



Catalogue no. 12-001-XIE

Survey Methodology

December 2002



How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to: Business Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 (telephone: 1 800 263-1136).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our website.

National inquiries line	1 800 263-1136
National telecommunications device for the hearing impaired	1 800 363-7629
Depository Services Program inquiries	1 800 700-1033
Fax line for Depository Services Program	1 800 889-9734
E-mail inquiries	infostats@statcan.ca
Website	www.statcan.ca

Information to access the product

This product, catalogue no. 12-001-XIE, is available for free. To obtain a single issue, visit our website at www.statcan.ca and select Our Products and Services.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service that its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136. The service standards are also published on www.statcan.ca under About Statistics Canada > Providing services to Canadians.



Statistics Canada
Business Survey Methods Division

Survey Methodology

December 2002

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2006

All rights reserved. The content of this publication may be reproduced, in whole or in part, and by any means, without further permission from Statistics Canada, subject to the following conditions: that it is done solely for the purposes of private study, research, criticism, review, newspaper summary, and/or for non-commercial purposes; and that Statistics Canada be fully acknowledged as follows: Source (or "Adapted from", if appropriate): Statistics Canada, name of product, catalogue, volume and issue numbers, reference period and page(s). Otherwise, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopy, for any purposes, without the prior written permission of Licensing Services, Marketing Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

January 2006

Catalogue no. 12-001-XIE
ISSN 1492-0921

Frequency: semi-annual

Ottawa

Cette publication est disponible en français sur demande (n° 12-001-XIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Strategies and Approaches for Small Area Statistics

Gordon J. Brackstone¹

Abstract

National statistical offices are often called upon to produce statistics for small geographic areas, in addition to their primary responsibility for measuring the condition of the country as a whole and its major subdivisions. This task presents challenges that are different from those faced in statistical programs aiming primarily at national or provincial statistics. This paper examines these challenges and identifies strategies and approaches for the development of programs of small area statistics. The important foundation of a census of population, as well as the primary role of a consistent geographic infrastructure, are emphasized. Potential sources and methods for the production of small area data in the social, economic and environmental fields are examined. Some organizational and dissemination issues are also discussed.

Key Words: Small area statistics; Census; Geography.

1. Introduction

The mandate of most national statistical offices (NSO) focuses on the monitoring of social, economic, and environmental conditions at the national level, and for the major administrative units (provinces, states, major metropolitan areas) within the country. However, the demand for data at lower geographic levels is always present, especially from local governments and from businesses needing to make investment, marketing, and location decisions that depend on knowledge of local areas. We will use the term “small area statistics” to mean statistics for areas below the level of state, province, or major metropolitan areas – a broad spectrum of areas from large towns, through urban neighbourhoods, to rural villages. In some circles the term “small areas” is used more broadly to refer to any small subgroup or domain of the population, but here we are talking strictly about small geographic areas.

The extent of an NSO’s responsibility for small area statistics depends on the division of governmental responsibilities within a country. For example, in some countries local governments are the creation of provinces and the responsibility for supporting their statistical needs may rest with provincial governments. But in many countries, whatever the formal division of powers, it is, *de facto*, the NSO that is expected to respond to the need for small area statistics, either within its own resources or in cooperation with other levels of government. At the very least, it is the NSO that must set the standards and framework for small area data if these are not to become a mishmash of uneven and overlapping statistics incomparable across the country.

With limited budgets an NSO is faced with the difficult trade-off between investment in national statistics and provision of small area detail. How should it choose between covering more subject areas, or existing subject areas in more detail, at the national and provincial levels, and, on the other hand, providing more small area detail for

subject areas it is already covering nationally? There is no formula for resolving this problem. The balance struck in any country will be largely a function of national needs, relative powers, and historical tradition, with perhaps some statistical considerations on the margin. Nevertheless, there is a series of measures and approaches that a NSO can consider to maximize the degree to which it can satisfy demands for small area statistics within a limited budget.

