

2020 Census: Administrative Record Usage

National Academy of Science Panel to Evaluate the Quality of the 2020 Census
Date: July 20, 2021

Thomas Mule
Special Assistant to the Chief,
Decennial Statistical Studies Division

Data presented were approved for dissemination by the Census Bureau Disclosure Review Board
(CBDRB-FY20-ACSO003-B0011 and CBDRB-FY21-DSSD007-0012)

Shape
your future
START HERE >

United States[®]
Census
2020

Administrative Record Usage in the 2020 Census

Major initiative area of research and development this decade included

- Administrative record modeling to support the reduction of contacts in the Nonresponse Followup Operation
- Assigning a census ID to self response without one (NonID operation)
- Self-Response and Nonresponse Followup Quality Assurance

This presentation will focus on how the administrative record modeling was modified based on the delayed start in the Nonresponse Followup Operation and the extension of the Internal Revenue Service tax filing deadline from April 15th to July 15th.

Mulry, Mule, Keller and Konicki (2021) and Keller, Mule, Morris and Konicki (2018) provide more information about the methodology originally developed and implemented.

The presentation will highlight measures taken to assess and help ensure data quality of using administrative records to reduce contacts during the NRFU operation.

This presentation will also include some information on the off-campus student record initiative

Original 2020 Census Plan: Identifying Vacant and Nonexistent Addresses

Can we determine if address is vacant or does not meet our definition of a housing unit?

Example sources for AR

- United States Postal Service information
 - USPS Undeliverable-as-Addressed (UAA) reasons for census mailings made around April 1
 - Delivery Sequence File information

Internal Revenue Service (IRS) 1040 filings

IRS 1099 information returns

Centers for Medicare and Medicaid Services Medicare Enrollment database

Indian Health Service Patient database

Veterans Service Group of Illinois (VSGI) third-party national files

Census Bureau Master Address File

ACS Area-level estimates: % vacancy, % poverty, % Hispanic, etc.

