Survey Methodology, June 1994
Vol. 20, No. 1, pp. 3–22
Statistics Canada

Issues and Strategies for Small Area Data

M.P. SINGH, J. GAMBINO and H.J. MANTEL[1]

ABSTRACT

This paper identifies some technical issues in the provision of small area data derived from censuses, administrative records and surveys. Although the issues are of a general nature, they are discussed in the context of programs at Statistics Canada. For survey-based estimates, the need for developing an overall strategy is stressed and salient features of survey design that have an impact on small area data are highlighted in the context of redesigning a household survey. A brief review of estimation methods with their strengths and weaknesses is also presented.

KEY WORDS: Sample design strategy; Design estimates; Model estimates.

1. INTRODUCTION

For decades, administrative records and censuses were the main sources of data used for policy and planning for both large and small areas. These are still the richest source of statistical data at small area levels in most countries. During the forties and fifties, however, as the reliance on sample surveys increased, survey based estimates complemented the traditional sources because they provide more timely and cost efficient statistical data in a variety of subject matter fields. Although designed to provide reliable estimates primarily at larger area levels such as national and provincial, increasingly such surveys are being used to meet the growing demands for more timely estimates for various types and sizes of domains. No technical problem arises as long as these domains are large enough (*e.g.*, age-sex groups, larger cities and sub-provincial regions) to yield estimates of acceptable reliability. If data are needed for small domains, however, particularly if such domains cut across design strata, special estimation problems arise and several methods have recently been proposed to deal with such problems.

The main message of this paper is to emphasize the need to look at the problem of small area data in its entirety. Small area needs should be recognized at the early stages of planning for large scale surveys. The sampling design should include special features that enable production of reliable small area data using design or model estimators. The handling of this growing challenge to statistical agencies at the estimation stage should be viewed as a last resort.

In section 2, we discuss data needs and the three main sources of socio-economic data in the Canadian context, namely, the census, administrative records and surveys. Section 3 identifies some technical issues regarding the three sources of data and highlights the problems of quality measures and their interpretation. Then a need for developing an overall strategy that includes the planning, designing and estimation stages in the survey process is highlighted in section 4. Two aspects of the design, namely, clustering in a multi-stage sample design and sample allocation are discussed. In section 5, we present some sample design options being incorporated during the current redesign of the Canadian Labour Force Survey, the largest monthly household survey conducted by Statistics Canada, with a view to enhancing the survey capacity to provide better quality small area data. The purpose of section 6 is to review the many different approaches to estimation for small areas. We also suggest some new estimators and provide comments on the strengths and weaknesses of various domain estimators. A cautious approach towards the use of model estimators is stressed.

2. INFORMATION NEEDS AND DATA SOURCES

As the country's national statistical agency, Statistics Canada plays an integral role in the functioning of Canadian society. While guaranteeing the confidentiality of individual respondents' data, the agency's information describes the economic and social conditions of the country and its people. Its economic, demographic, social and institutional statistics programs produce reliable data on many aspects of life at the national, provincial, and sub-provincial levels for use by federal and provincial governments, private institutions, academics and the media. With increases in the planning, administration and monitoring of social and fiscal programs at local levels, there has been increasing demand for more and better-quality data at these levels. Three major sources of social, socio-economic and demographic data with emphasis on small area statistics are briefly discussed below.

[1] M.P. Singh, J. Gambino and H.J. Mantel, Statistics Canada, 16th Floor, R.H. Coats Building, Tunney's Pasture, Ottawa, Ontario, Canada K1A 0T6.

Census of Population: The quinquennial census of population provides benchmark data and serves as the richest source of information, available every five years, for small areas and for various characteristics/domains/target groups of policy interest such as ethnic minorities, disabled persons, youth and aboriginal peoples.

Administrative Records: Administrative records are an increasingly important source of statistical data. These are extensively used in the demographic field by statistical agencies to produce local area estimates (Schmidt 1952, Verma and Basavarajappa 1987). In certain areas, such as vital statistics, administrative records are the only source of information for production of statistics at various levels of aggregation. In others, the relative merits of administrative records compared to censuses or surveys as data sources in terms of timeliness and quality of data determine the manner and the extent to which these data sources are used. In addition to direct tabulations, administrative records are used in a number of programs as a source of supplementary information for use in improving the quality of survey-based estimates. They are also being used in the construction of sampling frames for conducting surveys. Examples at Statistics Canada include the Business Register and the Address Register of residential dwellings.

Like the census of population, administrative records are very rich in geographical detail, making them a useful source of information for small area statistics. They are available more frequently and, due to recent technological advances, they are becoming a more cost-effective data source. However, administrative data are based on definitions made for programmatic rather than statistical purposes and their content is limited. Details of a Statistics Canada program for integration and development of an administrative records system to produce statistical outputs are given by Brackstone (1987a, 1987b). Experiences in the use of administrative records in other countries are included in the conference proceedings edited by Coombs and Singh (1987).

Household Surveys Program: Household surveys have long been an important source of economic and social statistics at Statistics Canada. Surveys under this program may be placed in three groups, namely, (i) the Labour Force Survey, (ii) Special Surveys and Supplementary Survey Programs and (iii) Longitudinal/Cyclical Surveys. These surveys are briefly introduced below indicating the scope for small area statistics in general.

Starting as a quarterly survey in 1945, the Canadian Labour Force Survey (LFS) became a monthly survey in 1952. The information provided by the survey has expanded considerably over the years and currently it provides a rich and detailed picture of the Canadian labour market. In addition to providing national and provincial estimates the survey regularly releases estimates for subprovincial areas. Regular estimates of standard labour market indicators are also in great demand for small areas such as Federal Electoral Districts, Census Divisions and Canada Employment Centres. These estimates are used by both federal and provincial governments in monitoring programs and allocating funds and other resources among various political and administrative jurisdictions.

Because of cost considerations, the LFS is heavily used as a vehicle for conducting *ad hoc* and periodic surveys at the national and provincial levels in the form of supplementary or special surveys. In the case of supplements, the LFS respondents themselves are asked additional questions, whereas for special surveys a different set of households is selected using the LFS frame. Both special and supplementary surveys are usually sponsored by other government departments and are conducted on a cost-recovery basis. For these surveys, the demands for small area statistics differ greatly from survey to survey, and generally the demands seem to be less pressing than those from the LFS itself.

Statistics Canada conducts a General Social Survey (GSS) annually to serve, in a modest way, the growing data needs on topics of current social policy interest. The GSS program (Norris and Paton 1991) consists of five survey cycles, each covering a different core topic, repeated every five years. Because of the limited size of sample (10,000 households nationally) the focus of the GSS is on estimates at the national level and on analytical statistics.

Longitudinal/panel surveys are new in the Canadian context. Statistics Canada has started two longitudinal surveys that will enrich the household survey program greatly, namely, the Survey on Labour and Income Dynamics and the National Population Health Survey. Both are large scale panel surveys and they are already creating expectations for data at sub-provincial and local area levels.

3. ISSUES IN DOMAIN ESTIMATION

There are numerous policy and technical issues that need to be addressed in the provision of small area statistics. The seriousness of these issues may vary from agency to agency and from one application to the next within the same agency depending on data quality and release policies. These issues are relevant for national and provincial estimates, but they assume higher significance in the context of small area statistics. As Brackstone (1987a) notes "on the issue of small area data evaluation, it is worth noting that error in small area estimates may be more apparent to users than error in national aggregates... at a local area level, there will be critics quick to point out deficiencies... it is true that for small areas, where estimation is more difficult, scrutiny of estimates is also more intensive". Several research and developmental studies on small area estimation are described in two volumes, one edited by Platek *et al.* (1987), and the other by Platek and Singh (1986). For a

recent overview of small area estimation techniques currently being used in United States federal statistical programs see U.S. Statistical Policy Office (1993).

Use of Administrative Records: Federal and provincial government policies are the prime factors that influence the supply as well as the demand for small area data in most situations. On the supply side, government program driven administrative records contain a wealth of statistical information that can be used to produce local area data. Examples of files being used in the Canadian context are: Family Allowance, Unemployment Insurance, Income Tax, Health, Education, Old Age Security. Income-related statistics are produced at the local area level on a regular basis. Any **change** in government policy and associated programs can have immediate impact, for better or worse, on the coverage, availability, timeliness or quality of statistics derived from the corresponding administrative records. On the demand side, as noted earlier, governments need local area data for planning, implementing and monitoring their policies.

Conceptual issues: Quite frequently, conceptual and definitional issues in a data series are confounded with sampling and estimation problems. For example, consider the Unemployment Insurance (UI) system in Canada. UI regulations stipulate different qualification and requalification periods depending on the unemployment rate in a given region such that regions with higher unemployment rates require shorter qualifying periods of continuous employment. The estimates of regional unemployment rates derived from the LFS are used in determining the eligibility for an individual to receive benefits. These local area estimates are thus continually under close scrutiny by the public and the media. Such scrutiny however refers more often to **conceptual** issues rather than estimation issues per se; aspects such as the treatment in the survey questionnaire of discouraged workers, lay-offs and job search methods are questioned.

Use of Models and Related Quality Measures: Domain estimates are produced for virtually all large scale surveys, and as long as design estimators, *i.e.*, approximately design-unbiased estimators are of acceptable quality, no problem arises. We consider two classes of design estimators. Following Schaible (1992), **direct** estimators refer to estimators which use values of the study variable only for the time period of interest and only from units in the domain (*e.g.*, the regression estimator with slope estimated using only data from the domain). Such estimators may, and often do, use information on one or more auxiliary variables from other domains or other time periods, and are design unbiased or approximately so. The second class of design estimators, **modified** direct estimators, may use information from other domains on both the auxiliary and the study variable but still retain the property of design unbiasedness or approximate unbiasedness (*e.g.*, the regression estimator with slope estimated using the whole sample). There is a growing literature on **indirect** (or **model**) **estimators**, that is, estimators which use information on both the study and auxiliary variables from outside the domain and/or the time period of interest without any reference to their design unbiasedness properties.

