Exploring the Utility of Linking Data on Domestic Migration: Preliminary matching of the 2010 American Community Survey and 2009 National Change of Address File[1]

Megan Benetsky, Alison Fields, William K. Koerber,
Caleb Miller, Amy O'Hara, and Melanie A. Rapino[2]
U.S. Census Bureau, 4600 Silver Hill Road, Washington, DC 20233

Abstract

This paper discusses the utility of the National Change of Address (NCOA) database in providing supplemental geographic specificity to the self-reported movers in the American Community Survey (ACS). In the analysis, migration flow pairs (moves between residence origins and destinations) were constructed for records in both sets of data and then individuals were matched through the Person Identification Validation System (PVS). Findings reported here suggest that the complexity of the move match process between datasets is highly resource intensive. Using the NCOA to supplement missing geographic information for matched movers during normal ACS processing would not be feasible at this time.

Key Words: Administrative Records, Data Linkage, Migration, American Community Survey, National Change of Address

1. Introduction

The Census Bureau has been working on several projects that look at the integration of person-level micro data from household surveys with administrative data from several sources. These linked data sets are primarily research vehicles that are not publicly available due to confidentiality concerns. Only a few, such as the SIPP Synthetic Beta Data Product (SSB), have advanced to modeling variables on the person-linked records, to preserve the covariate relationships, and allowing the public to have access to such a rich data source.

This project is the first major attempt to link individual National Change of Address (NCOA) database records to individual mover records in the American Community Survey (ACS). Missing data exists in both household interviews and administrative records. A long-term goal of the project is to determine whether information collected in the NCOA can supplement the ACS edit and implementation process for the migration question when records are confidently matched. Supplementation could occur if there was an NCOA record for a specific move reported in the ACS survey, for which complete address information of the previous residence was not collected.

This paper uses data from the 2010 one-year ACS file and the 2009 NCOA file to provide only preliminary results of an assessment of the comparability of the two data sources. [3] 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 of matched records.

2. Background

Domestic migration in the United States has been measured using Decennial Census long form data since 1940. After Census 2000, the American Community Survey (ACS), a

---

[3] The estimates in this paper are based on responses from a sample of the population. As with all surveys, estimates may vary from the actual values because of sampling variation and other factors. All comparisons made in this paper have undergone statistical testing and are significant at the 95-percent confidence level unless otherwise noted. For information on confidentiality protection, sampling error, non-sampling error, and definitions see: http://www.census.gov/acs/www/methodology/methodology_main/.

monthly survey that collects social, economic, and household data released annually, replaced the long form data collection. One major difference between Census and ACS migration data is the time reference for the question. The question on the Decennial Census asked where a person lived five years ago with a reference date of April 1, while the ACS asks where a person lived one year ago. The move reference period for the ACS migration question was changed to reflect the on-going data collection of the survey, and 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). This data provides the longest yearly migration trend collected with migration rates since 1945. While the CPS ASEC provides the longest trend in migration rates, the ACS one-year migration question is the only source for nationally representative detailed sub-state information on migration. Other data that can proxy annual migration estimates include 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 the dynamics of domestic migration in the United States is not well understood.

Migration data are used by governments, city planners, and the transportation community, to decide where to place items such as of hospitals and schools. They are also used as data inputs for federal funding programs, so the data need to be as complete and accurate as possible to meet the needs of these programs – often at very low levels of geography.

3. Data and Methodology

Data

This research uses two main sources of data: the 2010 one-year ACS file and the 2009 NCOA records database. The ACS is a nationwide survey of about three million housing unit addresses designed to collect and produce economic, social, demographic, and housing information annually. ACS data are collected continuously using independent monthly samples through mail-out/mail-back questionnaires with telephone and personal interview nonresponse follow-up. It is conducted in all U.S. counties and Puerto Rico municipios. The sample from Puerto Rico was excluded for this analysis.

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.