Four potential sources of small area statistical data either individually or in combination, account for most production of small area data by statistical agencies. Censuses or complete enumerations of populations are the traditional source. Administrative records, including national registers, that cover all, or almost all, of a defined population are in many respects equivalent to a census. National sample surveys are rarely large enough to produce small area data directly but they do represent a valuable current source of information that can be used, under certain assumptions and in combination with other sources, to produce small area data. And finally, local studies focused on particular small areas will produce small area data, but not for complete sets of small areas. Sources such as satellite imaging or aerial photography can be thought of as censuses or local studies depending on their coverage.

In this paper we first review the important role of the Census of Population, with or without a population register, in the provision of small area socio-economic data (section 2), and then emphasise the fundamental role of an up-to-date geographic infrastructure to support any production of small area statistics, including especially the census of population (section 3). We then examine approaches to providing small area data on individuals and families between censuses (section 4), on business activities (section 5), and on environmental issues (section 6). We conclude with some general observations about the dissemination of small area statistics and the management of small area statistics within an NSO.

1. Gordon J. Brackstone, Informatics and Methodology Field, Statistics Canada, Ottawa, Ontario, K1A 0T6. E-mail: bracgor@statcan.ca.

2. Census of population

The census of population, in most countries, plays the central role in the provision of small area data about people, families and households. Based on a complete enumeration of the population (at least for basic characteristics), its estimates are free of the sampling error that limits the ability of sample surveys to produce small area estimates. Provided the individual households are geographically coded to a fine level (*e.g.*, a block or block face), direct tabulations of households can produce statistical aggregates for any geographic area that can be defined, or approximated, in terms of the lowest level of geographic coding.

However, censuses have their drawbacks. They are costly, and therefore they are infrequent. Data from the last census may provide a poor representation of a small area that is undergoing rapid development. In many countries, sampling is utilized in the census for many of the questions. While this introduces sampling error into estimates from the census, these samples are still huge compared with those in a typical sample survey. Furthermore, the samples are typically spread through every enumeration area of the country, so the ability to produce small area estimates is maintained, even though the small areas will need to be somewhat larger than in a true census.

Potentially more serious, with respect to accuracy, are nonsampling errors such as coverage error and response bias. Most censuses miss some people, or count some people twice, and it has been repeatedly shown that those miscounted are generally not typical of the population as a whole. Census estimates may therefore be biased against certain sub-groups of the population. If these subgroups (*e.g.*, certain immigrant groups) tend to be geographically clustered, this can have a serious impact on estimates for some small areas. Response bias arises if a census question is systematically misunderstood by many respondents. Both small area and large area estimates would be affected by such errors.

Countries that maintain a population register have the potential to produce census-like data for small areas more frequently than the traditional 5–10 year cycles of a census. Up-to-date residence registration is clearly a requirement for accurate small area data from such registers. The breadth of data available from a register system may be less than that available through a conventional census, since the former is limited to the characteristics maintained in linkable administrative registers. In some countries the population register may be used as the basis for a census that collects the necessary additional characteristics not available within existing registers Redfern (1989) provides a useful description of practices within Europe in this regard.

Since the Census has the potential to produce estimates for very small areas, rules to protect against direct or residual disclosure of individual data have to be in place. These can include imposing a minimum population on areas for which data will be released, random perturbation of data,

suppression of data, or other techniques (Jabine (1993), Zayatz, Steel and Rowland (2000)). NSOs have also to be concerned about privacy issues arising with the publication of small area census data that, while not disclosing any individual responses, do reveal dominant characteristics of an area (*e.g.*, that 90% of the families received unemployment benefits). Such findings cannot be withheld, but they can be selected and presented with sensitivity.

Though a census, with or without a population register, is a source of direct small area data as of census day, the value of such data declines as time passes. However, the role of census data in the provision of small area statistics goes well beyond the direct use of the results from each periodic enumeration. Inter-censally, census data may be used as a benchmark, a sampling frame, or as auxiliary information to be used with other sources of data that are available between censuses. These usages are pursued in section 4. An innovative alternative to the traditional census is described in section 4.4.

