

1-1-2006

# Using the Census Bureau's Public Use Microdata for Migration Analysis

Mark Salling


Cleveland State University, [m.salling@csuohio.edu](mailto:m.salling@csuohio.edu)

Ellen Cyran

Cleveland State University, [e.cyran@csuohio.edu](mailto:e.cyran@csuohio.edu)

## How does access to this work benefit you? Let us know!

Follow this and additional works at: [https://engagedscholarship.csuohio.edu/urban\\_facpub](https://engagedscholarship.csuohio.edu/urban_facpub)

 Part of the [Categorical Data Analysis Commons](#), [Databases and Information Systems Commons](#), [Geographic Information Sciences Commons](#), [Longitudinal Data Analysis and Time Series Commons](#), and the [Urban Studies Commons](#)

---

## Repository Citation

Salling, Mark and Cyran, Ellen, "Using the Census Bureau's Public Use Microdata for Migration Analysis" (2006). *Urban Publications*.  
0 1 2 3 418.

[https://engagedscholarship.csuohio.edu/urban\\_facpub/418](https://engagedscholarship.csuohio.edu/urban_facpub/418)

This Conference Proceeding is brought to you for free and open access by the Maxine Goodman Levin College of Urban Affairs at EngagedScholarship@CSU. It has been accepted for inclusion in Urban Publications by an authorized administrator of EngagedScholarship@CSU. For more information, please contact [library.es@csuohio.edu](mailto:library.es@csuohio.edu).

**Mark Salling, Ph.D., GISP**

**Ellen Cyran, GISP**

Northern Ohio Data & Information Service (NODIS)

Maxine Goodman Levin College of Urban Affairs

Cleveland State University

## **USING THE CENSUS BUREAU'S PUBLIC USE MICRODATA FOR MIGRATION ANALYSIS**

**Abstract:** The paper reports on the use of the Census Bureau's Public Use Microdata Sample (PUMS) to analyze migration to and from Cleveland-Akron-Lorain Metropolitan Area. Discussion of the PUMS database, its geographic components, and the results of the migration analysis are presented. Thus the purpose is two-fold – to inform the reader about the usefulness of the data and to illustrate its use with a brief descriptive analysis of migration to and from Northeastern Ohio.

PUMS data enable the researcher to calculate custom cross-tabulations and summary statistics of population and housing. Among the data available in PUMS is the location of the person five years earlier; for the 2000 Census the data provide the 1995 place of residence and for the American community Survey (ACS) the data provide the residential location one year ago. Thus with the 2000 PUMS data the user can generate the characteristics of persons who moved to a region between 1995 and the 2000 census, whether from other parts of the nation or from abroad. In addition, the user can identify people who lived elsewhere in the United States at the time of the census but reported that they lived in the region in 1995. Thus the researcher can compare movers to and from the region. This database is a rich source of information about where a region draws migration from and where its migrants move to - important knowledge for regional, community, and economic development planning.

### **INTRODUCTION**

Micro-level data files on persons, households, and housing provided by the U.S. Census Bureau are a valuable resource for research that is not possible with the standard summary-level tabulations that are more commonly understood and used. This paper provides an overview of the data and an example of its use in research on migration to a from the Cleveland-Akron-Lorain Metropolitan Area.

We provide some sources of data and more information about these data at the end of the paper.

### **What is Census Public Use Micro Data?**

The Public Use Microdata Sample (PUMS) data uses actual survey responses from the decennial census long form or, more recently, the American Community

Survey (ACS). The data are edited to protect the confidentiality of individuals. PUMS data have many of the housing and population characteristics available in the decennial census and ACS survey summary tables.

PUMS data enable the researcher to calculate custom cross-tabulations and summary statistics of the population. Statistics estimated from a sample are subject to sampling error. Small numbers and small differences in numbers are subject to sampling error to a greater extent and are less reliable in representing the population than are larger numbers and differences. We provide a brief discussion of confidence intervals for PUMS data at the end of this paper.

PUMS data in particular are based on a relatively small sample (compared, for example to the national 17 percent sample for the 2000 census) and largely for this reason the geographic information in PUMS data are limited. Small areas such as census tracts and even medium to small cities are not identifiable for the sampled subjects. Instead, PUMS data identify the Public Use Micro Areas (PUMAs) for each subject sampled. PUMAs are areas comprised of 100,000 or more population and are combinations of census tracts, cities, townships, villages, or counties.

In addition, the Census Bureau combines PUMAs to whole counties (one or more) to create Migration PUMAs for reporting where migrants moved from. Thus, for each PUMA we know the Migration PUMA from which people moved. For the migration analysis discussed below we aggregate the region's PUMAs and Migration PUMAs to the eight-county metropolitan area. The PUMA and MigPUMA geography for northeast Ohio is shown in Map 1.