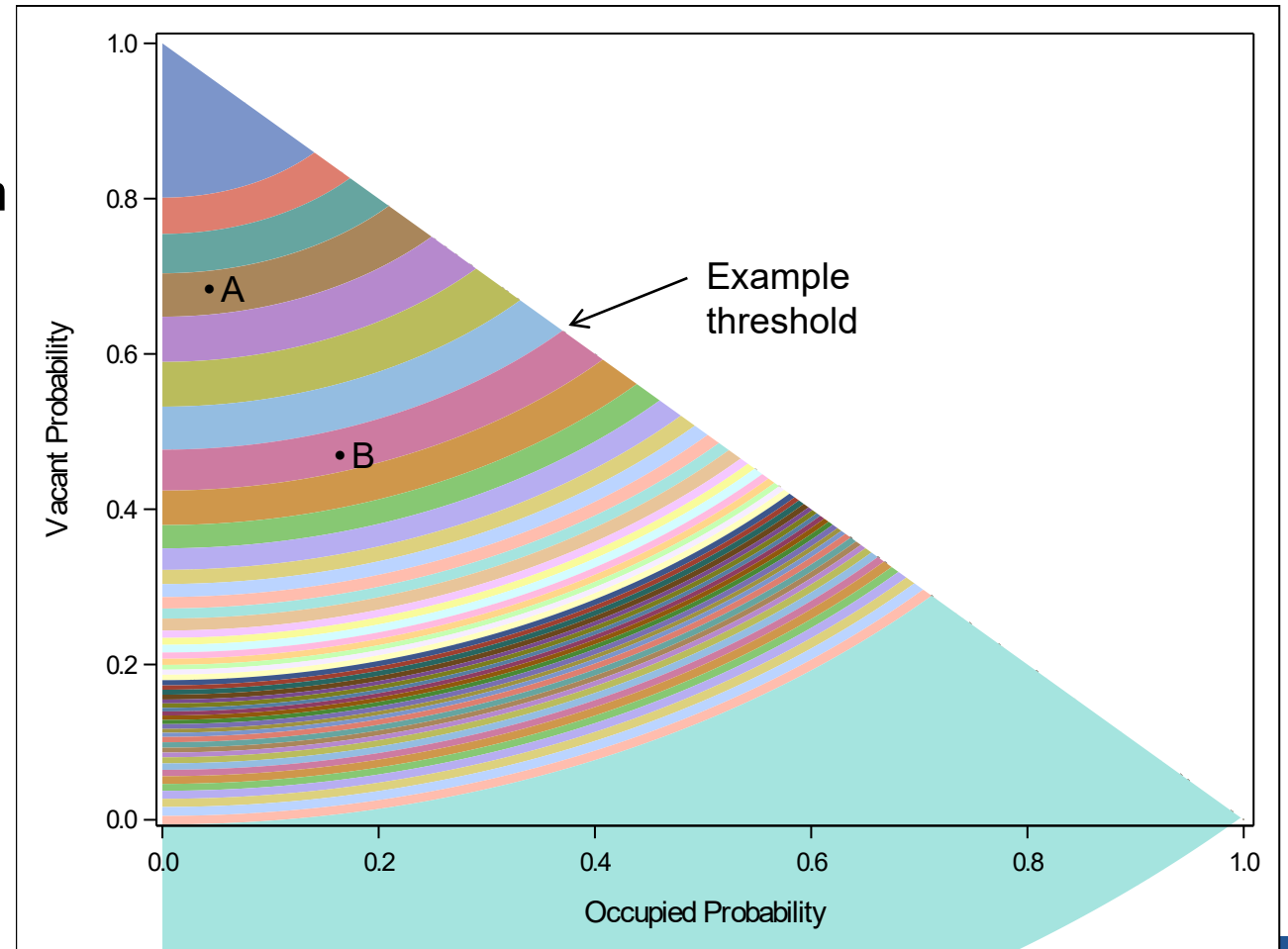
Original 2020 Census Plan: Identifying Vacant and Nonexistent Addresses Distance Function

The distance function can be visualized as successive bands of cases emanating from the point (0,1) in the top left corner

Each successive band represents an additional amount of the NRFU workload

In this example, unit A is identified as initial AR vacant while unit B is not

Similar approach implemented for Non-Existent or addresses that need to be deleted



Original 2020 Census Plan: Identifying Vacant and Nonexistent Addresses Operational Flow

March 2018						
Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
			1	2	3	
4	5	6	April 2018			
11	12	13	Sunday	Monday	Tuesday	Wednesday
18	19	20	1	2	3	4
25	26	27	8	9	10	11
Notes:			15	16	17	18
			22	23	24	25
			29	30		

May 2018						
Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
6	7	8	9	16	17	18
23	24	25	26	31	1	2
30	31					

Use administrative records to determine possible vacant and nonexistent address

Address has to have at least one UAA in TEA1 or TEA 6 mailings

Send mailing to address about 6 weeks after Census Day

Address receives one field visit
 Mail undelivered and no sign of occupancy

Field work resolution or self-response

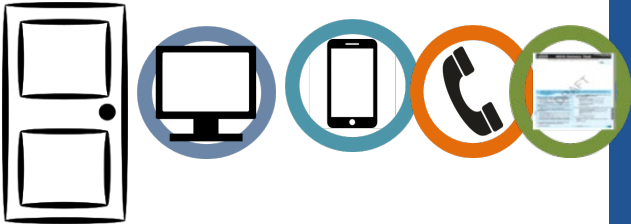
Administrative record vacant

Administrative record nonexistent address

Address receives one field visit
 Mail delivered or Mail undelivered but sign of occupancy

Address has opportunity to self-respond

Address receives full NRFU contact strategy



VACANT



Shape
your future
START HERE >

United States[®]
Census
2020

Original 2020 Census Plan:

Using Administrative Records to Enumerate NRFU Housing Units

Can we reduce the number of contact for 101 Main Street, Anytown USA?