3. Geographic infrastructure

To enable a national census to produce accurate data for small areas, a geographic infrastructure of boundaries and mapping capacity covering the whole country is a prerequisite. Such an infrastructure requires that each dwelling be associated with a precise geographic location on the ground, where the degree of precision determines the fineness with which small areas can be defined. Though modern global positioning technology makes it possible to pinpoint each dwelling to a specific pair of coordinates, it is usually sufficient for statistical purposes to associate each dwelling in an urban area with a block face (*i.e.*, one side of a street between two intersections), or a building in the case of high-rise buildings. In rural areas, the chosen degree of precision will depend on local administrative and natural boundaries, though maximum flexibility is preserved by using precise coordinates for each dwelling.

While necessary for a census, a geographic infrastructure is equally required for the provision of small area statistics from other sources. Essentially each data point, from whatever source, has to be associated with a geographic location at a level detailed enough to allow aggregation into any small areas of statistical interest. For example, if the data source is an administrative register, or a business register, the address in each record must be convertible into a pair of geographic coordinates, or at least into a small area within which the address falls. Since administrative registers often use mailing addresses, a file that converts postal codes into geographic locations is a valuable tool in the development of small area data.

The availability of an accurate up-to-date geographic infrastructure, whether maintained by the NSO or obtained from outside, is essential if a program of small area statistics is to have flexibility in the choice of areas for which statistics are produced.

4. Small Area Statistics on Persons and Households – Between Censuses

We turn now to the issue of producing small area data for persons or households inter-censally. Clearly the existence of a current population register makes a fundamental difference to what is possible, and how it can be done. We will confine ourselves to the case where no regularly updated population register exists.

In such circumstances, there are three main classes of approach. The first is to utilize census-like files that come from administrative systems and purport to cover the whole of a well-defined population. The second is to exploit sample survey data and, through additional model assumptions, produce estimates for smaller (though still not very small) areas than is possible through direct survey estimation. The third category is the combination of one or both of these first two approaches with the use of data from the most recent census. In the following paragraphs we review some of the characteristics of these approaches.

4.1 Administrative Files

An example of an administrative file with small area statistical potential is the annual file of individual income tax returns. Other examples, with narrower population definitions, might be drivers' licences, employment insurance recipients, or health insurance records. In the case of tax data, if each record contains a residential address that can be associated with a geographic point or small area, then data can be tabulated directly for small areas, with due regard for confidentiality (as with census data). The characteristics available would generally be restricted to demographic and income variables, and the coverage would be limited to taxfilers. Nevertheless, such a file represents a rich source of annual data for quite small areas. Population coverage can be improved through the imputation of dependents "claimed" on the tax record. In Canada, the coverage of such imputed files is approaching that of the census as coverage increases among low income earners who need to file tax returns to obtain social assistance benefits.

With administrative data in general, the statistician has to take what is available (though some influence on content may be possible in the longer term), reconcile any differences in concepts, definition or coverage between the administrative file and the statistical objectives, and assess any issues of reporting or coding accuracy in the records. Subject to these precautions, administrative data can provide a geographically rich potential source of small area data (Brackstone 1987).

4.2 Sample Survey Data

The problem with sample survey data as a source of small area statistics is sample size. There are frequently insufficient sample cases in the small area to allow a reliable direct estimate to be produced, and sometimes none at all. In

large national sample surveys it may be possible to devise sampling strategies that ensure an acceptable level of precision for planned small areas, such as sub-provincial regions, without significantly degrading the reliability of estimates at higher levels (Singh, Gambino and Mantel 1994). But for smaller areas, or for areas of similar size not taken into account during design, reliable estimation will not be possible. Larger samples help, and may allow direct estimation for some of the larger small areas, but budgets usually constrain this approach as a general solution. If no other data sources are available, statisticians can only resort to model-based methods which involve making assumptions about how data for a small area relate to other data. These methods are often described as "borrowing strength" since they borrow information from elsewhere in the sample survey to augment the number of units that contribute to the estimate for a given small area. The borrowing can be from other time periods, from sample units outside the given small area, or from other variables measured on the same sample unit. Some examples follow. Most of these examples will allow some expansion of the range of small area estimates that can be produced from sample surveys with relatively large samples. They cannot magically convert small sample surveys into rich sources of small area data.

