DE GRUYTER OPEN

Journal of Official Statistics, Vol. 32, No. 2, 2016, pp. 295–300, http://dx.doi.org/10.1515/JOS-2016-0016

Discussion

Jeffrey M. Gonzalez[1] *and John L. Eltinge*[2]

1. Introduction

The authors have produced a very insightful and valuable contribution to the literature on improved design of integrated surveys through the use of constrained optimization. This discussion provides some general context for this approach, and then suggests some complements to, and possible extensions of, the article.

1.1. Context: Survey Design Viewed as a Form of Constrained Optimization

As noted by Ioannidis et al., National Statistical Offices (NSOs) are exploring a wide range of innovative design options to improve overall data quality and to reduce costs and respondent burden. To provide a general framework for discussion of these options, and in keeping with Eltinge et al. (2013), let Q represent a vector of data-quality measures, potentially including components of a standard total survey error model, as well as more qualitative components like timeliness and relevance, as outlined in Brackstone (1999) and others. In addition, let C be a vector of measures of cost and respondent burden, for example, costs associated with instrument and systems development, data collection, production, and dissemination; and the temporal, operational and cognitive burden imposed on respondents. One may characterize the outcomes of the aforementioned improvement efforts through schematic models for overall survey quality

$$Q = g_Q(Z_D, Z_O, \gamma_Q) + e_Q \tag{1.1}$$

and costs

$$C = g_C(Z_D, Z_O, \gamma_C) + e_C \tag{1.2}$$

where Z_D is a vector of all survey design features under the control of the NSO (e.g., sample design and collection methods); Z_O is a vector that describes features of the population and survey environment that are observed but not controlled by the NSO (e.g., specific subpopulations' accessibility and willingness to respond to a given survey instrument); $g_Q(.\,,\,.\,,\,.)$ and $g_C(.\,,\,.\,,\,.)$ are parametric functions; γ_Q and γ_C are parameter vectors; and e_Q and e_C are deviation terms associated with other factors that are neither controlled nor observed by the NSO. In keeping with the cautionary comments in Section 6

[1] Office of Survey Methods Research, U.S. Bureau of Labor Statistics, PSB 1950, 2 Massachusetts Avenue NE, Washington, DC 20212, U.S.A. Email: Gonzalez.Jeffrey@bls.gov
[2] Office of Survey Methods Research, U.S. Bureau of Labor Statistics, PSB 1950, 2 Massachusetts Avenue NE, Washington, DC 20212, U.S.A. Email: Eltinge.John@bls.gov

of the main article on approximations to optimal designs, we emphasize that Models (1.1) and (1.2) should be viewed as schematic, since in most or all practical cases one would not have sufficient information to develop a rigorous assessment of the exact functional forms of all dimensions of $g_Q(. , . . , .)$ and $g_C(. , . . , .)$, nor consistent estimators of all elements of γ_Q and γ_C.

Subject to that cautionary note, one may characterize many survey design approaches as forms of constrained optimization centered on four steps:

a. Definition of a class $\mathbf{Z}_D$ of design specifications Z_D. Considerations of operational feasibility often lead to substantial constraints imposed directly on the class $\mathbf{Z}_D$, for example, the admissibility restrictions on instrument composition discussed in the main article.
b. Imposition of additional (indirect) constraints on $\mathbf{Z}_D$ through restrictions on some components of Q or C, for example, the variance and coefficient-of-variation bounds considered by Ioannidis et al.
c. Identification of one dimension of Q or C (other than those considered in Step (b)) as the objective function of primary interest.
d. Conditional on the constraints identified in (a) and (b), determination of the vector Z_D within $\mathbf{Z}_D$ that optimizes the objective function from (c), or the expectation of that objective function (evaluated over the sources of random variability associated with Z_O, e_Q and e_C).

1.2. Three Groups of Questions for Improvement of Integrated Modular Designs

For integration of multiple surveys through modular design, the ideas and results in Ioannidis et al. provide interesting and important insights into Steps (a) through (d), with special emphasis on operational constraints imposed through Step (a), precision requirements imposed through Step (b), selection of a relatively simple cost criterion in Step (c), and use of simulated annealing and simplex methods to carry out the optimization for Step (d).

The remainder of this discussion highlights some questions that could lead to extensions of the main article within the framework defined by Steps (a) through (d). Section 2 explores some quality dimensions that may be important for modular designs. Section 3 outlines some extensions of the modular design options considered in the main article. Section 4 provides additional comments on constraints and on the optimization methods used by Ioannidis et al.