Most producers and users of survey data are accustomed to design estimators and the corresponding design-based inferences. They interpret the data in the context of repeated samples selected using a given probability sampling design, and use estimated design-based **cv**s (coefficients of variation-square root of design variance divided by the design estimate) as the measures of data quality. For situations where either domains are too small or the sampling design did not foresee production of small area estimates, the design estimates may lead to large design **cv**s and model estimates may be the only choice if the survey-based estimates have to be provided for individual domains. A major challenge for statisticians is how to estimate, compare and explain to the users the relative precision of estimates from a survey that produces a large number of estimates at the national, subnational and large and small domain levels, most using design estimators but a few using model estimators. The model-based **cv**s (square root of design variance of model estimate divided by the model estimate) may convey a completely different message and may be several times lower than the corresponding design-based **cv**s for the same small area and in many cases, lower than the design-based **cv**s for much larger areas.

For model estimators, it is usually straightforward to derive expressions for the corresponding mean square errors (*i.e.*, design variance + square of the design bias). Estimation of these expressions, with an adequate degree of reliability, is a different matter. If we follow the argument that the data (*e.g.*, sample size) for such domains are inadequate for producing design estimates, it is unlikely that they would be adequate for producing design estimates of the corresponding variances and biases. As the estimation of bias is relatively more difficult, some authors seek design consistent model estimators, implying perhaps that bias can be ignored. However, if the sample size within the domain is sufficiently large to make the model estimator consistent, then the design estimator itself should give reliable estimates for the domain. For model estimators, suggestions have been made to use estimates of average mean square error computed over all domains. As the need for estimates for different domains usually arises because these domains are thought to be different from each other, a challenging task is to explain why estimates from all such domains are given the same degree of reliability. Another possibility is to construct indirect model-based estimates of the variance and bias of the model estimators for **individual** domains. Finding suitable methods of estimating mean square error for individual domains should be a research priority. Another serious concern for survey practitioners is how to guard against model failures. This

suggests a need for research into model validation for complex survey situations. Further, for model estimators that use data on study variables for periods other than the time period of interest, estimates of **change** over different time periods would be of questionable quality; see Schaible (1992). Also, model estimators that borrow strength from other domains in the larger area will suffer a similar drawback when comparing differences in the two domains within the large area.

Issue of Privacy: In order to construct rich data bases for providing small area statistics, it is sometimes necessary to combine census, survey and/or administrative records. This necessitates linkage of records obtained from different sources. However, given the public's concern about privacy, record linkages should be carried out only after careful examination of all their implications. Under the Statistics Act, Statistics Canada may have access to administrative records of other departments for statistical purposes. But even for statistical purposes, as Fellegi (1987) notes, "we should have rigorous and auditable review procedures to ensure that we only carry out record linkage where the resulting privacy invasion is clearly outweighed by the public good from the new statistical information".

4. NEED FOR AN OVERALL STRATEGY

Even though large scale surveys are designed primarily for national and provincial estimates, it is rare that the estimates from such surveys relate only to the national/provincial populations as a whole. That is, invariably, such surveys are used to produce estimates for various **cross-classified** domains and in some cases for **areal** domains (*e.g.*, subprovincial) as well. In many cases, no special attention is paid to achieving a desired level of precision at the domain level either at the design or the estimation stage as long as the reliability is (believed to be) within reasonable limits. Problems arise when the cross-classified domain refers to a rare subpopulation or when the areal domain refers to a small area in which case either no estimates are possible/available or the estimates are of questionable quality. In a number of cases, this may happen simply because not enough attention was paid to these needs at the start of the survey planning process. If small area data needs are to be served using survey data then there is a need to develop an overall strategy that involves careful attention to meeting these needs at the planning, sample design and estimation stages of the survey process. For discussion of the design and estimation aspects, we will classify domains into the following two types:

Planned domains: In sampling terms these are individual strata or groups of strata for which desired samples have been planned. In the Canadian context these are typically subprovincial regions, such as Economic Regions, Unemployment Insurance Regions, and Health Planning Regions. In other cases, such domains could be larger counties, districts or similar subprovincial regions.

Unplanned domains: These are areas that were not identified at the time of design and thus may cut across design strata. Such domains can be of any size and they may create special estimation problems.

Planning: As noted earlier, the data demands from continuing periodic surveys such as the LFS are relatively much higher than from *ad hoc* surveys. In the case of periodic surveys that are redesigned every five or ten years, a suitable strategy can be developed during survey redesigns, since, in such cases, statistical agencies are usually in a much better position to project future small area data needs based on past demands. For *ad hoc* surveys, designers should include the establishment of such needs as an integral part of objective setting for the survey. Thus, in both cases, survey designers should establish the desired degree of precision, not only for national and provincial level estimates, but also for the domains of interest.

The first step of a strategy, in terms of the provision of small area data, will depend on the extent to which domains are identified in advance so that they can be treated as planned domains at the time of the design (or redesign) of the survey. If budgetary considerations do not permit reliable estimates for certain very small domains, then the option of either collapsing domains, pooling estimates over different surveys or not providing the estimates at all should be given serious consideration by survey designers in discussions with the survey sponsors. Some domains cannot be determined in advance. These unplanned domains should be handled through special estimation methods.

Sample design: In practice, it is rare that a design is optimal either for the national or provincial levels or for a single subject matter of interest. Usually varying degrees of compromise are introduced at different stages of sampling and the data collection process to satisfy theoretical and operational constraints. Depending on the data needs, estimates for domains should also form an integral part of this compromise. We will discuss two ways of taking small area data needs into account at the design stage, namely, sample allocation and the degree of clustering of the sample.

Allocation Strategy: In general, an optimum allocation strategy for national level estimates allocates samples to provinces approximately in proportion to their population. The reliability of estimates for smaller provinces in such cases suffers. Therefore a compromise allocation is usually preferred. There are different ways in which this compromise can be achieved depending on the emphasis placed on subnational estimates. Small reductions in sample sizes for larger provinces usually have little effect or the reliability of data for such provinces (or the national level data) but the corresponding sample increase in smaller provinces has significant impact on the reliability of their data.

The same principle holds for planned domains within the provinces. This is because optimum allocations in most situations are flat and the designers can exploit this feature by reallocating sample from the larger areas to planned domains that are smaller in size.

Clustering: Large scale household surveys usually involve stratified multistage designs with relatively large primary sampling units in order to make the design cost-efficient for national and provincial statistics. Such designs are thus highly clustered and, therefore, detrimental to the production of statistics for unplanned areal domains in the sense that, due to chance, some domains may be sample-rich while others may have no sample at all. Given the importance of domain estimates, attempts should be made to minimize the clustering in the sample. The following factors are important in this context: choice of frame, choice of sampling units and their sizes, number and size of strata and stages of sampling. The goal should be to make the design effects as low as possible given the operational constraints.

Estimation: No matter how much attention is paid to domain estimates at the early stages of planning and designing a particular survey, there will always be some smaller domains for which special estimation methods will be required for producing adequate estimates. Recently, synthetic estimators, which borrow strength from domains that resemble the domain of interest, have attracted a good deal of attention. However, since synthetic estimators are very sensitive to the assumption that domains resemble each other, even a small departure from the assumption can make the design bias high and put their use in question. Probability samplers, conscious of design bias, have suggested combinations of direct and synthetic estimators, with a view to addressing the design bias problem while trying to retain the strengths of the synthetic estimator. Empirical Bayes and similar techniques have been used to assign a weight to each component in the combined estimators. A brief review of these developments is given in section 6 on estimation.

5. SAMPLE DESIGN CONSIDERATIONS

5.1 Introduction

The small area problem is usually thought of as one to be dealt with via estimation. However, as was noted in the previous section, there are opportunities to be exploited at the survey design stage. This section uses the Canadian Labour Force Survey (LFS) to illustrate this.

The current LFS design: The Canadian Labour Force Survey is a monthly survey of 59,000 households which are selected in several stages using various methods. The ultimate sampling unit, the household, remains in the sample for six months once it is selected and is then replaced. Higher stage units (primary sampling units (PSU), clusters) also rotate periodically. Each of Canada's ten provinces is divided into economic regions (ER) which the LFS further divides into self-representing areas (medium and large cities) and non-self-representing areas (the rest of the ER). Stratification and sample selection take place within these areas, and the number of stages of sampling as well as the units of sampling differ between these two types of area. For example, in areas outside cities, there are three stages of sampling, whereas there are only two in the cities. For a detailed description of the current LFS design, refer to Singh *et al.* (1990).

5.2 Sampling Stages and Sampling Units

Area frames are usually associated with clustered sampling, *i.e.*, the first-stage units of selection are typically land areas containing a number of second-stage units. If a list of the second-stage units becomes available, then sampling directly from the list becomes possible, leading to a less clustered sample. This will result not only in improved estimates (due to lower design effects) but also in better small area estimates for unplanned domains. The latter holds since, by spreading the sample more evenly, it is more likely that an unplanned areal domain will contain some selected units. In contrast, in a clustered design we are often faced with a situation where one domain has sufficient sample because it happens to contain sampled clusters while a similar domain happens to have too few or no sampled clusters to produce good estimates.

To reduce clustering in the LFS we investigated two options: (i) the possibility of replacing the area frame (with its two stage design) in the larger cities with a list frame using the Address Register and (ii) reducing the sampling stages in rural areas and smaller urban centres. The Address Register, created to improve the coverage of the 1991 Canadian census (Swain, Drew, Lafrance and Lance 1992), consists of a list of addresses, telephone numbers and geographical information for dwellings by census enumeration area (EA). One option involved the selection of a stratified simple random sample of dwellings from the Address Register frame. This sample could then be supplemented with a sample selected from a growth frame which comprises a set of dwellings that are not in the post-censal address register. Handling of growth became the major stumbling block in pursuing option (i) as no cost-efficient method could be devised and tested in time for the current redesign. However, an updating strategy for the post-censal Address Register is still being investigated for future censuses and surveys.

