

REDESIGN OF THE CANADIAN AGRICULTURE SURVEYS

Rebecca Morrison, Suzelle Giroux and Claude Julien¹

ABSTRACT

Every five years, the agricultural commodity surveys are redesigned following the Canadian Census of Agriculture. Since the previous redesign, significant changes to the target population and the survey program itself have occurred. Firstly, the number of farms in Canada has decreased and, secondly, in order to alleviate respondent burden, the commodity surveys will no longer interview 'small' farms. In addition, it was decided to investigate a further reduction of the sampling population. As a result, the survey designs have undergone many changes; the most significant of which is the adoption of a simpler design with a more dynamic frame to ensure good coverage until the next redesign in 2007. The focus of the presentation will be the design used to achieve the multiple objectives of the survey program in light of the new challenges that the program now faces.

KEY WORDS: Coverage, Redesign, Response burden

RÉSUMÉ

Chaque cinq ans, à la suite du recensement agricole canadien, les enquêtes agricoles sont remaniées. Depuis le dernier remaniement, des changements considérables à la population cible et au programme d'enquêtes se sont produits. Premièrement, le nombre de fermes au Canada a diminué, et deuxièmement, dans le but de diminuer le fardeau de réponse, il a été décidé de ne plus enquêter les «petites» fermes. De plus, il a été décidé d'examiner l'impact d'une réduction encore plus grande de la population à échantillonner. En conséquence, beaucoup de changements au plan de sondage ont été effectués; le plus considérable est l'adoption d'un plan de sondage plus simple avec une base de sondage plus dynamique pour s'assurer une bonne couverture jusqu'au prochain remaniement. Le sujet principal de la présentation sera le plan de sondage adopté pour atteindre les objectifs multiples du programme d'enquêtes.

MOTS CLÉS : couverture, fardeau de réponse, remaniement

1. INTRODUCTION

Every five years following the Census of Agriculture, Statistic Canada's main agriculture surveys are redesigned. Survey frames are updated based on the latest census data, sample designs are evaluated and any necessary changes or improvements to the survey designs are made. It is very important to have a robust survey design that ensures stability of the estimates over time, as it will stay in effect for the duration of the next five year survey cycle. Following the 2001 census, the Semi-annual Livestock Survey, the Panel of Crop Surveys and the Semi-annual Survey of the Atlantic Provinces were completely redesigned.

The agriculture survey program's goal is to collect timely information on all crop and livestock inventories in Canada. The Semi-annual Survey of the Atlantic Provinces is a multi-purpose survey that collects information on both crop and livestock farms in the Atlantic Provinces. The remaining six provinces, Québec to British Columbia are covered by the Semi-annual Livestock Survey and the Panel of Crop Surveys which collects livestock and crop data respectively. The focus of this paper will be the newly redesigned Semi-annual Livestock Survey.

The Semi-annual Livestock Survey is comprised of two survey occasions conducted in January and in July. The two occasions make use of the same frame and stratification plan but the January sample is a subsample of the previous July sample. During the January and July survey occasions, information is collected on a wide variety of livestock, of which

¹ Rebecca Morrison (rebecca.morrison@statcan.ca), Suzelle Giroux (suzelle.giroux@statcan.ca) and Claude Julien (claudio.julien@statcan.ca), Statistics Canada, 17th floor, R. H. Coats Building, Tunney's Pasture, Ottawa, Canada, K1A 0T6,

the most important commodities are cattle, sheep and pigs. It is most important to have reliable provincial level livestock estimates, especially for these three main commodities of interest.

2. SEMI-ANNUAL LIVESTOCK SURVEY

2.1. Previous Livestock Survey Design

The livestock survey was last redesigned in 1996. Under that redesign, the livestock survey made use of a list frame comprised of virtually all census farms, livestock farms and non-livestock farms alike. The frame remained for the most part static throughout the survey cycle. Given the multiple objectives of the survey, a multivariate stratification plan was implemented. A multivariate clustering method was used to form the strata independently for each province. This design gave very precise estimates for all the variables of interest, however, this method led to very complex stratum definitions making it very difficult interpreting what a stratum represented. This and other design features were identified for improvement under the 2001 redesign.

