Discussion: Accounting for Incomplete-Data Issues to Improve the Quality of Statistical Information Products and Services

John L. Eltinge

WSS Workshop on Survey Nonresponse

April 3, 2017

Acknowledgements and Disclaimer

The discussant thanks the workshop organizers and three presenters for the opportunity to review the papers presented in this workshop.

The views expressed in this discussion are those of the author and do not necessarily represent the policies of the United States Census Bureau.

Overview

Hearty thanks to the NSF-MMS program:

All three projects are relevant to the dual mission of

- *intellectual merit*: scientific rigor

and

- *broader impacts*: applications, software, education and workshops (today)

Overview (Continued)

Hearty thanks to all three presenters

- Very interesting material

- Helpful contributions to the research and methodology missions of federal statistical agencies and partners

  1. Expansion of statistical information products and services

  2. Improvement of the quality/risk/cost profile

Overview (Continued)

This discussion:

1. Unifying Themes

2. Extension: Empirical Assessment of:
   - Estimation Bias
   - Impact of Bias on Stakeholder Value

3. Extension: Integration of Data from Surveys and "Non-Designed" Sources

I. Unifying Themes

A. Using statistical tools to address incomplete-data issues, with emphasis on:

1. Evaluating properties (mostly bias & variance) of a class of data collection and estimation methods

2. Suggesting improved methods (and evaluate realistic degree of improvement achieved)

I. Unifying Themes (Continued)

B. Papers Today: Incomplete Data and Measurement Error

1. Brenner: Mode effects – coverage, nonresponse issues

2. Peytchev et al: Partitioned designs (split questionnaires): reduce burden & mitigate risks from NR, meas error

3. Kott and Liao: Expanded use of auxiliary vars to improve calibration wts, mitigate effects of incomplete data

I. Unifying Themes (Continued)

C. Illustrate a broader class of design questions that warrant extensive study: Given a space of candidate data sources subject to missingness problems:

1. Feasible (and nominally optimal) ways to use for specified inferential goals? - *Large body of methodological work*

2. Worthwhile to use that source at all, after accounting for multiple dimensions of quality, risk and cost, *plus constraints* ?

II. Extension: Empirical Assessment

A. Estimation Bias

1. Conceptual quality criteria (e.g., Brackstone, 1999):

Transparency and reproducibility

Qualitative or “user defined” (timeliness, relevance, comparability, coherence and accessibility)

Quantitative – “accuracy” (total survey error)

II. Empirical Assessment (Continued)

B.2. "Accuracy" component: TSE decomposition

(Estimator) – (True value)

= (frame error/coverage)
+ (sampling error)
+ (nonresponse effects)
+ (measurement error)
+ (processing effects)

Large literature, including extensions to non-survey data

II. Empirical Assessment (Continued)

C. Idealized View:

1. Well-identified mean and covariance function models for all (dominant?) TSE components of survey observations

2. Use (C.1) to:

   a. Identify targets for improved design

   b. Possibly adjust point estimation, inference

II. Empirical Assessment (Continued)

C.3. Broad acknowledgement (e.g., Biemer et al., 2014, *JOS*)

a. Empirical information on TSE components often piecemeal and expensive

b. Ex: Is quality surface

- Dominated by a few main effects, low-order interactions (cf. search for "powerful predictors" of response, per Groves & Couper, 1998)?

- Very "bumpy" - allowing only local exploration?

II. Empirical Assessment (Continued)

C.4. Design modifications and adjustments focus on error components that are:

- Believed to be relatively large
- Reasonably subject to control
- Have applicable auxiliary information

C.5. Each presentation today has an empirical assessment of bias (or mitigation thereof) qualitatively aligned with (C.4)

II. Empirical Assessment (Continued)

C.6. Brenner:

a. Goal: Assess biases arising from modes of:
- recruitment mode (telephone/mail)
- collection (telephone – IVR & CATI; FTF)

b. Method: Split-sample randomization of 3000 units

c. Restriction: Five Boston-area neighborhoods/suburbs

d. Extension: Embed analyses into general linear models incorporating dominant TSE elements, including (C.6.c) and related interactions

II. Empirical Assessment (Continued)

C.7. Peytchev et al: Use split-questionnaire designs to reduce burden, mitigate nonresponse bias, meas error effects

a. Dominant predictor variables in assessment of full path:

Content/length ⇨ Perceived Burden
⇨ Decreased Cooperation ⇨ Reduced quality?

b. Optimal (improved?) design, accounting for sampling and nonsampling error components (cf. Ioannidis et al, 2016)?