Figure 1: One-year Migration question from the American Community Survey

The United States Postal Service (USPS) maintains the National Change of Address (NCOA) registry. It is a database of national change of address forms that were completed by individuals or families to have their mail forwarded to a new address, typically because of a change in residence. The NCOA file contains records for forms completed at a post office (free of charge) or online electronically (at a $1.00 surcharge). Businesses, families, and individuals use it to inform the 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.

Figure 2: Examples of National Change of Address (NCOA) paper and Internet Completion forms.

Methodology

The ACS universe was carefully defined for this preliminary work. First, the ACS mover sample was restricted to the following:

1) *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;*

2) *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 own 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.

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. NCOA records with move effective dates from January through December 2009 are aggregated for this analysis.
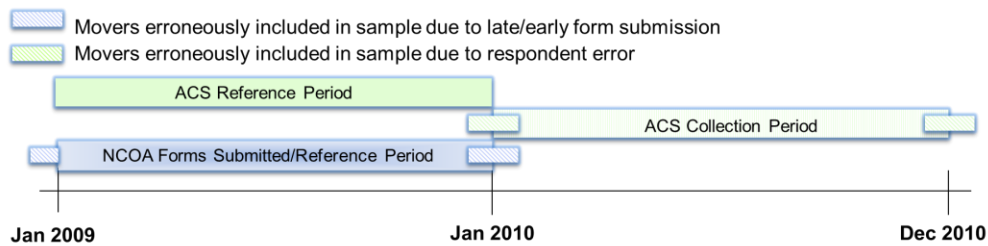
Figure 3: ACS 2010 and NOCA 2009 collection and reference periods.

For about the same reference period, the 2010 ACS estimates that 45, 665, 324 people in the US lived in a different residence than their current residence sometime in 2009; about 15% of all Americans were movers. The NCOA database for 2009 had 35,023,094 records with a move effective date in 2009. Less than 1 percent of the records in NCOA reported that it was a temporary move and 36 percent of the forms designated family move. In the case of a family move, one form was submitted for all persons at one address. Families and individuals can appear as multiple NCOA records representing potentially multiple moves per individual in the year.

To compare 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). These keys are assigned using the Person Identification Validation System (PVS). Name and address data from each dataset 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, on each data set, with validated information. This process works optimally when input data sets contain full name, address, date of birth, and sex (NORC 2011). Not every person in each data set received a validated PIK, and the rate of PIKing varied across the two. NCOA records contain only first and last name and address. The PIK rate is lower for NCOA than many other federal files because of the limited amount of available personal identifiers.

All records in either dataset without PIKS were excluded from the analysis. Additional records that received multiple PIKs in both ACS and NCOA were also set aside for future analysis. Duplicate PIKS can occur for several reasons. In NCOA, the low level of personal identifiers could mean that people are given multiple PIKs incorrectly. There were many "real" duplicate PIKs in the NCOA file as well. These are people with multiple moves during the year, or who were "consolidating" or "splitting" residences. These examples all create legitimate multiple records with duplicate PIKs 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.

4. Results and Analysis

ACS Movers without NCOA Records

We found that the PIK rate for movers in ACS is high (94 percent) for our defined universe nationally. However, the rate of PIKS in the top 10 counties for movers (as defined by the ACS 2005-2009 County-to-county migration flows) in 2009 varies. In Table 1, we found that all but one of the top destination counties in 2009 fell below the national PIK rate. This is likely explained by the fact that people of Hispanic origin are most likely to move to these counties but are less likely to be PIKed.

Table 1.

PIK Rates for Top Destination Counties in 2009

	County	State	Percent PIKd	Above or Below National Average (94.39%)
1	Los Angeles County	California	89.74%	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	93.96%	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

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. The final dataset for the analysis consisted of a set of matched 91,755 (unweighted) ACS movers to a NCOA record by PIK.

Table 2 shows the PIK rates for NCOA by month. Here we see that each successive month is associated with a greater likelihood of not being PIKed. January through July are below the average 70% PIK rate, and August to December are higher than the national PIK rate. This indicates a possible seasonal trend in NCOA. We further examine the distribution of PIKed cases in the ACS later in the paper, however any seasonality research on ACS is not recommended.[4]

---

[4] See King et al., 2009
http://www.census.gov/acs/www/Downloads/library/2010/2010_King_01.pdf

Table 2.