With regard to option (ii), in keeping with the idea that less clustering is better for small area estimates, changes in the units and reduction in the stages of sampling were investigated for the areas outside the cities. Due to the changes that have taken place in data collection techniques,

namely, from face-to-face interviewing to telephone and computer assisted interviewing, the cost-variance analyses from the past are no longer relevant. More than 80 percent of LFS interviews are now conducted by telephone. With the increase in telephone interviewing and the resulting decrease in travel, it became feasible in almost all cases to eliminate the current PSU stage and to sample EAs directly.

5.3 Stratification

One approach to stratification, similar in spirit to the above discussion on PSU size, is to replace large strata by many small ones. The hope is that a redefined domain or an unplanned domain will contain mostly complete strata. This will make the sample size in the domain more stable.

There may be several overlapping areas for which estimates are required. For example, each Canadian province is partitioned into both Economic Regions (ER) and Unemployment Insurance regions (UIR). One way to deal with this situation is to treat all the areas created by the intersections of the partitions as strata. In the Canadian case, for example, the 71 ERs and 61 UIRs yield 133 intersections, a manageable number. In some cases, however, the number of intersections may be too large to handle effectively. In addition, some of the intersections may have very small populations, making them unusable as strata.

By combining decreased clustering with smaller strata, we hope to have a design which is better able to meet small area needs. For example, the design should provide more flexibility in satisfying both ER and UIR requirements efficiently and in dealing with future changes in the definition of regions.

5.4 Allocation

If the definitions of small areas are known in advance, we may be able to treat them as planned domains and take them into account when designing the survey. The survey designer may endeavour to allocate sufficient sample in each small area to make the production of reliable estimates feasible. For large surveys such as the Canadian Labour Force Survey, this approach can, at least in theory, make the production of a great many small area estimates feasible. With a monthly sample of 59,000 households, and assuming that, say, 100 households per month are needed to produce reliable quarterly estimates, the country can be divided into about 600 non-overlapping areas, each guaranteed to have sufficient sample. Unions of such areas will also have enough sample to produce reliable monthly estimates.

Various sample allocation strategies are possible. In a top-down approach, once a provincial sample size is determined, the sample is allocated among the sub-provincial regions. However, it may turn out that it is not possible to satisfy the requirements for the reliability of sub-provincial estimates for the given provincial sample size. In a bottom-up strategy, the sample would be allocated to sub-provincial regions first in such a way that reliability objectives for each region are satisfied. As a result, we would expect comparable sample sizes in each sub-provincial region. This approach may result in a provincial sample size that is bigger than the one specified in the top-down approach. Regardless of which of the two strategies is used, adjustments to the initial allocations will usually be required. The resulting allocation will likely resemble a compromise between proportional allocation and equal allocation. In practice, the survey designer must perform a complex juggling act among provincial reliability requirements, sub-provincial requirements for one or more sets of regions, total survey costs and in-the-field details.

The approach taken in the current LFS redesign may be useful in other surveys as well. The sample was allocated in two steps: first, a core sample of 42,000 households was allocated to produce good estimates at the national and provincial levels; then the remaining sample was allocated to produce the best possible sub-provincial estimates. The resulting compromise allocation will produce reliable estimates for almost all planned domains. The compromise resulted in only minor losses at the provincial level and substantial gains at the subprovincial level. For example, the expected CVs for 'unemployed' for Ontario and Quebec are 3.2 and 3.0 per cent, respectively, instead of 2.8 and 2.6. The corresponding figures for Canada are 1.51 and 1.36. Optimizing for the provincial level yields CVs as high as 17.7 per cent for UI regions. With the compromise allocation, the worst case is 9.4 per cent.

Sample redistribution: There is usually some scope for moving sample from one area to another. For example, reducing the sample size by 1,000 households in a large province and making a corresponding increase in a small province will cause a marginal deterioration in the quality of provincial estimates in the former but will improve the estimates in the latter significantly. Similar movements of sample can be attempted within province.

5.5 Other Considerations

Change in definitions of small areas: Survey designers are faced with the fact that the definitions of planned domains may change during the life of a design and they may then have to treat the new domains as unplanned domains. For example, it is quite possible that the definitions of Unemployment Insurance Regions will change two or three years after the new LFS design is introduced in 1995. To deal with this at the design stage, the best that the survey designer can do is to choose as building blocks areas which are standard (*e.g.*, census-defined areas whose definitions are fairly stable) and hope that the redefined regions are unions of these standard areas. This is the approach that was taken in the current LFS redesign.

An alternative is to adopt an update strategy. This entails a reselection of units, doing it in such a way that the overlap between the originally selected units and the newly selected ones is maximized. By taking this approach, the number of new units that have to be listed is minimized. This also minimizes other field disruptions such as the need to hire new interviewers.

6. ESTIMATION

The purpose of this section is to review some of the different approaches to estimation of totals for small areas. No attempt is made to provide an exhaustive review; the discussion indicates the trend of developments in small area estimation research. For a detailed review, see the recent paper by Ghosh and Rao (1993). To facilitate this review we will classify small area estimation methods into two types. This is just one of many possible classification schemes. The first class of estimators we call design estimators, *i.e.*, (approximately) design unbiased estimators, which includes direct and modified direct estimators. As noted earlier, design estimators are often unsatisfactory, having a large variance due to small sample sizes (or even no sample at all) in the small areas. The second class we call indirect (or model) estimators, and it includes synthetic and combined estimators. Some of these estimators are compared empirically in an earlier version of this paper by Singh, Gambino and Mantel (1992).

6.1 Design Estimators

Direct Estimators: Direct small area estimators are based on survey data from only the small area, perhaps making use of some auxiliary data from census or administrative sources in addition to the survey data. The simplest direct estimator of a total is the expansion estimator,

$$\hat{Y}_{e,a} = \sum_{i \in s_a} w_i y_i, \tag{6.1}$$

where s_a is the part of the sample in small area a and w_i is the survey weight for unit i. This estimator is unbiased; however, it may have high variability due to the random sample size in area a.

If the population size N_a is known then a post-stratified estimator,

$$\hat{Y}_{pst,a} = N_a \sum_{i \in s_a} w_i y_i \Big/ \sum_{i \in s_a} w_i = N_a \hat{Y}_{e,a}/\hat{N}_{e,a} = N_a \bar{y}_{e,a}, \tag{6.2}$$

may be used. This estimator is more stable than the expansion estimator; however, there may be some ratio estimation bias in complex surveys.

If the sampling scheme is stratified and the $N_{h,a}$ are known, where $N_{h,a}$ is the population size in stratum h and small area a, an alternative post-stratified estimator is $\hat{Y}_{st,pst,a} = \sum_h (N_{h,a} \sum_{i \in s_{h,a}} w_i y_i / \sum_{i \in s_{h,a}} w_i) = \sum_h N_{h,a} \hat{Y}_{h,e,a}/\hat{N}_{h,e,a} = \sum_h N_{h,a} \bar{y}_{h,a}$. The strata may also be post-strata instead of design strata.

Ratio estimation is similar to post-stratified estimation, the difference being that another auxiliary variable is used in place of the population counts N_a and $N_{h,a}$. For example, if x is a covariate for which the small area totals, X_a, or the stratum small area totals, $X_{h,a}$, are known then we may define the ratio estimators

$$\hat{Y}_{r,a} = X_a \hat{R}_a \quad \text{and} \quad \hat{Y}_{st,r,a} = \sum_h X_{h,a} \hat{R}_{h,a}, \tag{6.3}$$

where $\hat{R}_a = \hat{Y}_{e,a}/\hat{X}_{e,a}$ is an estimate of the ratio Y_a/X_a and $\hat{R}_{h,a} = \hat{Y}_{h,e,a}/\hat{X}_{h,e,a}$.

A regression estimator attempts to account for differences between small area subpopulation and subsample values of the covariates via an estimated regression relationship between the variate of interest, y, and the covariates, x. An advantage of regression type estimation is that it is easily extended to vector covariates. The estimator is given by

$$\hat{Y}_{reg,a} = \hat{Y}_a + \hat{\beta}_a (X_a - \hat{X}_a), \tag{6.4}$$

where $\hat{Y}_a$ may be an expansion or post-stratified estimator, $\hat{X}_a$ must be calculated in the same way as $\hat{Y}_a$, and $\hat{\beta}_a = \sum_{i \in s_a} v_i^{-1} w_i y_i x_i' \{ \sum_{i \in s_a} v_i^{-1} w_i x_i x_i' \}^{-1}$ where v_i are given weights for the regression. Note that $\hat{\beta}_a = \hat{R}_a$ when x is scalar and $v_i = x_i$. When $\hat{Y}_a$ and $\hat{X}_a$ are expansion estimators this estimator is also called the generalized regression estimator. Approximate design unbiasedness of this estimator follows from that of $\hat{Y}_a$ and $\hat{X}_a$.

As with the ratio type estimators, regression type estimation may also be applied within design strata or post-strata.

Modified Direct Estimators: Modified direct estimators may use survey data from outside the domain; however, they remain approximately design unbiased. By a modified direct estimator we mean a direct estimator with a synthetic adjustment for model bias; since the adjustment would have approximately zero expectation with respect to the design, the modified estimator is approximately design unbiased if the direct estimator is. An example is obtained by replacing $\hat{\beta}_a$ in (6.4) by a synthetic estimator $\hat{\beta} = \sum_{i \in s} v_i^{-1} w_i y_i x_i' \{ \sum_{i \in s} v_i^{-1} w_i x_i x_i' \}^{-1}$; we will denote this estimator by $\hat{Y}_{sreg,a}$. $\hat{\beta}$ would generally be more stable than $\hat{\beta}_a$; the choice between them would depend on the size of the variance of $\hat{\beta}_a$ relative to the variation in the $\beta_a s$ over areas a. A compromise is to take a weighted average $\lambda_a \hat{\beta}_a + (1 - \lambda_a)\hat{\beta}$ where λ_a is suitably chosen;

options for the choice of λ_a are discussed under combined estimators in Section 6.2. A second example is obtained by replacing $\hat{\beta}_a$ in (6.4) by $\hat{R} = \hat{Y}_e/\hat{X}_e$; note that $\hat{R}$ is a special case of $\hat{\beta}$ where x is scalar and $v_i = x_i$.

