DE GRUYTER OPEN

Journal of Official Statistics, Vol. 32, No. 2, 2016, pp. 301–305, http://dx.doi.org/10.1515/JOS-2016-0017

Rejoinder

Evangelos Ioannidis, Takis Merkouris, Li-Chun Zhang, Martin Karlberg, Michalis Petrakos, Fernando Reis, and Photis Stavropoulos

We would like to thank all the discussants for their thoughtful and encouraging comments. Not surprisingly, there is some overlap among the issues raised. Below we organise our response in three parts, in correspondence with the discussion, and note the connections and overlaps across the parts when and where we consider appropriate.

1. Chipperfield

In his first comment, the author points out that the estimation method that we assumed in designing an integrated survey does not exploit correlations between variables collected in different modules to improve the accuracy of estimates. Thus he suggests that this improvement, achievable by suitable estimation methods, could be factored into the design problem, to make the design more efficient. Gonzalez and Eltinge make a similar comment in the last paragraph of (their) Section 2.

Our approach to designing an integrated survey assumes a baseline estimation procedure, involving standard Horvitz-Thompson (HT), estimators for items surveyed in a single instrument and simple but efficient composite HT estimators combining data on common items surveyed in different instruments. This general approach, requiring only estimates of design effects, is applicable to any setting of an integrated survey. As is the norm in the case of a single survey, we do not factor into the design of an integrated survey the effect of a regression or calibration estimator which might be used. Design-based estimation methods (cited in our article) that exploit correlations between variables for improved accuracy are in fact special calibration procedures, whereby estimates of the same totals from different instruments are aligned. Such methods can be used profitably in our setting, but the design we propose is not contingent on the use of any of them for the following reasons.

Firstly, an exact theoretical quantification of the correlation effects is intractable for the type of surveys under consideration, involving complex sampling designs plus multiple instruments of varying composition and periodicity of modules, different production timetable for various statistics, etc. Furthermore, it is not at all obvious how to factor such (variable-dependent) effects into the sampling design, in a manner analogous to the design-effect scalar adjustments. This would essentially require a measure of compound design-regression-correlation effect that accounts for the interaction between the three components, ideally for a number of important items. Devising such a measure seems to be an extremely challenging task.

Secondly, factoring correlations into the design depends on the particular process of exploiting them, and its interaction with customary calibration. But such estimation/

calibration procedures may be optional or subject to revision over time, and thus it is not sensible to embed their effect into the fixed survey design.

In his second comment, the author raises the issue of using data from a survey for analytic purposes, in addition to the descriptive purposes served by traditional survey designs. While he points out that our design for an integrated survey can accommodate the needs of analysts via ‘enforcing crossings’, which provide the necessary information on various interactions between variables, he wonders “whether measures of accuracy for a broad class of analysis could be incorporated into the design, as they are for population means”.

For the typical setting of SQD discussed by the author, with specialized survey requirements, an explicit incorporation of such measures of accuracy into the design, within an analytic framework of modelling methods, may well be considered (see Chipperfield and Steel 2011). But for an integrated survey with wide-ranging and primarily descriptive requirements, and an already complex optimization algorithm involving multiple constraints, such an expanded design encompassing modelling considerations might not be practical.

In his third point, the author suggests assigning instruments to respondents with a probability that depends upon the respondent’s characteristics, and cites a particular application of this idea (Chipperfield et al. 2013). Gonzalez and Eltinge (2008) also proposed an adaptive assignment of subsampling probabilities based on data (e.g., demographic) from the first interview in a panel consumer expenditure survey with a split-questionnaire design. See also the second “adaptive design” option suggested by Gonzalez and Eltinge in Section 3 of their discussion, in the context of rare populations and low prevalence characteristics.

Such a procedure of assigning instruments to respondents could be well adopted in our setting for increased design efficiency, if the instruments are administered to subsamples from an initial sample that collects the necessary information: a process resembling two-phase sampling. As a design feature, the resulting increase in design efficiency would be factored into the design effect, although at increased design complexity. Akin to this theme is the author’s consideration, in his concluding remarks, of the possibility of using administrative data to determine a respondent’s assignment to a particular module, to enhance the efficiency of the design. This possibility is worth considering when designing an integrated survey. See also below our discussion of “adaptive design options” as considered by Gonzalez and Eltinge.

2. Gonzalez and Eltinge

Gonzalez and Eltinge discuss a number of potential “complements, and possible extensions” to our approach. We agree that many of them seem worth studying in greater detail in the future. Here, we comment specifically on two of them.

Firstly, Gonzalez and Eltinge express a very relevant concern for estimation in “multiple” or even “a large number of domains” (Section 2, second last paragraph). Insofar as the *same* instrument is to be administered to every sample unit, a practical solution could be to use in our Equations (3) and (4) the overall sample size, corresponding to a sampling design that appropriately balances between the national and multi-domain estimation purposes, and to control for the desired domain sample sizes in sample selection.

Secondly, an interesting feature of the "schematic" model (1.1) and (1.2) in Section 1 is the "deviation terms" e_Q and e_C. The survey Quality and Cost are thereby made random, instead of being completely determined at the design stage. We find this a plausible and potentially useful perspective, in order to accommodate the "adaptive design options" (Section 3), in the spirit of the MAR-SQD approach considered by Chipperfield in his third comment.