Rate of Failure to PIK in NCOA 2009		
Month of Effective Move	**Records**	**PIK Failure Rate**
January	2,914,734	26.50%
February	2,770,250	26.90%
March	2,933,480	26.90%
April	2,725,238	26.50%
May	2,887,165	26.10%
June	3,266,201	27.30%
July	3,272,296	29.10%
August	3,410,425	31.20%
September	3,025,848	34.10%
October	2,894,504	35.80%
November	2,547,450	37.20%
December	2,375,503	39.40%
	35,023,094	

Geographic distribution of PIK rates for ACS and NCOA

Figures 4 and 5 show the distribution of counties that were above or below the national average for the respective dataset. As you can see, the ACS data had many counties whose movers were all PIKed (navy blue). NCOA had no counties that had a 100% PIK rate. Otherwise, the above/below distribution looks relatively similar across datasets: high PIK rates in the Midwest and lower PIK rates in the south and southwest. One region that looks different between the ACS and NCOA is the northeast, where it appears that PIK rates are more likely to be above the average for NCOA, but below the average for the ACS. This may be explained by the higher average PIK rate in the ACS, however.

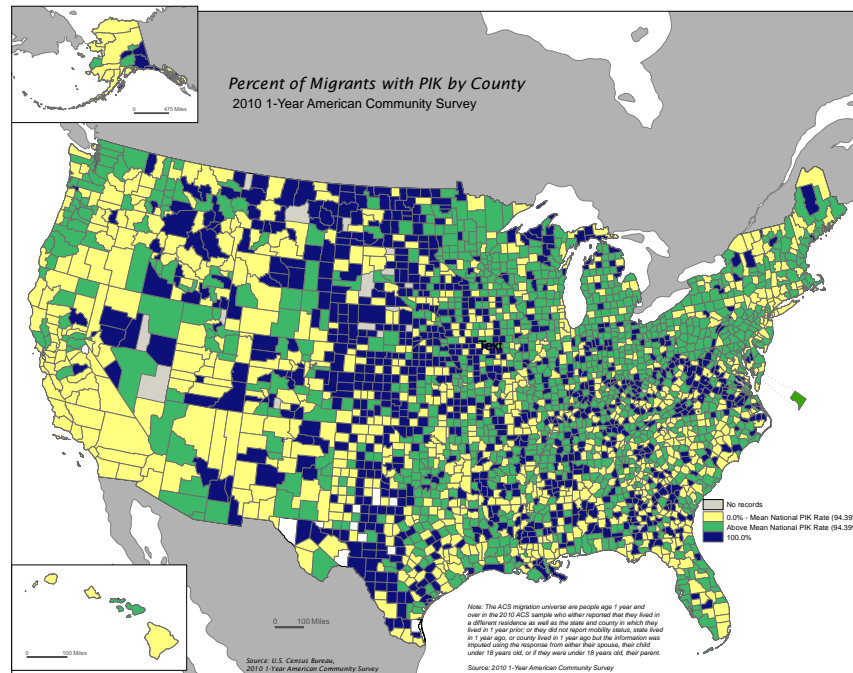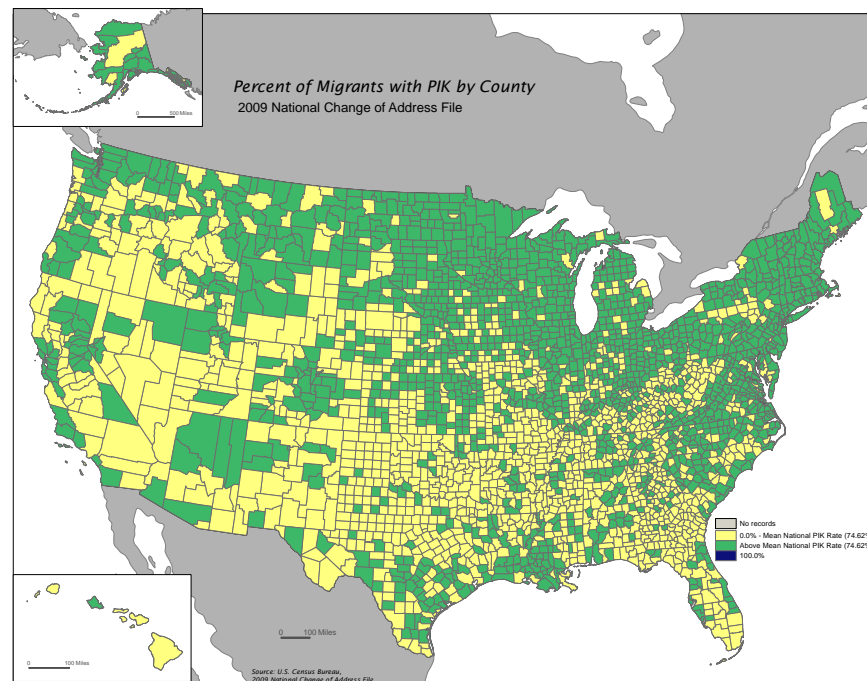**Figure 4. Percent of migrants with PIK by county, ACS**