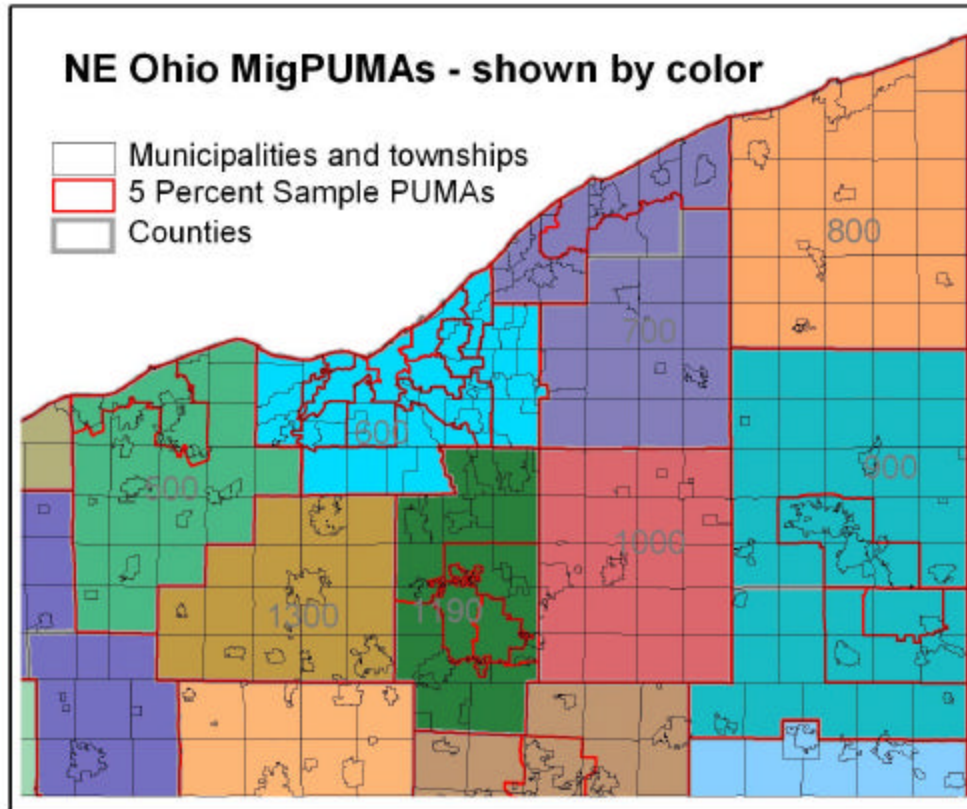
### **How is Migration Analysis Possible Using PUMS Data?**

In regard to migration analyses, the five percent PUMS of the 2000 Census of Population and Housing provides the location of the resident five years earlier, that is in 1995. There are migration data fields for state, MSA/CMSA, and migration PUMAs which are counties or groups of counties. Many PUMAs cross MSA/CMSAs so analysis for all MSAs is not directly possible with PUMS. Starting with 2005 ACS, census data also provide migration fields by state and migration PUMA for the previous year. The ACS 2005 PUMS is only a one percent sample based on the annual ACS survey. The ACS data also differ from the decennial census data in that the survey is taken throughout the year rather than at a point in time (April) as is the case with the 2000 decennial census.

The PUMS data allow analysis of migrants to a certain region and also those who left a region. We can compare the demographic, socioeconomic, and housing characteristics of different groups of migrants, e.g., domestic migrants into and out of a region and foreign migrants into a region. Foreign migrants out of a region are not available from PUMS since the census does not enumerate population in other countries. Foreign migrants to a region can be compared to other foreign migrants to other place or the nation as a whole, and to domestic

migrants to and from the region. In addition to comparing the migration groups, one can also compare the regions to (within the United States) and from which a region's migrants move.

MAP 1: PUMA AND MIGPUMA GEOGRAPHY IN NORTHEAST OHIO



#### **How does the ACS PUMS Data Differ from the 2000 Census PUMS?**

In 2005, PUMS data from the ACS became available for PUMA level geography. Prior to this it was only available at the state level. ACS PUMS data constitute a one percent sample and, like the American Community Survey summary tables, do not include group quarters population. Inclusion of group quarters population is planned for the 2006 ACS, however. For migration, instead of asking about one's residential location five years ago as in the 2000 Census, the ACS asks 'Did this person live in this house or apartment 1 year ago?'

