CONSISTENCY OF OUTPUT FROM REMOTE ACCESS SERVERS, MICRODATA, AND TABULAR DATA RELEASE: THE NEED FOR COMMON PERFORMANCE METRICS
DATA GATHERED BY GOVERNMENT AGENCIES ARE AN IMPORTANT SOURCE OF INFORMATION TO THE GENERAL PUBLIC.
It is likely that all three types of access will be employed ... at least in the short run.
CONSISTENCY OF OUTPUT

- Necessary in order to
  - Ensure that output from one type of access does not compromise data in another access type of access
  - Convince the public that the statistical disclosure methods employed are effective both in providing meaningful data for analysis purposes and in preventing disclosure of confidential information.
Access method, disclosure risk, and information loss

- Development of global measures applicable across all methods of access
- Difficult problem as evidenced by the extensive literature on the tabular data/microdata comparisons
Data Characteristics

- The underlying characteristics of the data may dictate our selection of the access provided and the adoption of the disclosure risk and information loss measures.
- One size fits all is unlikely to succeed.
Input versus Output masking

- This age old question still remains unanswered
  - Input masking for microdata
  - Output masking for tabular data
  - End result – inconsistencies

- Both are feasible for remote analysis ..... but which one is preferable?
Flexibility

Let us consider the combination that offers the most flexibility in terms of both access and masking.
Remote Analysis Systems

- A single system that responds to all queries
  - tabular queries, statistical queries, perhaps even microdata access (if necessary)
**FLEXIBLE MASKING MECHANISM**

- **Input or source masking**
  - Mask original data and use the masked data to respond to all queries
    - Response consistency
    - Ability to assess information loss and disclosure risk characteristics prior to release
  - Unsuitable for data that changes frequently
Data administrator’s have to decide on the appropriate mode of access and data masking mechanism to ensure that they provide the public with the information that is accurate, safe, and easy to access.

Policy issues (including legal requirements) are likely to play a greater role in this decision than just technical issues.
An integrated system with a single data source offers the best solution to protect the confidentiality of the sensitive data while simultaneously providing the users with analytical results that are meaningful and consistent.
THANK YOU

Comments, suggestions, questions?