Disclosure Avoidance Methods and Research at the U. S. Census Bureau

April 21, 2015

Billy Wisniewski, Amy Lauger, and Laura McKenna, Center for Disclosure Avoidance Research (CDAR), U.S. Census Bureau

Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Outline

- Overview

- Demographic data

- Economic data

- Remote access

- Research

Census Bureau Goals

- Disseminate high quality data to the public

- Uphold the legal and ethical obligations to protect respondents' confidentiality
  - Title 13
  - Title 26
  - Confidential Information Protection and Statistical Efficiency Act (CIPSEA)

United States™ **Census** Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

3/19

Confidentiality

- Public identification of or harm to the respondent is prohibited

- Data are immune from legal process

- Data are not divulged to others in ways that are inconsistent with the understanding of the original disclosure

        - OMB Statistical Policy Working Paper 22

United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

4/19

Disclosure Review Board

- Consists of at least nine members, representing Census's demographic, decennial, and economic directorates, as well as the RDCs, research and policy areas

- Establishes disclosure avoidance policies

- Ensures consistency in the disclosure review of all publicly released Census Bureau data products

United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

5/19

Types of Data

- Demographic
    - Decennial Census on Population and Housing
    - American Community Survey (ACS)
    - Many current surveys on people and households
- Economic
    - Economic Census
    - Many surveys on businesses and governments
    - Monthly leading Economic Indicators
- Commingled
    - Longitudinal Employer-Household Dynamics

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

6/19

Demographic Data

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Data Swapping

- Used for Decennial Census and ACS
- For a small number of records, data from one household is swapped with data from another household in a different geographic area
- "Risky" households are targeted
- Households must match on a few key characteristics to be swapped
- Estimates for counties and larger areas are minimally affected
- Some alterations: swapping routine, use different variables in determining risky HHs, increase the percentage of HH that are swapped

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

8/19

Tables

- Data swapping

- Use of geographic thresholds

- Use of cell size thresholds

- For survey data, usually only weighted estimates are published

- Synthetic data sometimes used

United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

9/19

Public Use Microdata

- Removal of names, addresses, etc.
- Limited geographic detail
- For Decennial and ACS: subsample of survey respondents
- Coarsening of categories
    - Each variable category must have at least 10,000 weighted people/households in U.S.
- Random noise added to some variables for some records
- Rounding
- Topcoding / bottomcoding

Economic Data

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Cell Suppression

- Traditional method for protecting magnitude data
- A risky cell is one for which an user can closely estimate a particular respondent's value (via p% rule)
- Such cells are suppressed
- Many complementary cells must also be suppressed
- New linear programming software is used to accomplish this

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

12/19

Alternative Methods

- Noise Infusion
    - Respondents' data are altered by a small amount prior to tabulation
    - Riskier values have proportionately more noise added
- Synthetic Data
    - Original values are replaced by "synthetic" values using statistical models
    - Synthetic datasets reflect the statistical properties of the original dataset
- Both methods allow all cells to be published

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

13/19

Remote Access

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Research Data Centers

- Remote locations used to perform analysis on confidential microdata

- Access requires an approved research proposal, security clearance, and a pledge to protect confidentiality

- 19 RDCs nationwide

- http://www.census.gov/ces/rdcresearch/

Microdata Analysis System

- Web-based tool allowing users to request customized tabulations and analyses

- Requests tabulated from confidential microdata

- Results returned if request passes real-time disclosure avoidance checks

- Variances calculated dynamically

- Pilot using ACS available soon

Research

United States Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Research Topics

- Re-identification studies

- Data swapping

- Noise infusion

- Synthetic data

- Cell Suppression

Questions?

- william.wisniewski@census.gov

- 301-763-2683

- Amy Lauger, Billy Wisniewski and Laura McKenna. "Disclosure Avoidance Techniques at the U.S. Census Bureau: Current Practices and Research (CDAR2014-02)" http://www.census.gov/srd/CDAR/cdar2014-02_Discl_Avoid_Techniques.pdf