#### **How are the Data Structured in Regard to Housing and Population Records?**

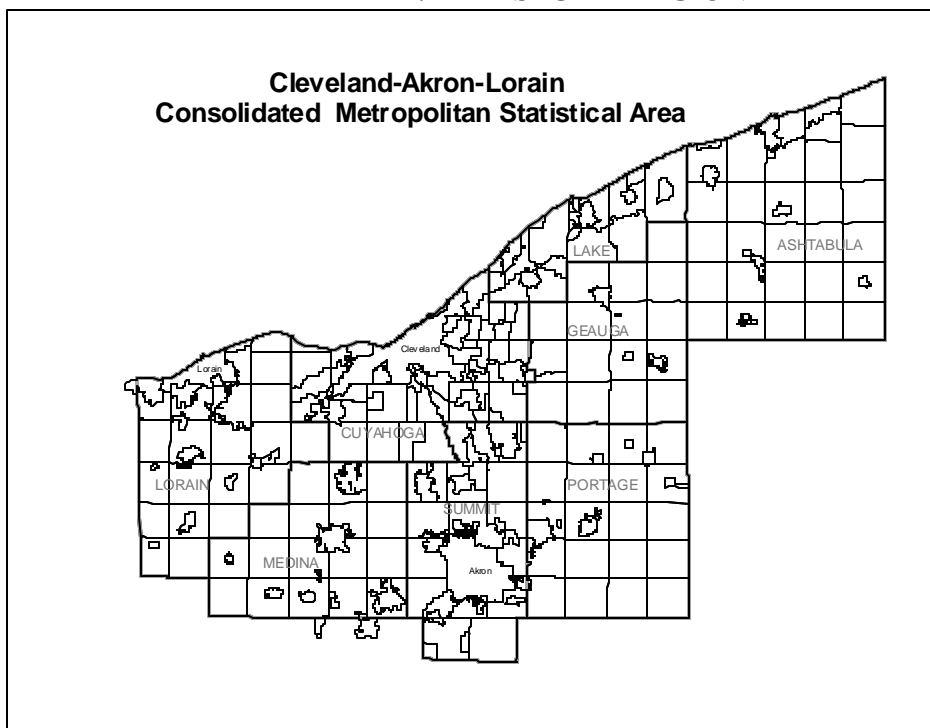
Both 2000 Census PUMS and ACS PUMS have separate housing and population records. The housing records provide data for occupied and vacant housing units and the population records provide household population data and, in the case of the 2000 PUMS, also data about group quarter population. The housing records each have a housing weight, as do the population records, which have different values for each member of the household.

The relationship to the householder for all persons is available and each person record includes a household identifier. Thus the housing data can be joined with the population data to facilitate programming and analysis.

## **MIGRATION TO AND FROM THE CLEVELAND-AKRON-LORAIN CMSA**

Our studies have mainly focused on the Cleveland-Akron-Lorain (CAL) Consolidated Metropolitan Area (CMSA) using the 2000 PUMS. The CAL CMSA includes Ashtabula, Cuyahoga, Geauga, Lake, Lorain, Medina, Portage, and Summit Counties within Northeast Ohio.

MAP 2: THE STUDY REGION



### **Data and Methodology**

Among the data available in the 2000 Census PUMS is the location of the person five years earlier, that is in 1995. Thus we can generate the characteristics of persons who moved to the CAL between 1995 and the 2000 census, whether from other parts of the nation or from abroad. In addition, we can identify people who lived elsewhere in the United States (including Puerto Rico) at the time of the census but reported that they lived in the CAL in 1995. Migrants reported in PUMS are therefore at least five years old in the 2000 census.

Software. SAS® software was used for processing of the data. Results were produced using macros for repeated processing with different universes, different variables, and to produce confidence intervals. The main statistical functions used were frequencies, means, and medians using PROC FREQ and PROC UNIVARIATE. The SAS® output delivery system (ODS) was used to quickly export multiple tables to Microsoft® Excel worksheets.

Migration Datasets. The out-migrants were first extracted for Ohio from files from all 50 states and Puerto Rico based on migration state or residence five years ago. The in-migrants were extracted from the Ohio PUMS file based on their residence in 1995 or five years ago being Ohio. Both of these datasets are for persons age five and older. After the Ohio data were extracted, the data were subset for the Cleveland-Akron-Lorain CMSA.

For foreign migrants, the data were subset from the in-migrants file for those who's migration country was not a U.S. state or Puerto Rico. Foreign in-migrants for the whole U.S. were also extracted from files for all 50 states and Puerto Rico based on the same criteria.

Foreign and Domestic Migration. Domestic migrants include all migrants within the United States or Puerto Rico. For analysis of the foreign migrant population, the population included anyone moving from outside the United States. Foreign migrants include both U.S. born and foreign-born populations that lived in another country in 1995.

Though a comparable set of data concerning those persons who moved abroad during this same period would be useful, detailed information about foreign out-migrants is not available from the PUMS census data. However, we estimate migration of 22,185 persons moving from the CAL to other nations.<sup>1</sup>

---

<sup>1</sup> We estimate the region's emigrants to other nations by using the net international migration estimate for 2000 that is provided by the Census Bureau (See <http://www.census.gov/popest/estimates.php> for information about these data; download the entire dataset at <http://www.census.gov/popest/counties/files/CO-EST2004-ALLDATA.csv>). We multiplied the Census Bureau's estimate of 4,748 migrants from the CAL to other countries in 2000 by five to estimate the five-year total from 1995 to 2000. Since this estimate of 23,740 would include children born between 1995 and 2000, we also subtract out the estimated number of children by applying the percentage of the general population of the region that is in that age cohort (6.55%).

Weighting. The unweighted and weighted counts for various populations used in our migration analysis are provided in Table 1.

