



# Overview of New Legislation Protecting Confidentiality of Statistical Information and Statistical Disclosure Limitation Methodologies

*Nancy Kirkendall*

*Director of Statistics and Methods Group*

*Energy Information Administration (EIA)*



BTS Confidentiality Seminar Series, Feb. 2003



# Main Topics for this Seminar

- ◆ New law
- ◆ How it affects statistical agencies
- ◆ How my agency, the Energy Information Administration, is reacting
- ◆ Statistical disclosure limitation methodologies to protect confidential information



# Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA)

- ◆ Title V of E-Government Act of 2002, Public Law 107-347
- ◆ Signed into law December 17, 2002
- ◆ Entire E-Gov Act is 72 pages; CIPSEA is on last 9 pages
- ◆ Available at

[http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=107\\_cong\\_public\\_laws&docid=f:publ347.107.pdf](http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=107_cong_public_laws&docid=f:publ347.107.pdf)



# CIPSEA

- ◆ Subtitle A, Confidential Information Protection
  - Offers a consistently high level of protection to all statistical data collected under a pledge of confidentiality
- ◆ Subtitle B, Statistical Efficiency
  - Directed toward sharing of business data by Census, BEA, and BLS; this subtitle has no direct effect on other agencies



# CIPSEA Subtitle A, Confidential Information Protection

- ◆ An agency may collect information under a pledge of confidentiality for exclusively statistical purposes
- ◆ Such information may not be disclosed in identifiable form for any nonstatistical purpose without the informed consent of a respondent
- ◆ Such information is also exempt from release under the Freedom of Information Act (FOIA)



# Statistical and Nonstatistical Purposes

- ◆ Statistical purposes include using information to describe or make estimates about whole or subgroups of the economy, society, or the environment
- ◆ Nonstatistical purposes include using information for administrative, regulatory, law enforcement, judicial, or other purposes that may affect the rights, privileges, or benefits of a respondent





# CIPSEA Benefits for Federal Statistical Agencies

- ◆ Most agencies did not have specific laws ensuring confidentiality of information
- ◆ Agencies can now better protect data collected for exclusively statistical purposes
- ◆ Higher level of confidentiality may encourage respondents to participate in surveys
- ◆ Agencies can avoid disputes about withholding from release under FOIA



# CIPSEA Effects on Agencies

- ◆ An agency may designate information as being for exclusively statistical purposes
  - Information collected under CIPSEA
    - Cannot be shared for nonstatistical purposes
    - Can be shared for statistical purposes by entering into special written agreements, agent bound to provide same level of protection
  - In EIA's case, CIPSEA overrides existing laws that required EIA to share for official purposes which could be nonstatistical
- ◆ A statistical agency must clearly explain to respondents before any information is collected if it is to be used for nonstatistical purposes



# EIA View of Survey Confidentiality Options



- ◆ CIPSEA - Confidential and for exclusively statistical purposes
- ◆ Confidential, but not for exclusively statistical purposes; agency may withhold from public release using other laws such as FOIA and the Privacy Act
- ◆ Not confidential and may be publicly released in identifiable form



# EIA Actions

- ◆ Consult with OMB and DOE/OGC
- ◆ Create a team to examine EIA surveys and determine confidentiality appropriate to each
- ◆ In particular, what data/information should be included into the new CIPSEA confidentiality category?
  - Likely possibilities end-user and other sample surveys
  - Inclusion in this new category precludes any future sharing of information for nonstatistical purposes (e.g., DOE/Policy, FERC, EPA, DOJ)



# EIA Adoption of CIPSEA Confidentiality for Surveys

- ◆ Develop wording for all pledges of confidentiality
- ◆ Discuss with OMB, obtain clearance
- ◆ Notify respondents (by mail for on-going surveys, in instructions for upcoming surveys)



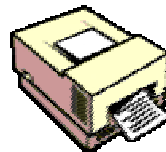
## Other EIA Actions

- ◆ Training for EIA staff on CIPSEA
  - Surveys covered
  - Additional procedures for protecting data
  - CIPSEA fines and penalties (Class E felony with prison up to 5 years and/or \$250,000 fine for willfully disclosing such information to a person or agency not entitled to receive it)



# Confidential Survey Information May Be In Different Formats

- ◆ Completed survey forms
- ◆ Electronic files and printouts
- ◆ Information products such as printed publications and web site information
- ◆ Public-use microdata files (information about individual survey respondents)





