Using IPUMS US Microdata to Study Immigration and Migration
Grace Cooper and Steph Richards
Overview

• Migration/Immigration in microdata collections
• Collection comparability
• Highlights of each collection with research examples
• Additional resources
IPUMS US Microdata Collections

- IPUMS USA
- IPUMS CPS
- IPUMS ATUS
- IPUMS NHIS
- IPUMS MEPS
Comparing Migration/Immigration Across Microdata Collections
Measuring Immigration

- These data capture information about the current resident US population.
- Cover place of birth.
- Parents' places of birth and origins are available for some collections.
- Cannot capture information about persons born in the US who have emigrated to other countries.
Measuring Migration

Some data collections include information about recent moves:

- **USA ACS/PRCS**: place of residence 1 year ago and 5 years ago
- **Linked Historical Census**: place of residence, tracked across multiple decennial censuses
- **CPS**: recent past state of residence and migration status, reason for moving, Hurricane Katrina evacuation variables
Foreign-born estimates by collection

Percent Foreign Born by IPUMS US Microdata Collection

Collection:
- ATUS
- CPS
- MEPS
- NHIS
- USA
Commonalities between data collections

• Showcase a representative sample of the US population*

• Reasonably similar measures of the US-born population and years lived in the US

• Include common socio-demographic traits (age, race, sex, education, occupation, family composition)

* Institutionalized populations are only captured in USA data files. Historical full-count census data are also available.
Comparability between data collections

• Internal migration only available in some data collections

• Different levels of detail in demographic and family relationship variables
  – Varying definitions of family composition

• Variation in temporal coverage of data collection & immigration-related variables
  – Historical coverage in USA only
Availability of immigration variables

- USA
- CPS
- ATUS
- NHIS
- MEPS

Legend:
- Nativity
- Timing of Migration
- Mobility
Differences between data collections

Topical coverage varies from collection to collection. Covariants of interest may include:

- Geographic location and metropolitan status
- Internal migration
- Second-generation immigrant status
- Health conditions, outcomes, and expenditures
- Detailed occupation and household income
- Living conditions and facilities
- English language fluency
Key Features of Different IPUMS US Collections
Overview of IPUMS USA

Public Use Microdata Samples
- U.S. decennial census, 1850-2010
- American Community Survey and Puerto Rico Community Survey, 2000-present

Full-Count Historical Census data
- 19th Century: 1850 - 1880
- 20th Century: 1900 - 1940
- 1950 available soon!

Notable Features of USA
- Large sample sizes (ACS is 3+ million records)
- Broadest temporal coverage of all data collections
- Detailed geographic information, PUMA (Public Use Microdata Area) is smallest identifiable unit
- Wide array of migration topics
- Comprehensive covariant list
- IPUMS MLP offers the opportunity to study migration cross-generationally
Example Research Question

How is language influenced by multigenerational living?

Number of Spanish speakers who speak English summarized by generationals in household

<table>
<thead>
<tr>
<th>Generations in Household</th>
<th>Speak English</th>
<th>Do not speak English</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>1 Single adult</td>
<td>4,384,990</td>
<td>85</td>
</tr>
<tr>
<td>2 Parent and child</td>
<td>10,585,390</td>
<td>87</td>
</tr>
<tr>
<td>3 Grandparent, parent, and child</td>
<td>2,377,153</td>
<td>82</td>
</tr>
</tbody>
</table>

Data analysis in R

data %>%
  filter(YRSUSA1>0 & SPEAKENG>0 & MULTGEN>0 & LANGUAGE==12) %>%
  mutate(english = case_when(SPEAKENG %in% c(1) ~ "No",
                            SPEAKENG %in% c(3,4,5,6) ~ "Yes")) %>%
  group_by(english, MULTGEN) %>%
  summarize(n = sum(PERWT)) %>%
  spread(english, n)
Information about when an individual immigrated to the United States or moved to their current dwelling can be combined with foreign birthplace variables to examine patterns of destination choice and migration rates (Zimran, working paper).