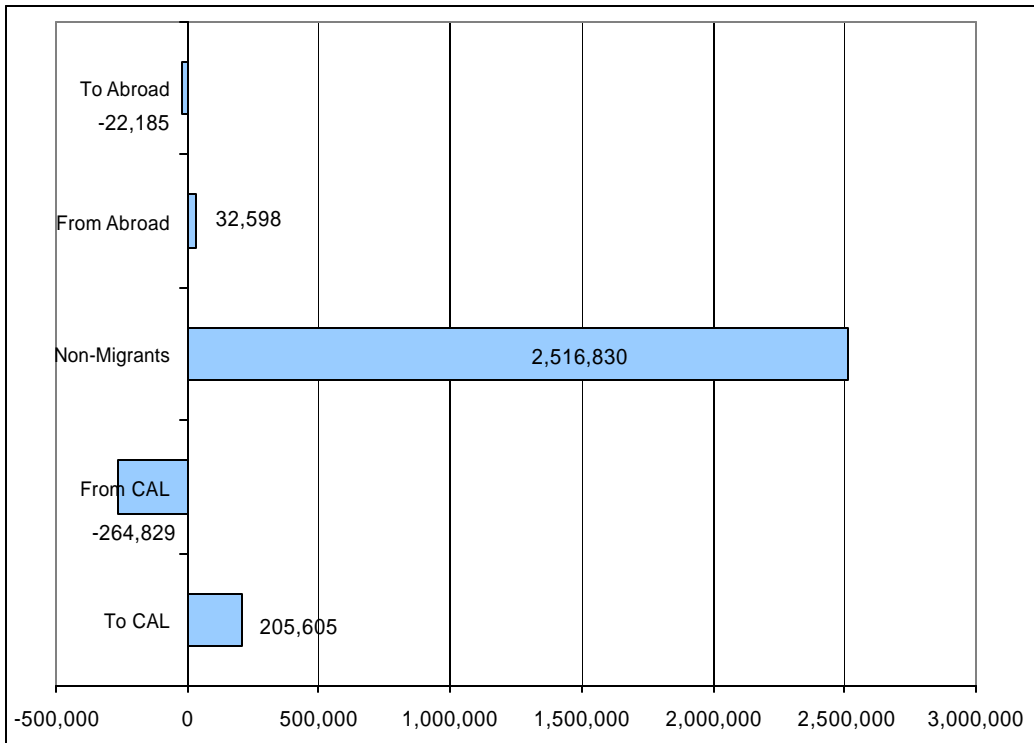
TABLE 1: WEIGHTED AND UNWEIGHTED PUMS RECORDS

Census 2000	Unweighted	Weighted	Percent
Ohio Person Records	569,795	11,353,531	5.0%
CAL Person Records	139,640	2,948,392	4.7%
CAL Out-migrants Person Records	12,660	264,829	4.8%
CAL Foreign In-migrants Person Records	1,481	32,598	4.5%
CAL Domestic In-migrants Person Records	9,379	205,605	4.6%
<b>ACS 2005</b>			
Ohio Person Records (without Group Quarters)	117,251	11,146,050	1.1%

**What Did We Find?**

With an estimate from PUMS of 32,600 moving to the CAL from abroad and an estimated 22,200 moving abroad, the region had a net gain in international migration of an estimated 10,400 persons during this period. Thus while the region lost approximately 59,200 in net migration with the remainder of the nation, some of that loss was mitigated by positive net international (see Figure 1). As a result, the region lost approximately 48,800 through total net migration.

FIGURE 1: NUMBER OF MIGRANTS



## Foreign Migration

For this discussion we focus on the foreign migrants, though we also compare them to domestic migrants as well.

The summary findings from our foreign migration analysis include the following:

1. With a net increase of more than 23,000 persons from abroad, foreign migration between 1995 and 2000 helped mitigate the 59,000 person net loss from domestic migration to approximately 48,800 persons lost through migration in the five-year period.
2. The largest single group of migrants to the region from outside the fifty states was Puerto Rico.
3. Aside from this population, the CAL's foreign migrants from this period were largely Eastern European or Asian.
4. Demographically, they were more likely to be older, married, and in families with children than foreign migrants to the rest of the nation. The CAL's foreign migrants were evenly split in gender, which was different from the pattern of mostly male foreign migrants to the state and nation.
5. In terms of housing, the CAL's foreign migrants were largely housing renters; and while those that owned their homes had, on average, housing valued higher than the region's average, they were less valued than those owned by movers to the region from other parts of the country.
6. Unemployment and poverty rates were higher than those of non-migrants in the CAL, though their poverty rate was essentially the same as that of the state's and nation's groups from this period.
7. Despite the higher unemployment rate, they were more likely to be in technical and higher skilled occupations, such as in computer and mathematical, education, science, and engineering categories, than either the region's non-migrants or foreign migrants to the nation as a whole.
8. Though the region did not benefit from large numbers of migrants from abroad (compared to the rest of the nation), it did receive a generally more educated foreign population. They had higher percentages of persons with a bachelor's degree or higher than did the region's non-migrants, domestic migrants to or from the region, and other foreign migrants to the U.S. They were also more often attending college in 2000 than the general population of the region and the other foreign migrants to the nation.

