

Coverage of the 2011 Canadian Census of Agriculture

David MacNeil

Statistics Canada, 100 Tunney's Pasture Driveway, Ottawa, Canada, K1A 0T6,
dave.macneil@statcan.gc.ca

Abstract

The collection methodology for the Canadian Census of Agriculture will change in 2011, moving from a primarily drop-off/mail-back approach to a primarily mail-out/mail-back one. To ensure that the new collection methodology does not have a negative impact on farm coverage, modifications to the existing methodology are required. Prior to the census, improvements are needed to both the quality and coverage of the Farm Register to ensure a good mailing list; a redesigned survey to identify new farming operations will help to achieve this. During census collection, the strategy to prioritize and select non-responding farms for telephone follow-up will play a key role in maintaining coverage. Following census collection, imputation will be used to help maintain coverage. In addition to covering these topics, the methodology for evaluating farm coverage will be reviewed.

Key Words: Census of Agriculture, Coverage, Register, Non-response follow-up

1. Introduction

The Canadian Census of Agriculture is conducted every five years concurrently with the Census of Population. Historically, questionnaires were dropped off by field enumerators: every dwelling would receive a Census of Population questionnaire, and if it was determined that someone in the household operated an agricultural operation then a Census of Agriculture questionnaire would also be left. More recently, there was a move to mail-out questionnaires: in 2006, approximately two thirds of the Census of Population dwellings were covered by mail-out as opposed to enumeration. The majority of these dwellings were in urban areas while the majority of agricultural operations are located in rural areas. For this reason, the impact on the Census of Agriculture was not significant as only about 6% of agricultural operations were not covered by field enumeration. However, in 2011, the Census of Population will increase the coverage of its mail-out strategy to 80%, including more rural areas. This shift in the collection methodology for the Census of Population meant that the Census of Agriculture had to review its collection methodology.

2. Collection Strategy of the 2011 Census of Agriculture

The review of the Census of Agriculture collection methodology involved an evaluation of its strengths, weaknesses, as well as the opportunities and risks associated with new approaches. In the end it was determined that the best option for the Census of Agriculture was to go with a 100% mail-out strategy. Since the Census of Population targets dwellings rather than specific individuals within dwellings, the Census of

Agriculture will send questionnaires to specific farm operators independently from the Census of Population mail-out. However the two censuses will continue to be conducted at the same time and thus the Census of Agriculture will continue to benefit from the Census of Population publicity campaign: the 2009 Census Tests revealed that the Census of Agriculture response rates were much higher in areas where the Census of Population was also testing. Another advantage for the Census of Agriculture is the fact that the Census of Population contains a question asking if anyone living in the dwelling is a farm operator. The benefits of this question to coverage of agricultural operations will be explained in section 4.

3. Statistics Canada Farm Register

The mailing list for the 2011 Census of Agriculture will be the Statistics Canada Farm Register (FR), a repository of all known agricultural operations in Canada. It contains approximately 290,000 agricultural operations (230,000 of which are currently in business) linked to 400,000 operators. The FR is the frame for the Agricultural Survey Program. There are more than 30 surveys with a combined annual sample size of 200,000 operations (100,000 of which are unique farm operations). However, not all farms on the FR are contacted during the 5 year intercensal period because the sample overlap from year to year is high. The Agricultural Survey Program is still a significant source for updates to administrative information and the operating status of existing agricultural operations on the FR, but historically the Census has been the main source for births.

Over the past few censuses, the FR has been used more and more to ensure the coverage of agricultural operations, as well as to help measure any remaining undercoverage. In addition, it was successfully used as the mailing list for the 6% of agricultural operations that were not covered by enumerators in 2006. However, it will obviously play a much more significant role in the 2011 Census of Agriculture. Therefore, the key coverage activities in advance of the Census will involve improving both the content and coverage of the FR.

3.1 Improving Farm Register Content

The most important improvement to the content of the FR is related to the quality of the address fields. Historically, these fields have not been given high priority due to the fact that the majority of surveys are collected via computer-assisted telephone interviews and the Census was primarily drop-off/mail-back. The quality of the existing address fields was evaluated by passing them through a Statistics Canada software which standardizes address fields and validates or assigns (where possible) a postal code. Initially, 79% of addresses were validated and 14% of addresses were imputed (i.e. the municipality name was assigned using the postal code part of the address), leaving 7% of addresses that were considered to be invalid. In order to improve these results, other administrative sources (Census of Population, telephone directories, and tax records) were used to update the address fields. Using these additional sources, 92% of the addresses can be validated and 6% can be imputed, leaving only 2% invalid. The invalid addresses will be resolved manually, and a sample of the imputed addresses will be checked manually. In addition, the quality of the address fields can be further validated through the Farm Register Update, which will be explained in section 3.2.