Inform policy-makers about who will benefit from pending legal status and permanent residence programs by describing potential beneficiaries (Kerwin et al, 2022).
Understand the mechanisms in place that exacerbate the **upward mobility gap** experienced by Black families using information about race in conjunction with out-migration variables at the county level (Derenoncourt, 2022).

---

**Can You Move to Opportunity? Evidence from the Great Migration†**

By Ellora Derenoncourt§

This paper shows that racial composition shocks during the Great Migration (1940–1970) reduced the gains from growing up in the northern United States for Black families and can explain 27 percent of the region’s racial upward mobility gap today. I identify northern Black share increases by interacting pre-1940 Black migrants’ location choices with predicted southern county out-migration. Locational changes, not negative selection of families, explain lower upward mobility, with persistent segregation and increased crime and policing as plausible mechanisms. The case of the Great Migration provides a more nuanced view of moving to opportunity when destination reactions are taken into account. (JEL H75, H76, J15, J62, K42, N32, R23)
IPUMS USA Multigenerational Longitudinal Panel (MLP)

- Links millions of records across available full-count census data files, 1850-1940
- Trace individuals and families over time
- Pre-linked data files available from IPUMS USA website
- Facilitate research on outcomes across generations or with regard to early-life or family circumstances
- USA MLP site: https://mlp.ipums.org
Overview of IPUMS CPS

**Labor force survey with topical supplements**
- Basic Monthly Surveys, 1976 - present
- Annual Social Economic Supplement (ASEC), 1962 - present
- Other topical supplements (variably available)
  - Education
  - Voter
  - Fertility and Marriage
  - Veterans
  - Volunteer
  - Agricultural Worker
  - Civic Engagement
  - Computer and Internet Use
  - Contingent Work
  - Disability
  - Displaced Worker
  - Food Security
  - Immunization
  - Job Tenure
  - Public Arts
  - Tobacco Use
  - Under-Banked
  - Work Schedules
  - Unemployment
  - compensation

**Notable Features of CPS**
- Samples include approximately 70,000 households
- Short rotating panel
- Relatively long temporal coverage
- Sub-state identifiers included
- Source of detailed information about work and income
- Can be linked to ATUS data
The effect of COVID-19 on employment, examined by immigrant status and race/ethnicity. Liao and Villareal look at potential explanations for employment disparities during the pandemic by jointly considering immigration and ethnoracial identities in their analysis.

Unequal effects of the COVID-19 epidemic on employment: Differences by immigrant status and race/ethnicity

Kristin Tianqi Liao, Andrés Villarreal
Published: November 15, 2022 • https://doi.org/10.1371/journal.pone.0277005
How is culture transmitted across generations? This study used the panel structure of CPS to analyze cohorts and found that *intergenerational links* are weaker among young cohorts. Age at first birth and nonmarital fertility are influenced by the strength of intergenerational links (NoghaniBehambari et al, 2020).

Overview of IPUMS ATUS

American Time Use Survey (ATUS), 2003 - present
- Structured narrative time-diary accounts of 24-hour intervals
- One person completes the diary per household
- Captures activity patterns for the general population
- Information about what people were doing, where they were, and with whom
- Occasionally-fielded modules on Eating & Health, Well-Being, and Job Leave

Notable Features of ATUS
- Subsample of CPS, sample sizes 8k-20k depending on year
- Smallest geographic unit is county, but not all available
- Family origin and citizenship status available
- Custom time-use variables summarize time spent in activities meeting certain criteria
- Additional detail about caregiving activities
Quantify time spent on informal childcare. Older women are more likely to migrate than men to care for grandchildren, in light of drastic increases in childcare costs. This study found that, among women who provide childcare, older women who recently immigrated spent more time providing informal (unpaid) childcare than native-born women.