Percent of Migrants with PIK by County
2010 1-Year American Community Survey

No records
0.0% - Mean National PIK Rate (94.39%)
Above Mean National PIK Rate (94.39%)
100.0%

Note: The ACS migration universe are 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 in 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.

Source: 2010 1-Year American Community Survey

Source: U.S. Census Bureau,
2010 1-Year American Community Survey

Figure 5. Percent of migrants with PIK by county, NCOA

Percent of Migrants with PIK by County
2009 National Change of Address File

No records
0.0% - Mean National PIK Rate (74.62%)
Above Mean National PIK Rate (74.62%)
100.0%

Source: U.S. Census Bureau,
2009 National Change of Address File

Sociodemographic differences in match likelihood

Given the relatively small number of ACS movers who were retained for analysis once matched to an NCOA record, we explored some basic demographic differences between ACS movers with and without a PIK-matched NCOA record. In each of the following figures, *Matched* refers to ACS movers with a PIK that could be matched to a single PIK on a NCOA record. *Unmatched* refers to ACS movers with a PIK, for whom a NCOA record with the same PIK was not found. 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.

Figures 6 and 7 show the distributions by race and Hispanic origin. Somewhat surprisingly, the distributions were not significantly different by ability to match. Both groups were disproportionately non-Hispanic White. The explanation for this lies in the restriction to examining ACS movers with PIKs. The ACS movers, whom were not able to receive a PIK in the PVS system, were significantly different from movers with PIK. They were more likely to be Hispanic and non-White.
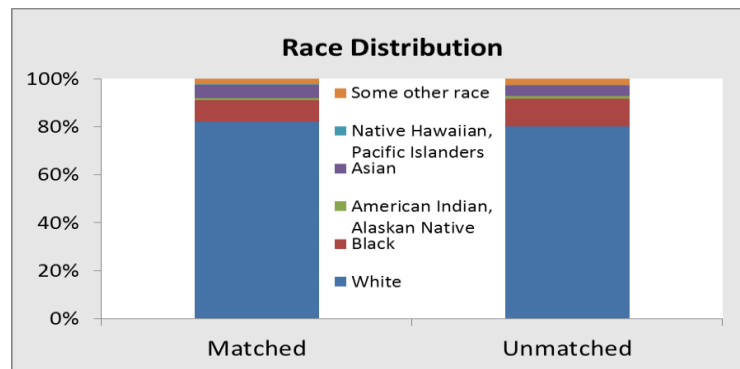