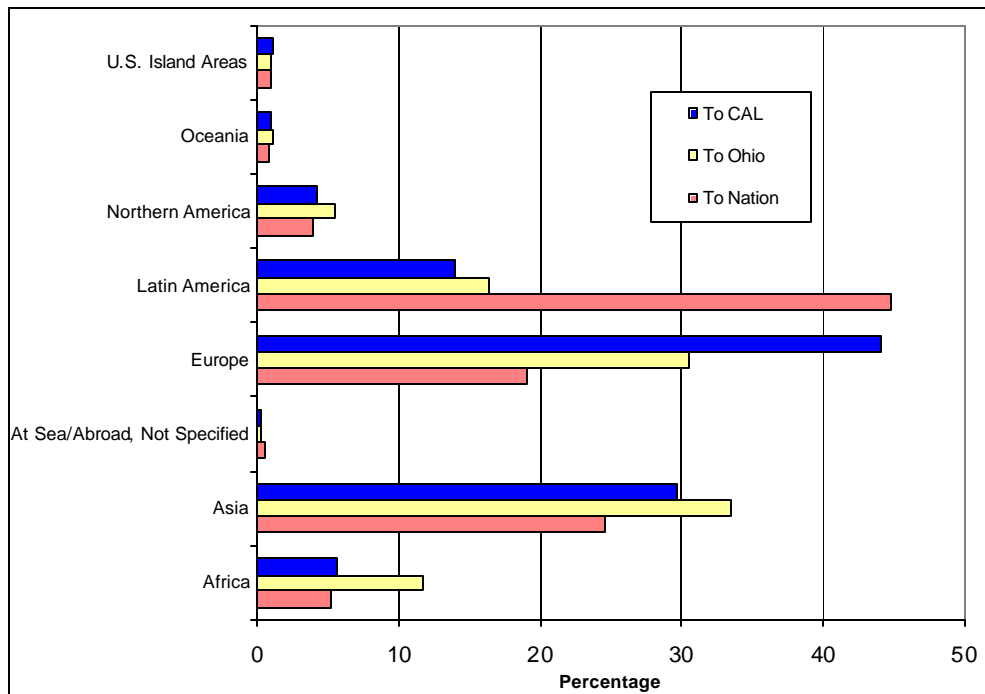


## Discussion

The initial set of characteristics used in analyzing the foreign migration were place of origin, age, race, Hispanic ethnicity, nationality/immigration status, gender, family/household type, educational attainment, employment status, industry of employment, occupation, income, housing tenure, cost of housing, type of housing structure. A selection of the analysis is presented here.

A significant portion (15.9 percent) of the migrants to the region were from abroad (see Figure 2). In fact, one in 100 (1.1 percent) of the region's total population in 2000 were foreign migrants. More than 2.7 percent of the nation's population had migrated to the U.S. since 1995. Thus the region had a lower rate of foreign migration than the nation. Most of this difference is due to the relatively large influx of migrants from Latin American countries into the Southwest and Southeast regions of the country.

FIGURE 2: CONTINENTS OF FOREIGN MIGRATION

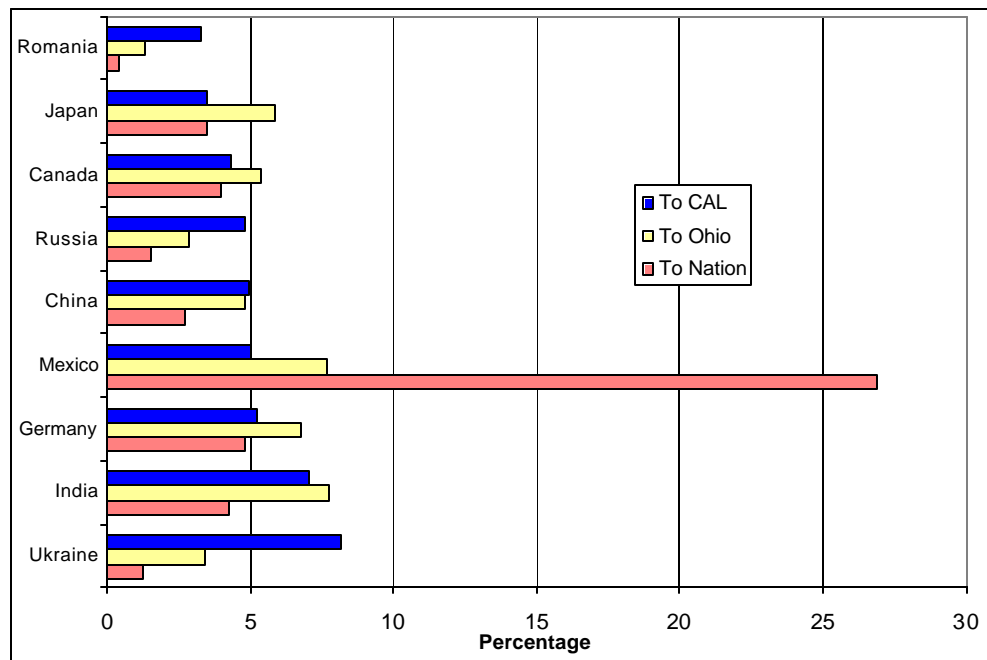


The countries with the largest number of emigrants to the CAL were the Ukraine (2,663) and India (2,303), together accounting for 15.3 percent of the migrants from abroad. The CAL also attracted a large proportion of persons from Russia and Romania relative to the nation and state (see Table 2 and Figure 3).