How are parenting norms influenced by acculturation for first- and second-generation immigrant families? Time use patterns show quantitative and qualitative differences in parenting involvement across different generational groups.
Overview of IPUMS NHIS

**National Health Interview Survey, 1963 - present**
- More than 17,000 variables
- Extensive information on health behaviors and conditions, family health history alongside rich socio-demographic characteristics
- Representative of civilian, non-institutionalized population
- Immigration variables: global region of birth and years living in the US
- Complex survey design; samples one adult and one child per household

**Notable Features of NHIS**
- Samples approximately 45,000 households
- Less geographic detail (smallest unit is region)
- Offers dichotomous nativity and ethnicity variables
- Health conditions, health care access and utilization
- Socioeconomic and demographic correlates of health
- COVID-19 variables
- Can be linked to the National Death Index
Combine region of birth and demographic information with different measures of health to examine a wide range of topics, including the outcomes of public policy on immigrant well-being.

**THE LANCET Public Health**

Volume 2, Issue 4, April 2017, Pages e175-e181

**Articles**

Health consequences of the US Deferred Action for Childhood Arrivals (DACA) immigration programme: a quasi-experimental study

Dr Atheendar S Venkataramani MD a, S, Sachin J Shah MD a, Rourke O’Brien PhD b, Prof Ichiro Kawachi PhD c, Alexander C Tsai MD d

**International Journal of Environmental Research and Public Health**

Article


Georgiana Bostean 1,*, Annie Ro 2 and Nancy L. Fleischer 3
Overview of IPUMS MEPS

**Medical Expenditure Panel Survey (MEPS), 1996 - present**

- Longitudinal panel about people’s interactions with the health care system, health conditions, and medical expenditures
- Five rounds of interviews over two years (expanded panels during COVID-19 pandemic)
- Linkable to the NHIS, offering broad-spectrum demographic data and detailed information on individuals and families over time
- Panel follows individuals; records people entering and exiting survey households
- Data checked against or imputed from health provider records

**Notable Features of MEPS**

- Subsample of NHIS, approximately 12,000 households
- Short longitudinal panel
- Limited geographic detail
- Useful for much more than expenditures (e.g., exploring transactions alongside demographic or health changes)
- Can be linked to the NHIS (in an FSRDC)
Additional Data Resources
Overview of IPUMS NHGIS

➢ Aggregate U.S. census & survey data (1790 to present)
  ○ Summary tables
    ■ All geographic levels down to census blocks
    ■ Nationwide extent (mostly levels)
  ○ Time series tables
    ■ Align geographic units across time (nominally integrated)
  ○ Geographic crosswalks
    ■ Compare U.S. census summary data at different geographic units
  ○ GIS files
    ■ Boundary files
    ■ Place points (locations of incorporated, unincorporated, and census places for 1900 to 2015)

Notable Features of NHGIS
➔ Larger sample size
➔ More spatial detail
➔ Tables for many geographic levels (tracts/block groups/etc.)
➔ Migration-related tables
  ◆ Nativity
  ◆ Residence 1 or 5 years ago
  ◆ Year moved in
  ◆ Year naturalized
  ◆ Etc.
American Community Survey
Migration Flows Data

➢ Migration flows between counties, minor civil divisions, and metropolitan areas using ACS 5-year data
   ○ Starting with the 2006-2010 5-year ACS
   ○ Distributed by the Census Bureau (not IPUMS)

➢ Crossed with demographic & socio economic characteristics
   ○ Age, Sex, Race, Hispanic Origin
   ○ Relationship to householder
   ○ Employment status & Occupation
   ○ Place of birth
   ○ Educational attainment,
   ○ Personal & household income
Conclusion

• All US microdata collections are appropriate for identifying immigrant populations
• Select between data collections based on topical and temporal coverage, and level of detail available in key measures for your research application
Questions
## USA Migration Variables