1. Build a roster from most recent administrative record sources

TY 2019 Internal Revenue Service Individual Tax Returns 1040

TY 2019 Internal Revenue Service Informational Returns

Centers for Medicare and Medicaid Services Medicare Enrollment database

Indian Health Service Patient Database

Census Bureau Household Composition Key File

2. Check that multiple sources indicate the household lives at an address

This uses wider set of sources than those listed above.

3. Evaluate the roster

How likely is it that we are counting all of the people rostered in the right place?

How likely is it that the household composition of the rostered family matches the Census?

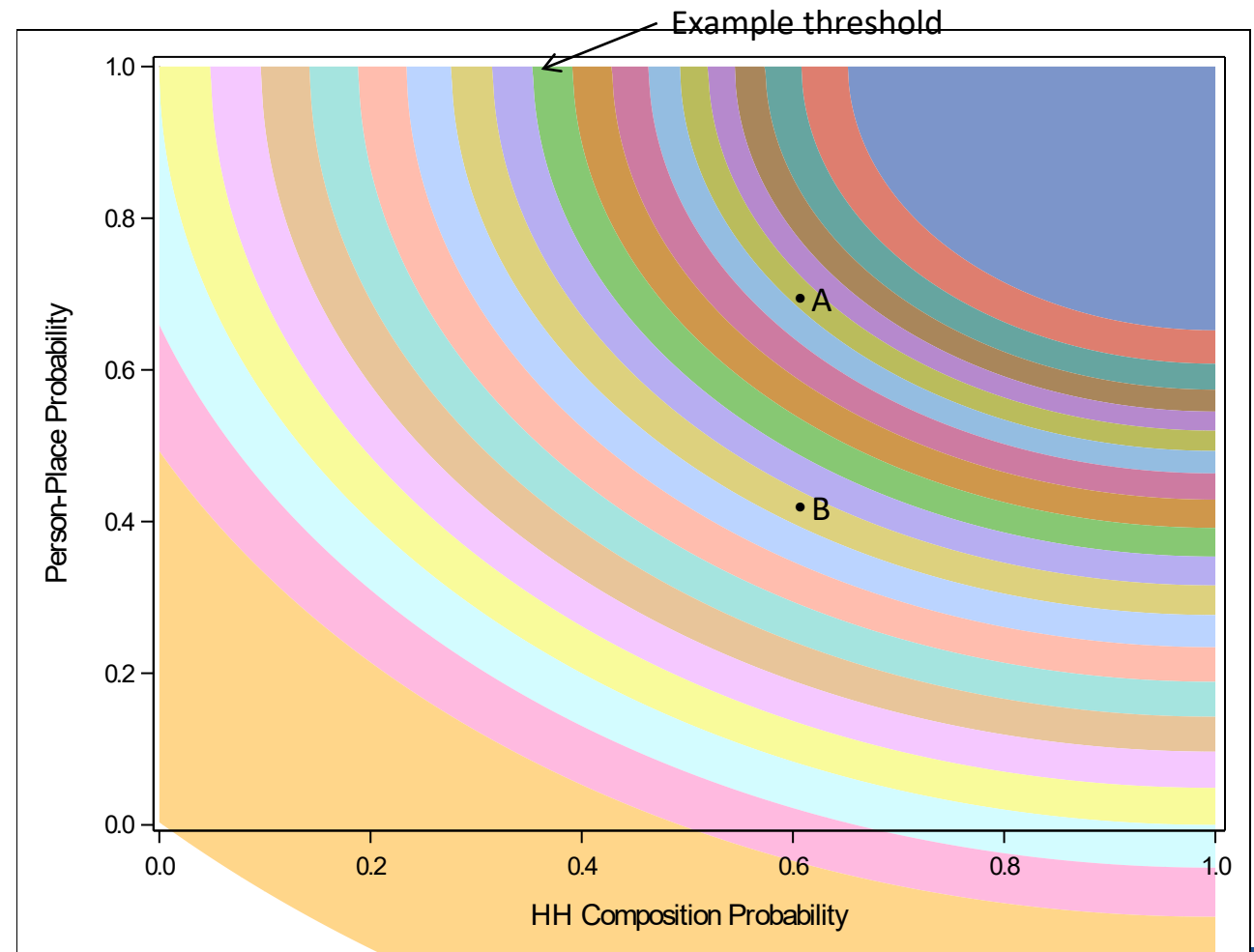
4. Decision whether to use administrative record data for 101 Main Street

