

# Quality of Administrative Records as Source Data

Mike Berning and Dave Sheppard  
Data Acquisition and Curation Area  
U.S. Census Bureau

December 1, 2017

# Agenda

- About the Data Acquisition and Curation Area
- What are Administrative Records Source Data
- Increased Emphasis on Quality Measures
- Quality Framework
- Data Quality Branch and Experience
- Data Quality Review
- Data Quality Priorities

# The Data Acquisition and Curation Area

- October 2017 Reorganization:
  - Transferred administrative records production activities from Research and Methodology Directorate to Economic Directorate
  - Administrative records acquisition, processing and data management support
  - Added data ingest responsibilities
  - Established a Data Quality Branch

# Data Acquisition and Curation Area - continued

- Portfolio of more than 150 interagency agreements completed or under development
- FY17 - Processed more than 16,000 administrative records and survey files that added more than 1,100 research datasets to the administrative records data warehouse

# Administrative Records Source Data

- Data originating from federal, state or third party sources
- Stand-alone data as received - prior to linkage or other processing

# Increased Emphasis on Quality

- More sources of administrative records data
  - Understanding distinction between signal and noise
- More users of administrative records data
  - Decennial and Surveys
    - Cost
    - Response Rates
  - Research

# Data Quality Framework

- Qualitative Factors<sup>\*</sup>
  - Relevance – meets user requirements/fitness for use?
    - Replicability
  - Accessibility – can we acquire the data?
  - Coherence – consistent over time, geography?
  - Interpretability – can we understand it?
  - Accuracy – close to known values?
  - Institutional Environment – quality of source?
- Quantitative Factors
  - Metrics
  - Cost Effectiveness

<sup>\*</sup>Data Quality Assessment Tool for Administrative Data;  
Iwig, Berning, Marck and Prell; Feb 2013

# Data Quality Branch

- Develop processes and procedures to standardize ingest and quality clearance
  - Better define data request to provider
  - Minimize internal file handoffs
  - Eliminate duplication of effort
- Formalize ingest and quality clearance into repeatable processes
- Make quality results accessible to users



# Data Quality Experiences

Prior experience exploring data quality

- Contact frame creation and evaluation
- Sample file evaluations as part of RFPs
  - Names, addresses, phones, emails
  - Teacher rosters from schools
- Comparing ARs to survey response data
  - American Community Survey

# Checking Source File Quality

- First time reviews uncover more issues where we work out what we should be receiving
  - Receive and review test data
  - Discuss concerns and obtain clarifications
- Subsequent deliveries follow a more streamlined review process

# Data Quality Review

1. Can we process the delivery?
2. Are all the parts of the delivery consistent?
3. Are the data reasonable?
4. Did we meet the customer's requirements?

# 1. Can We Process the Delivery?

- Do we have the agreement in hand?
- Are the files encrypted?
- Were data and metadata provided?
- Are the file formats acceptable?

## 2. Are all the Parts of the Delivery Consistent?

- Agreement details
- Confirm record counts
- Confirm record lengths
- Match data elements to those listed in the layout/data dictionary
- Confirm valid values from metadata

# 3. Is the Data Reasonable?

Are the data elements well formed?

- Reasonableness of the values
  - Review fields that are empty, partially filled fields, or filled with non-meaningful content
  - Identify the truncation of data fields or presence of hidden characters
- Ensure special variable formats are correct
  - Dates, SSNs, phone numbers, email addresses, state abbreviations, etc...

# 4. Did We Meet the Customers Requirements?

- SAS datasets created
  - Sensitive variables dropped or separated
  - Linking variables added
- Metadata made available
- Naming conventions followed
- Sometimes custom requirements are requested

# Data Quality Priorities

- First, establish a documented repeatable process to keep the data flowing
- Work toward further ensuring and communicating data quality
  - Work closely with providers
  - Work closely with data users
    - Potential to expand data quality review, consider comparing source data to benchmark statistics
  - Share more of the metadata more widely



# Thanks!

Questions?

[Michael.A.Berning@census.gov](mailto:Michael.A.Berning@census.gov)

[David.W.Sheppard@census.gov](mailto:David.W.Sheppard@census.gov)