

DATA QUALITY CONCERNS WITH SUB-ANNUAL BUSINESS SURVEY FRAMES

N. Laniel and H. Finlay, Statistics Canada
N. Laniel, 11-E, R.H. Coats Bldg, Ottawa, K1A 0T6, CANADA

KEY WORDS: Misclassification, Coverage Deficiency, Error Measurement, Frame Updating

1. INTRODUCTION

Sub-annual business surveys are used by Statistical Agencies to provide key economic indicators for policy analysis. These indicators should reflect the current economic changes. This need for current information, expressed in terms of frame requirement, means that a survey list of businesses must be kept as up-to-date as possible. This is not an easy task since the business universe is dynamic in nature.

Statistical Agencies usually make efforts to produce a coherent and integrated set of economic statistics coming from many business surveys. A natural tool to help meeting this requirement of integration is the use of a common Business Register. That is a unique list of all businesses in the country which serves as the basis to construct all the sub-annual survey frames.

This paper deals with the data quality concerns involved with the use of the Statistics Canada (STC) Business Register (BR) by sub-annual surveys. These concerns are, in summary, the extent of deficiencies on the BR and their impact on the survey estimates. In regards to these concerns, actions must be taken such as: keeping the BR up-to-date, making sure that the BR resources are allocated where it counts, updating regularly survey frames and adjusting survey estimates for deficiencies. This paper concludes with work that needs to be done on methodologies for using a BR.

2. FRAMES

In this section the design of the Business Register is briefly described, then its relationship with the survey sampling frames is outlined and, finally, the typical design of an economic survey is described.

2.1 Business Register

The Business Register is a list frame mainly built with two administrative sources from Revenue Canada Taxation (RCT): the Tax returns of both corporations (T2's) and individuals (T1's) and the Payroll Deductions accounts (PD's) (Colledge, 1987). The PD's are accounts used by employers to remit monies to RCT for Pension Plans, Unemployment Insurance contributions and other deductions.

The BR is composed of two sets of units. First, the complex or large businesses with a revenue above specific thresholds, defined by 2-digit Standard Industrial Classification (SIC) code and province. The Tax returns and their corresponding PD's are linked for these large businesses. These links are useful for purpose of integration of the economic statistics produced by annual and sub-annual surveys.

The second set of units is composed of the remaining small businesses. In the case of these units, which are in a much larger number, the Tax returns and the PD accounts are not linked for operation cost reasons. This leads to the existence of two lists of small businesses: a Tax-based list and

a PD-based list. The PD-based list of small businesses along with the list of large businesses is used as the basis for constructing sub-annual business frames. This is because PD accounts provide a current source of information to update the BR, for example, they can be opened or closed by an employer at any point in time. The disadvantage with this source is that the non-employer businesses are not covered.

At any point in time, there exists a set of accounts at which sub-annual surveys do not have access. That set of inaccessible accounts is composed of PD's either out-of-scope (OOS) to economic surveys or in-scope but that cannot be made available to surveys for various reasons. Typically, OOS PD's are: accounts without remittances for the last twelve months, household accounts, foreign accounts, accounts owned by businesses which have ceased their economic activities, government special work program accounts, pension plan accounts or accounts used to solve a succession (estate of). The inaccessible in-scope PD's are essentially accounts, owned by active Canadian businesses, for which a proper SIC or Standard Geographical Classification (SGC) code is not known.

The BR keeps a list of all the units which are to be accessed by sub-annual surveys. This list is called hereafter the Statistical list. It contains the frame data necessary for economic surveys to delineate their in-scope population, for example, the SIC code, SGC code, Gross Business Income and number of employees.

2.2 Survey Frames

The economic sub-annual surveys are using the BR Statistical list data in order to build their sampling frames, to draw their samples and to produce their estimates.