Original 2020 Census Plan: Identifying Occupied Addresses Distance Function

The distance function can be visualized as successive bands of cases emanating from the point (1,1) in the top right corner

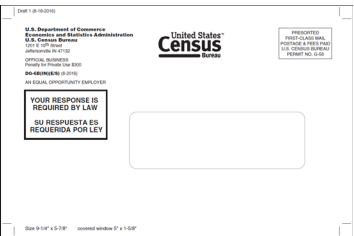
Each successive band represents additional addresses of the NRFU workload that could reduce contacts

In this example, unit A is identified as AR occupied while unit B is not

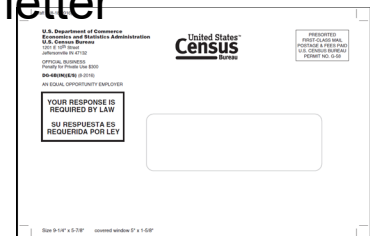


Original 2020 Census Plan: Using Administrative Records to Enumerate NRFU Housing Units

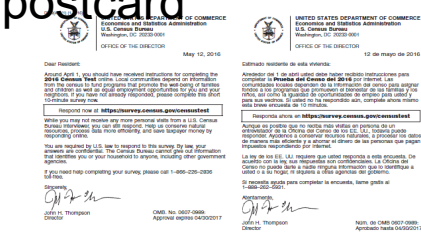
#1 Initial letter



#2 Reminder letter



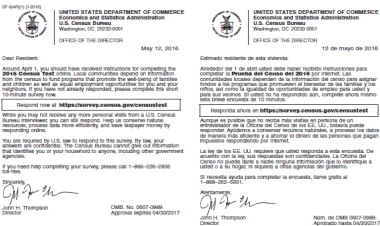
#3 Reminder postcard



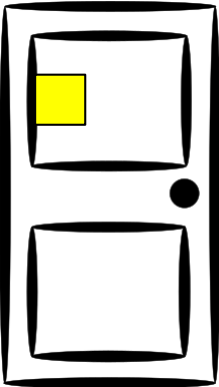
#4 Questionnaire



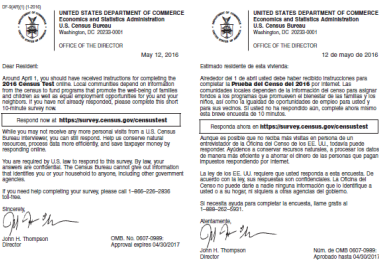
#5 Not too late postcard



#6 First visit by enumerator and notice of visit



#7 Final postcard about one week after visit



Receive a self-response return – we use the respondent provided data

No return received – we use Administrative Records Data



This example is for TEA1. For Update/Leave TEA, #1 to #5 are replaced with enumerator leaving questionnaire packet at door and two reminder mailings.

Changes to the IRS Tax Filing Deadline and NRFU Operation

Internal Revenue Service (IRS)

- On March 21, 2020, IRS announced that the tax filing deadline was extended from April 15, 2020 to July 15, 2020
 - While there was a delay in the deadline, the Census Bureau continued to receive monthly deliveries of processed 1040 records.