<table>
<thead>
<tr>
<th>Topic</th>
<th>Variable</th>
<th>Description</th>
<th>Time Period</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nativity</strong></td>
<td>BPL</td>
<td>Birthplace</td>
<td>All samples</td>
</tr>
<tr>
<td></td>
<td>MBPL</td>
<td>Mother’s birthplace</td>
<td>1880-1970</td>
</tr>
<tr>
<td></td>
<td>FBPL</td>
<td>Father’s birthplace</td>
<td>1880-1970</td>
</tr>
<tr>
<td></td>
<td>NATIVITY</td>
<td>Foreign birthplace or parentage</td>
<td>1870-1970</td>
</tr>
<tr>
<td></td>
<td>CITIZEN</td>
<td>US Citizenship Status</td>
<td>1870, 1900-present</td>
</tr>
<tr>
<td><strong>Timing of Migration</strong></td>
<td>YRIMMIG</td>
<td>Year of immigration</td>
<td>All samples</td>
</tr>
<tr>
<td></td>
<td>YRSUSA1</td>
<td>Years in USA (continuous)</td>
<td>All samples</td>
</tr>
<tr>
<td></td>
<td>YRSUSA2</td>
<td>Years in USA (intervalled)</td>
<td>All samples</td>
</tr>
<tr>
<td></td>
<td>MOVEDIN</td>
<td>When occupant moved into residence</td>
<td>All samples</td>
</tr>
<tr>
<td><strong>Migration Status &amp; Mobility</strong></td>
<td>MICCOUNTY1</td>
<td>County of residence, 1 year ago</td>
<td>2005-2021 ACS</td>
</tr>
<tr>
<td></td>
<td>MICPUMA1</td>
<td>PUMA of residence, 1 year ago</td>
<td>1990-2000</td>
</tr>
<tr>
<td></td>
<td>MIGRATE1</td>
<td>Migration status, 1 year ago</td>
<td>1950, 2000-2021 ACS</td>
</tr>
<tr>
<td></td>
<td>MICPLAC1</td>
<td>State of country of residence, 1 year ago</td>
<td>1950, 2000-2021 ACS</td>
</tr>
<tr>
<td></td>
<td>MICTMT1</td>
<td>Metropolitan area of residence, 1 year ago</td>
<td>1950, 2000-2011 ACS</td>
</tr>
<tr>
<td></td>
<td>MIGTYPE1</td>
<td>Metropolitan status, 1 year ago</td>
<td>1950, 2005-2021 ACS</td>
</tr>
<tr>
<td></td>
<td>CNTY</td>
<td>Country (to connect with IPUMS-I and PR samples)</td>
<td>All samples</td>
</tr>
<tr>
<td><strong>Puerto Rico Migration</strong></td>
<td>YRIMMIGPR</td>
<td>Year of immigration to PR</td>
<td>PRCS, 1910-1930 &amp; 1970-2000 decennial</td>
</tr>
<tr>
<td></td>
<td>YRSRPR</td>
<td>Years in PR (continuous)</td>
<td>PRCS, 1910-1930 &amp; 2000 decennial</td>
</tr>
<tr>
<td></td>
<td>YRSRPR2</td>
<td>Years in PR (intervalled)</td>
<td>PRCS, 1910-1930 &amp; 1980-2000 decennial</td>
</tr>
<tr>
<td></td>
<td>USACTIV</td>
<td>Main activity in US during stay</td>
<td>1970-1990 PR decennial</td>
</tr>
<tr>
<td></td>
<td>USADURR</td>
<td>Length of last stay in US</td>
<td>1970-1990 PR decennial</td>
</tr>
<tr>
<td></td>
<td>USARESID</td>
<td>Residence in the US for six months or more</td>
<td>1970-1990 PR decennial</td>
</tr>
<tr>
<td></td>
<td>USAYRRET</td>
<td>Year of return to PR from US</td>
<td>1970-1990 PR decennial</td>
</tr>
</tbody>
</table>
# CPS Migration Variables

