

# The Microdata Analysis System (MAS): A Tool for Data Dissemination

presented by Michael Freiman  
Center for Disclosure Avoidance Research  
U.S. Census Bureau

*Disclaimer:* The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

# Predecessor: Advanced Query System

- Advanced Query System (AQS) developed as part of American FactFinder after Census 2000
- Allowed users to create tables dynamically
- Typical users included beta testers, academics, researchers, SDCs, CICs
- AQS discontinued after contract expired

# Microdata Analysis System

- A successor to the Advanced Query System developed in-house
- Free to the public
- Part of the Census Bureau's DataWeb system; can be accessed using DataFerrett or possibly a new web-based interface
- More surveys available
- More capabilities for different types of statistical analyses in addition to tables
- Similar to systems being developed by other statistical agencies

# Microdata Analysis System

- Will allow users access to custom analyses based directly on the internal, unpublished data
- Will allow some tables and other analyses not currently available using AFF or public-use microdata
- Estimates will be consistent with AFF, except for rounding

# Source Data

- Starting with most recent American Community Survey (ACS) 5-year internal microdata
- More ACS datasets to be added later
- Decennial Census, other demographic surveys, and economic data also to be considered

# Universes

- Users create a “universe” or subset of observations for their tables/analyses
- Users can subset data based on up to 3 variables and a geography
- Geographies can go down to lower level than the public-use microdata
- Based on the universe, user can request a cross-tabulation of variables of the user’s choosing

# Table Rules

- User may construct 1-, 2- or 3-way tables
- Tables must pass disclosure checks to ensure tables are not too sparse
- Table cells are rounded according to the rules for special tabulations

# Variance Considerations

- Tables/analyses will include measures of variance
  - Margins of error for ACS
- Variance methodology used for official estimates will be employed
  - Replicate weights used for most ACS estimates

# Recodes

- Categorical variables often recoded into coarser categories
- Numerical variables categorized into predetermined ranges
- Often several versions of a recode with differing detail are offered
- Recodes developed with the program areas to ensure maximum utility and consistency with other data products
- When universe is smaller, only requests with less detailed recodes are likely to be returned

# Regression/models (planned)

- Ordinary least squares, logistic regression
- Rules on construction of universes to prevent a universe based on sparse data
- Limits on interactions, transformations and goodness of fit for disclosure purposes

# Other capabilities under consideration

- Regression diagnostic plots (residual plots and Q-Q plots)
- Histograms and scatterplots
- Means and medians
- Statistical testing

# Comments? Questions?

Amy Lauger  
amy.d.lauger@census.gov

Michael Freiman  
michael.freiman@census.gov

Tiffany Julian  
tiffany.a.julian@census.gov