1. In a monthly survey, it may be possible to combine data for a small area over a period of consecutive months to produce direct estimates of a multi-month moving average for the area. For example, quarterly estimates may be possible where monthly ones were not.
2. One may be ready to assume that means or proportions estimated for a larger area apply equally to a smaller component area within it. If the size of the small area is known, an estimate can be obtained by multiplying by the assumed mean or proportion. This assumption may be more realistically made within subgroups of the population (*e.g.*, age groups), rather than for the population as a whole. In this case, if the size of each sub-group is known for the small area, a synthetic estimator can be built up by multiplying the sizes by the assumed means and aggregating.
3. If additional related variables are available from the survey, more elaborate models may be set up relating the variable being estimated to these auxiliary variables. The parameters of the model may be estimated at a higher geographic level where there is sufficient sample to estimate them reliably. The model is then applied with the estimated parameters to the data for the given small area.

All of these approaches suffer from the lack of reliable baseline data for each small area. If such data are available,

for example from a recent census or from administrative records, then the data may be used in combination to produce more reliable estimates than from either source alone.

4.3 Combined Sources

Methods that combine census or administrative information from the recent past with current sample survey data are borrowing strength from outside the survey. They still require model assumptions. However, these can often be weaker (since they involve assumptions about change from the benchmark, rather than about absolute levels of each small area) and so more acceptable, or more plausible, than in the case of sample survey data alone.

A wide variety of estimation methods (which we won't attempt to describe here) have been developed to handle this situation. Some of these methods can be thought of as estimating change since the most recent benchmark, others as distributing reliable current sample survey estimates among component small areas based on benchmark data, and yet others as recalibrating old benchmark figures to new current estimates. In essence, they all involve some kind of balancing of three kinds of estimates: (a) high variance but unbiased direct current survey estimates for the small area in question; (b) low variance current survey estimates for some surrounding or comparable larger area; and (c) census-type estimates for the same small area from recent administrative data, or a past census, which may contain unknown bias due to the source and the time lag. Any available auxiliary data can be incorporated to improve the accuracy of each component estimate. The way in which these three types of estimates are combined is determined by the choice of model and model parameters.

In summary, the methods of this and the previous section essentially reduce variance by making use of more data, but at the expense of introducing potential bias due to model assumptions that will never be exactly correct. It is very important to analyse the performance of these methods before their use, for example by carrying out the estimation process in a census year when direct estimates are available for comparison, and periodically thereafter. Model checking is becoming an area of increased research activity (Bayarri and Berger 2000). For more detailed descriptions of available methods in this class see, for example, Purcell and Kish (1979); Fay and Herriott (1979); Ghosh and Rao (1994); Singh *et al.* (1994); Schaible (1996); Rao (1999) and Gambino and Dick (2000).

4.4 Rolling Censuses

An innovative alternative to the census is being investigated in at least two countries. The method of producing small area data based on a large rolling sample has long been advocated by Leslie Kish as an alternative to the traditional census (Kish 1990, 1998). The sample survey "rolls" in the sense that over a long period (*e.g.*, a decade) each of the smallest areas for which estimates are required

would be included once in the sample so as to provide a direct estimate for that area once each period. Successively larger areas (aggregates of the smallest areas) would be represented more often in the sample, allowing either more reliable or more frequent estimates for those areas. For even larger areas, including provinces and the whole country, the accumulated sample would be sufficient to provide reliable annual, or more frequent, estimates at certain levels of detail. The approach may be considered with or without a periodic census to collect basic demographic data against which to calibrate the inter-censal survey estimates.

The rolling census avoids the need for the assumption of models, but presumes that unbiased estimates of multi-year averages, or asynchronous estimates for different areas of the country, are satisfactory alternatives to the simultaneous point-in-time estimates of the traditional census. Relative cost is also a key factor, especially in the situation where a basic census is also carried out. On the other hand, by producing reliable annual estimates for many of the larger areas, and with much of the content detail of a census, this approach could effectively address the issue that census estimates can be up to 12 years old before the next ones appear. It also responds to mounting concerns over increasing difficulties and costs associated with the conduct of a traditional census.

This approach is being tested in the United States under the name of the American Community Survey (Alexander 1999, 2002) and in France where it is referred to as the "recensement continu" (Isnard 1999; Durr and Dumais 2002).