<table>
<thead>
<tr>
<th>Topic</th>
<th>Variable</th>
<th>Description</th>
<th>Time Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nativity</td>
<td>BPL</td>
<td>Birthplace</td>
<td>1994-2023 BMS &amp; ASEC</td>
</tr>
<tr>
<td></td>
<td>MBPL</td>
<td>Mother's birthplace</td>
<td>1994-2023 BMS &amp; ASEC</td>
</tr>
<tr>
<td></td>
<td>FBPL</td>
<td>Father's birthplace</td>
<td>1994-2023 BMS &amp; ASEC</td>
</tr>
<tr>
<td></td>
<td>NATIVITY</td>
<td>Foreign birthplace or parentage</td>
<td>1994-2023 BMS &amp; ASEC</td>
</tr>
<tr>
<td></td>
<td>CITIZEN</td>
<td>US Citizenship Status</td>
<td>1994-2023 BMS &amp; ASEC</td>
</tr>
<tr>
<td>Migration Status &amp; Mobility</td>
<td>COUNTRY</td>
<td>Country of previous residence</td>
<td>1996-2006 ASEC</td>
</tr>
<tr>
<td></td>
<td>WHYMOVE</td>
<td>Reason for moving</td>
<td>1999-2022 ASEC</td>
</tr>
<tr>
<td></td>
<td>MIGSTA1</td>
<td>State of residence 1 year ago</td>
<td>1982-1984 &amp; 1986-2022 ASEC</td>
</tr>
<tr>
<td></td>
<td>MIGRATE1</td>
<td>Migration status, 1 year ago</td>
<td>1963-2022 ASEC (approx.)</td>
</tr>
<tr>
<td></td>
<td>MIGRATE75</td>
<td>Migration status, 1975 to this year</td>
<td>1978-1980 ASEC</td>
</tr>
<tr>
<td>Hurricane Katrina</td>
<td>KATEVAC</td>
<td>Evacuate due to Katrina, household</td>
<td>2006 ASEC</td>
</tr>
<tr>
<td></td>
<td>KATEVAC2</td>
<td>Evacuate due to Katrina, individual</td>
<td>2006 ASEC</td>
</tr>
<tr>
<td></td>
<td>KATPRIOR</td>
<td>State of residence prior to Katrina</td>
<td>2006 ASEC</td>
</tr>
</tbody>
</table>
# ATUS Migration Variables

<table>
<thead>
<tr>
<th>Topic</th>
<th>Variable</th>
<th>Description</th>
<th>Time Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nativity</td>
<td>BPL</td>
<td>Birthplace</td>
<td>All samples</td>
</tr>
<tr>
<td></td>
<td>MBPL</td>
<td>Mother's birthplace</td>
<td>All samples</td>
</tr>
<tr>
<td></td>
<td>FBPL</td>
<td>Father's birthplace</td>
<td>All samples</td>
</tr>
<tr>
<td></td>
<td>CITIZEN</td>
<td>US Citizenship Status</td>
<td>All samples</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>HISPAN</td>
<td>Hispanic origin</td>
<td>All samples</td>
</tr>
<tr>
<td></td>
<td>SPHISPAN</td>
<td>Hispanic origin (spouse or partner)</td>
<td>All samples</td>
</tr>
<tr>
<td></td>
<td>ASIAN</td>
<td>Asian origin</td>
<td>All samples</td>
</tr>
</tbody>
</table>
## NHIS Migration Variables