Most surveys have a target population which corresponds to an industry division (e.g. Retail Trade, Manufacturing, etc...) or a subset of it. Few surveys cover more than one industry (e.g. the Canadian Survey of Employment, Payroll and Hours covers all except Agriculture, Fishing and Trapping). The sub-annual surveys usually use a three-way stratification: industry by geographic area by a measure of size. The first two variables are ordinarily used to satisfy user data requirements while the third one is used for optimal sample allocation. Also, it is typical of economic surveys to select with certainty the units in the large size strata. For small size strata where units are selected with probability, the sample design is often based on a rotation scheme where units are kept a minimum number of months (say 12 or 24) in the sample and, ordinarily, at least 12 months out of the sample after they rotate out. The stratification variable values initially assigned to units on the sampling frame are usually frozen for a number of occasions. That is, they are not immediately updated as stratum changes occur on the BR Statistical list. As a consequence, some surveys use a "domain estimation" approach when units change strata.

3. FRAME DEFICIENCIES

It is well known that perfect sub-annual business frames do not exist. Deficiencies are essentially due to the difficulty of

keeping it up-to-date (Tupek, Copeland and Waite, 1988). Definitions and sources of frame deficiencies are given below followed by their impact on economic survey estimates. The deficiencies being described in this section are not peculiar to Statistics Canada business surveys but also common with those of other countries such as the U.S. survey frames (Konschnik, 1988). Raj (1972) gives general definitions for defects found on frames.

3.1 Misclassification

A unit is misclassified when its SIC code or SGC code is the wrong one. This can be the result of the use of improper information for coding or a coding error by clerical staff. Another reason for misclassification can also be that it is unknown that the unit has changed its classification since the last time it was asked for.

A misclassification on the BR introduces biases at the stratum level when it causes a unit to be wrongly assigned to a stratum within the right survey population. It introduces coverage biases at the population level when the error is such that the unit is assigned to the wrong survey. A unit misclassified on the survey sampling frame but correctly classified on the BR introduces an increase in variance with the "domain estimation" approach but a bias otherwise.

3.2 Extraneous units

Extraneous units are units that should not be present on the BR Statistical list. Typically these are businesses which are out-of-scope for surveys.

If extraneous units are not identified as such by the BR or the survey, then such units in sample get data imputed and introduce a positive bias in the estimates. A unit on the survey sampling frame and known to be extraneous introduces an increase in variance.

3.3 Missing units

Missing units are businesses that should, in principle, be present on the BR Statistical list but are not. These are, for example, insufficiently classified PD remitters and units wrongly categorized as out-of-scope, RCT active PD's not yet on the BR lists and non-employer businesses.

The missing units cannot be included in survey sampling frames and thus introduce an undercoverage bias in the estimates.

3.4 Duplication

Duplicates are units which are represented more than once on the BR Statistical list. These are units in the small business list which are unknown to be already represented in the large business list and therefore should not be present in the smaller business list. They are also portions of large businesses unknown to represent the same business activity.

Like the extraneous units, the duplicates can increase bias and variance in the survey estimates. These units also introduce a burden for the respondents when duplicates of the same business activity are in the same survey sample.

3.5 Erroneous size measure

An erroneous size measure for a unit is a measure which is far from its true value for that unit.

The impact of erroneous values for a measure of size on the survey estimates can be serious. For example, loss of efficiency due to a weak stratification by size or presence of outliers can be observed. Also, the sample size determination, for targeting a specific precision, can be affected.

3.6 Inaccurate contact information

The contact information for the businesses on the BR may be inaccurate for an economic survey to effectively collect data. This happens because the contact data available on the BR often comes from the RCT administrative contact data which are not always appropriate for survey purposes.

The problem with inaccurate contact information is that it can lead to non-response when the appropriate contact person cannot be reached on time. Then the survey needs to impute the respondent data which in turn introduces a bias, although smaller than without imputation, and possibly increase the variance in the survey estimates.

4. FRAME UPDATING

This section discusses updating procedures to address the frame data quality concerns stated in the introduction. First, the elements for maintaining the BR Statistical list are presented, then, the updating of the survey frames is discussed.

4.1 BR Maintenance Elements

Two types of sources maintain the BR lists for sub-annual surveys as up-to-date as possible. These are the RCT administrative data and the BR surveys collecting frame data. As will be seen, the BR maintenance elements present some similarities with the sources of updates used by the U.S. Bureau of Census for their central list of establishments described in Detlefsen et al. (1985). In the following paragraphs, each of the BR sources of updates is described and the types of updates they provide are pointed out.

