

# Benefits and Costs of Transparency: Views from One Agency -- the National Center for Health Statistics

**Jennifer H. Madans, Ph.D.**

Associate Director for Science,  
National Center for Health Statistics



Workshop on Transparency and Reproducibility in  
Federal Statistics Workshop

CNSTAT, June 21-22, 2017

National Center for Health Statistics



# Transparency Tradition

- **Long tradition of maximizing transparency in all NCHS activities**
- **Defines NCHS as part of the federal statistical system**
- **Costs to the agency are substantial affecting**
  - the scope and amount of data that can be collected,
  - the scope and number of products that can be released, and
  - the speed of both collection and release
- **Benefit to the agency are also substantial as transparency has been seen as critical to maintaining agency credibility which**
  - maximizes utility and use of information
  - protects agency independence

# **Data Collection Objectives**

- **Focus is on releasing as much data as possible for use by the public (public use files and restricted access)**
- **NCHS' publication series support data releases and highlight key findings**
- **Limited or no work in projections and modeling**

# **Major NCHS Data Collection Systems**

- **National Health Interview Survey**
- **National Health and Nutrition Examination Survey**
- **National Survey of Family Growth**
- **National Vital Statistics System**
- **National Health Care Surveys**

# **Activities Related to Data Collection**

- **Planning**
- **Development**
- **Field Work**
- **Data Preparation, Editing and Imputation**
- **File Release and Documentation (public use and restricted access)**
- **NCHS published reports and papers published by others**

# NCHS' Transparency Scores

- Planning \*\*
- Development\*\*
- Field Work \*\*\*
- Data Preparation, Editing and Imputation \*
- File Release and Documentation (public use and restricted access) \*\*\*
- NCHS published report\*\* and papers published by others\*

# Transparency and Reproducibility

- Is transparency a goal only in that it supports reproducibility (same data, same methods)?
- If so, where is the starting point?
  - NCHS documentation of public use data files and analyses are designed so that analyses based on public use files can be reproduced.
  - Is this sufficient?
  - What is needed about how the PUF are created?
  - What is needed about analysis specific decisions?
  - Should this apply only to key indicators or to the content of all NCHS publications? What about reports/paper not published by NCHS?

# Transparency and Reproducibility

- Complete transparency should result in reproducibility but...
  - Isn't transparency in decisions about what to collect and how to collect it required even if not related to reproducibility?
  - Isn't transparency a goal (to support agency credibility and trust in agency products) independent of reproducibility?
    - For this purpose, how much transparency is enough and in what areas?
  - Is transparency in all aspects of the data collection process more important for replication?



# Transparency and Reproducibility

- **What are acceptable ways to document processes?**
  - Must everything be publically available even if the user community for some data elements is small?
  - Is it sufficient to maintain information and make it available on request?
- **Must everything that is known about a data element be reported whenever that data element is used in an analysis or is it sufficient if documented in some way?**

# Conclusion

- **NCHS maintains a commitment to transparency so that others can reproduce what we do AND so that others can understand what we do and why.**
- **The costs associated with transparency impact the scope of our data collection, data release and publication programs.**
- **The remaining question is to better define the boundaries of what we need to be transparent about and hopefully how to do it in a more efficient way.**