<table>
<thead>
<tr>
<th>Topic</th>
<th>Variable</th>
<th>Description</th>
<th>Time Period</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Nativity</strong></td>
<td>USBORNR</td>
<td>Born in the United States</td>
<td>1997-2021</td>
</tr>
<tr>
<td></td>
<td>CITIZEN</td>
<td>US Citizen</td>
<td>1998-2021</td>
</tr>
<tr>
<td></td>
<td>REGIONBR</td>
<td>Global region of birth</td>
<td>2000-2018</td>
</tr>
<tr>
<td></td>
<td>MOMORIGIN</td>
<td>Mother's national origin</td>
<td>1987, 1992</td>
</tr>
<tr>
<td></td>
<td>MOMCSONBR</td>
<td>Mother's country or state of birth</td>
<td>1987, 1992</td>
</tr>
<tr>
<td></td>
<td>PPOORIGIN</td>
<td>Father's national origin</td>
<td>1987, 1992</td>
</tr>
<tr>
<td></td>
<td>POPCSONBR</td>
<td>Father's country or state of birth</td>
<td>1987, 1992</td>
</tr>
<tr>
<td></td>
<td>HISPROMB</td>
<td>Hispanic country or state of birth</td>
<td>1987, 1992</td>
</tr>
<tr>
<td></td>
<td>CITIZEN</td>
<td>US Citizenship Status</td>
<td>1987, 1992</td>
</tr>
<tr>
<td><strong>Timing of Migration</strong></td>
<td>YRSINUSC</td>
<td>Years in the US, grouped</td>
<td>1989-2021</td>
</tr>
<tr>
<td></td>
<td>YRSINUSC</td>
<td>Years in state of present residence</td>
<td>1989-1996</td>
</tr>
<tr>
<td><strong>Migration Status &amp; Mobility</strong></td>
<td>MOVEMILE</td>
<td>Miles moved to get to present address</td>
<td>1979-1981</td>
</tr>
<tr>
<td></td>
<td>MOVEMILE</td>
<td>Reasons (first and second) for moving to current address</td>
<td>1979-1990</td>
</tr>
<tr>
<td></td>
<td>MOVEMILE</td>
<td>Times ever moved</td>
<td>1988</td>
</tr>
<tr>
<td></td>
<td>MOVEMILE</td>
<td>Moved in past 3 years</td>
<td>1979-1990</td>
</tr>
<tr>
<td></td>
<td>MOVEMILE</td>
<td>Times moved in past 3 years</td>
<td>1979-1980</td>
</tr>
<tr>
<td></td>
<td>MOVEMILE</td>
<td>Moved to this address: Calendar Year, Month</td>
<td>1988</td>
</tr>
<tr>
<td></td>
<td>LIVRELNOW</td>
<td>Now living with relatives</td>
<td>1979-1980</td>
</tr>
<tr>
<td></td>
<td>LIVRELTP</td>
<td>Coresident relatives now, previously</td>
<td>1979-1990</td>
</tr>
<tr>
<td></td>
<td>LIVPEROPRE</td>
<td>Number of people person lived with at previous address</td>
<td>1979-1980</td>
</tr>
<tr>
<td></td>
<td>TIMEADD</td>
<td>Time at this address: Years, Months, Since birth, Time Period</td>
<td>1981, 1988</td>
</tr>
<tr>
<td></td>
<td>LIVADDCOM</td>
<td>State and county residence changes</td>
<td>1979-1980</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td>HISPEHT</td>
<td>Hispanic ethnicity</td>
<td>1976-2021</td>
</tr>
<tr>
<td></td>
<td>HISPTYN</td>
<td>Hispanic ethnicity, dichotomous</td>
<td>1997-2021</td>
</tr>
</tbody>
</table>
## MEPS Migration Variables

<table>
<thead>
<tr>
<th>Topic</th>
<th>Variable</th>
<th>Description</th>
<th>Time Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Timing of Migration</td>
<td>YRSINUSC</td>
<td>Years in US, continuous</td>
<td>2007-2012</td>
</tr>
<tr>
<td></td>
<td>YRSINUSG</td>
<td>Years in US, grouped</td>
<td>2002-2003, 2012-2020</td>
</tr>
<tr>
<td>Migration Status</td>
<td>REGIONMEPS</td>
<td>Census region as of 12/31 of the survey year</td>
<td>1996-2020</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>HISPETH</td>
<td>Hispanic ethnicity</td>
<td>1996-2020</td>
</tr>
<tr>
<td></td>
<td>HISPYN</td>
<td>Hispanic ethnicity, dichotomous</td>
<td>1996-2020</td>
</tr>
</tbody>
</table>