2.2. Redesign Highlights

The objectives of the redesign are to implement any necessary changes or improvements to the previous survey design. Compromise was necessary to meet the multiple and varied objectives of the survey. The main improvements or changes required under the current redesign are as follows:

i. Improve Coverage

In most surveys coverage is a key issue and the livestock survey is no exception. Over previous survey cycle, from 1998 to 2002, a gradual deterioration of the survey frame was observed that varied between 5% and 10% when compared with the 2001 census totals. A strategy is therefore necessary to address the coverage issue under this redesign.

ii. Reduce Respondent Burden

The agriculture survey program has a well developed infrastructure and already has many good practices in place to reduce respondent burden. The largest agriculture surveys use collocated sampling to minimise the overlap between samples, also the very large or complex structure farms receive special treatment to ensure strict control over the burden suffered by that part of the population. This paper will only highlight the new practices put in place under this redesign; above and beyond the already good practices in place.

iii. More Easily Interpretable Strata

As already mentioned, the previous livestock stratification plan led to very complex stratum definitions. The data validation stage of survey processing became difficult at times due to the fact that it was often difficult to interpret what a stratum represented. To help facilitate survey processing and data validation it was decided to implement a more simple stratification plan.

iv. Reduced Number of Strata

The number of strata under the previous design varied from 20 to 37 per province. There were some problems with stratum jumpers towards the end of the survey cycle. It is very easy for livestock farms to undergo changes in commodities which would make them better suited for another stratum with another weight but left in their current strata, these farms lead to outlying values. In order reduce the occurrence of stratum jumpers, it was decided to reduce the number of strata.

v. *Maintain a Similar Level of Precision*

Ultimately, the implemented design should be one that achieves a similar level of precision with respect to all the variables of interest when compared to the previous design.

2.3. Semi-annual Livestock Survey Design

2.3.1. Livestock Survey Frame

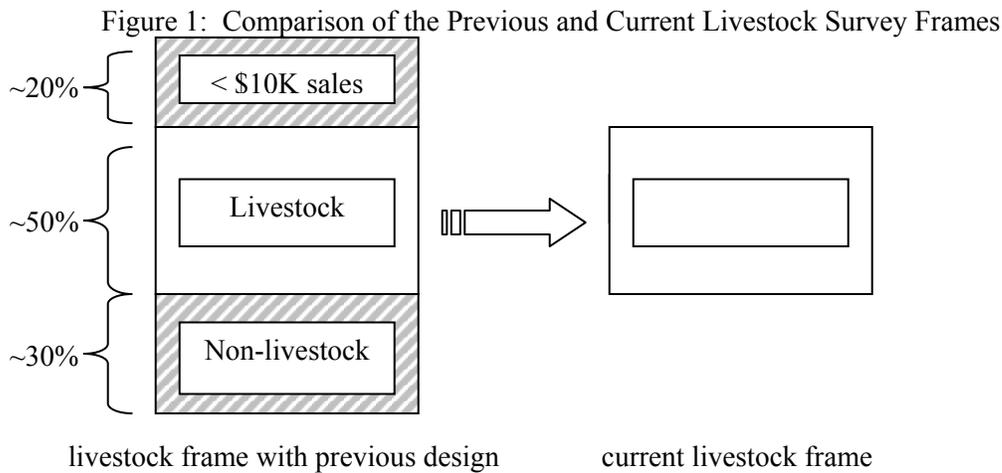


Figure #1 illustrates the major changes to the livestock survey frame which took place under this redesign. The shaded areas represent the parts of the previous frame no longer found on the current frame. Clearly there was a significant reduction to the survey population between this redesign and the last. Previously, the survey population was all farms in Canada, now the survey population is all livestock farms in Canada. These new exclusions as well as the reason behind the exclusions will be discussed in the remainder of this document.

In order to reduce the burden suffered by the smaller farms it was decided to exclude all farms reporting less than \$10,000 in sales at the 2001 census (shown in Figure 1 as '< \$10K sales'). Their exclusion should not have a significant impact on the overall estimation of the commodities covered by the survey as their contribution to the total commodities is minimal. Had the survey frame remained similar to the previous one, these farms would represent slightly less than 20% of the frame.