6.2 Indirect Estimators

Synthetic Estimators: Synthetic estimation methods are based on an assumption that the small area is similar in some sense to another area, often a larger area which contains it. Estimates for the other area would generally be more reliable than those for the small area. The resulting synthetic estimator would then have small variance, though it may be badly biased if the underlying assumption is violated.

One of the simplest synthetic estimators arises from the assumption that the small area mean is equal to the overall mean. This leads to the mean synthetic estimator

$$\hat{Y}_{syn,m,a} = N_a \sum_{i\in s} w_i y_i \Big/ \sum_{i\in s} w_i = N_a \bar{y}. \tag{6.5}$$

A more common synthetic estimator is based on stratification or post-stratification,

$$\hat{Y}_{syn,st,m,a} = \sum_h N_{h,a} \sum_{i\in s_h} w_i y_i \Big/ \sum_{i\in s_h} w_i = \sum_h N_{h,a} \bar{y}_h.$$

As with direct estimators, ratio synthetic estimation may be based on other auxiliary data besides the population counts N_a or $N_{h,a}$. For example, the common ratio synthetic estimators based on a covariate x are defined as

$$\hat{Y}_{syn,r,a} = X_a \hat{Y}_e/\hat{X}_e \quad \text{and} \quad \hat{Y}_{syn,st,r,a} = \sum_h X_{h,a} \hat{Y}_{h,e}/\hat{X}_{h,e}, \tag{6.6}$$

where $\hat{Y}_e = \sum_{i\in s} w_i y_i$ is the expansion estimator of the population total for y and $\hat{Y}_{h,e} = \sum_{i\in s_h} w_i y_i$. $\hat{X}_e$ and $\hat{X}_{h,e}$ are similarly defined. These estimators have been studied by Gonzalez (1973), Gonzalez and Waksberg (1973) and Ghangurde and Singh (1977, 1978), among others.

Singh and Tessier (1976) suggested an alternative ratio synthetic estimator, using X instead of $\hat{X}_e$, defined as

$$\tilde{Y}_{syn,r,a} = X_a \hat{Y}_e/X. \tag{6.7}$$

Both $\hat{Y}_{syn,r,a}$ and $\tilde{Y}_{syn,r,a}$ have the same synthetic bias and the ratio bias in $\hat{Y}_{syn,r,a}$ will be negligible for large samples. The choice between these two estimators depends on ρ, the correlation of $\hat{Y}_e$ and $\hat{X}_e$. It can be shown that for large samples $V(\hat{Y}_{syn,r,a}) \le V(\tilde{Y}_{syn,r,a})$ if $\rho \ge 0.5 c_x/c_y$, where c_x and c_y are the coefficients of variation of $\hat{X}_e$ and $\hat{Y}_e$, respectively. In most cases, when ρ is high or the population is skewed, $\hat{Y}_{syn,r,a}$ would be preferred; however, when c_x is high and the correlation is only moderate, $\tilde{Y}_{syn,r,a}$ may be the better choice.

In some situations information on a second auxiliary variable (z) in addition to x may be available. Then a bivariate ratio synthetic estimator may be constructed:

$$\hat{Y}^{(2)}_{syn,r,a} = \gamma_a X_a \hat{Y}_e/\hat{X}_e + (1 - \gamma_a) Z_a \hat{Y}_e/\hat{Z}_e, \tag{6.8}$$

where γ_a is suitably chosen. Extensions to a multivariate ratio synthetic estimator may be considered following Olkin (1958).

Regression synthetic estimation is similar to ratio synthetic,

$$\hat{Y}_{syn,reg,a} = \hat{\beta} X_a,$$

$$\hat{\beta} = \sum_{i\in s} v_i^{-1} w_i y_i x_i' \left\{ \sum_{i\in s} v_i^{-1} w_i x_i x_i' \right\}^{-1}. \tag{6.9}$$

Again, regression synthetic estimation may also be applied within design strata or post-strata. Royall (1979) suggested a slight variation, $\hat{Y}_{syn,Roy,a} = \sum_{i\in s_a} y_i + \hat{\beta}(X_a - \sum_{i\in s_a} x_i)$, where the sum of y-values for only units not included in the sample is estimated synthetically.

Remark: The examples of modified direct estimators presented in Section 6.1 can also be considered to be ratio or regression synthetic estimators with a design-based adjustment to correct for bias. For example, we may write $\hat{Y}_{sreg,a} = \hat{Y}_{syn,reg,a} + (\hat{Y}_a - \hat{\beta}\hat{X}_a)$ where $\hat{Y}_a - \hat{\beta}\hat{X}_a$ is an estimate of the bias of $\hat{Y}_{syn,reg,a}$. Similarly, $\hat{Y}_{sreg,a}$ can also be written as the Royall estimator, $\hat{Y}_{syn,Roy,a}$, with a design-based adjustment for bias.

Purcell and Kish (1980) discuss another type of synthetic estimation which they call SPREE (structure preserving estimation) for small area estimation of frequency data. Detailed historical counts, perhaps from a census, are combined with less detailed current survey estimates to produce detailed estimates of current counts. The assumption here is that certain relationships among the detailed counts are stable over time.

Combined Estimators: By a combined estimator we mean a weighted average of a design estimator and a synthetic estimator,

$$\hat{Y}_{com,a} = \lambda_a \hat{Y}_{des,a} + (1 - \lambda_a) \hat{Y}_{syn,a}, \tag{6.10}$$

where λ_a is suitably chosen. The aim here is to balance the potential bias of the synthetic estimator against the instability of the design estimator. There are three broad approaches which may be used to define the weights λ_a in (6.10); they may be fixed in advance, sample size dependent, or data dependent.

The first and simplest approach to weighting is to fix the weights in advance, for example, to take a simple average. However, this does not make any allowance for

the actual observed reliability of the design estimator. For some realized samples the design estimator for small area a is more reliable than for other realized samples. The weight given to the design estimator should reflect this.

The second general approach to weighting of the design and synthetic parts is called sample size dependent, in which the weights are functions of the ratio $\hat{N}_{e,a}/N_a$. Another possibility, not considered here, is to base the weights on the realized sample values of a covariate x; for example, the weight could be a function of $\hat{X}_{des,a}/X_a$ or of $S^2_{x,a}/\sigma^2_{x,a}$ where $S^2_{x,a}$ is the realized variance of $\hat{X}_{des,a}$, conditional on $\hat{N}_{e,a}$ or some other relevant aspect of the realized sample, and $\sigma^2_{x,a}$ is the unconditional variance of $\hat{X}_{des,a}$.

Some specific estimators in this class have been proposed earlier. Drew, Singh, and Choudhry (1982) proposed the sample size dependent estimator

$$\hat{Y}_{ssd,r,a} = \lambda_a \hat{Y}_{r,a} + (1 - \lambda_a)\hat{Y}_{syn,r,a}, \qquad (6.11a)$$

where

$$\lambda_a = \begin{cases} 1 & \text{if} \quad \hat{N}_{e,a} \geq \delta N_a \\ \hat{N}_{e,a}/\delta N_a & \text{otherwise} \end{cases} \qquad (6.11b)$$

and δ is subjectively chosen to control the contribution of the synthetic component. Särndal (1984) suggested

$$\hat{Y}_{ssd,reg,a} = \lambda_a \hat{Y}_{sreg,a} + (1 - \lambda_a)\hat{Y}_{syn,reg,a}, \qquad (6.12)$$

where $\lambda_a = \hat{N}_{e,a}/N_a$. Rao (1986) suggested a modification to this in which λ_a would be taken to be 1 whenever $\hat{N}_{e,a} \geq N_a$. Särndal and Hidiroglou (1989) refined Rao's suggestion by taking $\lambda_a = (\hat{N}_{e,a}/N_a)^{h-1}$ when $\hat{N}_{e,a} < N_a$, where h is chosen judgementally to control the contribution of the synthetic component.

It is the bias of the synthetic component that is of concern when using these sample size dependent estimators in practice. The weight associated with the synthetic component should be such that the bias is kept within reasonable limits. For example, the sample size dependent estimator of Drew, Singh and Choudhry (1982), with generalized regression estimation replacing the ratio estimation and $\delta = 2/3$, is currently used in the Canadian Labour Force Survey to produce domain estimates. For a majority of domains the weight attached to the synthetic component is zero as the direct estimator itself provides the required degree of reliability. For other domains the weight attached to the synthetic component is about 10% on average and never exceeds 20%. Depending on the risk of bias that one is willing to take, δ may lie in the range [2/3,3/2] for most practical situations.

The third approach to weighting we call data dependent. The optimal weights for combining two estimators generally depend on the mean squared errors of the estimators and their covariance. These quantities would generally be unknown but may be estimated from the data. For our combined estimators this would usually require some modelling of the bias of the synthetic part. An early and well known example of this approach is due to Fay and Herriot (1979). They model the biases of the synthetic estimators for the small areas as independent random effects with an unknown but fixed variance. To be more specific, if $\hat{Y}_{des,a}$ is the design estimator then they consider the model $Y_a = X_a\beta + \alpha_a$ and $\hat{Y}_{des,a} = Y_a + \epsilon_a$ where $\alpha_a \sim (0,\sigma^2)$, $\epsilon_a \sim (0,\nu_a^2)$, and α_a and ϵ_a are independent and uncorrelated over a, σ^2 is unknown and ν_a^2 are assumed known (in practice they would need to be estimated). For a given value of σ^2 the optimal weights for combining $\hat{Y}_{des,a}$ and $X_a\hat{\beta}$ can be calculated. An estimate of σ^2 is obtained by the method of fitting constants and substituted into the optimal weights. Some protection against model mis-specification is obtained by truncating the resulting estimate if it deviates from the direct estimate by more than a specified multiple of ν_a. Schaible (1979) and Battese and Fuller (1981) also consider empirically estimated optimal weights λ_a in (6.12) based on similar random effects models for the small area totals.

Prasad and Rao (1990) provide an estimator of the mean square error of the Fay-Herriot estimator which makes allowance for the estimation of the variance components. Kott (1989) proposes a design consistent estimator of the mean square error, but finds it to be very unstable.