2. Quality Dimensions

Ioannidis et al. place principal emphasis on quality functions defined by the sampling error variances, or coefficients of variation, for standard survey estimators of finite-population parameters under simple random sampling for each instrument, and use design-effect adjustments to extend their approach to account for complex designs. Moreover, Section 6 of the main article mentions nonsampling error issues, but expresses concern about the availability of applicable empirical information. We share that concern, and it leads naturally to design extensions in which in-depth empirical assessment of nonsampling error effects would be a preliminary step in defining the quality function to be optimized. For that purpose, there would be a special interest in extending "total survey error" methods from, for example, Andersen et al. (1979), Groves (1989), Weisberg (2005), Biemer and Lyberg (2010), and references cited therein. As an extension of comments on "thematic blocks" that are "logically interrelated" in Subsection 2.4.2.b of the main article, one particularly interesting example arises from potential context effects. Briefly defined, context effects in surveys occur when responses to questions are affected by prior items administered in the questionnaire because these prior items provide cognitive cues to the respondent (Johnson et al. 1998). Consequently, forms of instrument composition that include appropriate respondent cues may enhance the response process, and thus result in higher data quality. On the other hand, as noted in Section 6 of the main article, use of multiple instruments may in some cases increase the risk of confusion that leads to degradation of data quality.

In addition, Expression (2.3) of the main article considers precision requirements for estimation of multiple parameters, for example, the means of several variables in a given module. For many large-scale surveys, there would be a strong interest in additional exploration of this problem in the context of modular design, and in the related problem of precision requirements for estimation in multiple domains. For example, for many social surveys in the U.S., data users have strong preferences about the precision of design-based estimators of means or proportions for each of a large number of domains defined by demographic and geographical classifications.

Finally, one could also consider quality functions based on the variances of estimators computed through the integration of survey data and auxiliary data, for example, poststratification (e.g., Little 1993), regression estimation (e.g., Särndal et al. 1992), calibration weighting (e.g., Subsubsection 2.4.1 of the main article, and Chang and Kott 2008), or imputation (Little and Rubin 2002) of values that are missing due to the use of modular designs. For some of these cases, one could obtain appropriate quality functions through relatively simple scalar adjustments of the standard simple-random-sampling variances in a way that is closely analogous to the design-effect adjustments used by Ioannidis et al. In other cases, estimation of the appropriate quality function would be more complicated, and would need to account for components of variance associated, respectively, with the randomization design and the underlying models.

3. Adaptive Design Options

The main article places principal emphasis on "the dual design problem of instrument composition and sample size allocation" within the context of simple random sampling for each instrument. It would be of interest to explore the degree to which the proposed approach could be extended to three cases in which one may wish to assign modules adaptively, based on responses to certain items observed for all units. These cases would be of interest primarily for surveys that are administered electronically, so that the adaptation can take place in a way that is relatively seamless for the respondent. The resulting designs could be viewed as hybrids of the modular designs considered by Ioannidis et al., and responsive or adaptive designs considered by, for example, Groves and Heeringa (2006) and Beaumont et al. (2014). As in other adaptive-design work, care would be required to ensure that weighting or other adjustment steps are employed to make certain that the resulting estimators are approximately unbiased.

First, some NSOs seek to reduce costs and respondent burden by linking survey units with administrative records. Conditional on receiving consent to link (where required) and achieving a successful link, the NSO may then omit certain questions that are covered by the administrative data. If a given administrative source contains all items needed for one module, then one could omit that module for the consenting-and-linked sample units. This in turn would produce some reduction in the burden measure aggregated over all sample units. Linkage with administrative or commercial data may lead to bias issues (e.g., Sakshaug and Huber 2016, and references cited therein), and thus may require consideration of an expanded quality function as mentioned in Section 2 above.

Second, in keeping with comments in Subsection 2.4.1 on "certain modules related to rare population groups or low-prevalence characteristics of interest," one could consider the incorporation of two-phase sample design features into the modular approach presented by Ioannidis et al. Specifically, one could administer to all sample units some initial screening questions on membership in rare populations. A module of in-depth items relevant to the rare population would then be assigned with probabilities dependent on the responses to the screening questions.

Third, field personnel often note that an instrument of "typical length" can be problematic for some sample units that are time-constrained or reluctant to participate in a survey. For cases in which paradata can identify such units early in the survey process, one could adaptively assign those units to special instruments that contain fewer modules. This would be somewhat analogous to ad hoc "basic question procedures" often used in field operations (e.g., as discussed in Bethlehem and Kersten 1985), but assignment probabilities and module structure could be aligned to reduce the risks of bias that are incurred through purely ad hoc approaches.

4. Constraints and Optimization Techniques

Section 2 of the main article directs careful attention to a wide range of constraints that can be important for the proposed module-based integration of multiple surveys, and which led to the use of simulated annealing and simplex methods to identify the optimal design. As a complement to that development, one could consider three additional questions.

First, the first section of Ioannidis et al. describes the broad goal of "survey integration that includes all the social surveys managed by a national statistical agency." In some cases, that type of comprehensive integration may be feasible. In other cases, integration of all surveys may be problematic; however, the integration of certain groups of surveys might be feasible, and might in itself produce substantial reductions in cost and burden. Consequently, it would be of interest to identify criteria that could be used to guide the formation of these feasible groups.