TABLE 2: MAJOR COUNTRIES OF MIGRANTS, 1995 TO 2000

Country	Number moving to:			Percent of Total Foreign Migrants to:		
	Nation	Ohio	CAL	Nation	Ohio	CAL
Ukraine	93,764	3,887	2,663	1.3	3.4	8.2
India	309,095	8,810	2,303	4.2	7.8	7.1
Germany	351,432	7,711	1,706	4.8	6.8	5.2
Mexico	1,963,155	8,770	1,636	26.9	7.7	5.0
China	196,524	5,481	1,618	2.7	4.8	5.0
Russia	112,487	3,229	1,554	1.5	2.8	4.8
Canada	289,293	6,134	1,402	4.0	5.4	4.3
Japan	253,385	6,649	1,122	3.5	5.9	3.4
Romania	28,643	1,507	1,056	0.4	1.3	3.2

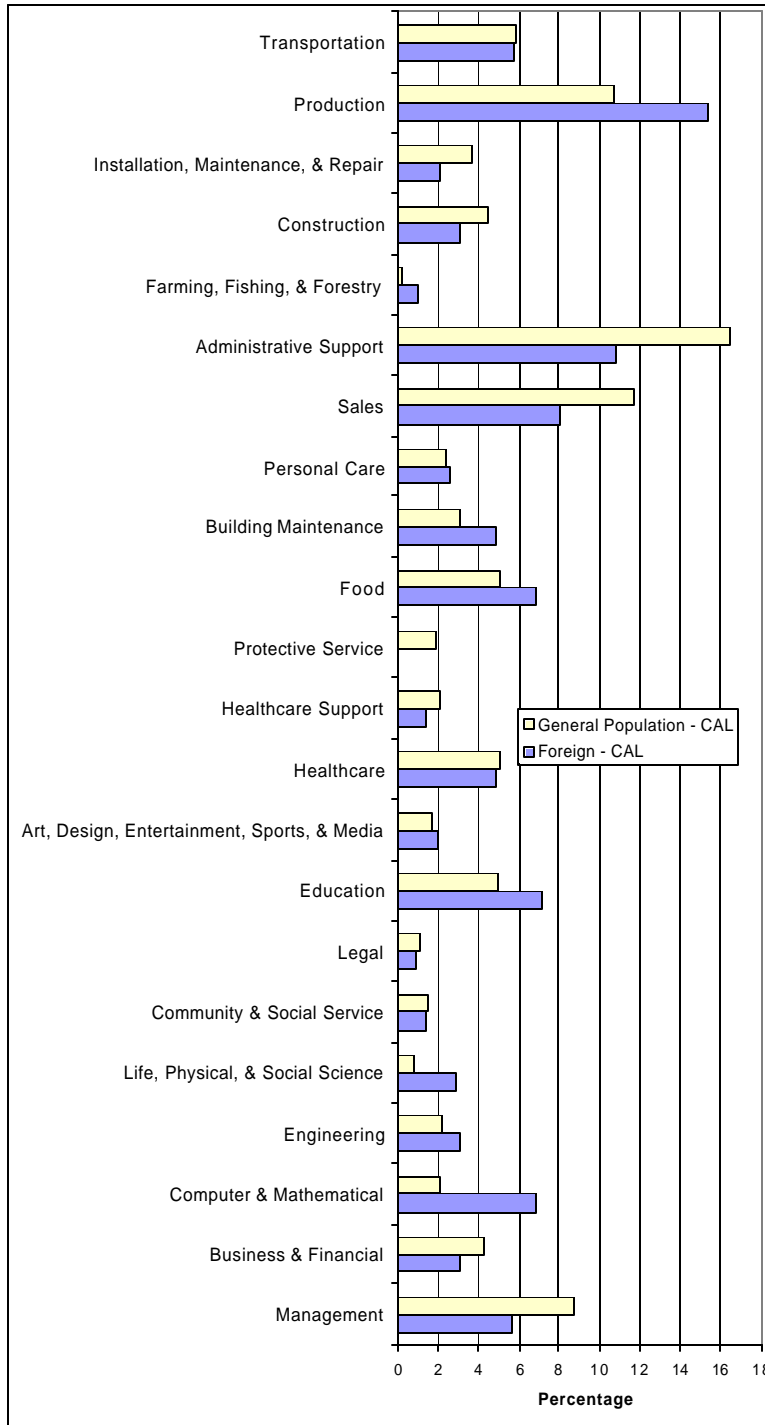
FIGURE 3: MAJOR COUNTRIES OF MIGRANTS



The occupations among foreign migrants to the CAL approximate the major types of occupations of the general population of the region, although there are some important differences as well (see Figure 4). The largest major category among employed civilian foreign migrants to the CAL was production at 15.3 percent -- higher than that of the general population, which was 10.5 percent. Among the more specific occupations in this category were metalworks, assemblers, machinists, and electrical assemblers. These workers were also more concentrated

in education, computer and mathematical, engineering, and science categories. On the other hand, the foreign migrants to the CAL were less likely to be among the managerial, administrative support, sales, construction, and repair occupations than the general population.