4.1.1 RCT Administrative Data

There are two major RCT's sources of updates based on PD accounts: (i) the PAYDAC file for births, deaths and name and address changes and (ii) the PD20 forms for initial classification data.

The PAYDAC file is an exhaustive list of all the PD accounts opened by employers. It is updated on a daily basis by RCT. That is, as soon as an employer requests the opening of an account, a new record is added to that file. Also, when an employer ceases his economic activities and sends all monies due, then the account is closed and the corresponding record deleted from the PAYDAC file in the month of January that follows at least 12 months after closure, providing no reactivation occurred.

Once a month STC receives the latest version of the PAYDAC file. It is then matched with the accounts belonging to the set of large businesses. Accounts that match the large units are assigned to that set and are used to signal changes in legal, operating or accounting structures of these businesses as well as potential births or deaths of units within these structures. The new units that do not match the set of large businesses are either assigned to the list of small businesses as births (if they remitted in the last twelve months and full classification information is available), or to the set of insufficiently classified accounts (if they remitted in the last twelve months but the classification information is insufficient) or the set of OOS units (if they did not remit in the last twelve months or the unit meets one of the exclusion criteria).

In principle, every month employers remit monies to RCT which then credit their accounts for the dollars received. That information is kept on the PAYDAC file made available to STC on a monthly basis. The remittances are then used to update the Gross Business Income, via a model, for all the small

accounts for which the difference between the old and new values is larger than some tolerance limits. Those accounts which did not remit for twelve months get an estimated Gross Business Income of zero. All the accounts represented on the BR Statistical list which get an estimated Gross Business Income of zero are removed (deceased) from that list. The monthly remittances will also be used to update the number of employees when a model will have been developed for the small businesses.

Each time an employer remits monies to RCT, he is sent a PD7AR receipt. A portion of the PD7AR form may be separated and returned to indicate any changes of name or address. Such changes are incorporated on the PAYDAC file.

When an employer request's to open a PD account is received, RCT sends him a PD20 form to fill. This form essentially asks for legal and operating data (amongst which the expected number of employees and a description of his industrial activity) on the employer which owns the new account. It can then take a few days or many months before that form is sent back to RCT.

Copies of the PD20 forms received by RCT are sent weekly to Statistics Canada. These forms are then captured and used to initialize the frame data for the small units only. This constitutes the main source of initial SIC codes and SGC codes for most births. When the copy of the PD20 form of an account is received and captured at STC and if the classification information on it is sufficient then that account is birthed on the Statistical list of the BR.

Another administrative source of updates for the large businesses is the annual Tax returns. Every year STC get copies of the Tax returns from RCT. These reports signal potential changes in business structures that have not been identified through the PD accounts.

4.1.2 BR Surveys

The BR conducts surveys which are used to obtain frame data on businesses when some pieces of data are missing, to confirm a change detected, or to refresh the data. For a discussion of some points considered to decide upon the BR surveys needed for updating refer to Colledge, Estevao and Foy (1987).

For small businesses, one BR survey is the Business Activity Report. It is a mail survey conducted for remitting employers for which the PD-20 data was not made available to the BR within 90 days from the opening of an account or for which the PD-20 business activity description could not be properly SIC coded. When the BR does not get a response from the business owning the PD account using the Business Activity Report, then it gets followed-up via the Classification Survey which is a telephone survey held in the Regional Offices. The Classification Survey also collects legal, contact and administrative information.

Another survey currently conducted by the BR for the small businesses is the Promotion Survey. This survey primarily consists of contacting all the small businesses showing an estimated Gross Business Income larger than the revenue thresholds in order to confirm that their revenue is effectively of large size. When the revenue of a business is confirmed to be of large size, then that business is to be promoted to the large business portion and no frame data is collected at that point. Such data is to be collected via a reaction profiling activity which is described below. The promotion of a unit to the set of large units generates the death of a small unit and the birth of a large unit on the BR Statistical list. If the revenue is identified as small size, then the business stays in the set of small

businesses and the same frame data as for the Classification Survey is collected on the spot.

The New Entrant Survey is a survey conducted by the BR for both the large and small businesses, on the BR Statistical list, which enter into economic survey samples. It collects the same frame data as the Classification Survey.