Second, Subsection 2.4.3 of the main article discusses "hard" and "soft" approaches to limiting questionnaire size. One could expand on this approach by considering limits on other measures of burden imposed on a given respondent. Given the diverse set of response tasks possibly required by the various surveys within a NSO, it is worth distinguishing between two perspectives on incorporating burden constraints into the optimization problem formulation. In some cases, NSOs can attempt to balance the cognitive or operational burden across the instruments so that each respondent would be similarly tasked. This most closely approximates standardized interviewing, in which every respondent has a similar interviewing experience. Alternatively, NSOs could constrain instruments to include only similar response tasks, to the extent feasible. To illustrate, one instrument may contain modules with questions requiring consultation of personal records (e.g., financial statements or health insurance explanation of benefits statements) while another instrument would contain no modules with questions requiring these tasks.

Third, constraints are often incorporated into survey designs via formal mathematical optimization techniques. Approaches to constrained optimization of survey designs vary depending on the nature of the survey operation and include: (1) exact approaches (e.g., LaGrange multipliers for optimal allocation for stratification); (2) approximate or iterative approaches (e.g., nonlinear programming procedures as in Valliant et al. 2014), and, (3) simulation-based approaches as highlighted in Ioannidis et al. The main article provides some general guidance on circumstances under which simulation-based approaches are necessary, due to the size and complexity of the proposed optimization method. It would be of interest to study this topic in additional detail within the context of survey optimization, and to have diagnostics that could help to provide guidance on tuning the optimization procedures to the dominant features of the design space $\boldsymbol{Z}_D$, constraints and objective function under consideration.

5. References

Andersen, R., J. Kasper, M.R. Frankel, and Associates. 1979. *Total Survey Error: Applications to Improve Health Surveys*. San Francisco: Jossey-Bass.

Beaumont, J.-F., C. Bocci, and D. Haziza. 2014. "An Adaptive Data Collection Procedure for Call Prioritization." *Journal of Official Statistics* 30: 607–621. Doi: http://dx.doi.org/10.2478/jos-2014-0040.

Bethlehem, J.G. and H.M.P. Kersten. 1985. "On the Treatment of Nonresponse in Sample Surveys." *Journal of Official Statistics* 1: 287–300.

Biemer, P.P. and L. Lyberg, eds. 2010. "*Special Issue: Total Survey Error*." *Public Opinion Quarterly* 74(5).

Brackstone, G. 1999. "Managing Data Quality in a Statistical Agency." *Survey Methodology* 25: 139–149.

Chang, T. and P. Kott. 2008. "Using Calibration Weighting to Adjust for Nonresponse under a Plausible Model." *Biometrika* 95: 555–571. Available at: http://www.jstor.org/stable/20441486 (accessed March 2016).

Eltinge, J.L., P.P. Biemer, and A. Holmberg. 2013. "A Potential Framework for Integration of Architecture and Methodology to Improve Statistical Production Systems." *Journal of Official Statistics* 29: 125–145. Doi: http://dx.doi.org/10.2478/jos-2013-0007.

Groves, R.M. 1989. *Survey Errors and Survey Costs*. New York: Wiley.

Groves, R.M. and S. Heeringa. 2006. "Responsive Design for Household Surveys: Tools for Actively Controlling Survey Errors and Costs." *Journal of the Royal Statistical*

Society, Series A 169: 439–457. Doi: http://dx.doi.org/10.1111/j.1467-985X.2006.00423.x.

Johnson, T., D. O'Rourke, and E. Severns. 1998. "Effects of Question Context and Response Order on Attitude Questions." Proceedings of the Section on Survey Research Methods, American Statistical Association, 857–860.

Little, R.J.A. 1993. "Post-stratification: A Modeler's Perspective." *Journal of the American Statistical Association* 88: 1001–1012.

Little, R.J.A. and D.B. Rubin. 2002. *Statistical Analysis with Missing Data*, 2nd ed. Hoboken, NJ: Wiley.

Sakshaug, J.W. and M. Huber. 2016. "An Evaluation of Panel Nonresponse and Linkage Consent Bias in a Survey of Employees in Germany." *Journal of Survey Statistics and Methdology* 4: 71–93.

Särndal, C-E., B. Swensson, and J. Wretman. 1992. *Model-Assisted Survey Sampling*. New York: Springer.

Valliant, R., F. Hubbard, S. Lee, and C. Chang. 2014. "Efficient Use of Commercial Lists in U.S. Household Sampling." *Journal of Survey Statistics and Methodology* 2: 182–209.

Weisberg, H.F. 2005. *The Total Survey Error Approach: A Guide to the New Science of Survey Research*. Chicago: University of Chicago Press.