

# Exploring the Utility of Linking Data on Domestic Migration

## Preliminary matching of the 2010 American Community Survey and 2009 National Change of Address File

Megan Benetsky, Allison Fields, Ph.D., William Koerber, Caleb Miller, Amy O'Hara, Ph.D., and Melanie A. Rapino, Ph.D. | Center for Administrative Records, Research, and Applications and the Social, Economic, and Housing Statistics Division

### Introduction

Domestic migration in the U.S. has been measured using Decennial Census long form data since 1940. After Census 2000, the long form data collection was replaced by the American Community Survey (ACS), a monthly survey that collects social, economic, and household data that is released annually. One major difference between Census and ACS migration data is the time reference for the question. Decennial data asked where a person lived 5 years ago, while the ACS asks where a person lived 1-year ago. The time period was changed to reflect the on-going data collection of the ACS, to allow an annual estimate of geographical mobility.

There are other Census data sources of yearly migration, most notably the Annual Social and Economic Supplement to the Current Population Survey (CPS ASEC). Other data exist that can proxy annual migration estimates, including administrative records from the Internal Revenue Service (IRS) and Medicare program. Another major source of administrative records data recording on-going residential change is the National Change of Address (NCOA) database from the United States Postal Service (USPS). However, much like the other administrative data, its utility in measuring US domestic migration is not well understood.

### Purpose

This project is the first major attempt to link individual NCOA records to individual mover records in the ACS. This analysis will provide only preliminary results, attempting to assess the comparability of the two data sources. It examines issues of reference period, specificity of the linkage process, and a brief analysis of the results of the merged data with respect to similarity of origin and destination geographies and basic demographic characteristics.

### Data

**2010 American Community Survey**  
The ACS is an ongoing survey conducted annually by the U.S. Census Bureau that captures changes in the socioeconomic, housing, and demographic characteristics of communities across the United States and Puerto Rico. The ACS questions related to migration ask if the respondent lived in the same home 1-year ago and if not to provide the complete address of that previous residence. The full address of a mover is geocoded to the place-level, and the block-level where possible. The ACS allows only one move to be reported by a respondent during the reference period.

**2009 NCOA**  
The National Change of Address (NCOA) file contains records for forms completed at a post office (free of charge) or online electronically (at a \$1.00 surcharge). It is used by businesses, families, and individuals to inform the United States Postal Service (USPS) of a pending or existing change of residence. This facilitates accurate mail delivery during the course of a move. Completion of an NCOA form is not mandatory. NCOA may record multiple moves for any given respondent during the reference year.

The file available to the Census Bureau is part of a data sharing agreement with the USPS. The 2009 data used in this project is an aggregate of monthly NCOA data deliveries with records of individuals or households advising the USPS to deliver their mail to a specific address beginning on a certain date. The dataset acquired by the Census Bureau includes name, origin address, destination address, and effective move date. Two flags are also available that identify whether it is a permanent or temporary move or an individual, family or business move.

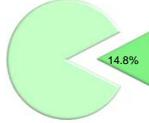
### ACS Questionnaire and NCOA Form Types



### The Basics

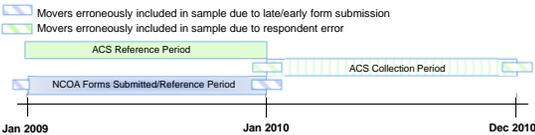
#### 2010 ACS

45,665,324 people in the U.S. reported living in a different house sometime in 2009; about 15% of all Americans were movers.



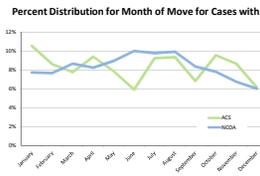
#### 2009 NCOA

35,023,094 NCOA records were acquired with a move effective date in 2009.  
 ✓ Less than 1 percent were temporary moves  
 ✓ Only 36 percent of the forms were completed as family moves, meaning one form for many individual moves with same origin and destination  
 ✓ Families and individuals can appear multiple times in the records representing multiple moves per individual in the year.