Another BR survey, which is conducted for both large and small businesses in economic survey samples, is the Survey Frame Feedback. It is conducted for any unit for which an economic survey has signalled a discrepancy between the frame data and the information given by the respondent while economic data was being collected. This BR survey provides the same frame data as the Classification Survey.

There is an activity called profiling which is a survey of large businesses collecting data on their, often complex, operational, legal and accounting structures. Farrall and Demmons (1987) discuss issues involved in profiling businesses. It is planned to use two types of profiling activities. First, the reaction profiling which consists of, for example, doing a mini-profiling exercise triggered by a change in the data of the set of PD accounts owned by a large business or a change in the information on its Tax returns. Clark and Lussier (1987) discuss in some details the use of administrative data in reaction profiling. The other type is named cyclical profiling and is meant to be an extensive and periodic profiling exercise of large businesses with the goal to maintain the information on these businesses as up-to-date as possible.

4.2 Updating Survey Frames with the BR

With the goal of integration of economic statistics it is natural that all economic surveys update their sampling frame using the BR Statistical list. These updates are the addition of births, the identification and removal of deaths and stratum changes. They are discussed below from a survey point of view. The handling of sampling frame updates is also discussed in Hidiroglou and Srinath (1991) and in Colledge and Lussier (1987). Raj (1972) gives general approaches to deal with defects found on sampling frames.

4.2.1 Births

There are essentially two types of births: new units on the BR Statistical list and units that were on that list but shown a change of their SIC code. There is no problem in adding births to a survey frame. The difficulty with births is to include them in the survey sample probability strata in a representative way. For example, the population age distribution should be well represented in the sample when births are added. This can be achieved by methods which give to every new units the same probability of inclusion as for the older units. Hidiroglou and Srinath (1991) discuss methods to select and add births in a sample.

4.2.2 Deaths

Deaths are units which have been removed from the BR Statistical list either because they became OOS or were found to be duplicates. They are also units which are still on the BR Statistical list but are now part of another survey population.

For a stratum where the units are selected in the sample with certainty, deaths can be removed as they are identified from the BR Statistical list without causing any problem from a sampling point of view. Its simply helps reduce the bias in the estimates.

For a stratum where the units are selected in the sample with probability, it is not obvious how to update a survey frame with deaths on the long run, given that their number increases

as time goes on. One way of handling deaths is to use domain estimation. That is in-sample dead units are assumed to report zero as the value for the variables of interest to the survey. A somewhat similar approach is used for the Canadian Monthly Retail Trade Survey (MRTS) and the Canadian Monthly Wholesale Trade Survey (MWTS) (Hidiroglou, 1988). Now as time goes on deaths are cumulated on the sampling frame. In the long run many deaths may be present in the economic survey sample which leads to estimates with undesirably large variances. Before that situation occurs, one solution is to remove all the deaths found at some point. As a result, the cleaned-up sample is composed of active units and the out-of-sample units are composed of both active units and unidentified dead units. This is due to the fact that the information on units in samples is more up-to-date than for out-of-sample units due to the New Entrant Survey and Survey Frame Feedback contacts. In other words, the cleaned-up sample is no more representative of its population. As a consequence one has no other choice than drawing a new sample. The disadvantage with this solution is that such a new sample will introduce a discontinuity in the series of estimates which is better to avoid since it is required to measure the current movements in the economy.

To handle the problem of the increasing sampling variance with time, as a result of the cumulation of deaths, without introducing discontinuities another solution can be used. It requires an "unbiased" and "regular" source for the identification of deaths. An unbiased source is a source for which the probability of identifying a death in the survey population is the same for both in-sample and out-of-sample units. By regular it is meant that the source regularly identifies deaths such that from one sub-annual period to the next it is likely that the number of deaths identified are in the same order of magnitude. When such an unbiased and regular source is used then deaths identified with that source can be removed from the survey frame without introducing a bias (conditional on the size of the cleaned-up sample) in the estimates and potentially without undesirably increasing the variance of the estimates of period-to-period change. That solution has been adopted by the US Bureau of the Census (USBC) for some of their economic surveys (Detlefsen, Konschnik and Monsour, 1985). At Statistics Canada, a number of sources can be labelled as unbiased and regular and could be used to remove deaths. The major such source is the set of Payroll Deduction accounts.