Livestock and non-livestock farms have traditionally been included on the frame for coverage purposes. In fact, a significant number of farms on the previous livestock frame did not have any of the three main commodities of interest, namely cattle, pigs or sheep. The reason the non-livestock are kept on the frame is that farms not reporting any livestock at the time of the census can begin livestock production afterwards. Livestock inventories can and do change drastically over short periods of time. Again, had the survey frame remained similar to the previous one these farms would represent roughly 30% of the frame (shown in Figure 1 as 'Non-livestock'). Keeping these non-livestock farms on the survey frame marginally increases coverage but is costly in terms of collection and processing resources. Historically, only a very small portion of the non-livestock farms, at the previous census, ever actually report livestock during future survey occasions and, of these, very few actually report a significant amount of livestock.

The main difference with regards to the treatment of the non-livestock farms compared with the previous design is that the survey will only target those which are likely to report livestock in the future. If a farm reported at least one of the three main commodities of interest during the 1996 census or afterwards on a survey then it will remain on the new livestock frame, otherwise it will be excluded from the frame.

From a response burden point of view this is very beneficial. Previously the livestock and crop surveys used a common frame (with different stratification plans). As the samples are drawn randomly, it is possible and quite likely that farms will be contacted by surveys for which they are not in scope. What this means is that some livestock farms could be contacted under the crop survey and therefore needlessly burdened and vice versa. By only targeting only farms that are

in-scope for the respective surveys, the agriculture surveys are moving towards respecting the survey populations of each survey.

As previously noted, the survey frame deteriorated throughout the previous survey cycle. The survey frame remained static in the past because the information necessary to add to farms to the frame was unavailable or incomplete. A mechanism was needed to find and include new or missed farms not already on the survey frame. After much consideration, it was decided to go with an entirely dynamic frame. This is the first such occasion where the livestock survey has adopted this approach. A new annual survey, called the Farm Update Survey (FUS), was developed to find new farms. This and other sources will be used to make regular updates to the frame. Livestock farms not already included on the frame will be added on a regular basis. The new survey design incorporates a very simple stratification plan that makes updating the frame easier and requires only a minimal amount of information about the new farms. As described in the following section, only the total number of head of cattle, pigs and sheep will be necessary to be able to add new farms onto the frame in the appropriate stratum.

2.3.2. Livestock Stratification Plan

As previously mentioned, there were some problems with stratum jumpers towards the end of the survey cycle. It is possible for livestock farms to undergo changes in commodities which would make them better suited for another stratum with another weight but left in their current strata, these farms lead to outlying values. For this reason it was decided to reduce the number of strata compared with the previous design. In addition, it was decided to implement a more simple stratification plan when compared with the previous design to help facilitate survey processing and data validation.

For the livestock survey each province is stratified independently. First, certain specified farms that generate very large sales and or have a complex structure are grouped into a take-all stratum. Next, the sigma-gap rule is applied to identify farms that are significant for a certain variable of interest. Six different livestock variables were used, total cattle, dairy cows, beef cows, total pigs, total sows and total sheep. The sigma-gap rule can be summarized as follows: if we let X_i , be the variable of interest X for unit i , ($i = 1, 2, \dots, N$), and that X is sorted in ascending order for all the $X_i > 0$ and we let k , if it exists, be the smallest number, where $X_k > MED$ and $X_k - X_{k-1} > \sigma$, where σ represents the standard error of the $X_i > 0$ and MED is the median of the X_i 's. Then the farms having $X_i > X_k$ are said to be sigma-gap specified. Any farm identified above the sigma-gap rule for any of the above mentioned commodities is also placed into a take-all stratum. These farms are removed from further stratification procedures.

The non-livestock farms which are likely to report livestock will also be grouped into their own stratum and will be sampled at a fixed rate of 10%. By keeping these farms in their own stratum there is the built in benefit of being able to exclude them from further survey occasion should the need arise.