Another alternative is to use historical data to calculate the weights; this has the advantage that the weights may be more stable than if they are estimated from current survey data; however, there is an underlying assumption that the optimal weights are stable over time.

Remark: In sample size dependent estimation the weights are allowed to depend on the observed size of the subsample s_a, but not on the values of the variate of interest. This non-dependence of the weights on the variate of interest has advantages and disadvantages. An advantage is that the same weights would be used for estimation of totals for all variates of interest; they need to be calculated only once. More importantly, the estimate of the sum of two variables is the sum of the estimates of the two variables. A disadvantage is that the weights do not directly take account of either the reliability of the design estimator for the variate of interest or the likely magnitude of the bias of the synthetic estimator.

Combining data over time: For repeated surveys pooling of data over survey occasions to increase the reliability of estimates is a common practice. Depending on the rotation pattern used for such surveys, significant gains in reliability can be achieved. This pooling or averaging over time is thus of particular interest in the context of domain estimation where reliability is usually low. For domain

estimation in the Canadian Labour Force Survey it is normal practice to use a sample size dependent estimator based on three month average estimates of employed and unemployed. Due to the six month rotation scheme used, as noted earlier, averaging over three months increases the sample size by one third. If samples completely overlap between periods then averaging does not result in any gain in efficiency. For other rotation patterns the sample size for domain estimates could be more than doubled through this process. There is, however, a conceptual problem with pooled estimates, in that such estimates refer to an average of the parameter of interest (*e.g.*, unemployment) over a period of, say, three months.

In composite estimation the current design estimator is combined with the composite estimator for the previous period, updated by an estimate of change based on the common sample. This idea was used, though not in the context of small area estimation, by Jessen (1942), and Patterson (1950), among others. Binder and Hidiroglou (1988) provide a review. The weights for the combination are typically estimates of the optimal weights under the assumption that these weights are time stationary. These data dependent weights have the disadvantage that they lead to inconsistency of estimates for different characteristics and their sums.

A recent development in small area estimation techniques is the use of time series methods for periodic surveys. The relationship between parameters of interest for different time periods is modelled and this model is exploited to improve the efficiency of the estimates for the current occasion. In most cases some allowance must also be made, through modelling or otherwise, for the non-independence of samples for different survey occasions due to the sample rotation scheme. Some references for this time series approach are Choudhry and Rao (1989), Pfeffermann and Burck (1990), Singh, Mantel and Thomas (1994) and Singh and Mantel (1991). All of these are generalizations of the Fay-Herriot model which allow the regression parameters, small area effects, and survey errors to evolve over time according to various time series models. The vector of small area estimates that results from this approach can be written as a weighted average of the vector of design estimates and a vector of synthetic estimates which are based on past data and the current values of covariates; however, the matrix of weights would not generally be diagonal so that the estimator for any single small area would generally depend also on the design estimates and synthetic estimates for other small areas.

7. CONCLUSION

To produce adequate survey-based domain estimates that are timely and up to date, sample designers must face several challenging tasks. The first is to convince the sponsors/program managers that some small area data needs cannot be met as a by-product of a system designed optimally for national/sub-national estimates. Significant gains, which may vary from survey to survey, can be achieved at the domain level at a marginal reduction in reliability at higher levels. There is a need to develop an overall strategy that incorporates desired reliability for the planned domains as well as for higher levels through compromise allocations, and reduced clustering to help improve estimates for unplanned domains. It should be noted that many of the planned domains at design time may become unplanned (revised) over time in the context of continuous surveys.

The overall strategy should also include consideration of both design estimators for larger domains and model estimators for small domains. A model estimator should be preferred over a design estimator only if its mean square error (design variance + bias^2) is estimable and it is sufficiently smaller than the corresponding variance of the design estimator. We should have estimates of mean square error for each of the individual domains. An option that statistical agencies can exercise is to pool similar domains or pool estimates over different time periods for the same domain. They may even suppress estimates for some domains on account of data reliability or privacy concerns.

The second challenging task for statisticians is to explain to users the different types of measures of reliability for different sets of estimates from the same survey. It is hoped that with more research on model validation and better estimates of mean square errors, designers will get more confidence in using model estimators for small domains. In the meantime model estimators should be used with caution even if they have significantly smaller coefficients of variation.

Censuses, supplemented by data from administrative records, are likely to remain the primary source of small area socio-economic data, especially for countries having a quinquennial census of population and housing. Also, concerns about problems with conceptual issues in the context of data for administrative records are likely to continue until statistical agencies are given an opportunity to influence the development of the forms used to collect such data. Until then, this immensely rich data source cannot be fully exploited for statistical purposes and more so for domain estimation.

ACKNOWLEDGMENT

We are grateful to Jon Rao for handling this paper as an Associate Editor and to the referees for many constructive suggestions.

REFERENCES

BATTESE, G.E., and FULLER, W.A. (1981). Prediction of county crop areas using survey and satellite data. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 500-505.

BINDER, D., and HIDIROGLOU, M.A. (1988). Sampling in time. *Handbook of Statistics*, (Eds. P.R. Krishnaiah and C.R. Rao). New York: Elsevier Science, 187-211.

BRACKSTONE, G.J. (1987a). Small area data: Policy issues and technical challenges. *Small Area Statistics*, (Eds. R. Platek, J.N.K. Rao, C.-E. Särndal and M.P. Singh). New York: Wiley and Sons, 3-20.

BRACKSTONE, G.J. (1987b). Statistical uses of administrative data: Issues and challenges. *Proceedings: Symposium on Statistical Uses of Administrative Data*, (Eds. J.W. Coombs and M.P. Singh), Statistics Canada, 5-16.

CHOUDHRY, G.H., and RAO, J.N.K. (1989). Small area estimation using models that combine time series and cross-sectional data. *Proceedings: Symposium on Analysis of Data in Time*, (Eds. A.C. Singh and P. Whitridge), Statistics Canada, 67-74.

COOMBS, J.W., and SINGH, M.P. (Eds.) (1987). *Proceedings: Symposium on Statistical Uses of Administrative Data*, Statistics Canada.

DREW, J.D., SINGH, M.P., and CHOUDHRY, G.H. (1982). Evaluation of small area estimation techniques for the Canadian Labor Force Survey. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 545-550.

FAY, R.E., and HERRIOT, R.A. (1979). Estimates of income for small places: An application of James-Stein procedures to census data. *Journal of the American Statistical Association*, 74, 269-277.

FELLEGI, I.P. (1987). Opening Remarks. *Proceedings: Symposium on Statistical Uses of Administrative Data*, (Eds. J.W. Coombs and M.P. Singh), Statistics Canada, 1-2.

GHOSH, M., and RAO, J.N.K. (1993). Small area estimation: An appraisal. To appear in *Statistical Science*.

GHANGURDE, P.D., and SINGH, M.P. (1977). Synthetic estimates in periodic household surveys. *Survey Methodology*, 3, 152-181.

GHANGURDE, P.D., and SINGH, M.P. (1978). Evaluation of efficiency of synthetic estimates. *Proceedings of the Social Statistics Section, American Statistical Association*, 53-61.

GONZALEZ, M.E. (1973). Use and evaluation of synthetic estimators. *Proceedings of the Social Statistics Section, American Statistical Association*, 33-36.

GONZALEZ, M.E., and WAKSBERG, J. (1973). Estimation of the error of synthetic estimates. Presented at the first meeting of the International Association of Survey Statisticians, Vienna, Austria.

JESSEN, R.J. (1942). Statistical investigation of a farm survey for obtaining farm facts. *Iowa Agricultural Station Research Bulletin*, 304, 54-59.

NORRIS, D., and PATON, D. (1991). Canada's General Social Survey: Five years of experience. *Survey Methodology*, 17, 227-240.

OLKIN, I. (1958). Multivariate ratio estimation for finite populations. *Biometrika*, 45, 154-165.

PATTERSON, H.D. (1950). Sampling on successive occasions with partial replacement of units. *Journal of the Royal Statistical Society*, Series B, 12, 241-255.

PFEFFERMANN, D., and BURCK, L. (1990). Robust small area estimation combining time series and cross-sectional data. *Survey Methodology*, 16, 217-237.

PLATEK, R., RAO, J.N.K., SÄRNDAL, C.-E., and SINGH, M.P. (1987). *Small Area Statistics*. Invited Presentations. New York: Wiley.

PLATEK, R., and SINGH, M.P. (1986). Small Area Statistics, An International Symposium' 85 (contributed papers), Technical Report Series of the Laboratory for Research in Statistics and Probability, Carleton University, University of Ottawa, Canada.

PRASAD, N.G.N., and RAO, J.N.K. (1990). The estimation of mean squared errors of small-area estimators. *Journal of the American Statistical Association*, 85, 163-171.

PURCELL, N.J., and KISH, L. (1980). Postcensal estimates for local areas (or Domains). *International Statistical Review*, 48, 3-18.

RAO, J.N.K. (1986). Synthetic estimates, SPREE and best model based predictors. *Proceedings of the Conference on Survey Research Methodology in Agriculture*, American Statistical Association and National Agricultural Statistics Service, USDA, 1-6.

ROYALL, R.M. (1979). Prediction models in small area estimation. In *Synthetic Estimates for Small Area*, NIDA Research Monograph Series 24, U.S. Department of Health, Education and Welfare.

SÄRNDAL, C.-E. (1984). Design-consistent versus model-dependent estimators for small domains. *Journal of the American Statistical Association*, 79, 624-631.

SÄRNDAL, C.-E., and HIDIROGLOU, M.A. (1989). Small domain estimation: A conditional analysis. *Journal of the American Statistical Association*, 84, 266-275.

SCHAIBLE, W.L. (1979). A composite estimator for small area statistics. *Synthetic Estimates for Small Area*, NIDA Research Monograph Series 24, U.S. Department of Health, Education and Welfare, Library of Congress catalogue number 79-600067, 36-53.