There is another solution to the updating of deaths which can as well present a reduction of variance in comparison to the cumulating-cleaning-redrawing approach while avoiding discontinuities. It is called "death adjustment" and used by the Canadian Survey of Employment Payroll and Hours. That solution is discussed in details in Schiopu-Kratina and Srinath (1986). It consists of removing all deaths from the sampling frame as they are identified without consideration of the source of update. At the same time a death adjustment (decimal) number is kept to represent deaths not yet identified. The adjustment number is equal to the number of deaths necessary in the sample to represent the estimated number of deaths out of the sample not yet identified. The calculation of that number is such that the sampling weight is kept constant. That adjustment number is then used at the time of estimation.

4.2.3 Stratum Changes

Changes of stratum are due to changes of SIC code, SGC code and measure of size within a given survey population. These changes of stratification variable values

happen on the BR Statistical list due to the different sources of update. The SIC code changes considered here are only those for which units do not change of survey population. In fact, as defined in the two previous sub-section, from an economic survey point of view a change of SIC code to (from) another survey population is considered a death (birth).

For a stratum where the units are selected in the sample with certainty, stratum changes on the survey frame can be made as they are identified on the BR Statistical list. It simply helps reduce the bias in the estimates.

For a stratum where the units are selected in the sample with probability, one way of handling stratum changes is to use domain estimation. For example, to produce an estimate for a given group of SIC codes one has to sum over all the survey strata where there are units with the SIC code of interest while assuming that the units which do not have an SIC code of interest report zero as the value for the variables of interest to the survey. That approach is used for the Canadian MRTS and MWTS (Hidiroglou, 1988). It is impossible to avoid units with stratification variable values different from their stratum being cumulated on the sampling frame as time goes on. In the long run many units with a stratum change may be present in the economic survey sample which leads to estimates with potential undesirably large variances. To overcome that situation, one solution is to make effective all the stratum changes found at some point. But, then the sample is composed of units in the appropriate stratum and the out-of-sample units are composed of both units in the appropriate stratum and unidentified units to be in the wrong stratum. The cause of this difference in composition and its consequences are the same as in the case of deaths discussed in 4.2.2.

To handle the problem of the increasing sampling variance with time, as a result of the cumulation of stratum changes, an other solution can be used. It requires an "unbiased" and "regular" source for the identification of stratum changes as in the case of deaths. At Statistics Canada a number of sources can be called unbiased and regular for some stratum changes. For SGC code and size changes the major sources are the Tax returns and the Payroll Deduction accounts. For SIC code changes, the major source for the large businesses is the Tax returns. In the case of SIC code changes for the small businesses, there is no major large source but only the other economic survey sample updates.

There is another solution to the updating of stratum changes which can as well present a reduction of variance in comparison to the non-updating approach while avoiding discontinuities. That solution, "stratum change adjustment", would be a modified version of the "death adjustment" discussed in section 4.2.2. The modification would be to make the changes effective for all the out-of-sample units belonging to another stratum and as well as the corresponding units in the sample except for a number equal to the adjustment number, which will then have to be an integer.

Another way of handling stratum changes in the case of a probability stratum is to ignore them. The advantage is that there will be no increase in variance as time goes on but the disadvantage is that biases will then be introduced in the survey estimates when the stratum changes are due to changes in SIC or SGC codes. Also, the presence of outlying values may cause the sample estimates to have a larger variance than planned when the stratum changes are due to size changes. If the negative impact of cumulating all the stratum changes becomes too important one can always redraw a new sample from the changed survey population but again that is likely to introduce a discontinuity in the series of estimates. As for the

domain estimation approach, the problem of discontinuity can be greatly reduced by using either an unbiased and regular source to modify the content of strata or by using a "stratum change adjustment".

This approach of ignoring changes, changing the population and then redrawing a sample is used by the USBC in the case of SIC code changes for some of their economic surveys (Detlefsen, Konschnik and Monsour, 1985). The SIC code changes are applied after every quinquennial economic census.

5. MEASURES OF DEFICIENCIES

In this section, the measurement of frame deficiencies as well as means of improving survey estimates from these measures are discussed to further address the data quality concerns enunciated in the introduction.