5. Business Statistics

The problems of producing small area data for businesses are different in many important respects from those encountered for data on persons or households.

Whereas the association of each individual with a "usual place of residence" is, for the vast majority of the population, a fairly clear and unambiguous concept (though perhaps becoming less clear with the growth of second residences, the incidence of prolonged absences away from the snow, and more flexible living arrangements), for businesses the question of where, geographically, to attribute various characteristics of a business is less clear in many situations. For single establishment businesses where all the activity takes place in a single location there is no conceptual problem, though there may still be a practical problem if the source of information is an administrative file that provides, say, an accountant's address rather than the place of business. For some variables, such as employment, there may be no major conceptual problem even for larger businesses (except perhaps for those working in the transportation industry, or certain service industries). However, for variables such as revenues and profits there can be real questions about how these should

be allocated geographically in multi-establishment businesses. The larger the geographic area the smaller the problem – location within a province doesn't matter if one is only interested in provincial totals. But, in general, geographic attribution rules have to be determined before small area estimates for business activity can be considered, and for some aspects of business activity small area estimates may not make conceptual sense.

While for household surveys the main obstacle to the production of small area estimates is sample size, for business surveys considerations of confidentiality usually constitute the major barrier. The smaller the area, the greater the chance that a particular industry will be dominated by one or a few major companies, thus precluding the provision of estimates for that area due to disclosure risk. Methods for checking statistical output on businesses to recognize potential disclosure risks are fairly well developed (Federal Committee on Statistical Methodology 1994) but require constant attention on the part of the NSO. The confidentiality problem is less of an issue in those industries characterized by small units – which may be the same industries in which the conceptual problems of the previous paragraph are not so severe. In those industries, considerations of sample size may indeed be the limiting factor, in which case the families of methods described in the previous section are available.

A third area of contrast with data on individuals, at least for countries that do not maintain a population register, is the existence of a relatively up-to-date list frame of businesses. This not only provides a base for sampling and a source of some auxiliary data for estimation, but also constitutes a potential source of direct estimates of business demography, at least annually. In many countries the currency of the business register is maintained by receiving transactions from the business tax system, which itself provides an annual census-like source of administrative data on business activity. However, use of tax data still requires careful consideration of the conceptual, geographical and confidentiality issues raised above.

6. Environment Statistics

Environment statistics provide yet different challenges for the production of small area statistics. While some environmental issues are national or even global in scope, many are by their nature local. Many sources of pollution are typically localized with their impacts being felt most severely in the neighbourhood of a plant or accident. The socio-economic impacts of broader environmental problems (*e.g.*, loss of fish stocks) are frequently felt in small and often isolated resource-based communities.

Some environment data are collected from households or individuals (*e.g.*, recycling practices, fuel use) and their potential as a source of small area data is subject to the considerations already described in section 4. Other

environment data (*e.g.*, waste generation, environmental protection expenditures, use of natural resources) come from businesses and would be governed by the considerations of section 5. However, a great deal of environment data is obtained from physical surveys (*e.g.*, geological, physiographic, hydrographic), from instrument measurement (*e.g.*, temperature, air quality, water quality, ozone layer thickness), and from direct observation (*e.g.*, land use). Different considerations govern the relation of these data sources to small area data.

Because environment data are no respecters of administrative boundaries, the need for a flexible geographic infrastructure, emphasised in section 3, is especially important here. Small area geographic identification is needed to regroup data to geographical units that are more suitable for environmental analysis. For example, the production of waste attributable to a certain type of agricultural activity might be aggregated for all of the producers within a river basin. Environmental geographic units are either predefined (ecozones, drainage basins) or dictated by special events (areas covered with different thicknesses of ice, land areas flooded by heavy rains or spring thaws). In some cases, the area studied could be a very small site such as a park.

Physical quantity or quality data can be difficult to aggregate or summarize. In some cases, point source data such as air quality measures cannot be considered representative of any larger geographic unit. Water quality may be summarized or compared by using an indicator, such as the number of days beaches are open for swimming, but not simply as an aggregate or average of water quality readings. For many measures, the focus of interest may be on change over time rather than small area comparisons. In other cases, sampling and estimation techniques may need to make use of spatial analysis techniques such as contouring or interpolation.