II. Empirical Assessment (Continued)

c. Evaluate: criterion validity & covariance structure for BMI:

- Extend: comb effects of model predictors, design vars

- Similarly for “policy analysis” modeling: beyond simple availability of relevant predictors

II. Empirical Assessment (Continued)

C.8. Kott and Liao: Application - RECS National Pilot

a. Report % differences in est means across 26 variables

- Inherent interest in the 26 variables

- Perhaps implicit: Error structure for the 26 vars spans a space of other estimators, and approximations thereof?

II. Empirical Assessment (Continued)

b. Extend the work of Chang and Kott (2008), Kott (2014), and Kott and Chang (2010), Kott and Liao (2017 a, b) to develop formal diagnostics for (C.8.a), based on relationships among X, Z and candidate Y variables?

II. Empirical Assessment (Continued)

B. Impact on Stakeholder Value

B.1. Schematic Model for Stakeholder Value:

$$V = f(\text{Quality}, \text{Risk}, \text{Cost}) + \text{error}$$

integrated over time, product lines and stakeholder groups (linkage with "public goods" literature)

II. Empirical Assessment (Continued)

B.2. Application: For specific key uses, spell out:

a. How estimates are used – important "differences"

b. Explicit or implicit accounting for quality (esp bias)

c. Willing to incur risks from covering only X% of population?

d. Bayesian methods: elicitation of utility functions, priors?
O'Hagan et al (2007); Garthwaite (2013)

III. Extension: Integration of Data from Surveys and "Non-Designed" Sources

A. Background:

1. Recent attention: "Big data" (or "organic data" "non-designed data" or "alternative data" Groves, 2012; Couper, 2013; Citro, 2014; Lohr and Raghunathan, 2017) – e.g., administrative records

2. Use these data sources to:

   a. Improve current quality/risk/cost profile?

   b. Expand product lines?

III. Data Integration (Continued)

A.3. Non-designed sources often:

a. Have only partial population coverage (“patchwork”)

b. Require record linkage at unit level (imperfect links)

c. Are “unit rich and variable poor” (“tall and thin”)

A.4. Question: Extend approaches of the three presentations to applications with multiple data sources?

III. Data Integration (Continued)

B. Brenner

1. Empirical assessment of TSE-type errors in non-designed sources: Same challenges as in surveys, writ large.

2. Per Brenner's case, practical decisions on error-source assessment (randomized vs. convenient) will balance:
   - Anticipated error magnitudes (w/prior literature)
   - Feasibility constraints

III. Data Integration (Continued)

3. Conceptual models for multiple error sources:

   - not fully estimable,

   - but provide useful qualitative guidance on resource allocation

III. Data Integration (Continued)

C. Peytchev et al: Assess split-questionnaire approach:

1. Imputation of non-observed values, with modeling via:
   - Observed survey and frame data (per slide 31)
   - Additional variables captured through record linkage

2. Which non-designed sources are worth linking?
   - Burden, incomplete-data, specification-err, other effects

3. Per slide 29, evaluate fixed and variable cost components from std design & split-quest and record-linkage variants

III. Data Integration (Continued)

D. Kott and Liao – extend calibration weighting approaches to integration of non-designed sources with

1. Surveys for non-covered subpopulations
   - Identified in the survey frame
   - Not identified in the frame (extensions of multiple-frame literature)

2. "Bridge surveys" to account for errors in non-designed sources: unit problems; reference period; other specification effects

IV. Closing Remarks

A. Hearty thanks to the workshop organizers and presenters, as well as NSF-MMS

B. Very interesting work, with potential extensions to:

1. Broader assessment of estimation bias (and mitigation thereof) and related impact on stakeholder value

2. Integration of data from surveys and “non-designed” sources

Additional Literature References

Citro, C.F. (2014). From Multiple Modes for Surveys to Multiple Data Sources for Estimates. *Survey Methodology* **40**, 137-161. http://www.statcan.gc.ca/pub/12-001-x/2014002/article/14128-eng.pdf

Ioannidis, E., T. Merkouris, L.-C. Zhang, M. Karlberg, M. Petrakos, F. Reis and P. Stavropoulos (2016). On a Modular Approach to the Design of Integrated Social Surveys (with discussion). *Journal of Official Statistics* **32**, 259-305. DOI: https://doi.org/10.1515/jos-2016-0013

Lohr, S.L. and T.E. Raghunathan (2017). Combining Survey Data with Other Data Sources. *Statistical Science.* To appear. http://www.imstat.org/sts/future_papers.html

National Academies of Sciences, Engineering, and Medicine (2017). *Innovations in Federal Statistics: Combining Data Sources While Protecting Privacy. Washington, DC: The National Academies Press. doi: 10.17226/24652*