5.1 Measurement

The specific deficiencies to measure are coverage errors (i.e. duplication, extraneous units, missing units and units classified in the wrong population) and the accuracy of the different frame data items (i.e. SIC code, SGC code, measure of size, name and address, etc). These should be measured in terms of counts of units to evaluate cost of correcting errors and in terms of measure of size to determine which deficiencies have the largest impact on survey estimates. These measures should also be periodic in order to be able to observe improvement or deterioration in the data quality and then be able to act in a timely and appropriate manner.

The measurement of a deficiency essentially depends on its nature as well as on the design of the BR. The latter enters into play because there are three distinct sets of units that can be used to make inferences on data quality. These are the large units, the small units and the in-scope active units not on the BR Statistical list.

5.1.1 Large units

For the large units, it is planned to use the results of cyclical profiling activities as the source of data to assess the accuracy of the frame data items and evaluate the amount of duplication between large units, of extraneous units and of misclassified units causing coverage errors. The profiling results could be analyzed and a model developed to provide measures.

5.1.2 Small units

For the small units, the results of a periodic sample survey will be used to assess the accuracy of the frame data items and evaluate the amount of duplication between small and large units, of extraneous units and of misclassified units causing coverage errors.

5.1.3 Units not on the Statistical list

Four set of units not present on the Statistical list can be distinguished: the active insufficiently classified PD accounts, the in-scope units erroneously categorized as OOS, the active employers not yet on the BR and the non-employer businesses. All of these four sets of units contributes to the undercoverage of the BR. The measurement of the amount of undercoverage due to these units is discussed below.

a) Insufficiently classified

As mentioned in section 4.1, these units are somehow all being surveyed either via the PD-20 form, the Business Activity

Report or the Classification Survey. In other words, these units are subject to a census. One approach here is to select the version of that population for a given reference period then, to wait a few periods to get some data collected and, finally, to poststratify by variables related to the response rate. Inferences will then be made by assuming that non-response occurs randomly within the poststrata.

b) Erroneously OOS

The second set of units not on the Statistical list are the OOS accounts erroneously categorized as such. To get an estimate of the number of these units, the straightforward approach is to draw a sample from the set of OOS units and to survey the sampled units (to assess their SIC and SGC codes and verify if they are active). The disadvantage with this approach is that it would require a large sample size to provide reliable estimates since the number of OOS units is large (way over a million) and the proportion of erroneously OOS units is expected to be small. It is not likely that one would have the resources for such a sample. Hence this should not be attempted.

It has been observed that some of the OOS units become in-scope after a certain time, the characteristics of these units should be studied to eventually find a way to model and predict their number in an efficient way.

c) Not yet on the BR

These units represent employer businesses which have started to operate before opening a PD account or which have opened an account but are being cumulated on the RCT PAYDAC file between two monthly transmissions to STC. For this third set of units, one could track them as they appear on the BR, then come up with a model which would predict the number of these missing units for a given period.

d) Non employers

Non-employer businesses are by definition excluded from the Business Register sub-annual list. In such a case the usual course of action is to supplement the list with an area frame (Raj, 1972). Such a supplemental frame is not used at STC for business surveys. This makes the estimation of the number of non-employer businesses difficult. A relatively cheap way that should be investigated would be to take a sample of Tax returns, then attempt to link that sample with the complete set of PD accounts, for the corresponding reference period. Finally, the unmatched returns could be used to provide measures.

5.2 Survey Adjustments

Two types of adjustments are briefly discussed in this section: (i) prior to drawing the sample and (ii) after the estimates have been produced.

If one knows the proportion of extraneous units before drawing the sample one can use this information to increase the sample size to get the desired sampling variances for the estimates. Latouche and Hidioglou (1987) provides a method of sample allocation taking into account the proportion of unidentified extraneous units.

Adjustments to survey estimates are more likely to be done for undercoverage deficiencies than for overcoverage ones. The reason is that it is likely that extraneous units and duplicates will be identified in the sample and then that information used in the estimation process.