The privacy and confidentiality concerns associated with environment data depend on their source. Data collected from households or businesses, even if they involve physical measurements, are protected by the same confidentiality rules as other data from those sources. Direct measurements of the stock of natural resources or the quality of the environment do not raise these concerns. Cartographic representation of spatial patterns may be one way to overcome some of the analytical frustrations of data suppression for small areas. Choropleth maps (maps which show the distribution of variables or characteristics by using colour or shading for ranges of the distribution) can explicitly represent the ranges implicit in rows or columns that would be suppressed in a published table.

Cross-border pollutant flows and their global effects make physical environment data an international issue. Cooperation between neighbouring countries is necessary to ensure that national boundaries do not impede analysis of the impact of physical processes that recognize no such boundaries.

In summary, the small area dimension is particularly important for environment data, not only because a locality is frequently the point of interest, but also because data must often be reaggregated to geographic areas more appropriate for environmental analysis such as ecozones or watersheds.

7. Organization and Dissemination Issues

Most NSOs are organized by subject-matter area. The production of small area estimates cuts across subject-matter areas, but requires support from Geography staff for geographic infrastructure, from Methodology staff for estimation and evaluation methods, and perhaps from other staff for analyzing and packaging data across subject areas. The question of how to organize small area estimation within an NSO therefore arises.

Requiring subject-matter areas to manage small area estimation in their areas, with support from methodology and geography staff as needed, is a natural choice since they should be most in touch with the data requirements and data limitations in their subject areas. More of an issue is how to package data for small areas for dissemination to users. Who should be responsible for pulling together data from different subject-matter areas for a particular small area? Should this be a regular program, or something that is done 'on demand'? Here there are different models to choose from – and Statistics Canada has tried most of them over the years.

At some periods in the past a division focussing on regional or urban statistics has existed to provide a regional focus for statistical data. At times, the census program, which is of course the richest source of small area data, has spearheaded the production of small area data profiles. At other times, an inter-divisional project has been used to manage a program of profiles for electoral districts or for other geographic areas. At the same time, regional office staff have played a key role in pulling together information for small areas in response to client requests. None of these arrangements has been ideal. The production of profiles has typically been a labour-intensive task requiring a broad subject-matter understanding and a lot of searching and manipulation of data. Despite the existence of standard geographic areas, the combination of data based on several different geographic bases is usually an issue. Ensuring that data for a large number of small areas are properly matched and collated can be an arduous quality assurance challenge.

Pre-planned profiles on paper were never overly successful. As a result, a strategy of maximizing responsiveness to client demands as they arose was preferred. With recent advances in technology, and broader coverage of small area data in the corporate database, a more automated approach is possible. A component of the Statistics Canada website (www.statcan.ca), called Community Profiles, and largely based on Census of

Population data, is our most recent attempt to make small area data more accessible and promises to be a precursor of future directions in this field. Some health data for health districts are already included, and certain other non-census sources of community data are under consideration.

8. Conclusions

The production of small area statistics by an NSO raises issues that are qualitatively different from those faced in its regular production of national, provincial or other large area data. The statistical theory that makes data based on sufficient individual measurements inherently reliable for large areas (ignoring bias for the moment) begins to break down for smaller areas. Unless a current census or administrative source with full coverage is available, this means that the NSO has to resort to some model-based help in order to provide estimates. Since alternative models can produce different estimates, a degree of arbitrariness is introduced into estimates, and this may be seen by some as undermining the objectivity of a NSO and its methods. The fundamental principle of openness and transparency about methods, including the choice of any models used and the impact of different assumptions, takes on even greater importance in the domain of small area estimation.

On top of this, an NSO should expect that small area estimates will come under more focused scrutiny than do many large area estimates. Though large area estimates receive broader attention, few individuals have the capacity to confirm or refute an estimate at the national level. But at the local level there will be many who think they know what is going on in their town. And typically small area estimation does not work uniformly well for all areas. The argument that a method works well on average will not quell criticism from those areas where it has not worked well – unless it has also worked to the local advantage! The NSO has to be prepared for the double jeopardy of weaker estimates under closer scrutiny.

