

Eight Questions for Work with Non-Designed Data Sources

John L. Eltinge

U.S. Bureau of Labor Statistics

Conference on Applied Statistics in Defense

October 22, 2014



Acknowledgements and Disclaimer

The author thanks Bill Barron, Paul Biemer, Moon Jung Cho, Don Dillman, Jeff Gonzalez, Erica Groshen, Rachel Harter, Anders Holmberg, Alan Karr, Wendy Martinez, Bill Mockovak, Manny Parzen, Polly Phipps, Clyde Tucker and Alyson Wilson for many helpful discussions of the topics considered in this paper.

The views expressed in this paper are those of the author and do not necessarily represent the policies of the U.S. Bureau of Labor Statistics

Summary

Expanding opportunities to work with large and complex datasets provide a good opportunity for the statistical community to:

- Increase its contributions to the scientific community and to society at large
- Reconsider eight longstanding questions in the application and management of statistical work

A. Information Needs of Primary Stakeholders

1. Understand those needs in a way that can best inform subsequent statistical work
2. Statistical responses to (1) that are firmly anchored in fundamental concepts, norms, standards and practices of the statistical community

Example:

Bayesian elicitation of priors and utility functions

B. Evaluation of Data Sources

1. Traditional “designed data” (surveys, experiments)
2. “Non-designed data” and “organic data”
(Groves, 2011; Couper, 2013)
 - Administrative and commercial records
 - Records of “transactions” (defined broadly)
 - Social media traces

 - Relationship with “big data”

C. “Value Added” from Sophisticated Statistical Concepts and Methods

1. Alignment of information needs with prospective data sources in a given application - value from:
 - a. Simple exploratory tools
 - b. More sophisticated analyses
 - c. Substantive narrative flowing from (a) and (b)
2. Distinguish between:
 - a. Value conveyed under standard conditions
 - b. “Option value” under extraordinary conditions

D. Inferential Risk

Complex exploratory inferences:

Align our stakeholders' information needs with standard measures of inferential risk

1. Traditional multiple inference issues
2. False discovery rates
3. Other approaches
4. Visualization tools for (1)-(3)

Challenge: “pointillist painting” vs. traditional decision-theoretic views of inference

E. Communication with Stakeholders

1. Communication on data quality (B) analytic value (C) and inferential risk (D)
 - in a way that resonates with our stakeholders, based on our understanding of their information needs from (A)
2. Understand cognitive load and training needs related to communication in (1)
 - For statisticians
 - For non-statisticians

F. Broader Managerial Topics

1. General observations:
 - a. Sophisticated statistical work is capital- intensive
 - b. Almost all intangible capital:
 - Human capital
 - Institutional capital
 - c. Need to understand value provided through ongoing investments in intangible capital
2. Continuing development of technical skills and leadership skills throughout statistical careers

F. Managerial Topics (Continued)

3. Effect of management structures, as well as individual and group incentives (explicit and implicit)
4. Relationships among statisticians, computer scientists and substantive area specialists

G. Intellectual Property Issues

1. Statistical community:
 - a. High priority on transparency
 - b. Some statistical areas (e.g., government):
Reinforced by view of statistical methods and statistical information as “public goods”
 - c. Intellectual property rights have focused on:
 - Acknowledge priority (theory & methods)
 - Acknowledge source (published data)

G. Intellectual Property (Continued)

2. Interface with some application areas:

Different views of intellectual property re:

a. Specific empirical results

b. Details of statistical methods
and related algorithms

c. Special restrictions in some areas of defense,
confidential microdata for surveys

3. Effects of (1) & (2) on research and dissemination?

H. Institutional and Intellectual Culture

Expectations, norms, standards and practices that can:

1. Guide our continued exploration of longstanding questions (A)-(G)
2. Enhance our long-term contributions to societal needs for high-quality information
3. Help to ensure that the statistical community will thrive

Contact Information

John L. Eltinge

Associate Commissioner

Office of Survey Methods Research

www.bls.gov/ore

202-691-7404

eltinge.john@bls.gov