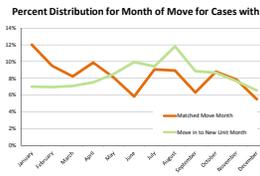


### Geography and Month of Move

After NCOA and ACS movers are linked by PIK, there is still the question of whether the data sources are describing the same move. We explore this by examining origin and destination and move dates. Though the files are linked by person identifiers, analyzing the geography requires more matching.



All ACS movers have a Master Address File identification codes (MAFID) for each interview address, otherwise described as the move "destination." This is the housing unit where the ACS interview took place. ACS "origins" are the responses to the previous residence questions; these addresses are processed through normal geocoding operations to obtain geocodes, possibly down to block-level geography. They are not assigned a MAFID.



In NCOA 2009, upon receipt of the monthly deliveries the Census Bureau attempts to put a MAFID on both the origin and destination address reported in the NCOA record. Using that stringent criteria only about 38 percent of the NCOA records obtain a MAFID for both the origin and destination addresses on a record. About 61 percent of NCOA records had address elements to match to the Master Address File for the destination address.

### Moving to the Future

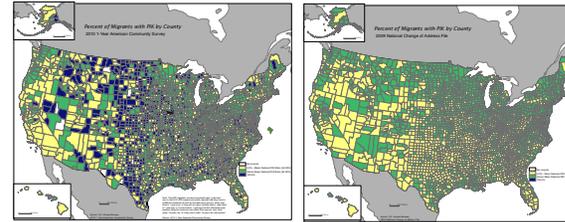
Continuing research will explore the records omitted from this analysis and assess the utility of NCOA for various survey operations. The linked file characteristics omitted the duplicate PIK records in NCOA. These included persons with multiple moves during 2009, or those with "consolidating" or "splitting" moves that have multiple records in the file during the year. A consolidating move occurs when a person requests the USPS to divert their mail from two or more locations to a new, third location. An example would be a college graduate who asks the USPS to forward their college dorm mail and mail from their parents' address to a new apartment. A splitting move occurs when a person requests the USPS to divert their mail from a single location to two or more new locations. This could happen when roommates move to separate units or when couples separate.

Using the PIKed file was convenient to identify duplicate records in NCOA and to facilitate linkage between NCOA and ACS. Linkage by PIK is not necessary; direct name matching could be employed. Direct matching within MAFIDs where move activity has been recorded in one or both sources would retain more records in the analytic universe. As the NCOA map indicates, the limited amount of name and address data in NCOA produced lower validation results than the ACS (which included name, address, date of birth and sex). Future research will explore combinations of optimal methods to match records between the files.

This analysis also treated family move forms as individual moves (for the person who completed the form). We know that assumption assigned many children to the Unmatched category in the characteristics analysis to the left. We need to conduct more research to determine which household members could be flagged as migrants when a family move is recorded in NCOA.

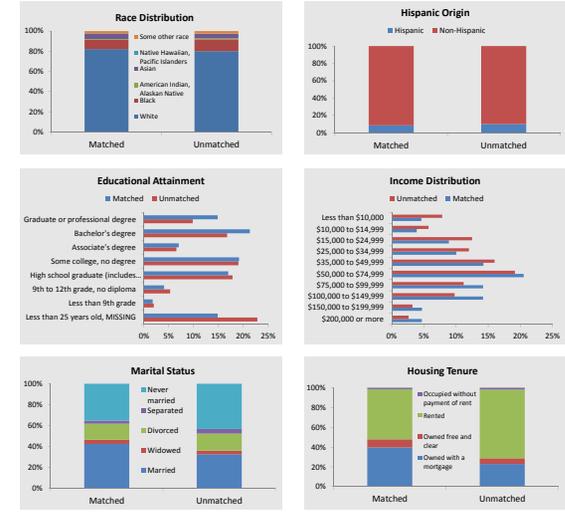
We will investigate ways in which NCOA would be used to improve survey data. Enhancements to autocoding, imputation, and benchmarking systems are possible. The extent to which NCOA records can be geocoded to appropriate geographic levels will be investigated accordingly, given operational requirements. Using NCOA to impute missing data will require greater understanding of how and when mail customers use the NCOA form, how name data can be matched without PIKs, and how to handle aligning move months to ensure that ACS and NCOA are referencing the same move.