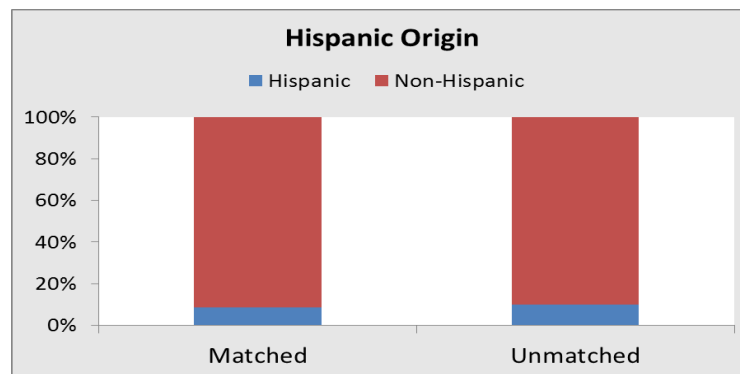**Figure 6: Distribution of ACS movers matched to NCOA records by race**

Figure 7: Distribution of ACS movers matched to NCOA records by Hispanic Origin

Even given the above selectivity of the group of ACS movers with PIK, significant differences were evident between the matched and unmatched groups for educational attainment, income, marital status and tenure. As mentioned earlier, 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 and young adults in the NCOA dataset is evident in both the larger share of never-married persons in the unmatched bar for marital status, and the under 25 universe without a value for completed educational attainment. Application of a flag on family moves associating the information from a single PIK to everyone in the household who was a part of the described move would include more young people in the pool of matched ACS movers to an NCOA record.

Educational attainment in Figure 8 shows a clear, positive relationship between higher education and the likelihood of being matched. Figures 9 and 11, or what may be seen as two outcomes of education, show those with higher income and those with a mortgage are more likely to be matched. Once someone has an annual income of $50,000 or more, they are more likely to be matched. And those who have a mortgage or who have finished paying their mortgage are more likely than renters to be matched across datasets as well.
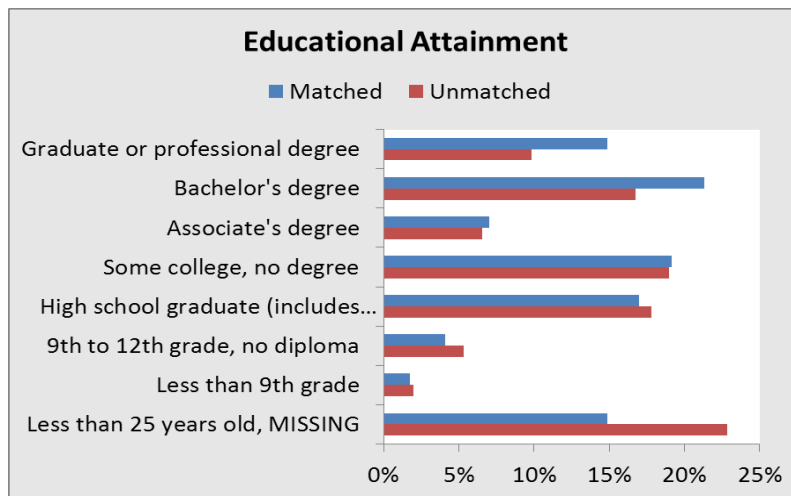

Figure 8: Distribution of ACS movers matched to NCOA records by Educational Attainment
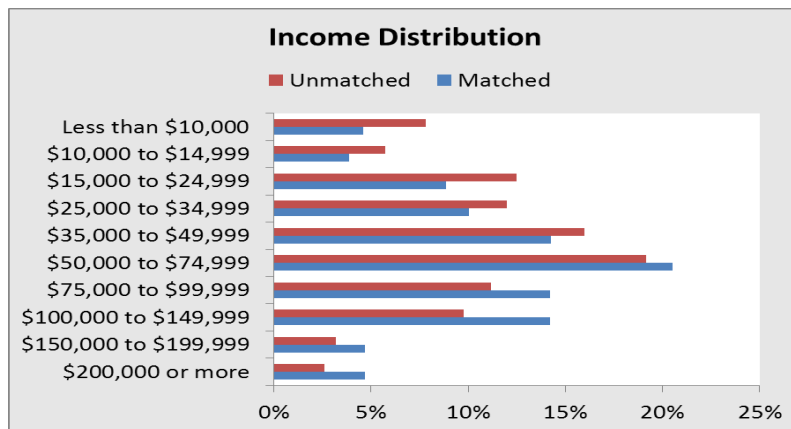