It can happen that the adjustment factors have a relatively high coefficient of variation in particular if they are estimated from a small sample. In such a case the application of an

adjustment factor to a survey estimate may increase its variance up to the point where it becomes less reliable, from a mean square error point of view, than the biased unadjusted estimate. To avoid that situation a simple minimum Mean Square Error approach can be used, for example, in conjunction with a composite estimator to reduce the undercoverage bias.

6. CONCLUSIONS

From the above discussion on the updating of frames and the measures of deficiencies, it is clear that more than one solution can be used for dealing with the dynamic nature of the business universe. For instance, in the case of deaths occurring in probability strata, one can distinguish three possible solutions to handle them. All three use domain estimation but make use of a different strategy for removing deaths from the survey frame.

The first solution consists of cumulating deaths on the survey frame up to the point where the sampling variance gets to a predetermined level. Then a sample redraw is made from the cleaned-up population. The second solution is to wait for deaths to be identified by an unbiased and regular source to clean-up the survey frame. The third solution consists of cleaning-up all deaths identified and keeping a sample adjustment factor to represent the deaths not yet identified in the out-of-sample portion of the survey population. It is not obvious which solution is the best for the different surveys.

In the case of stratum changes, one can think of at least six solutions to cope with them. The first three ones make use of domain estimation but the three last ones do not. The first three and the last three solutions are analogous to the three approaches for the deaths. It is not clear which one of these stratum change solutions is preferable for the current economic surveys.

One major problem with births is the time it takes for the new units to be made available to the surveys. This delay causes undercoverage. The approach proposed here for that issue is to use a composite adjustment estimator. The benefits of the use of such an estimator to compensate for undercoverage should be investigated. Furthermore, methods for estimating undercoverage and other deficiencies need to be developed and evaluated. A few have been proposed in this paper.

From the discussion above, it is easy to see that the handling of the dynamic nature of the business universe when maintaining survey frames is an area of survey methodology needs more research.

REFERENCES

- Clark, C. and Lussier, R. (1987), "The Use of Administrative Data for Initial and Subsequent Profiles of Economic Entities", Proceedings of the Section on Survey Research Methods, American Statistical Association
- Colledge, M., Esteveao, V. and Foy, P. (1987), "Experiences in the Coding and Sampling of Administrative Data", Proceedings of the Section on Survey Research Methods, American Statistical Association
- Colledge, M. and Lussier, R. (1987), "Updating Survey Frames", Internal document, Business Survey Methods Division, Statistics Canada, Ottawa
- Colledge, M. (1987), "Use of Administrative Data in the Business Survey Redesign Project", Proceedings of the Statistics Canada Symposium on Statistical Uses of Administrative Data, Ottawa
- Detlefsen, R.E., Kongschnik, C.A. and Monsour, N.J. (1985), "Constructing and Maintaining Frames and Samples for Business Surveys", Proceedings of the Section on Survey Research Methods, American Statistical Association
- Farrall, K. and Demmons, P. (1987), "Profiling for Statistics Canada's Central Frame Data Base", Business Survey Redesign Project Working Paper, Statistics Canada, Ottawa
- Hidiroglou, M.A. (1988), "Methodology for the Monthly Wholesale-Retail Surveys", Business Survey Methods Internal Report, Statistics Canada, Ottawa
- Hidiroglou, M.A. and Srinath, K.P. (1991), "Problems associated with designing sub-annual business surveys", Proceedings of the Section on Business Statistics, American Statistical Association
- Kongschnik, C.A. (1988), "Coverage Error in Establishment Surveys", Proceedings of the Section on Survey Research Methods, American Statistical Association
- Latouche, M. and Hidiroglou, M.A. (1987), "Sample Size Determination and Allocation for the Monthly WRTS", Business Survey Methods Internal Report, Statistics Canada, Ottawa
- Raj, D. (1972), "The Design of Sample Surveys", Chapter 8: Sampling from Imperfect Frames, pp126-138, McGraw-Hill
- Schiopu-Kratina, I. and Srinath, K.P. (1986), "The Methodology of the Survey of Employment, Payroll and Hours", Working paper no. 86-010E, Methodology Branch, Statistics Canada, Ottawa
- Tupek, A.R., Copeland, K.R. and Waite, P.J. (1988), "Sample Design and Estimation Practices in Federal Establishment Survey Research Methods, American Statistical Association, 298-303