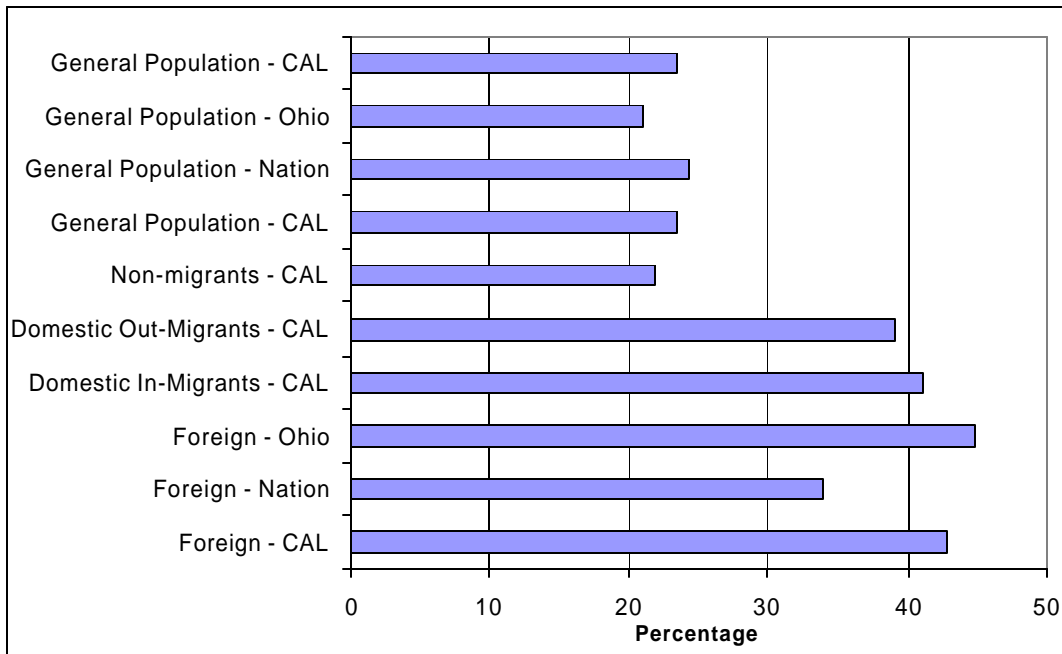
FIGURE 3: OCCUPATIONS



The CAL generally scored well in regard to educational attainment of its population (see Figure 4).<sup>2</sup> Ohio and the CAL had a greater number of foreign migrants age 25 and older with bachelor's degrees or higher attainments compared to the nation. Forty-five (45 percent) of Ohio's foreign migrants age 25 and older had a bachelor's degree or higher, while 43 percent of those coming to the CAL and 34 percent of those migrating to the nation had a bachelor's degree or higher.

In addition, the foreign migrants were slightly more educated than other migrants to the CAL. Forty-one percent of domestic migrants to the CAL had a bachelor's or higher degree, and out migrants were even less educated with 39 percent having a bachelor's degree or higher. Meanwhile, all these migrant groups were much more educated than the region's non-migrants, as only 22 percent of them had a bachelor's degree or more education.

FIGURE 4: WITH A BACHELOR'S DEGREE OR MORE (AGE 25+)