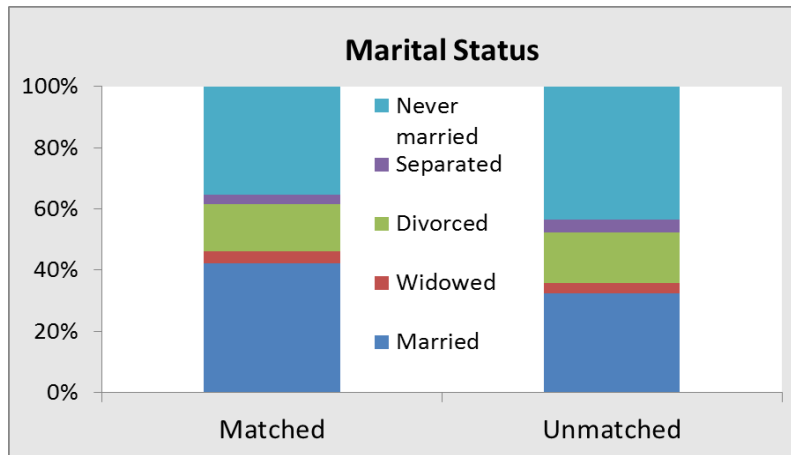**Figure 9: Distribution of ACS movers matched to NCOA records by Income**

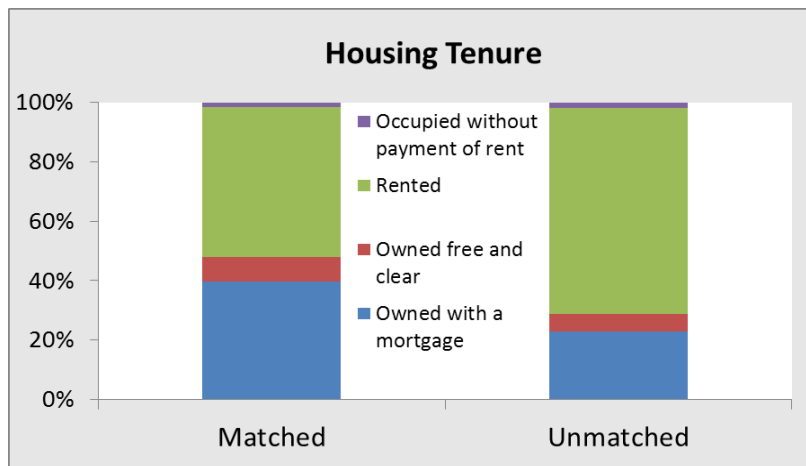**Figure 10: Distribution of ACS movers matched to NCOA records by Marital Status**

Figure 11: Distribution of ACS movers matched to NCOA records by Housing Tenure

Geography

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.

For 2009 NCOA, upon receipt of the monthly deliveries the Census Bureau attempted to put a Master Address File identification code (MAFID) on both the origin (moving from) and destination (moving to) address reported in the NCOA record. Using those stringent criteria, only about 38 percent of the NCOA records obtained 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.

All ACS movers have a 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. In any given year, the supplied addresses are processed through normal geocoding operations to obtain a set of census geocodes, typically down to census block-level geography. ACS "origins" were not assigned a MAFID in the geocoding process.