Several other FR variables will also need to be cleaned-up in advance of the Census. Since the farm name and operator name will appear on the mail-out label, these fields

will need to be validated to ensure that there are no obviously invalid values (e.g., an operator name of ‘Mickey Mouse’). Since many of the smallest farms have not been contacted for any surveys since the last Census in 2006, there is work to be done to validate the telephone numbers, which will be essential for the follow-up of non-respondents, as well as for matching and removing duplicates. It is also important that every effort be made to update the gender and date of birth fields, as these also help the matching processes.

3.2 Improving Farm Register Coverage

Since there will be no field enumeration in 2011, improving the coverage of the FR is more important than ever to the success of the Census of Agriculture. In order to address these issues, existing processes for finding births and identifying duplicates have evolved into a continuous frame update process with two components: an undercoverage component called the Farm Register Update, and an overcoverage component called the Farm Register Administrative Clean-up. The former has a quarterly sample size of about 4,000 while the latter has a quarterly sample size of about 1,000. The combined sample sizes will increase as the Census approaches in order to ensure the best possible coverage just before the Census mail-out.

3.2.1 Undercoverage Component

The goal of the undercoverage component is obviously to increase the coverage of the FR. The main target population is potential new farms identified through tax records and external lists from the provinces and commodity groups. Although external lists were used to help improve coverage of the FR before the 2006 Census, their role in improving the coverage of the FR has been greatly expanded in preparation for the 2011 Census.

Each list is matched to the FR to determine which records from that list are already on the FR. Those records that do not match the FR form the population of potential new farms to be sampled. The processing of external lists is an ongoing process. Each quarter, a sample of the available lists is selected to determine if the lists have enough value to warrant contacting all of their units. In subsequent quarters, the remaining units on the sampled lists are contacted if it is determined that the list has a high probability of identifying active farm operations. It is possible that there could be hundreds of external lists available, so given that it may not be possible to contact all units on all lists, we do not want to spend resources contacting units on lists where only a very small percentage end up being true farms.

The questionnaire is very short, collecting only administrative data and a few key stratification variables. The collection is done via mail-out/mail-back, with a telephone follow-up for non-respondents. After collection, the completed units are matched to the FR a second time using the updated administrative information that was collected. If the unit still does not match, it is then birthed to the FR.

Existing farms on the FR with partially missing or invalid addresses or telephone numbers can also be targeted in order to update the information in advance of the Census. In addition, farms that are not contacted by any surveys during the intercensal period may also be contacted to confirm their status and update their administrative information. This would provide the added benefit of being able to assess the quality of the address fields on the FR.

3.2.2 Overcoverage Component

The objective of the overcoverage component is to identify and remove duplicate records on the FR. Potential duplicates are identified through survey feedback as well as by flagging units on the FR that have operators in common. Collection is done via telephone interview: an interviewer contacts the operator and asks a series of questions in order to determine how many distinct farms that operator is responsible for. In order to keep the interview brief, only administrative data is collected. If it is determined that there is indeed a duplicate farm on the FR, it is removed. If it is determined that the operator does actually operate more than one farm, then this information is saved in order to ensure the operator is not contacted again and asked the same questions.

The knowledge that an operator has multiple farms can help with many other Census processes. The intention is to contact the entire population of potential duplicates. At the beginning of 2009 there were approximately 8,000 potential duplicates on the FR that needed to be contacted, with an additional 1,000 per year anticipated. Improved quality is not the only benefit of removing duplicate records from the FR - cost savings can be realized by reducing census follow-up and unduplication efforts. There are also cost savings to be realized by the on-going survey programs.

3.3 Farm Register Snapshot

Just before census collection, a snapshot of the FR will be taken. This will be used to create the mail-out list and to monitor the progress of collection. All in-business farms on the FR will be included in the snapshot in addition to some out-of-business farms (it is possible that they have resumed farming). However, out-of-business farms with operators that are linked to an in-business farm are not included because there is always the concern of creating duplicates when mailing multiple questionnaires to the same operator.

4. Coverage Activities in 2011

Now that the main pre-Census activities related to coverage have been discussed, the focus of the paper will shift to three of the key coverage activities that will occur during Census processing, after the mail-out in May 2011. These key activities are: the Missing Farms Follow-up, Whole Record Imputation and the Coverage Evaluation.

