Integrating Survey and Administrative Data Across Sources and Agencies to Create Statistical Products: A Case Study from Education

Sarah Grady
Study Director of the National Household Education Surveys,

National Center for Education Statistics

Emily Isenberg
Senior Researcher,

American Institutes for Research

IES Institute of Education Sciences

Case study from the National Household Education Survey

- NHES basics

- Goals of NHES geocode files

- Typical NHES data file contents

- Additional geocode file variables

- Process and challenges

- Lessons learned

IES Institute of Education Sciences

National Household Education Survey basics

- Sponsored by the National Center for Education Statistics within the U.S. Department of Education's Institute of Education Sciences.

- Collects cross-sectional, nationally representative data on education topics that are not easily collected through institutions, such as young children's care arrangements, homeschooling, school choice, homework, adults' work training, and adults' nondegree credentials.

- In 2016, NHES used an address-based sample and was conducted by mail.

- Conducted in two stages: screener and "topical" survey.
    - The Early Childhood Program Participation survey
    - The Parent and Family Involvement in Education survey
    - The Adult Training and Education Survey

IES Institute of Education Sciences

Evidence-based policymaking goals

Recommendations for Improving Secure, Private, and Confidential Data Access from the Commission on Evidence-Based Policymaking https://cep.gov/report/cep-final-report.pdf

Facilitate access to data for evidence building

Privacy and transparency in how those data are used

Goals of NHES supplementary geocode files

- Provide additional data to data users without increasing respondent burden

- Leverage data already collected by NCES and other statistical agencies

- Increase research possibilities of NHES survey data

Supplementary restricted-use geocode files

Purpose:
Provide additional geographic identifiers and geography-based measures for use in conjunction with the NHES:2016 restricted-use data files for the three surveys. The additional variables leverage administrative data and survey data from NCES and other federal agencies to expand the analytic utility of NHES:2016 data. Variables were chosen to enhance information about sample members' neighborhoods, communities, transportation options, local labor markets, access to job training, access to private schools, charter and magnet schools, and access to certain types of preschool options.

- Data are still nationally-representative.

- Can analyze by type of area: students living in low-population density areas with high employment rates across the U.S.

- Cannot analyze for unique pieces of geography: schooling experiences of students in a specific rural area.

- NCES restricted-use license required to access these new data files.

- Data files must be merged onto previously-released NHES:2016 restricted-use data files.

Institute of
Education Sciences

Typical NHES data file contents

System variables like ID and questionnaire receipt date.

Variables directly from the questionnaire.

Derived variables.

Zip code and zip code tabulation area variables from American Community Survey 5-year files.

For survey about K-12 students, school-level information from the Common Core of Data or Private School Universe Survey.

Variables about the household members surrounding the sampled child or adult.

Weights.

Imputation flags.

Institute of Education Sciences

Radius-based preschool access variables added to Early Childhood Program Participation file

- Private School Universe Survey (PSS) 2015-16 school year variables, from NCES
    - Number of nursery or prekindergarten programs that are part of private schools that offer one or more grades of K-12 within a 2.5-mile, 5-mile, or 10-mile radius of a sampled person's home (three variables).

- Head Start locations from Department of Health and Human Services 2017 variables
    - Number of Head Starts within a 2.5-mile, 5-mile, or 10-mile radius of a sampled person's home (three variables).

IES Institute of
Education Sciences

School district variables added to Parent and Family Involvement in Education file and Early Childhood Program Participation file

Children's school district identifiers from the 2015-16 school district boundaries from TIGER/Line data, created by the U.S. Census Bureau

- 7-digit NCES agency (LEA) identification number

- Education agency (LEA) name

- Available for all students, including homeschooled children and children attending private schools

- Available for all children not yet in kindergarten

Radius-based private school variables and district-level charter and magnet school variables added to Parent and Family Involvement in Education file

- Private School Universe Survey (PSS) 2015-16 school year variables, from NCES
  - Radius-based variables
  - Number of private schools within a 2.5-mile, 5-mile, or 10-mile radius of a sampled person's home (three variables)
- Common Core of Data (CCD) 2015-16 school year variables, from NCES
  - Based on district identifiers from district boundary files and home latitude and longitude
  - Percentage of schools in the student's school district's geographic area that are charter schools
  - Percentage of schools in the student's school district's geographic area that are magnet schools
  - Percentage of schools in a school district's geographic that are missing values for the magnet indicator

Radius-based job assistance variable added to Adult Training and Education Survey file