2020 Census

- Start of the NRFU operation delayed until August 9, 2020
 - Soft-launch of NRFU operation started on July 16, 2020 in selected ACOs

Revised Administrative Record Modeling

May Vacant and Delete Determinations

Determined vacant and delete addresses at end of May with post card in-home arrival of June 12th

Update/Leave addresses had only one mailing in early April

- Modified our vacant and delete modeling to use only one UAA determination

Update/Leave mailing in April included P.O. Box Only addresses

- Implemented rule to not allow determination if address was in P.O. Box only zip code
- 430,000 vacant and 20,000 delete addresses changed

Concern about identifying vacant or delete addresses in zip codes with high concentrations of UAAs

- Implemented rule to not allow determinations for the top 5th percentile of zip codes had a with rate of UAAs on the first mailing (57 percent or higher)
- 1,000,000 vacant and 20,000 delete addresses changed

Revised Administrative Record Modeling

June Administrative Record Update

Identified initial administrative record occupied cases

- Count agreement comparisons Adrec occupied determinations to where we had self-responses
- Preliminary analysis was showing 81 percent agreement and similar amounts of availability of age, sex, race and Hispanic origin as during tests.

Changing May vacant or delete addresses to full contacts

- May vacant or delete determinations to addresses in Early NRFU areas were changed to full contacts
 - 440,000 vacant and 200,000 delete addresses changed
- Distance function value in June modeling run was now outside the distance cutoff
 - 100,000 vacant and 8,000 delete addresses changed

Revised Administrative Record Modeling

Remaining Administrative Record Update

Early July Processing

- Additional one-visit occupied cases
- First batch of closeout occupied cases if a 1040 was returned for the address
- No closeout vacant or delete addresses in this run since we did not have the IRS delivery that included the July 15th submissions

Early August Processing

- Additional one-visit occupied and closeout occupied cases including additional sources and years used to support matching
- Additional one-visit vacant and delete cases that had over 90 percent probability of being unoccupied
- First batch of closeout vacant and delete addresses

September processing

- Additional one-visit occupied addresses
- Additional closeout occupied, vacant and delete addresses
- Closeout processing expanded to include sole source determination
 - This allowed closeout cases to use population count but no characteristics
- Introduced a model to determine administrative record occupied addresses on American Indian Reservations

Revised Administrative Record Modeling

AR Occupied modeling for American Indian Reservation

Original plan was for no administrative record modeling on reservations

To mitigate potential undercoverage, researched a similar approach to what we do for the rest of the country to determine occupied with a roster for self-response (TEA 1) and update/leave (TEA 6)

Developed a training modeling on 2010 American Indian Reservation addresses to apply to 2020 AIR address

Evaluation Statistics Using 2010 Census Data

Difference	AIR Models on AIR	NRFU Model on all NRFU*
AR count lower	14%	16%
Same count	67%	62%
AR count higher	19%	22%

*Results from Administrative Records Modeling Team CSAC March 2017 paper

Proportion of People in AIR AR Occupied MAFIDs with AR Demographic Information			
	Age	Sex	Race/Hispanic Origin
AIR Models on AIR	100%	100%	94%
NRFU Model on all NRFU*	100%	100%	90%

*Results from Administrative Records Modeling Team CSAC March 2017 paper

Administrative Record Enumeration

If an address does not self-respond by the end of data collection, we need to make response records. If it is occupied, we need to enumerate the roster and their available characteristics.

Original specification used roster rules that reflected May and June IRS 1040

Concern because we started identifying occupied addresses in June but 1040 returns could be received in August and September after the July 15th filing deadline

Modified our roster formation for enumeration

- Started by using the roster developed for the housing unit when the AR occupied status was determined (An address identified in June would start with the June roster, July would use July,...)
- If an occupied address did not have a 1040 return when determination was made then we would look at the all of the available 1040 returns for the address at that time. If all of the original people match to the 1040 return then we would add the additional 1040 people to the enumeration roster.