4.1 Missing Farms Follow-up

The Missing Farms Follow-up (MFFU) targets two distinct populations from which we expected to receive a completed Census of Agriculture questionnaire but have not: existing farms on the FR snapshot and potential farms identified by the Census of Population.

As the Census of Agriculture records come in, they are matched to the FR snapshot. At the end of the process, records on the FR snapshot that do not match to a Census of Agriculture record are targeted by MFFU.

In addition to the FR snapshot, the Census of Population question asking “do you operate a farm operation?” plays a very important role in maintaining coverage as a tool to identify potential farms. Households that indicate on the Census of Population questionnaire that they operate a farm are matched to both the Census of Agriculture base and the FR snapshot. Those that do not match to either base are also targeted by MFFU.

MFFU responses are collected via a computer-assisted telephone interview (CATI). The objective is to obtain a completed Census of Agriculture questionnaire. However, due to time and budgetary constraints, it isn't possible to follow up every record so only the largest existing farms and the potential farms most likely to be true farms are followed up.

A sample of farms not followed up by MFFU is contacted by the Coverage Evaluation Survey which will be explained in section 4.3.

It should be noted that some of the farms targeted by MFFU are not actually 'missing', it might just be that we have been unsuccessful in matching to their returned questionnaire, or that they are no longer farming – in which case they are out-of-scope and not missing.

4.1.1 Missing Farms Follow-up Prioritisation Strategy

The prioritisation strategy differs depending on whether the record is an existing farm on the FR, or a potential farm identified by the Census of Population.

For the existing farms, all of the work is done before the Census. There are many different data sources available to determine the importance of a record to a given commodity or geography, including historical Census data, recent survey data, tax data and external lists. The strategy is to ensure a record is given a high priority if it has an indication of importance coming from any data source. For example, if a record had a large gross farm income according to historical census data, but a small value according to recent tax data, it would still be given a high priority. The result of this strategy is that all of the important farms should be selected for follow-up. Since most records are followed-up anyway, it is not a problem if some less important farms receive a high priority – what must be avoided is important farms receiving a low priority. The prioritisation is done in advance of the Census, and therefore once the Census is up and running it is very easy to assess the return rates of the high priority records.

There is obviously no information available for the potential farms identified by the Census of Population in advance of the Census of Agriculture, and there is very little information available during Census of Agriculture processing. Complicating things is the fact that there are many false positives: respondents who indicate that they operate a farm that are not truly farm operators. The prioritisation strategy involves ranking these records based on the likelihood that that they are a true farm. Studies have shown that if the records can be linked to a tax return with gross farm income, that there is a very high likelihood that it is a farm. Other strong predictors are age, gender, and the number of farms in the geographic area.

4.2 Whole Record Imputation

When a respondent is known to operate an agricultural operation, but refuses to complete a questionnaire, Whole Record Imputation is used to account for its contribution and maintain farm coverage. The imputation method used is nearest neighbour donor imputation. Historical census data are used as matching variables to find the nearest neighbour. When a donor is found, the current year's data of the donor record is copied to the recipient record.

In addition to refusals, the coverage of other types of records is accounted for through Whole Record Imputation. Before the selection of farms for the MFFU, some existing

farms on the FR that have not completed a questionnaire can be excluded from follow-up and instead sent directly to imputation if they: a) have a very low priority, b) have recently responded to an agriculture survey, and c) have been linked to a Census of Population questionnaire indicating that they operate a farm. The benefit of this approach is that follow-up resources can be better allocated to larger, more important records or those records for which we are unsure of their operating status.

After the MFFU, there are still a number of unresolved cases. For existing farms on the FR, all refusals and non-respondents with an indication that they are active either from the Census of Population or from having responded to a recent survey are imputed – this is a potential source of overcoverage. Conversely, non-respondents without an indication of activity are not imputed – this is a potential source of undercoverage. This overcoverage and undercoverage is measured by the Coverage Evaluation discussed in section 4.3.

The strategy of imputing non-responding and refusal records with an indication of activity was done in 2006, and survey results since the Census were used to validate the strategy and determine if it should be retained for 2011. Of the records that were imputed during the 2006 Census and subsequently responded to a survey, 86.9% were still active on Jan 1, 2009.

4.3 Coverage Evaluation

The Coverage Evaluation measures the undercoverage coming from active farms that have not completed a Census of Agriculture questionnaire and that were either a) not followed up by MFFU or b) followed up by MFFU but not resolved during collection. The former consists of very small existing farms on the FR or potential farms from Census of Population that have a low probability of being true farms. The latter can consist of farms of any size, although additional measures are taken to ensure that all of the largest, most important farms are accounted for. The final coverage estimates are calculated by combining the contributions of the two sources and making some necessary adjustments.