SCHAIBLE, W.L. (1992). Use of small area estimators in U.S. federal programs. *Small Area Statistics and Survey Designs*, (Vol. 1), (Eds. G. Kalton, J. Kordos and R. Platek). Warsaw: Central Statistical Office, 95-114.

SCHMIDT, R.C. (1952). Short-cut methods for estimating county populations. *Journal of the American Statistical Association*, 47, 232-238.

SINGH, A.C., and MANTEL, H.J. (1991). State space composite estimation for small areas. *Proceedings: Symposium 91, Spatial Issues in Statistics*, Statistics Canada, 17-25.

SINGH, A.C., MANTEL, H.J., and THOMAS, B.W. (1994). Time series EBLUPs for small areas using survey data. *Survey Methodology*, 20, 33-43.

SINGH, M.P., DREW, J.D., GAMBINO, J.G., and MAYDA, F. (1990). *Methodology of the Canadian Labour Force Survey*. Catalogue No. 71-526, Statistics Canada.

SINGH, M.P., GAMBINO, J.G., and MANTEL, H. (1992). Issues and options in the provision of small area statistics. *Small Area Statistics and Survey Designs*, (Vol. 1), (Eds. G. Kalton, J. Kordos and R. Platek). Warsaw: Central Statistical Office, 37-75.

SINGH, M.P., and TESSIER, R. (1976). Some estimators for domain totals. *Journal of the American Statistical Association*, 71, 322-325.

SWAIN, L., DREW, J.D., LAFRANCE, B., and LANCE, K. (1992). The creation of a residential address register for coverage improvement in the 1991 Canadian Census. *Survey Methodology*, 18, 127-142.

U.S. STATISTICAL POLICY OFFICE (1993). Indirect Estimators in Federal Programs. Statistical Policy Working Paper 21. Prepared by the subcommittee on Small Area Estimation, Federal Committee on Statistical Methodology.

VERMA, R.B.P., and BASAVARAJAPPA, J.G. (1987). Recent developments in the regression method for estimation of population for small areas in Canada. *Small Area Statistics*, (Eds. R. Platek, J.N.K. Rao, C.-E. Särndal and M.P. Singh). New York: Wiley and Sons, 46-61.

COMMENT

W.A. FULLER[1]

The authors are to be congratulated on an excellent description of the design and estimation considerations associated with domains. The authors discuss estimation for planned domains, particularly situations in which domain membership can be identified in the frame, and estimation for unplanned domains including domains for which the domain membership cannot be determined from the frame. This is a fine contribution to the growing literature on domain estimation.

The authors give a particularly good description of the planning, data collection, and processing activities associated with surveys conducted by Statistics Canada. Included are the traditional design problems of balancing needs for domain estimation with desire for efficiency at higher levels, the importance of confidentiality in using administrative records in constructing domain estimates, and the importance of definitional compatibility in attempting to combine information from different sources.

The importance of considering domain estimation at the design stage is very well taken and is a point often ignored by authors concentrating on small area estimation. As the authors emphasize, careful design can often enable one to construct estimates for domains in a direct and design consistent manner. I am sure that those actually designing surveys have considered the importance of clustering when designing surveys that will be used for domain estimation, but it is pleasant to see an explicit discussion.

The authors describe several types of estimators for domains. Their classification emphasizes the number of alternatives available to the practitioner. It is possible to use the theoretical mean square errors to provide information on the relative merits of the estimators. As an example of such a comparison, assume a simple random sample of size n selected from a population divided into K domains. Assume that the domain sizes and the domain means of an auxiliary variable, X, are available. Consider the three regression estimators of the domain mean,

$$\hat{\mu}_{(1)yi} = \bar{y}_{i.} + (\mu_{xi} - \bar{x}_{i.})b_i,$$

$$\hat{\mu}_{(2)yi} = \bar{y}_{i.} + (\mu_{xi} - \bar{x}_{i.})b_.$$

and

$$\hat{\mu}_{(3)yi} = \bar{y}_{..} + (\mu_{xi} - \bar{x}_{..})b_.,$$

where

$$(\bar{x}_{..}, \bar{y}_{..}) = \sum_{i=1}^{k} N^{-1}N_i(\bar{x}_{i.}, \bar{y}_{i.}),$$

$$(\bar{x}_{i.}, \bar{y}_{i.}) = n_i^{-1} \sum_{j=1}^{n_i} (X_{ij}, Y_{ij}),$$

$$b_i = \left[\sum_{j=1}^{n_i} (X_{ij} - \bar{x}_{i.})^2 \right]^{-1} \times \sum_{j=1}^{n_i} (X_{ij} - \bar{x}_{i.})(Y_{ij} - \bar{y}_{i.}),$$

$$b_. = \left[\sum_{i=1}^{k} N^{-1}N_i n_i^{-1} \sum_{j=1}^{n_i} (X_{ij} - \bar{x}_{i.})^2 \right]^{-1} \times \sum_{i=1}^{k} N^{-1}N_i n_i^{-1} \sum_{j=1}^{n_i} (X_{ij} - \bar{x}_{i.})(Y_{ij} - \bar{y}_{i.}),$$

n_i is the number of observations in domain i, N_i is the population size of domain i, μ_{xi} is the population mean of X for domain i, and $\mu_{x.}$ is the grand population mean of X. In the authors' terminology, the first estimator is a direct regression estimator, the second is a modified direct estimator, and the third is a synthetic estimator. We have

$$\mathrm{MSE}\{\hat{\mu}_{(1)yi} | n_i\} = n_i^{-1}(1 + n_i^{-1}) V\{Y_{\ell j} - \beta_\ell X_{\ell j} | \ell = i\} + O(n_i^{-2}),$$

$$\mathrm{MSE}\{\hat{\mu}_{(2)yi} | n_i\} = n_i^{-1}(1 + n^{-1}) V\{Y_{\ell j} - \beta X_{\ell j} | \ell = i\} + O(n^{-2}),$$

$$\mathrm{MSE}\{\hat{\mu}_{(3)yi} | n_i\} = (1 + n^{-1}) \times \sum_{i=1}^{k} N^{-2}N_i^2 n_i^{-1} V\{Y_{\ell j} - \beta X_{\ell j} | \ell = i\} + (\mu_{xi} - \mu_{x.})^2 V\{b_.\} + [\mu_{yi} - \mu_{y.} - \beta(\mu_{xi} - \mu_{x.})]^2 + O(n^{-2}),$$

where $V\{b_.\} = E\{(b_. - \beta)^2\}$, $V\{a_\ell | \ell = i\}$ is the variance of the variable a for domain i,

$$\beta_i = [V\{X_{\ell j} | \ell = i\}]^{-1} C\{Y_{\ell j}, X_{\ell j} | \ell = i\}$$

[1] W.A. Fuller, Distinguished Professor, Statistical Laboratory and Department of Statistics, Iowa State University, Snedecor Hall, Ames, Iowa.

and

$$\beta = \left[\sum_{i=1}^{k} N^{-1} N_i V\{X_{\ell j} \mid \ell = i\} \right]^{-1} \times \sum_{i=1}^{k} N^{-1} N_i C\{Y_{\ell j}, X_{\ell j} \mid \ell = i\}.$$

The estimator $\hat{\mu}_{(1)yi}$ uses only information in the sample of n_i observations. Hence, all properties of the estimator are functions of n_i and of the domain parameters. The regression bias is order n_i^{-1} and the variance is order n_i^{-1}. The estimator $\hat{\mu}_{(2)yi}$ uses the domain means, but the entire sample to estimate the regression coefficient. Hence, the basic variance remains order n_i^{-1} and will be larger than the basic variance of $\hat{\mu}_{(1)yi}$ in those situations where $\beta_i \neq \beta$. However, the second order contribution to the variance is order $n_i^{-1}\, n^{-1}$ for $\hat{\mu}_{(2)yi}$ and is order n_i^{-2} for $\hat{\mu}_{(1)yi}$. Also, the regression bias for $\hat{\mu}_{(2)yi}$ is order n^{-1}. If the domains were strata, $\hat{\mu}_{(1)yi}$ might be called the separate regression estimator and $\hat{\mu}_{(2)yi}$ might be called the combined regression estimator.

The estimator $\hat{\mu}_{(3)yi}$ is a synthetic estimator and has a variance of order n^{-1} instead of the order n_i^{-1} variance of the first two estimators. The cost of this reduction in variance is that the bias is order one. Only if the regression line is the same for the domain as for the entire population will the bias be zero.

The average mean square error of the three estimators for any subset of small areas can be estimated. If the n_i are small, the estimated variances will provide only limited information for discriminating among estimators. Likewise, there is only one degree of freedom for bias squared for one particular domain. However, a large domain deviation, relative to the standard error, will lead one to reconsider the synthetic estimator.

In their discussion of models, the authors stress the importance of providing estimators of the reliability for small area estimators. They allude to the fact that the principal estimators of mean square error for model based procedures are estimators of an average mean square error. While this is true, it seems worth mentioning that components-of-variance procedures do not assume the mean square errors to be the same in each domain. Also, for the typical survey situation, the estimators of mean square error need not be constant over domains. For example, one of the terms in the mean square error estimator of the components of variance procedure is the estimator of the variance of the direct estimator. The estimated variance of the direct estimator will be a function of the domain sample size and can also be a function of the direct estimated variance of the direct estimator for that domain. See Battese, Harter, and Fuller (1988), Harville (1976), Prasad and Rao (1990), and Ghosh and Rao (1993).

In their discussion of designs, the authors explain that the variance function is often relatively flat in the vicinity of the optimum allocation to strata. A slight reallocation of sample among strata can markedly increase the efficiency of domain estimators for a relatively small decrease in the efficiency of the overall estimates. The same is true with respect to the combination of direct and synthetic estimators. Thus, if one has a relatively good idea of the variance component associated with small areas, either from a previous study on the same population or from a study on a similar population, and if one is under pressure to produce estimates in a brief time span, then it is reasonable to assign fixed weights to form the linear combination. The loss in efficiency is apt to be modest and the programming required for estimation construction considerably reduced. One estimator in this class, and the one adopted by many practitioners, is the synthetic estimator.