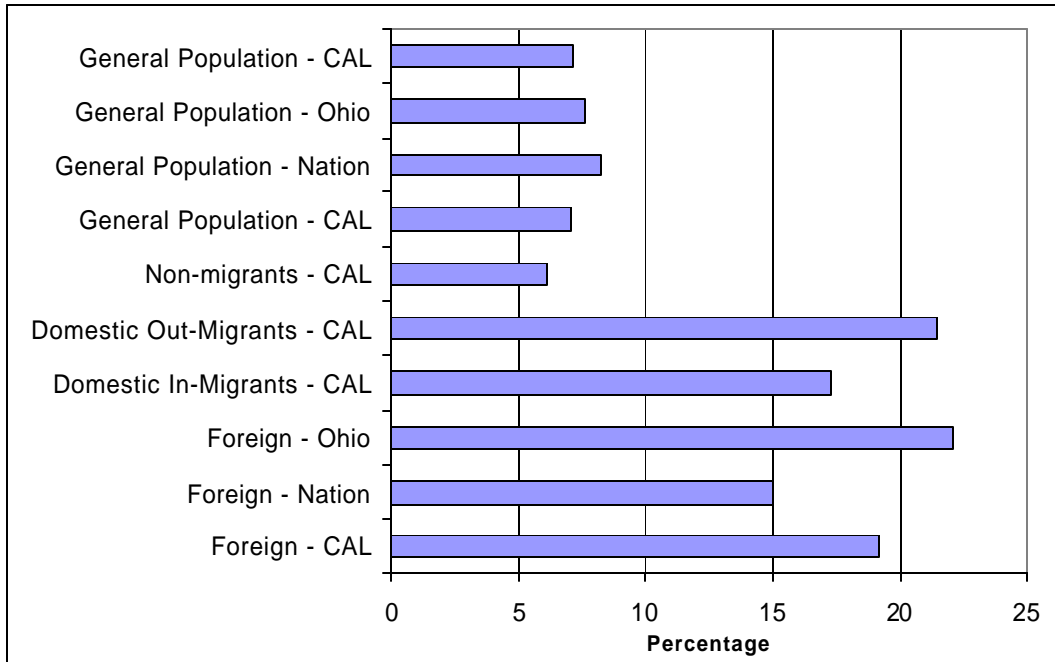


In addition, based on college enrollments in 2000, the CAL's foreign migrants continued to acquire more education than the general population -- 19 percent of these persons age 18 and older were in college compared to seven percent of the general population of that age in the CAL. A substantial number (17 percent) of persons moving to the CAL from somewhere else in the nation were also enrolled in college. However, foreign migrants to the CAL were less likely to be enrolled in college in 2000 than were domestic out-migrants from the CAL (19 percent versus 21.5 percent).

<sup>2</sup> Indeed, as can be seen in Figure 4 and contrary to assumptions by many in the region about a "brain drain" from the region, the region attracted a greater proportion of migrants to the region with a bachelor's degree than it lost,

In 2000, even though the state had a lower percentage of its population enrolled in college than the nation, Ohio attracted a higher proportion of foreign migrants pursuing a college education than either the nation or the CAL (22 percent versus 15 and 19 percent, respectively). Thus, the state and the region stood to benefit more than the nation from the educational aspirations of foreign migrants.

FIGURE 5: ENROLLED IN COLLEGE



## CONCLUSION

PUMS data provide a good analytical tool for a variety of research topics, particularly regional migration analysis. Characteristics of movers to and from a region can easily be compared to each other and to the population that did move out of the region.

In addition, though not presented here, the characteristics of the places to and from which the population moved can be compared. Does Northeast Ohio attract populations from older, industrial cities or rural and small town settings? Where are the region's out-migrants going? The jobs or warmer whether locations or other older industrial cities?

While the analysis discussed here uses the 2000 Census PUMS, the American Community Survey PUMS provides an opportunity to monitor changes in migration on a yearly basis, particularly once the survey reaches full implementation in the coming years.

## APPENDIX

### SOME SOURCES FOR CENSUS PUBLIC USE MICRODATA

The 2000 5% PUMS text files for each state can be obtained at <http://www.census.gov/Press-Release/www/2003/PUMS5.html>. The Census Bureau also provides DataFerrett for extracting data from the PUMS samples.

Other sources for PUMS data are Missouri State Data Center, which has SAS datasets for all states and Puerto Rico. These are available at <http://mcdc2.missouri.edu/cgi-bin/uexplore?/pub/data/pums2000>. These datasets can also be available for remote access directly through a local SAS session.

The Integrated Public Use Microdata Series (IPUMS)<sup>3</sup> at <http://www.ipums.org/> provides PUMS data across time. The series provides most, but not all fields available on the Census PUMS and the data is standardized for comparison across time.

Quick crosstabs of the ACS 2005 PUMS are available from UC Data at UC Berkeley at <http://ucdata.berkeley.edu:7101/ACS2005/>.

### CONFIDENCE INTERVALS

Confidence intervals are ranges of values that are likely to contain the true value. 90 percent confidence intervals are often used with the Census data. The summarized ACS data includes 90 percent confidence intervals with all data that is released. With the possibility of producing crosstabs with very small numbers the confidence intervals are very important in determining the possible range of the data.

The 2000 Census PUMS has two methods to produce confidence intervals. These are documented in the 2000 Census of Population and Housing Public Use Micro Sample (<http://www.census.gov/prod/cen2000/doc/pums.pdf>) in Chapter 4 on Accuracy of Microdata Sample Estimates.

The first method is to estimate the confidence intervals using tables provided by the Census Bureau in the technical documentation. This method though easier is not as accurate as producing the intervals directly from the sample. The PUMS data has 100 subsamples available for producing confidence intervals directly from the sample with the random group method provided in the documentation.

---

<sup>3</sup> Steven Ruggles, Matthew Sobek, Trent Alexander, Catherine A. Fitch, Ronald Goeken, Patricia Kelly Hall, Miriam King, and Chad Ronnander. *Integrated Public Use Microdata Series: Version 3.0* [Machine-readable database]. Minneapolis, MN: Minnesota Population Center [producer and distributor], 2004.