

# Balancing Confidentiality Requirements with Data Users' Information Needs

---

BLS Disclosure Review Board

Data Users Advisory Committee  
November 8, 2012



# Request for Guidance from DUAC on the Impact of Disclosure Limitation Methods on Data Users

---

1. The impact that disclosure limitation methods have on data use
2. The reasons for the impact identified in (1)
3. Additional steps BLS could take to explore the impact in (1) with the broader community of data users

BLS already reaches out to and receives input from the data user community in a number of ways, but, disclosure limitation topics are not typically addressed.

# Presentation Outline

---

1. Key aspects of the legal, regulatory, and reputational environment affecting disclosure/confidentiality
2. BLS publication practices
3. Cell suppression and other options for sensitive cells
4. Examples from some BLS programs
5. Feedback from the panel

# The Legal, Regulatory, and Reputational Environment Affecting Confidentiality

- Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA)
- The Privacy Act (5 U.S.C. 552a)
- Information Quality Act (P.L. 105-554 § 515)
- State Laws (for Federal/State programs)
- Federal Statistical Confidentiality Order
- **Various BLS Commissioner's Orders**
- Reputation: Respondent cooperation

BLS takes confidentiality very seriously.

# BLS Standard CIPSEA Pledge (to respondents)

---

The Bureau of Labor Statistics, its employees, agents, and partner statistical agencies, will use the information you provide for statistical purposes only and will hold the information in confidence to the full extent permitted by law. In accordance with the Confidential Information Protection and Statistical Efficiency Act of 2002 (Title 5 of Public Law 107-347) and other applicable Federal laws, your responses will not be disclosed in identifiable form without your informed consent.

# BLS Disclosure Review Board (DRB)

- Chartered in 1999 to provide guidance to BLS programs to safeguard sensitive information provided by respondents
- Provide guidance on the policies and practices **to BLS's varied statistical programs**
- Operates under the guidance of the BLS Disclosure Review Executive Committee (DREC) who rules on policy issues raised by DRB

# BLS Publication Practices

---

BLS establishment survey programs:

- Publish large amounts of data
- Primarily in tabular form
- Structure and volume are driven by:
  - ▶ Stakeholder requests
  - ▶ OMB/Congressional budget decisions
  - ▶ User input (very important part of publication decisions)

# Publication Structure

---

Tables often include many dimensions, but vary by program (e.g.):

- Geography
- Industry
- Occupation
- Product Classification
- Injury type

# Reasons Some Cells Are Not Published

---

- Insufficient precision
- Disclosure limitation (confidentiality)
- Resources

**Today's goal:**

Explore options and consequences related to disclosure limitation.

# Options for Sensitive Cells

---

What is a sensitive cell?

- Release of data for this cell:
  - ▶ Allows identification of one or more respondents, or
  - ▶ Discloses sensitive information for one or more respondents
- Typically, sensitive cells arise because:
  - ▶ Small number of respondents, or
  - ▶ One or two dominant respondents

# Cell Suppression and Cell Coarsening

---

For sensitive cells, it is not realistic to publish standard direct estimates.

Instead, three realistic options:

- Option 1: Do not publish data for the cell **at all (flag as “suppressed cell”)**
- Option 2: Combine cells, i.e., users are referred to published higher level/coarser cells
- Option 3: Cell perturbation (next page)

# Alternative to Cell Suppression: Cell Perturbation (“Added Noise”)

---

- Frequently proposed alternative:  
**Cell perturbation, or “added noise”**
- Added noise can reduce disclosure risk
- Has been used by other statistical agencies
- Significant research in statistical literature
- **Crucial factor for today’s discussion:**  
Potential impact on data users from cell  
**perturbation (“added noise”)**

# Practical Impact of Disclosure Limitation Options (1) through (3)

- Balance between disclosure protection and practical use of released data
- Option 1: Cell suppression - No data for that cell
- Option 2: Cell coarsening - Value depends on relevance of data from the “coarser” cell
- Option 3: Cell perturbation - Value depends on perceptions of “real data” vs. “perturbed data”

# Example from the Quarterly Census of Employment and Wages

The Quarterly Census of Employment and Wages (QCEW) Program has data on Establishment Counts, Employment, and Wages, by detailed Industry and Area

- Over four thousand areas
- Over two thousand industries
- 3.5 million active combinations
- Currently suppress 60 percent of the 3.5 million active cells

# Example from QCEW

## Clarke County Alabama (ND = Not Disclosed)

		2010	2010	2010	2011	2011	2011
NAICS	Industry	Number of Establishments	Employment	Average Annual Pay	Number of Establishments	Employment	Average Annual Pay
515	Broadcasting, except Internet	6	18	26755	6	ND	ND
5151	Radio and television broadcasting	2	ND	ND	2	ND	ND
51511	Radio broadcasting	2	ND	ND	2	ND	ND
515112	Radio stations	2	ND	ND	2	ND	ND
5152	Cable and other subscription programming	4	ND	ND	4	ND	ND

# Cell Suppression Example from QCEW (synthetic data)

## Raw data vs. Cell Suppression

	Raw Data			Cell Suppression		
Industry	Number of Establishments	Employment	Average Annual Pay	Number of Establishments	Employment	Average Annual Pay
Total	14	52	43,051	14	52	43,051
Industry A	7	30	42,212	7	ND	ND
Sub-Industry A1	3	6	18,317	3	ND	ND
Sub-Industry A2	4	24	48,186	4	ND	ND

# Cell Suppression Example from QCEW (synthetic data)

## Raw data vs. Noise Perturbed Data

	Raw Data			Noise Perturbed		
Industry	Number of Establishments	Employment	Average Annual Pay	Number of Establishments	Employment	Average Annual Pay
Total	14	52	43,051	14	53	43,093
Industry A	7	30	42,212	7	29	42,661
Sub-industry A1	3	6	18,317	3	7	16,485
Sub-industry A2	4	24	48,186	4	22	50,989

# Example from the SOII Program

The Survey of Occupational Injuries and Illnesses Program (SOII) estimates the number of work-related injuries and measures the rate at which they occur.

- Mandatory participation from sampled establishments
- Detailed industry and category incidence rates and case counts

# Example from the SOII Program

- Sample tailored to meet national and state needs and interests
- Both quality and disclosure standards affect decisions on publication for specific cells
- Complementary suppression mostly shows up as rolled up data

# Example from the NCS Program

---

The National Compensation Survey estimates wages and benefits for employees by occupation, for the U.S. and a number of areas

- Desire is for more detailed breakouts than currently released
- Release is limited by both quality and confidentiality standards

# Questions for DUAC Members (1)

---

1. What is the current or potential impact of disclosure limitation methods on data users?
  - Option 1: Cell suppression - No data for that cell
  - Option 2: Cell coarsening - Value depends on **relevance of data from the "coarser" cell**
  - Option 3: Cell perturbation - Value depends on **perceptions of "real data" vs. "perturbed data"**

# Questions for DUAC Members (2)

---

2. What are the primary reasons for the effects identified in (1)?
  - a) Concrete examples?
  - b) What subject-matter factors will dominate the trade-offs between Option 2 (coarser cells) and Option 3 (perturbed cell data)
  
3. Recommended additional steps for BLS to explore the impact of Options 1-3 with the broader community of data users?

# Contact Information

---

Michael Buso (today's presenter)

[buso\\_m@bls.gov](mailto:buso_m@bls.gov)

202-691-6499

BLS Disclosure Review Board

John Eltinge, Chair

[eltinge\\_j@bls.gov](mailto:eltinge_j@bls.gov)