Determining a geographic move match was restricted to the levels of coded geography that were available from each of the sets of data. The utility of MAFID was not high because it was not represented for both parts of the address pairs at a sufficient degree on the linked data. While not precise, a county-level code for both ACS and NCOA pairs proved to have the most utility for this analysis. Sixty-three percent of the linked ACS-NCOA move records shared the same origin and destination at the <u>county level</u>. Ten percent had the same origin but a different county destination; 14 percent had the same destination but a different origin; and for seven percent neither origin nor destination county matched.
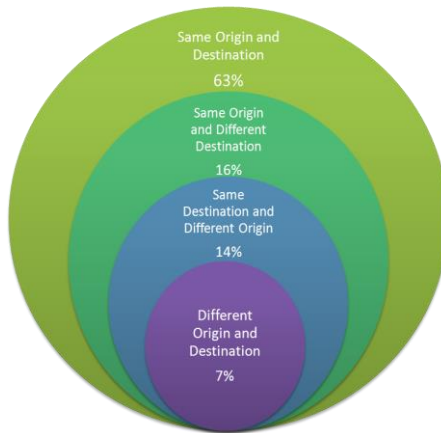

Figure 12: Match Rate on Origin and Destination County Geography for the ACS/NCOA Matched Movers

Month of Move

As mentioned, the month and year of the effective move date as reported on the NCOA form was designated as the NCOA move date for this analysis. For ACS, the move date was the month and year, one year prior to the interview month. Figure 13 then shows the distribution of the month of move for all singly PIKed move records in the NCOA and ACS. The rate of PIKed ACS cases appears to vary from month to month. NCOA's curve is much smoother, peaking in the summer months.
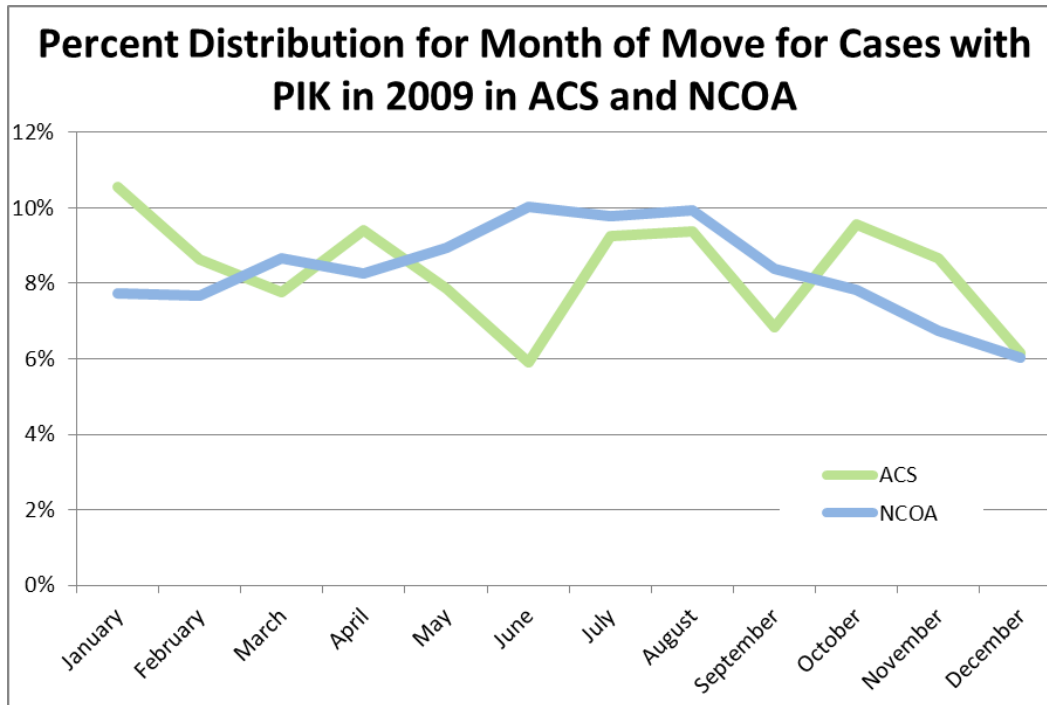
Percent Distribution for Month of Move for Cases with PIK in 2009 in ACS and NCOA

Figure 13: Month of Move for all single PIK mover records in ACS and NCOA

The variation in the months of move show typically higher rates in the summer months as have been evident in other analyses of survey data (Marlay and Mateyka, 2011; Marlay and Fields, 2010). The lower rates in NCOA for the fall and winter months may be an issue of when the NCOA form is completed. Since NCOA forms are not mandatory, it may be that fewer forms were completed during those time periods, or the characteristics of movers in those time periods were less likely to complete a form, or less likely to be able to receive a PIK. These were not elements that could be fully addressed in this preliminary analysis.