The authors briefly raise the question of internal consistency associated with the construction of small area estimates. As they say, if one uses a data dependent procedure, such as variance components, for each dependent variable, then one produces estimates that are not internally consistent. One option is to use multivariate procedures. See, for example, Fuller and Harter (1987) and Fay (1987). Another procedure suggested by Fuller (1990) is to construct components of variance estimators for a limited subset of variables and then use these estimates as control variables in a regression procedure. The regression procedure produces weights for the individual observations. Once the weights are constructed, any number of output tables can be constructed and all estimates are internally consistent.

It is my observation that the gains made in most practical domain estimation problems come primarily from the wise use of auxiliary information. Thus, effort directed towards obtaining quality auxiliary information is effort well spent. If we are able to find a variable x that is highly correlated with the variable y, then there is less variability remaining to be allocated between area to area variance and sampling variance.

ACKNOWLEDGEMENTS

I thank Jay Breidt for comments.

REFERENCES

BATTESE, G.E., HARTER, R.M., and FULLER, W.A. (1988). An error components model for prediction of county crop areas using survey and satellite data. *Journal of the American Statistical Association*, 83, 28-36.

FAY, R.E. (1987). Application of multivariate regression to small domain estimation. In *Small Area Statistics*. (Eds. R. Platek, J.N.K. Rao, C.-E. Särndal and M.P. Singh). New York: Wiley.

FULLER, W.A. (1990). Analysis of repeated surveys. *Survey Methodology*, 16, 167-180.

FULLER, W.A., and HARTER, R.M. (1987). The multivariate components of variance model for small area estimation. In *Small Area Statistics*. (Eds. R. Platek, J.N.K. Rao, C.-E. Särndal and M.P. Singh). New York: Wiley.

HARVILLE, D.A. (1976). Extension of the Gauss-Markov Theorem to include the estimation of random effects. *The Annals of Statistics*, 4, 384-395.

GHOSH, M., and RAO, J.N.K. (1993). Small area estimation: An appraisal. Unpublished manuscript. Carleton University, Ottawa, Ontario, Canada.

PRASAD, N.G.N., and RAO, J.N.K. (1990). The estimation of mean squared errors of small-area estimators. *Journal of the American Statistical Association*, 85, 163-171.

COMMENT

GRAHAM KALTON[1]

As Singh, Gambino and Mantel (SGM) indicate, there is a growing demand for surveys to provide domain estimates for domains of various sizes and types. This demand is being experienced in many countries throughout the world. In part it may simply reflect a natural growth in the sophistication of survey analysts, who once were content with national estimates and estimates for a few major domains, but who now want to compare and contrast estimates for many different types of domain. In part it results from the needs of policy makers, who require domain information in order to examine how current policies affect different domains, to predict what effects changes in policies might have, and for policy implementation. Information on administrative area domains (*e.g.*, provinces or states, counties, and school districts) is of particular interest for policy purposes (*e.g.*, for identifying low income areas for government support).

In some circumstances the need for domain estimates of adequate precision can be satisfied within the design-based inference framework that is standardly used in the analysis of survey data. This holds for large domains for which the sample sizes are adequate to give the precision required. It can also hold for small domains provided that they are identified in advance, and the sample design is constructed in a way that provides adequate sample sizes. Thus, for example, in the United States, the National Health and Nutrition Examination Survey and the Continuing Survey of Food Intakes by Individuals use differential sampling fractions by age, sex and race/ethnicity and by age/sex and low income status, respectively, in order to provide adequate samples for the domains created by the cross-classifications of these variables. The U.S. Current Population Survey employs differential sampling fractions across the states in order to be able to produce state-level employment estimates. The limitation of this approach is evident when there is a large number of small domains, in which case the sum of the required sample sizes for each domain produces an extremely large overall sample size. This situation occurs often with small administrative districts, such as counties, school districts, and local employment exchanges. In such cases, it may be necessary to discard the standard design-based inference approach in favor of a model-dependent approach that employs a statistical model in the estimation process to borrow strength from data other than that collected in the survey for the given small area. The model-dependent approach may also be required for unplanned small domains, where the need for oversampling had not been foreseen at the design stage.

In response to the demand for small area estimates, a sizeable literature has developed on model-dependent small area estimation methods. Little has, however, been written on the broader issues of small area estimation discussed in the SGM paper, issues that need more attention. Like the authors, I believe that a cautious approach should be adopted to the use of model-dependent small area estimators. I therefore welcome their discussion of methods to make small area estimates within the design-based framework.

From my perspective, the first approach to making small area estimates is to see whether estimates can be produced with adequate precision within the design-based framework. If the domains have been identified in advance, consideration should be given to designing the sample to meet the needs for small area estimates. This may involve ensuring that the small areas do not overlap strata, and ensuring a sufficient sample size for each small area. Another approach suggested by SGM is to minimize the amount of clustering. The smaller the amount of clustering, the less the sample size in each small area is subject to the vagaries of chance. In this regard I see the benefits of less clustering as mainly directed at providing the ability to produce estimates for small areas that were not identified at the design stage. When small areas for which estimates are planned are made into separate strata, the sample size in each small area should be under adequate control even with a clustered sample (provided that the measures of size used in the PPES sampling are reasonable). However, even with planned estimates, there will often be an issue of how to compute variance estimates for a small area from a clustered design, since the number of PSUs sampled in each small area is likely to be small. A variance estimate based on the PSUs within the small area will then be imprecise, with few degrees of freedom, and a generalized variance function approach may be preferred (*e.g.*, assuming that the national design effect applies for each small area). In other words, although the estimate itself may be a design-based estimate, the estimate of its variance may be an indirect one, borrowing strength from other areas. This consideration favors as unclustered a design as possible even for planned small area estimates. The need to model variances is, however, of lesser concern than the need to model the estimates themselves.

An integral part of the design-based framework is a recognition that auxiliary information available for the population may be used at the design stage, at the analysis stage, or at both stages. When information on auxiliary

[1] Graham Kalton, Westat, 1650 Research Blvd., Rockville, Maryland, U.S.A. 20850.

variables that are closely related to the survey variable is available, substantial gains in precision can accrue. The use of auxiliary information at the analysis stage, through such techniques as post-stratification and ratio, regression and difference estimation, has a special appeal for small area estimation. It should be emphasized that ratio and regression estimators may be motivated by assumptions about the model relating the survey variable (Y) and the auxiliary variables (X), but that the resultant estimators are design-consistent irrespective of the appropriateness of the model. The use of an appropriate model produces the greatest gains in precision, but the estimates are approximately unbiased whatever model is chosen. This may be seen in a simple case where variables $X_1, X_2, \ldots, X_p$ are known for every element in the population, and the linear combination $\tilde{Y}_i = B_0 + B_1X_{1i} + \ldots + B_pX_{pi}$ is used to estimate Y_i, the value of the Y-variable for population element i. Assume, for simplicity that the B's are determined from external data, not dependent on the sample. With $Y_i = \tilde{Y}_i + e_i$, the domain total is $Y_a = \sum_{i\epsilon a}\tilde{Y}_i + \sum_{i\epsilon a} e_i = \tilde{Y}_a + E_a$. Since $\tilde{Y}_a$ is known, the estimation problem is one of estimating E_a. From a sample of elements in domain a, E_a may be estimated by $\hat{E}_a = \sum_{j\epsilon s_a} e_j/\pi_j$, where π_j is the selection probability for element j in the sample. The estimator $\hat{E}_a$ is unbiased, independent of the validity of the model employed. The estimation procedure in fact translates the estimation problem from one of estimating Y_a directly to one of estimating E_a and adding on a known constant $\tilde{Y}_a$. To be effective, the procedure requires the domain variance of the e_i to be smaller than that of the Y_i. There is no requirement that $E_a = 0$. The general logic remains the same in the more usual situation where the B's are estimated from the sample. In this case, the estimate of Y_a is design-consistent, irrespective of the model adopted (Särndal 1984). Moreover, the B's may be estimated from the sample data only for the domain of interest, producing what SGM term a direct estimator, or from the total sample, producing a modified direct estimator. A key consideration in the choice between the direct and modified direct estimators in this case is whether the overall B's also apply for the domain. If not, interaction terms between the X's and the domain indicators are called for in the total sample model. With a full set of these interaction terms, the modified direct estimator in effect then reduces to the direct estimator.

The need for a model-dependent approach occurs when the design-based estimate lacks sufficient precision even after the auxiliary data available have been used in as effective a manner as possible. Indeed, in some cases the computation of a direct estimate may be impossible because there are no sample cases in the small area. In such situations, it becomes necessary to use a statistical model to borrow strength from other data, often data from other areas. Such models are built upon assumptions (*e.g.*, $E_a = 0$ in the above example), and the quality of the resultant small area estimates depends on the suitability of the assumptions made. The assumptions are inevitably incorrect to some degree, leading to biases in the small area estimates. Since indirect estimates are biased, the design-based mean square error (MSE) is widely used as the measure of their quality, where MSE $= V' + B^2$ and V' is the variance and B is the bias of the estimate.

The common way to compare the quality of a direct and an indirect estimate is to compare the variance, V, of the former with the MSE of the latter. However, reading the paper caused me to question whether the MSE is the appropriate measure of quality of an indirect estimator. In a practical setting the variance V of the direct estimate can be estimated whereas the design-based MSE of the indirect estimate cannot. In view of this situation, if $V =$ MSE, then the direct estimator would be clearly preferred. In fact, the direct estimator may tend to be preferred if the direct estimator has adequate precision, irrespective of the likely relative magnitudes of V and MSE. In other cases, if B is the expected bias, then the direct estimator may be preferred to the indirect estimator unless $V > V' + kB^2$, where k is a multiplier greater than 1 that allows for the fact that the unknown bias may be larger than expected.

The same argument can be applied to combined (or composite) estimators that employ a weighted average of a direct and an indirect estimator. Often the principle for choosing the weights is taken to be to minimize the mean square error of the combined estimator, leading to weights for the direct and indirect estimators that are inversely proportional to V and MSE, respectively. However, following the above argument, an alternative procedure would be to minimize the weight of the indirect estimator, subject to the condition that the combined estimator is sufficiently accurate. Alternatively, the weights could be determined on some maximum likely value of the MSE, rather than the expected MSE, to reduce the risk of serious bias in the combined estimator.