### Do we have the tools to match?



County	State	Percent Migrants with PIK	Above or Below National Average (94.39%)
1 Los Angeles County	California	82.74%	BELOW
2 Maricopa County	Arizona	94.08%	BELOW
3 Harris County	Texas	88.95%	BELOW
4 Cook County	Illinois	91.76%	BELOW
5 San Diego County	California	93.86%	BELOW
6 Riverside County	California	91.26%	BELOW
7 Dallas County	Texas	92.41%	BELOW
8 San Bernardino County	California	92.70%	BELOW
9 Orange County	California	92.17%	BELOW
10 King County	Washington	94.66%	ABOVE

### Are ACS Movers Matched to NCOA different?



### Methodology

The reference period for moves in the 2010 ACS is the year 2009. The NCOA form does not specify a move date, therefore the move effective date serves as a proxy, and NCOA records with move effective dates from January through December 2009 are aggregated for this analysis. To compare the migration information, the movers in ACS and NCOA need to be linked. This is done by matching persons across the two files using Protected Identification Keys (PIK). Name and address data from each data set are compared to a reference file containing name and address records from other Federal data sources. A unique PIK is then appended to each person record with validated information. This process works optimally when input data sets contain full name, address, date of birth, and sex. Not every person in each data set received a validated PIK, and the rate of PIK varied in the two datasets.

The universe was carefully defined for this preliminary work. First, the ACS mover sample for the analysis was restricted to those who reported: *people age 1 year and over in the 2010 ACS sample who either reported that they lived in a different residence as well as the state and county in which they lived 1 year prior; or they did not report mobility status, state lived in 1 year ago, or county lived in 1 year ago but the information was imputed using the response from either their spouse, their child under 18 years old, or if they were under 18 years old, their parent.* This created the cleanest set of ACS movers with reported responses.

Second, all records with multiple PIKs in both ACS and NCOA were set aside for future analysis. There were many "real" duplicate PIKs in the NCOA file. People with multiple moves during the year, or whom were "consolidating" or "splitting" moves will have multiple records in a NCOA file for a given year. Additional analysis is required to develop a process of adjudicating these duplicate PIKs both within and across datasets.

We demonstrate that the PIK rate for movers in ACS is high (94 percent) for our defined universe. However the rate of PIKs in the top 10 counties for movers in 2009 varies, and many fall below the national average PIK rate. For NCOA, just less than 70 percent of the records received a PIK. The process attempted to match NCOA person record data to a file based on the Social Security Numerical Identification file (NumIdent), and had less than optimal results for NCOA because of the limited amount of personally identifying information in the dataset. Additionally, the NCOA PIK rate shows an increasing level of degradation by move effective month, which will be explored in future research.

We matched 35,287 (unweighted) ACS movers to a NCOA record by PIK.

### Results

The charts to the left indicate how ACS movers who matched to a NCOA record differ from the ACS movers who did not match to a NCOA record. Note that all NCOA records with multiple moves are excluded from this preliminary analysis, and that family move forms are treated as individual moves for the person who completed the form. These unweighted results are presented as a proof of concept and to stimulate future research. The Matched and Unmatched data are similar for Race, Hispanic Origin and Tenure. Differences are observed for educational attainment, income and marital status. It is not mandatory to complete a change of address form when moving; more educated, higher income individuals – specifically adults in the households – are completing the forms. The lack of children in the NCOA dataset can be seen in the larger share of never-married persons in the Unmatched bar for Marital Status. Future analysis on flagging ACS household members as movers based on a Family move flag in NCOA will shift some children in to the Matched category.

**Universe Definitions**  
**Matched:** ACS Movers with PIK, that were matched to same PIK in NCOA record  
**Unmatched:** ACS Movers with PIK, that were not matched to NCOA record with PIK, can include spouses and children that were on a family NCOA form