Let us consider $\mathbf{n}*$ on the right-hand side of our Equation (4) as 'the minimum required number of ideal (i.e., complete and error-free) observation units' for each module. On the left-hand side, instead of the fixed sample sizes of all the instruments, let $\mathbf{n}$ be a matrix of the same dimension as $\mathbf{A}$, where n_{ij} is in general a *random* number of ideal observation units for module I arising from the sample of instrument j. These can be random because of the presence of adaptive design options, such as two-phase design subject to screening, possible substitution of survey questionnaire by administrative data, adaptive assignment of proxy/backup modules (Section 3), or MAR-SQD (Chipperfield, 3rd comment), etc. We can now for example replace Equation (4) with

$$\mathrm{E}\{\mathrm{Diag}(\mathbf{A}\mathbf{n}^{\mathbf{T}})\} \geq \mathbf{n}^*,$$

in which case one requires that the survey accuracy satisfy, *in expectation,* the minimum requirement. Or, we can for example, use instead

$$\Pr\{\mathrm{Diag}(\mathbf{A}\mathbf{n}^{\mathbf{T}}) \geq \mathbf{n}^*\} \geq \alpha_{m\times 1},$$

for chosen threshold α-values, provided it is possible to calculate these probabilities.

Similarly, the cost function can be made stochastic and the minimisation could be with respect to the *expected* cost instead of the fully deterministic one. Together, they could provide the starting point for a modular design approach that allows for adaptive design options.

3. Dolson

In our article, we focus on the dual design problem of determining the optimal instrument composition and appropriate sample sizes given a certain instrument composition. However, as pointed out by us and confirmed by Dolson, the application of these methods requires several other major elements.

Karlberg et al. (2015) provides a synoptic overview of the "Streamlining and integration of the European social surveys" project, and enumerates many such challenges. It is encouraging to see in Dolson's description how Statistics Canada has made headway regarding many of these components, such as (using the terminology of Karlberg et al. 2015) *harmonization of variables*, *definition of modules*, *harmonisation of sampling frames* and *IT infrastructure issues*. Still, as noted by Dolson, the additional challenges in the large-scale integration are substantial. In this connection, Karlberg et al. (2015) bring up regulatory and governance issues, user relations (eliciting user needs in terms of required precision rather than sample sizes) as well as issues triggered by the increase in the number and internal heterogeneity of instruments (going from a "one instrument – one survey" situation to a situation with multiple, multi-thematic instruments, and the challenges for interviewers that this would pose).

We can only agree with Dolson's conclusion, that "It will be interesting to observe statistical offices in their assessment of the benefits of large scale integration and using this technique and their choices in whether and how best to proceed." This will, to a large extent, depend on the political will to integrate and the path chosen towards integration. In developing countries building a statistical system from scratch, it would of course make sense to deploy an integrated system right away, but for advanced statistical systems, such as the one in Canada, it would not be advisable to go for a "big bang" approach. Still, as discussed by Gonzalez and Eltinge, "integration of certain groups of surveys might be feasible" (Section 4).

Karlberg et al. (2015) propose a gradual roll-out, in which the focus would be precisely on the issues where developments at Statistics Canada have already taken place. The first objective would be to achieve "**pooling maturity**", that is, a system that allows data to be pooled across surveys to provide more precise estimates. The key requirement here is that variables are harmonized across surveys and that the sampling frames are aligned; some attention also has to be given to complex indicators (such as the poverty rate). In all likelihood, Statistics Canada could already conduct pooling for surveys where concepts and frame have been harmonized (thereby obtaining increased precision "for free") – perhaps this is already done on an experimental basis, or even in a production setting?

Only then would one proceed to actually modify the design of the surveys. The subsequent step would be to reach "**reallocation maturity**", that is, a system which would allow the application of the simplex algorithm, as described in our article, to find a solution that is globally optimal taking into account that data would be pooled – with the major constraint that the existing survey instruments would remain unchanged. Technically, this step is trivial, as it mainly requires that the way the precision requirement is specified is harmonised between surveys. However, it could generate controversy in a "stovepipe setting", since surveys would need to accept a reduced sample size and to rely on other surveys in order to reach the total sample sizes needed for their required precision. This second step might only yield quite marginal gains in terms of cost, since excessive sampling would still take place for variables with low precision requirements. This would be the case when variables are administered in the same survey as variables with high precision requirements. As this step combines potential controversy with presumably marginal gains, the gains should be assessed before it is practically implemented. If the gains are marginal, it might be better to refrain from taking this step in isolation, and instead strive to achieve "**recomposition maturity**", that is, a system in which all technical, organisational and methodological challenges have been addressed, so that current survey questionnaires can be recomposed into modular instruments through the application of the optimization algorithms presented in our article.

4. References

Chipperfield, J.O. and D.G. Steel. 2011. "Efficiency of Split Questionnaire Surveys." *Journal of Statistical Planning and Inference* 141: 1925–1932. Doi: http://dx.doi.org/10.1016/j.jspi.2010.12.003.

Chipperfield, J.O., M. Barr, and D.G. Steel. 2013. "Split Questionnaire Designs: Are They an Efficient Design Choice?" In Proceedings of the 59th ISI World Statistics Congress, 25–30 August 2013, Hong Kong. 311–316. Available at: http://2013.isiproceedings.org/Files/IPS033-P1-S.pdf (accessed June 2015).

Gonzalez, J.M. and J.L. Eltinge. 2008. "Adaptive Matrix Sampling for the Consumer Expenditure Quarterly Interview Survey." American Statistical Association, In Proceedings of the Section on Survey Research Methods, Denver, Colorado, August 6, 2008. 3069–3075. Available at: http://www.amstat.org/sections/srms/proceedings/y2008/Files/301351.pdf (accessed April 2016).

Karlberg, M., R. Reis, C. Calizzani, and F. Gras. 2015. "A Toolbox for a Modular Design and Pooled Analysis of Sample Survey Programmes." *Statistical Journal of the International Association for Official Statistics* 31: 447–462.