4.3.1 Farms not followed up by MFFU

As explained earlier, the MFFU does not select all farms due to budgetary and time constraints. The lowest priority farms are not selected for the MFFU, but a sample of them are selected and contacted by the Coverage Evaluation Survey (CES) in order to calculate an estimate of their contribution. The population is stratified by geography and type of farm.

In previous censuses the CES used a separate, stand-alone CATI application for data collection. The questionnaire was brief and only a handful of key variables such as total land area and total sales were collected – completing a Census of Agriculture questionnaire was not the goal. With the exception of some very large farms discovered during collection, the units were not added to the census base due to the fact that it was a laborious manual process. This ad-hoc strategy of adding only some units to the base also complicated the calculation of the undercoverage estimates.

For 2011, the same application will be used for both MFFU and the CES. This means that all farms responding to the CES will be added to the census base, thus improving coverage. Their weighted responses will still be used to measure the undercoverage of

this population, but their weights will be adjusted so that their own contribution is removed. For example, assume that there are 100 units in the CES population and that 10 units are sampled. For the purposes of measuring the undercoverage of the population, the weight of each of the sampled units would be $100/10 - 1 = 9$. This does not take into account any non-response adjustments that may be necessary.

4.3.2 Farms followed up by MFFU but not resolved during collection

The undercoverage component coming from farms selected by MFFU but not resolved during collection is more complicated to calculate. As mentioned earlier, the decision to impute all non-resolved MFFU farms with an indication of activity and to not impute any non-resolved MFFU farms without an indication of activity results in both overcoverage and undercoverage.

The overcoverage is calculated by estimating the percentage of inactive and out-of-scope farms that were whole record imputed because they had an indication of activity, using the results from the resolved cases. For example, if 10% of resolved farms with an indication of activity are inactive or out-of-scope, then it is assumed that 10% of the contribution of whole record imputed farms is overcoverage.

To calculate the undercoverage resulting from the farms with no indication of activity, the same assumptions are made based on the results of the resolved cases. However, since these farms are not imputed, an imputation simulation must be run (mimicking the true imputation process) in order to estimate their contribution to the undercoverage.

The net result of this component should always be undercoverage, meaning that we should never ‘over-impute’.

4.3.3 Calculation of the final coverage estimates

The calculation of the final estimates is not as simple as just combining the components described in the previous two sections. During Data Validation, the step where subject matter analysts review the Census data at both aggregate levels and the individual questionnaire level, it may be determined that there are important farms missing that need to be contacted. If a questionnaire is completed for a farm such as this, we need to ensure that it is not already being accounted for in the coverage estimates. In fact, a check of the census base is needed for all records that were in the CES population but not selected. If after the sample selection they were received by any method (i.e. late mail return) then their contribution must be subtracted from the coverage estimates to avoid overstating the undercoverage estimates.

The final undercoverage estimates are produced and delivered to Data Validation before certification of the results, so that the analysts can use them to help explain any difference with other referential sources, or with previous censuses. In 2006, the estimated undercoverage rate was 3.4% at the national level in terms of number of farms, and less than 2% in terms of total farm area and gross farm receipts.

5. Summary

The Census of Agriculture faces a significant change in the collection methodology for 2011. However, the experiences of the 2006 Census of Agriculture leave us well positioned to react to the challenge.

The work done in advance of the Census to improve the Farm Register will be essential to maintaining coverage. During the Census, a very good non-respondent follow-up prioritisation strategy is key to ensure no important farms are missed, and thus coverage is maintained. A good strategy to evaluate the coverage is also essential to success.

It is important to keep in mind that good Census Coverage is not just important to the Census, it is crucial for the entire Agriculture Statistics Program because it will form the base for the surveys over the subsequent 5 years.

Acknowledgements

The author would like to thank Pierre-Olivier Julien for all his work on the improvements to the content and coverage of the Farm Register. The author would also like to thank Steven Danford and Peter Xiao for their comments on the content of this paper.

References

- Levesque I., MacNeil D. and Poirier C. (2008). *Design of the 2011 Census of Agriculture*. Internal document presented to the Advisory Committee on Statistical Methods, Ottawa, Canada: Statistics Canada.
- Lachance, M. (2005). *Design La couverture du Recensement de l'agriculture de 2006*. Article présenté au Colloque francophone sur les sondages 2005, Québec, Canada.
- Lessard, M. (2005). *Évaluation, amélioration de la qualité du Registre des fermes et les bénéfices retirés*. Article présenté au Colloque francophone sur les sondages 2005, Québec, Canada.