In comparison, Figure 14 shows the same data but for only the ACS movers linked to an NCOA record by PIK. Respondents' move month in ACS was more likely to match NCOA's move effective date in the beginning of the calendar year, from January to May. For the second half of the year, the ACS move month and NCOA move effective date were less likely to match.
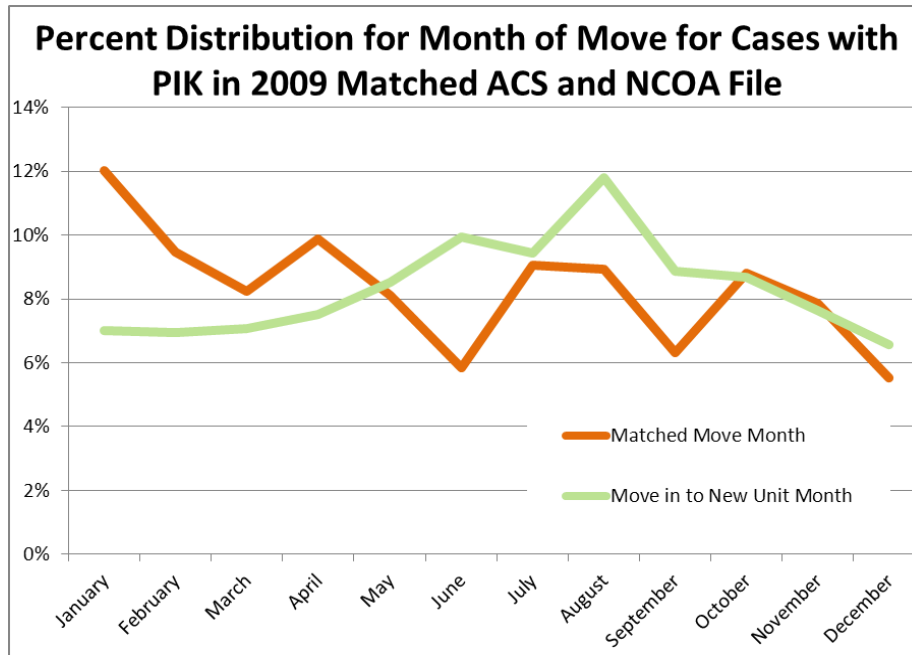
Figure 14: Month of ACS Move, and NCOA effective move date for linked ACS and NCOA file movers

5. **Conclusions, Future Work, and Other Considerations**

Continuing research will explore the records omitted from this analysis and assess the utility of NCOA for various other 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. However, 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. 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.

6. References

Benetsky, M. and William Koerber. (2012). How do the ACS five-year migration data compare to the 2000 Census migration data? Journey to Work and Migration Statistics Branch. Social, Economic and Housing Statistics Division. Working Paper Number 2012-13. http://www.census.gov/hhes/migration/files/PAA_2012_County-County.pdf

Marlay, Matthew C., and Peter Mateyka .(2011). The Seasonality of Moves: 2009. SEHSD Working Paper.http://www.census.gov/hhes/migration/files/2009-Seasonality.pdf

Marlay, Matthew C., and Alison K. Fields. (2010). Seasonality of Moves and the Duration and Tenure of Residence: 2004. Current Population Reports, P70-122. U.S. Census Bureau, Washington, DC.

NORC/University of Chicago. 2011. *Assessment of the U.S. Census Bureau's Person Identification Validation System*. Final Report presented to the U.S. Census Bureau. http://www.norc.org/Research/Projects/Pages/census-personal-validation-system-assessment-pvs.aspx.