# Disclosure Limitation Methodologies

- ◆ Statistical agency must have controls to ensure protection of confidential information
- ◆ Actions to protect the information
  - Internal procedures
  - Aggregate information used in agency products such as tables, charts, graphs, and text
  - Microdata; i.e., information about individual survey respondents





## Disclosure Limitation in Tables

- ◆ Ensure that aggregate data do not inadvertently disclose individually-identifiable confidential survey information
- ◆ For example, a data cell in a table may represent responses from only one or two respondents or the cell may be dominated by a small number of large respondents



# Disclosure Limitation Methods for Tables

- ◆ Cell suppression is most common
  - Do not release a cell if it may be used to estimate confidential information (called primary suppression)
  - May also require not releasing one or more other cells to ensure the sensitive cell cannot be determined (called complementary suppression)



## Coal Stocks at Other Industrial Plants by Census Division and State (Thousand Short Tons)

Census Division	9/30/2002	6/30/2002	9/30/2001	% difference 9/30/02 vs 9/30/01
Mid Atlantic Total	⇒ W Complementary	W	W	W
New Jersey	⇒ W Primary	W	W	W
New York	296	235	199	48.6
Pennsylvania	202	215	187	7.9

W = Withheld to avoid disclosure of individual company data.

Note: Total may not equal sum of components because of independent rounding.

Source: EIA's Quarterly Coal Report,

<http://www.eia.doe.gov/cneaf/coal/quarterly/html/t35p01p1.html>



## Alternative to Suppression

- ◆ New method being developed will use synthetic data to protect confidentiality
  - Add or subtract a small amount to cell value so respondents cannot use it to estimate value of other respondents too accurately.
  - May be implemented using rounding



# Primary Suppression Rules for Tables

- ◆ Rules for determining if a cell is sensitive and requires primary suppression
  - n, k rule focuses on number of respondents represented in a cell's value and the percentage contributed by the larger respondents
  - pq rule
  - p-percent rule
  - Combination



## Primary Suppression Rules (Cont'd)

- ◆ Recommend using one of the above rules, or a combination
- ◆ They are simple and have important mathematical properties (union of nonsensitive cells is not sensitive)
- ◆ Rules are described in detail in Statistical Policy Working Paper 22, Report on Statistical Disclosure Limitation Methodology  
<http://www.fcs.m.gov/working-papers/spwp22.html>





## Hint

- ◆ If a table has too many suppressions, data not useful
- ◆ Redesign, combining categories to make a table with fewer suppressions



# Disclosure Limitation Methods for Public Use Microdata Files

- ◆ Public use microdata files consist of records that contain individual information on persons, businesses, or other entities
- ◆ Used for analytical and research purposes
- ◆ Agency must ensure that confidentiality is maintained



# Disclosure Limitation Methods for Microdata Files Include

- ◆ Rounding
- ◆ Top and bottom coding
- ◆ Recoding
- ◆ Collapsing categories
- ◆ Data swapping
- ◆ Adding noise
- ◆ Suppressing individual records or certain variables from all records



# Responsibilities for Confidentiality

- ◆ Agency and its contractors are responsible for ensuring confidentiality of survey information
- ◆ Broken confidentiality promise has potential for severe negative consequences
  - Including 5 years in prison/\$250K in fines for willful disclosure





# Additional References and Background Materials

Federal Committee on Statistical Methodology (FCSM), Statistical Policy Working Paper 22, Report on Statistical Disclosure Limitation Methodology

<http://www.fcsm.gov/working-papers/spwp22.html>

ASA Committee on Privacy & Confidentiality is creating a Privacy, Confidentiality, and Data Security Training Website (available Spring 2003)

<http://www.amstat.org/Comm/index.cfm?fuseaction=commdetails&txtComm=CCNMS10>



## Background (continued)

- ◆ FCSM's Confidentiality and Data Access Committee (CDAC)

<http://www.fcsm.gov/committees/cdac/cdac.html>

- ◆ CDAC's web site includes materials on:

- Checklist on Disclosure Potential of Proposed Data Releases
- Confidentiality and Data Access Issues Among Federal Agencies
- Restricted Access Procedures
- Panel on Disclosure Review Boards of Federal Agencies
- Identifiability in Microdata Files





# Contact Information

- ◆ Nancy Kirkendall  
Statistics and Methods Group, EI-70  
Energy Information Administration  
U.S. Department of Energy  
Phone: 202-287-1706  
Fax: 202-287-1705  
E-mail: [Nancy.Kirkendall@eia.doe.gov](mailto:Nancy.Kirkendall@eia.doe.gov)