All other farms are ultimately stratified based on the dominant livestock commodity and farm size. Farms are first grouped based on dominant commodity; mostly cattle, mostly pigs or mostly sheep. Next, within each of these three groups there is a further division based on farm size in terms of head of cattle, pigs or sheep. Uniform stratum boundaries were used across all of the six provinces. Using uniform stratum boundaries diminishes the designs efficiency but allows for easier interpretation and comparison at the provincial level. Table # 1 below shows the stratum boundaries used within each dominant farm type group.

Table # 1: Livestock Survey Stratum Boundaries

Farm Type	Farm Size		
	Small	Medium	Large
Mostly Cattle	< 100	100 - 300	> 300
Mostly Pigs	< 1000	1000 - 4000	> 4000
Mostly Sheep	< 100	100 - 400	> 400

Table # 2: Livestock Stratification Plan

Description	Stratum		
very large sales/complex structure/sigma-gap	take-alls		
large	mostly cattle	mostly pigs	mostly sheep
medium	mostly cattle	mostly pigs	mostly sheep
small	mostly cattle	mostly pigs	mostly sheep
non-commodity	potential livestock farms		

The very simple stratification plan adopted by the livestock survey is illustrated in table # 2. The previous livestock survey stratification plan could never have been summarized so simply.

2.3.3. Sample Allocation

The sample sizes for the two survey occasions are based on budgets and the sample sizes used under the previous design. Allocation of the sample was done independently for all provinces using a United States Department of Agriculture (USDA) program that employs an optimal allocation algorithm (see Bethel, 1986). This program allocates a given sample to minimise a linear combination of the squared coefficients of variation (CVs) of the selected allocation variables (Denis 1996). The CVs mentioned here are theoretical CVs calculated from the survey frame. As was the case with the previous design, the following allocation variables were used: total cattle, total milk cows, total beef cows, total pigs, total sows, total sheep, total pasture area and total field crop area excluding hay. The milk cow variable was not used for the Prairie Provinces including the Peace River region of British Columbia. The allocation for the July and January survey occasions was done separately.

2.3.4. Coefficients of Variation

The new design is expected to achieve similar precision compared to the previous design. The same level of precision was attainable, despite the much simpler stratification plan, because the sample sizes are similar to the previous design and the sampling population underwent a major reduction. The following table, table # 3, presents the ranges of provincial level expected coefficients of variation (CV) for the three main commodities of interest under the previous and current designs.

Table # 3: Ranges of Provincial Level Expected CVs

Commodity	Previous CV	Current CV
Cattle	0.68% to 1.41%	0.60% to 0.93%
Pigs	1.01% to 2.21%	0.44% to 0.65%
Sheep	1.35% to 2.65%	1.41% to 2.57%

3. CONCLUDING REMARKS

The newly redesigned livestock survey has adequately achieved the redesign objectives it was intended to. The livestock survey has seen the adoption of a more dynamic frame that, when used in conjunction with the update survey, will help to ensure good coverage throughout the next survey cycle from 2003 to 2007. In addition, the formation of non-livestock strata will help to better control the burden those units are subjected to and we have been more selective about which non-livestock farms to include on the frame. This redesign saw the exclusion of small farms from the regular commodity surveys to reduce the burden suffered by that part of the population. As well, the livestock survey has a much more simplified stratification plan with a reduced number of strata. Thanks to the reduced sampling population, the newly-designed livestock survey has maintained a similar level of precision of the estimates of key variables compared to the 1996 design. This survey design is not the most efficient design, precision-wise, but it does meet the varied requirements of the survey program. This trade-off in efficiency gives the benefit of more simple data treatment and processing and allows for a more easily updateable frame.

ACKNOWLEDGEMENTS

The authors wish to express their sincere appreciation to the many members of the agriculture section of the Business Survey Method Division of Statistics Canada who have helped in the realization of this project. Particular thanks are due to Chantal Belley, Martin Lachance, Ann Lim and Sarah McKinnon for their significant contributions.

REFERENCES

- Bethel, J. (1986). *An Optimum Allocation Algorithm for Multivariate Surveys*. Statistical Reporting Service Staff Report, USDA, Number SF&SRB-89.
- Denis, J. et al. (1998). *1996 Redesign of Agriculture Commodity Surveys*. Internal methodology report, Statistics Canada.