If that is not enough already, confidentiality considerations loom larger at the small area level. The very fact that estimates are being produced for local areas highlights the potential for identification of individuals even though the NSO has taken sufficient precautions to prevent such disclosure. Some users of small area data for marketing purposes do not help the situation by implying in their advertizing that they can target mail to households based on individual or household characteristics, when they are actually using small area data to distinguish neighbourhoods. Some methods of small area estimation require record linkage which may also raise privacy concerns. Again a policy of openness and careful review of all such applications, at a senior level and before they begin, is necessary to ensure that the public benefit outweighs any privacy invasion.

Despite these potential difficulties, the demand for small area data remains high, technology offers new approaches to the management and dissemination of small area data, and methodological work on small area estimation is an active research area among statisticians. While small area data will generally not be an NSO's first priority, the relevance of its statistical programs will be magnified many times if it is able to cater to the most important small area data needs.

Acknowledgements

This paper was originally prepared to introduce a discussion on small area statistics among Heads of National Statistical Offices at the June 2001 Conference of European Statisticians in Geneva. The author wishes to acknowledge the contributions of many individuals at Statistics Canada who provided input to this paper, as well as helpful suggestions from a referee.

References

- Alexander, C.H. (1999). A rolling sample survey for yearly and decennial uses. *Proceedings of the 52nd Session of the International Statistical Institute*. Helsinki.
- Alexander, C.H. (2002). Still rolling: Leslie Kish's "Rolling samples" and The American Community Survey. *Survey Methodology*, 28, 1, 35-41.
- Bayarri, M.J., and Berger, J.O. (2000). *P* Values for composite null models. *Journal of the American Statistical Association*, 95, 452, 1127-1142.
- Brackstone, G. (1987). Issues in the use of administrative records for statistical purposes. *Survey Methodology*, 13, 1, 29-43.
- Durr, J.-M., and Dumais, J. (2002). Redesign of the French Census of Population. *Survey Methodology*, 28, 1, 43-49.
- Fay, R.E., and Herriott, R.A. (1979). Estimates of income for small places: an application of James-Stein procedure to census data. *Journal of American Statistical Association*, 74, 269-277.
- Federal Committee on Statistical Methodology (1994). Report on Statistical Disclosure Limitation Methodology (Statistical Policy Working Paper #22). Washington, D.C., Office of Management and Budget, Office of Information and Regulatory Affairs, Statistical Policy Office.
- Gambino, J., and Dick, P. (2000). Small area estimation practice at Statistics Canada. *Statistics in Transition*, 4, 597-610.
- Ghosh, M., and Rao, J.N.K. (1994). Small area estimation: An appraisal. *Statistical Science*, 9, 55-93.
- Isnard, M. (1999). *Alternatives to Traditional Census Taking: The French Experience*. Paris: INSEE.
- Jabine, T.B. (1993). Statistical disclosure limitation practices of united states statistical agencies. *Journal of Official Statistics*, 9, 2, 427-454.
- Kish, L. (1990). Rolling samples and censuses. *Survey Methodology*, 16, 1, 63-71.
- Kish, L. (1998). Space/time variations and rolling samples. *Journal of Official Statistics*, 14, 31-46.
- Purcell, N.J., and Kish, L. (1979). Estimation for small domains. *Biometrics*, 35, 365-384.
- Rao, J.N.K. (1999). Some recent advances in model-based small area estimation. *Survey Methodology*, 25, 2, 175-186.
- Redfern, P. (1989). European experience of using administrative data for censuses of population: the policy issues that must be addressed. *Survey Methodology*, 15, 1, 83-99.
- Schaible, W.L. (1996). (Ed.) *Indirect Estimators in U.S. Federal Programs, Lecture Notes in Statistics*. New York: Springer-Verlag, 108.
- Singh, M.P., Gambino, J. and Mantel, H.J. (1994). Issues and strategies for small area data. *Survey Methodology*, 20, 3-14.
- Zayatz, L., Steel, P. and Rowland, S. (2000). Disclosure limitation for Census 2000. *Proceedings of the American Statistical Association, Section on Government Statistics and Section on Social Statistics*, 67-71.