Off Campus Student Records

In May, the Census Bureau decided to contact universities and colleges to see if they could provide us students who were living off-campus in the spring 2020 semester

- Name, Date of Birth and Age
- Local off-campus address
- Alternative permanent address

Between June 18th and August 14th, Census Bureau contacted over 1,300 schools to see if they could participate.

Over 600 schools provided us with the information

- Assigning our Census Master Address File identifier (MAFID) to local off-campus addresses

Our research allowed us to determine how we can use this information in our post-processing operation to enumerate households that may be vacant, delete or unresolved after data collection is done.

- Develop rosters based on off-campus and the AR modeling roster sources that can be used for enumeration
- Designate addresses as “Occupied but population count unknown” if the off campus roster is incomplete.

Administrative Record Modeling

(Results from April 2021 Quality Metrics)

4.59 percent of addresses nationwide have been completed using high-quality administrative records

- 3.20 percent occupied
- 1.15 percent vacant
- 0.24 percent delete

13.64 percent of the NRFU workload

- 9.51 percent occupied
- 3.43 percent vacant
- 0.70 percent delete

Source: https://www2.census.gov/programs-surveys/decennial/2020/data/operational-quality-metrics/census-operational-quality-metrics-release_1.xlsx

Summary

- Changes to the administrative record modeling due to changes to the IRS tax filing deadline to July 15, 2020 and the delayed start of the 2020 Census NRFU operation
- To mitigate potential coverage error, changes in May and June modeling resulted in 2.2 million addresses receiving full contacts that would have otherwise been classified as AR vacant and AR delete
- To mitigate potential undercoverage, analysis and changes to one-visit and closeout occupied determinations were implemented in June and July to compensate for change in tax filing deadline
- Introduced an occupied model for American Indian Reservations
- Roster determination for AR enumeration if a household does not self-respond was adjusted to allow household members included on later filed tax returns to be included
- Closeout processing expanded to include sole-source determination where we could use population count-only information
- Implementation of the off-campus student initiative to help improve the census results of students whose local school address may have been determined to have been unoccupied.

References

Keller, Mule, Morris and Konicki (2018) A Distance Metric for Modeling the Quality of Administrative Records for Use in the 2020 U.S. Census. JOS, Vol. 34, No. 3, 2018, pp. 599–624, <http://dx.doi.org/10.2478/JOS-2018-0029>

U.S. Census Bureau (2017). "Administrative Records Modeling Update for the Census Scientific Advisory Committee." Paper presented at the Spring 2017 meeting of the Census Scientific Advisory Committee. Available at: <https://www2.census.gov/cac/sac/meetings/2017-03/admin-records-modeling.pdf>

Keller, A. 2016. "Imputation Research for the 2020 Census." Statistical Journal of the International Association of Official Statistics 32: 189–198. Doi: <http://dx.doi.org/0.3233/SJI-161009>

Mulry, M., Mule, T., Keller A. and Konicki, S. (2021) "Administrative Record Modeling in the 2020 Census" 2020 Census Program Memorandum Series 2021.10 <https://www2.census.gov/programs-surveys/decennial/2020/program-management/planning-docs/administrative-record-modeling-in-the-2020-census.pdf>

Morris, D., A. Keller, and B. Clark. 2016. "An Approach for Using Administrative Records to Reduce Contacts in the 2020 Census." Statistical Journal of the IAOS 32:177–188. Doi: <http://dx.doi.org/10.3233/SJI-161002>

Morris, D.S. 2017. "A Modeling Approach for Administrative Records Enumeration in the Decennial Census." Public Opinion Quarterly: Special Issue on Survey Research, Today and Tomorrow 81(S1): 357–384. Doi: <http://dx.doi.org/10.1093/pog/nfw059>

Morris, D.S. (2014). "A Comparison of Methodologies for Classification of Administrative Records Quality for Census Enumeration," in JSM Proceedings, Survey Research Methods Section. Alexandria, VA: American Statistical Association. 1729-1743.

END