I do not follow the rationale for the sample size dependent estimators described by SGM in equation (6.11) and (6.12) in general, but under certain assumptions they may be seen to fit in to the logic given above. With an equal probability sample design and $\delta = 1$, these estimators reduce to the direct estimator when the achieved sample size is greater than, or equal to, the expected sample size. If one assumes that the expected sample size gives adequate precision for the small area, this outcome accords with the above reasoning. If the achieved sample size is smaller than expected, the sample size dependent estimator takes a weighted average of a direct and an indirect estimator. If one assumes that the expected sample size is the minimum sample size to give the required precision, this outcome also accords with the above reasoning. If this indeed is the basis of the sample size dependent estimators, then it would seem useful to generalize them to situations where

the expected sample size is not the sample size that just gives the level of precision required.

As has been noted, auxiliary information plays an important role in the production of accurate small area estimates. Such information may be used for improving the precision of design-based estimates or it may be used in the models employed with the model-dependent approach. Ideally auxiliary information that is highly related to the survey variables involved in the estimates is required. The regular compilation of up-to-date auxiliary data for small areas from administrative and other sources can provide a valuable resource for a small area statistics program.

Although the paper mentions the more general problem of small domains, it focuses predominantly on small areas. This is in line with the general literature and the application of indirect estimation procedures. In part, this may be because the number of socio-economic and other small domains of interest (*e.g.*, age/sex domains) is usually relatively small, compared with the numbers of small areas, so that socio-economic domains can be handled by designing the sample to provide direct estimates of adequate precision for each of them. In part, it may be because the definitions of socio-economic and demographic domains are often chosen in the light of the feasibility of producing design-based estimates of adequate precision for them (*e.g.*, using wider age groupings for some domains); in the case of areal domains, however, the areas are predefined, and no collapsing of areas is acceptable. In part, it may be because there is a lack of auxiliary data to use in the statistical models for such domains. In part, it may also be because the analysis of socio-economic domains is often conducted to make comparisons between the domains. Such comparisons are distorted when the estimate for one domain borrows strength from other domains (see, for example, Schaible 1992). This issue brings out the general point that indirect estimates should not be uncritically used for all purposes.

In conclusion, I should like to express my support for the general approach of this paper. Where possible, samples should be designed to produce direct small area estimates of adequate precision, and sample designs should be fashioned with this in mind. Auxiliary data should be used, where possible, to improve the precision of direct small area estimates. When indirect estimates are called for, a cautious approach should be used. Models should be developed carefully, estimators that are robust to failures in the model assumptions should be sought, and evaluation studies should be conducted to assess the adequacy of the indirect estimates. Lacking good measures of quality for individual indirect estimates, such estimates need to be clearly distinguished from design-based estimators. Since indirect estimates are not universally valid for all purposes, users need to carefully assess whether the given form of indirect estimate will satisfy their particular needs.

REFERENCES

SÄRNDAL, C.-E. (1984). Design-consistent versus model-dependent estimators for small domains. *Journal of the American Statistical Association*, 79, 624-631.

SCHAIBLE, W.L. (1992). Use of small area estimators in U.S. Federal programs. *Proceedings of the International Scientific Conference on Small Area Statistics and Survey Design*, (Vol. 1), 95-114, Central Statistical Office of Poland, Warsaw.

RESPONSE FROM THE AUTHORS

We would like to thank Wayne Fuller and Graham Kalton for their stimulating comments, which we find to be quite complementary to the position developed in our paper. In many cases their comments make certain points clearer and strengthen the arguments presented. Encouraged with this kind of endorsement we would like to carry some of the points about survey design further, while responding to the main points made by the discussants.

There is no doubt that survey designers try to optimize the design under operational constraints to meet the stated objectives of a survey. There are usually several objectives to be met by major surveys and it is quite likely that designers have limited influence in the setting of priorities among the various competing objectives. Nevertheless, it is at this stage of priority setting that the case for small area needs should be made strongly, particularly for major continuing surveys.

During the sixties and seventies emphasis in most countries was placed on sub-national (state/provincial) estimates and certain compromises were made to the earlier designs that optimized national estimates. For example, different sampling fractions were used to ensure a minimum sample size for smaller states/provinces. With the demands for data at the sub-state/province level, such as, county, district and municipality, more compromises to the national optimum allocation become necessary, requiring differing sampling fractions among the administrative areas within states/provinces. For example, if the aim is to produce sub-provincial estimates of comparable quality, then provinces will likely receive sample roughly proportional to the number of subprovincial regions they contain. Such an allocation may not be the same as one using the relative population sizes of the provinces. As we discussed in section 5.4, the allocation approach should put more emphasis on a bottom-up strategy. Losses at higher levels and gains at lower levels would differ from survey to survey but it is likely that in many cases a minor loss in CV at the national level will lead to appreciable gains at small area levels.

Kalton stresses the importance of reduced clustering for variance estimation; it is advantageous to increase the degrees of freedom by having a large number of smaller clusters rather than a small number of larger clusters. We would like to emphasize that clustering has another drawback for estimation, and especially small area estimation, namely, a highly clustered design will lead to high design effects, even for planned small domains. The usual reason for resorting to clustered designs is to reduce survey costs. In light of the changes that continue to occur in the data collection process, such as decreased reliance on at-home interviews and increased use of computer assisted interviewing, a periodic review of the cost-variance models that underlie clustering decisions is necessary.

One other issue not addressed in our paper is the impact of sample rotation in continuous surveys. For a given time point, there may be insufficient sample in some small domains to produce reliable estimates. But, as units rotate out of the sample and are replaced, the accumulated or effective sample in the domains increases and may allow the computation of reliable, albeit time-biased, domain estimates. By judicious choice of rotation schemes, survey designers can maximize the cumulative sample size over some time period. For example, for quarterly estimates in a monthly survey, the optimal rotation pattern is $[1(2)]^k$, *i.e.*, repeat the sequence "one month in sample, two months out" k times. This thinking is in the same spirit as Leslie Kish's ideas on cumulation of samples over time.

Kalton clarifies and elaborates the cautious approach to the use of indirect estimators by suggesting a weighted mean squared error, which attaches a weight greater than 1 to the bias term, to allow for the fact that the bias of the indirect estimator may be larger than expected. There are two distinct reasons why the bias may be larger than what is expected from the model for small area effects: random variation within the model, and model breakdown. It is worth recalling here the suggestion of Fay and Herriot (1979) to constrain a combined estimate to be within one standard error of a design estimate; this approach makes allowance for the possibility of large bias in the model estimator for whatever reason. Kalton also reiterates our position that if a direct estimator is of acceptable quality, then in practice, one may decide to use this direct estimator even though its estimated mean squared error exceeds that of model-based competitors. Because there is always the possibility of model failure lurking in the background, this "better safe than sorry" approach is desirable, at least until some experience with particular indirect estimators in specific situations has been gained. This does not contradict the view that there arise situations in which it is necessary to throw caution to the wind.

In his remarks on the sample size dependent estimator, Kalton's comments imply that there is a risk in the strategy which gives the synthetic component zero weight if the observed sample size in the small domain exceeds the expected sample size there since the latter may be too small to yield adequate direct estimates. One option is to use a value $n_{\min}$ which is the size that produces direct estimates that are just barely acceptable. Note, however, that $n_{\min}$ as defined here is characteristic-dependent.

In his comments, Fuller briefly describes an approach to small area estimation that takes advantage of a variance components model and yet has fixed weights for internal consistency among estimators for different characteristics. Besides internal consistency of small area estimates for different characteristics, a second type of consistency that

is sometimes required is that estimates of totals for the set of small areas within a larger area should add up to the published direct estimate for the larger area. One way to achieve this is to benchmark the small area estimates to the direct estimate for the larger area using, for example, a simple ratio adjustment; however, if the ratio adjustment factors depend on the characteristic then this would destroy the first type of consistency. Both types of consistency could be achieved simultaneously if the direct estimators for the larger area are generalized regression estimators, $\hat{Y}_e + (X - \hat{X}_e)\hat{\beta}$, and the modified direct (Section 6.1 in the paper) estimators $\hat{Y}_{sreg,a} = \hat{Y}_{e,a} + (X_a - \hat{X}_{e,a})\hat{\beta}$ are used for small areas.

As Fuller notes, the average squared bias of an estimator for any subset of small areas can be estimated. Here we would like to stress again that the average bias over a set of small areas is not directly relevant for any particular small area. It is for this reason that we prefer to use, whenever possible, estimators that are approximately design unbiased. When use of a model estimator is unavoidable, serious attempts should be made to find appropriate covariates for which reliable auxiliary information is available in order to minimize the residual bias of the model estimator.

Perhaps due to the obvious timeliness problems associated with census data, neither of the discussants commented on censuses as a source of data for smaller domains. In this context it is worth mentioning that some form of ongoing major post-censal survey replacing or supplementing the decennial census long-form may be considered. Such a strategy, called rolling samples, is described by Kish (1990); a similar approach, called continuous measurement, is described by Alexander (1994). This approach provides a number of options which are worth investigating as potentially cost effective means of producing timely statistics for smaller domains.

Lastly, we would like to stress that the emphasis we put on keeping domain estimation in mind at the design stage, particularly for medium size domains, in no way undermines the important role of models in estimating for very small domains.

We hope that the general direction of the strategy proposed in the paper, supplemented by the fine points brought out by the discussants, particularly the support and cautions summarized by Kalton in his concluding paragraph, will be helpful to survey designers and researchers in finding solutions appropriate to the particular problems they are dealing with.

ADDITIONAL REFERENCES

ALEXANDER, C.H. (1994). A prototype continuous measurement system for the U.S. Census of Population and Housing. Document for presentation at the annual meeting of the Population Association of America, Miami, Florida, May 5, 1994.

KISH, L. (1990). Rolling samples and censuses. *Survey Methodology*, 16, 63-71.