- Number of American Job Centers within a 25-mile radius of a sampled person's home
- List of American Job Centers provided by CareerOneStop, from the U.S. Department of Labor, July 2017
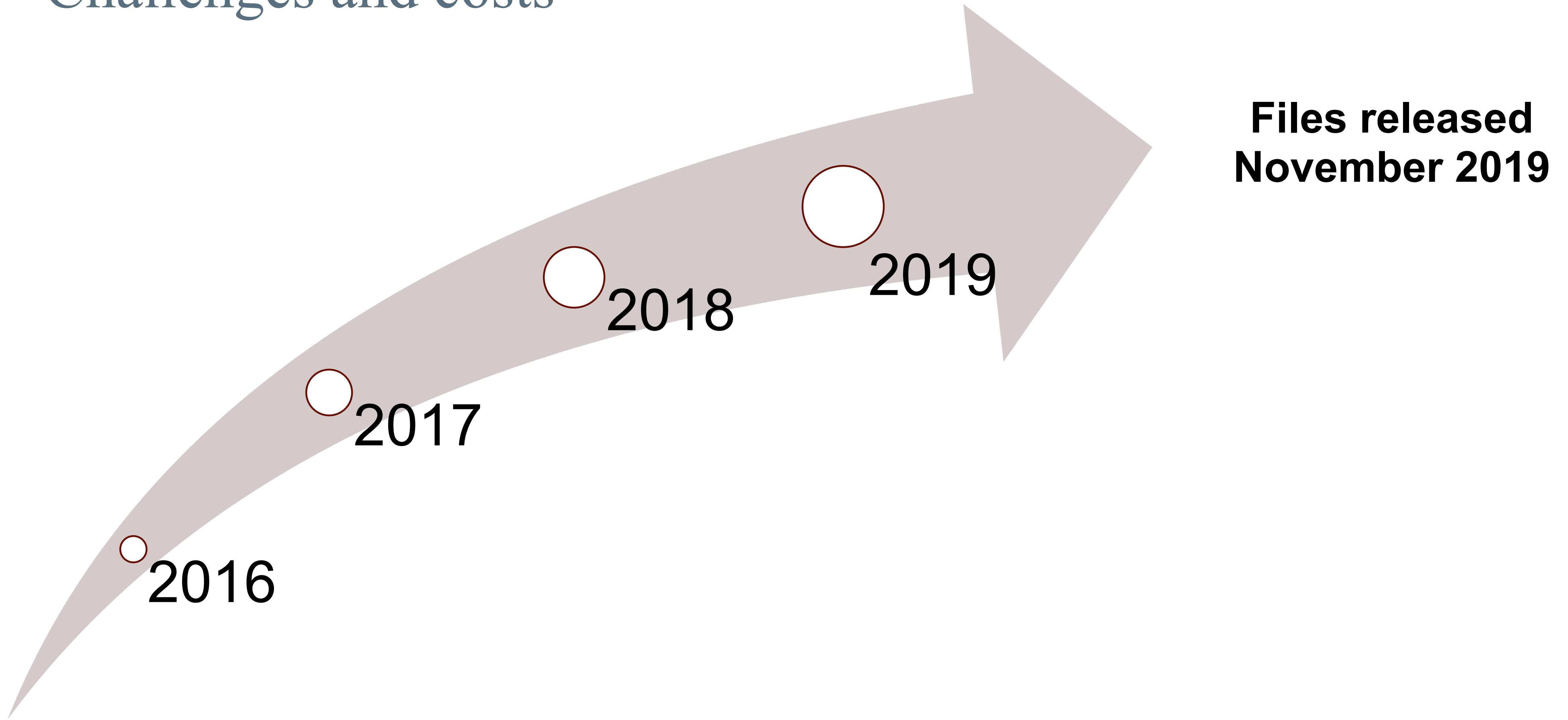
Additional variables added to all three NHES:2016 files

- Smart Location Database from the Environmental Protection Agency 2013
  - Census block group-level frequency of nearby public transit
- American Community Survey 5-year 2012-2016
  - Census tract-level commuting information
  - Census tract-level measures of income and poverty
  - County-level measures of percentage employed in different types of occupations
- Local Area Unemployment Statistics from the Bureau of Labor Statistics 2016
  - 2016 annual unemployment rate for counties, metropolitan and micropolitan statistical areas, and New England city and town areas (three variables)
- Quarterly Census of Employment and Wages from the Bureau of Labor Statistics 2016
  - 2016 average annual pay for counties and metropolitan and micropolitan statistical areas (two variables)

Institute of
Education Sciences

Challenges and costs

**Files released
November 2019**

2019

2018

2017

2016

Institute of
Education Sciences

Identifying and evaluating data sources to use

- Limited to subject matter expertise within NHES team
- Process challenges
  - Department of Homeland Security list of day care centers from https://gii.dhs.gov/hifld/
  - DHHS Head Start data availability changed
  - Local area labor market data and NECTAs

IES — Institute of Education Sciences

Understanding disclosure risk

- NHES:2016 files were already perturbed for disclosure prevention

- After considering various implications of providing these data, we defaulted to existing NCES restricted-use license provisions

Developing a user base for the files

- "NCES NewsFlash" email listserv announcements

- NCES blog

- Fund a statistical publication?

IES Institute of Education Sciences

Lessons learned

Adopting a model of combined survey and administrative data for statistical products for one survey program:

- Utilizes existing data security protocols within statistical agency
- Initially, takes several years for development, review, and release
- Requires staff time and program dollars for contractor time

IES Institute of Education Sciences

Questions?
sarah.grady@ed.gov
202-245-7063