*. Memorandum 2021,Washington, DC: U.S. Census Bureau. https://www2.census.gov/programs-surveys/decennial/2020/program-management/memo-series/2020-memo-2021_10.pdf.

Mulry, M. H. 2014. "Measuring Undercounts for Hard-to-Survey Groups." In R. Tourangeau, N. Bates, B. Edwards, T. Johnson, and K. Wolter (Eds.), *Hard-to-Survey Populations* (Chapter 3). Cambridge, England: Cambridge University Press. 37 – 57. DOI: 10.1017/CBO9781139381635.005.

Price, D. 1947. “A check on underenumeration in the 1940 Census.” *American Sociological Review*, 12(1), 44–49.

Rastogi, S. and A. O’Hara. 2012. “2010 Census Match Study.” 2010 Census Planning Memorandum No. 247. Washington, DC: U.S. Census Bureau. https://www.census.gov/library/publications/2012/dec/2010_cpex_247.html

Statistics Canada. 2007. *2006 Census Technical Report: Coverage*. Ottawa, Ontario: Statistics Canada. https://www12.statcan.gc.ca/census-recensement/2006/ref/rp-guides/rp/coverage-couverture/cov-couv_index-eng.cfm

U.S. Census Bureau. 2021. *Information Quality*. Washington, DC: U.S. Census Bureau. https://www.census.gov/quality/

U.S. Census Bureau. 2020a. *Data Acquisitions Frequently Asked Questions*. Washington, DC: U.S. Census Bureau. https://www.census.gov/content/dam/Census/about/about-the-bureau/adrm/data-linkage/Data%20Acquisitions%20Frequently%20Asked%20Questions.pdf

U.S. Census Bureau. 2019a. NRFU Detailed Operational Plan V2.0. Washington, DC: U.S. Census Bureau. https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/NRFU-detailed-operational-plan_v20.pdf

U.S. Census Bureau. 2019b. Detailed Operational Plan for: 18. Nonresponse Follow-up Operation (NRFU), Version 2.0, July 19, 2019. Washington, DC: U.S. Census Bureau. https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/NRFU-detailed-operational-plan_v20.pdf

U.S. Census Bureau. 2019c. *SAIPE Methodology*. Washington, DC: U.S. Census Bureau. https://www.census.gov/programs-surveys/saipe/technical-documentation/methodology.html

U.S. Census Bureau. 2019d. *SAHIE 2008 - 2018 Demographic and Income Model Methodology: Summary for Counties and for States.* Washington, DC: U.S. Census Bureau. https://www.census.gov/programs-surveys/sahie/technical-documentation/methodology/methodology-2008-2018.html

U.S. Census Bureau. 2017. "2020 Census Detailed Operational Plan for: 15. Group Quarters Operation (GQ)." Washington, DC: U.S. Census Bureau. https://www.census.gov/programs-surveys/decennial-census/2020-census/planning-management/planning-docs/GQ-detailed-op-plan.html#:~:text=The%202020%20Census%20Detailed%20Operational%20Plan%20for%20the,2020%20Census%20GQ%20operation%20and%20includes%20a%20summary

U.S. Census Bureau. 2016. "2020 Census Detailed Operational Plan for: 6. Geographic Programs Operation (GEOP) – 6-3. Geographic Data Processing Component (GEOP/GDP)." Washington, DC: U.S. Census Bureau. https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/GEOP_DP_detailed_operational_plan.pdf

U.S. Census Bureau Administrative Records Modeling Team. 2017. Administrative Records Modeling Update for the Census Scientific Advisory Committee. Washington, DC: U.S. Census Bureau. https://www2.census.gov/cac/sac/meetings/2017-03/admin-records-modeling.pdf

U.S. Office of Management and Budget. 2014. M-14-06: Guidance for Providing and Using Administrative Data for Statistical Purposes. [February 2017]. https://obamawhitehouse.archives.gov/sites/default/files/omb/memoranda/2014/m-14-06.pdf

Wagner, D. and M. Layne. 2014. "The Person Identification Validation System (PVS): Applying the Center for Administrative Records Research and Applications." CARRA Working Paper Series. Working Paper #2014-01. Washington, DC: Census Bureau. https://www.census.gov/content/dam/Census/library/working-papers/2014/adrm/carra-wp-2014-01.pdf

Walker, Shelley, Susanna Winder, Geoff Jackson, and Sarah Heimel. 2012. *2010 Census Nonresponse Follow-up Operations Assessment Report*, 2010 CENSUS PLANNING MEMORANDA SERIES No. 190. Washington, DC: U.S. Census Bureau. https://www.census.gov/content/dam/Census/library/publications/2012/dec/2010_cpex_190.pdf

Whitworth, E. 2001. Implementation and Results of the Internet Response Mode for Census 2000. Presentation at the meeting of American Association for Public Opinion Research, Montreal, Quebec, May 2001. In *Proceedings of the Annual Meeting of the American Statistical Association*, August 5-9, 2001. http://www.asasrms.org/Proceedings/y2001/Proceed/00298.pdf