Use of Respondent Burden Measurements at NASS

Wendy Scholetzky
USDA, National Agricultural Statistics Service,
1400 Independence Ave, S.W., Washington, DC 20250

KEY WORDS: burden measurements; respondent burden; annual cycle

The mission of the National Agricultural Statistics Service (NASS), a statistical agency under the United States Department of Agriculture (USDA), is to provide timely, accurate, and useful statistics in service to U.S. agriculture. Each year, NASS conducts hundreds of surveys covering many facets of U.S. agriculture. For most surveys, the agency uses a list of known farm operators and agribusinesses to define the sampled populations. The list frame contains names, addresses, telephone numbers, and other identifying information associated with the agricultural operation including the principal operator(s) and any partners. The list frame also contains quantitative information for agricultural commodities associated with the operation, which is referred to as control data. The control data are used to identify the populations of interest and select representative samples for surveys conducted by the agency.

Overview of the Annual Cycle

On an annual basis, NASS performs major updates to control data on the list frame and selects new samples for surveys. Control data on the list frame remain unchanged (i.e., frozen) while defining the sampled population and selecting the sample for each survey. In the spring, samples are selected for the major national surveys as well as most state surveys; samples for other surveys are selected at a different time during the year for various reasons. This paper will refer to this process as the annual cycle. There are three major steps in the annual cycle: updating the control data, defining the sampled population, and selecting the sample. At each step of the process, personnel within Headquarters and State Statistical Offices have specific responsibilities to ensure that each step is performed properly.

The first step of the annual cycle is to update control data on the list frame with new survey or administrative data obtained within the past year. The source and year of the updated control data are also stored on the list frame to serve as a measure of data quality in terms of origin and age. When updating the control data, different methodologies are used depending on the agricultural item. The resulting control data provide a general profile of agricultural commodities associated with the operation.

The second step of the process is to define the sampled populations for surveys in the annual cycle. The number of sampled populations varies within each state since all surveys are not conducted in all states. Using control data associated with the objective(s) of the survey, operations are included or excluded from the sampled population for each survey. As an example, only operations with positive control data for hogs are included in the sampled population for the Quarterly Hog Survey.

The third step of the process is to select the samples for the entire survey year for surveys in the annual cycle. The sample designs differ depending on the survey and are specified based on the objective(s) of the survey. Typically, stratified sampling is performed and the strata are formed based on size of the agricultural item(s) related to the survey. After all samples are selected, specific codes are assigned on the list frame to identify the records to be included in each survey.

In addition to the list frame, NASS also maintains an area frame which consists of the entire land area for the contiguous United States divided into segments. The agency uses the area frame to define sampled populations and to measure incompleteness in the list frame. For many surveys, the estimates are produced using a multiple frame sample design (i.e., the estimate from the list frame is combined with the estimate from the non-overlapping component of the area frame). However, the sample selection process for the area frame is not part of the annual cycle.

Impetus for Using Burden Measurements

For many surveys, the sampled populations are relatively small and highly skewed to the right (the mean is much larger than the median). In addition, operations in the distribution’s right tail are usually assigned a probability of selection that is close to, if not equal to, one. Given the population distribution and sampling methodology, there are quite a few records that are selected for multiple surveys; typically, these are large operations or operations with diverse agriculture. NASS is extremely concerned about the respondent burden imposed upon these operations. When possible, the State

Statistical Offices coordinate data collection activities for surveys with similar reference dates in an attempt to reduce respondent burden; however, additional strategies are necessary.

Because the list frame samples are selected for a majority of the surveys during the annual cycle, a measurement of respondent burden can be computed for the upcoming survey year (which runs from June through May) and this information can be utilized prior to the data collection phase. In fact, NASS calculates several burden measurements for an operation since a respondent's perception of burden can vary. These burden measurements consider the number of surveys, the number of contacts, and the total amount of time required by the operation to complete all of the survey questionnaires. Using these measurements, highly burdened operations are identified and special handling arrangements can be implemented for these potential respondents.

This paper discusses calculating burden measurements for records sampled for surveys in the annual cycle as well as records sampled for certain surveys outside of the annual cycle. Note that area frame records are excluded because operations in the area frame are not identified until the data collection phase is conducted. Thus, burden measurements are only calculated for records sampled from the list frame since the purpose of this endeavor is to utilize this information prior to the data collection phase.

Defining the Burden Measurements

Since there are different ways that these burden measurements can be defined, a brief description of each measurement is provided. The *number of surveys* is the total number of survey questionnaires that will be administered to the operation. A survey is counted several times when it is conducted multiple times during the survey year. For example, although it is referred to as one survey, the Quarterly Agricultural Survey is counted four times because it is conducted on a quarterly basis (June, September, December, and March).

The *number of contacts* is the total number of times that an operation will be contacted after the State Statistical Offices coordinate data collection activities for the surveys. This burden measurement is less than or equal to the *number of surveys*. When it is possible to administer multiple questionnaires during the same contact, the *number of contacts* is less than the *number of surveys*. For example, since the January Cattle Survey is typically coordinated with the January Milk Production Survey, an operation that is in sample for both surveys (and only these two surveys) would have the *number of surveys* equal to two and the *number of contacts* equal to one.

The *total time required* is the total amount of time in minutes required by the operation to complete all of the survey questionnaires that it will be administered. The amount of time to complete a survey questionnaire is defined as the time specified in the Office of Management and Budget (OMB) clearance package or the time specified in the respondent burden statement if OMB clearance is not necessary. The *total time required* provides a measurement that is independent of the *number of surveys*. For example, an operation with a *total time required* of 60 minutes could be administered a 5-minute questionnaire on a monthly basis or a 60-minute questionnaire once. In addition, the *total time required* is independent of the *number of contacts* since it remains the same when multiple questionnaires are administered during the same contact.

System Used to Perform the Calculations

The agency's original plan was to incorporate the burden measurements as part of a new respondent burden tracking and management system, but the system has not yet been developed due to insufficient staff resources. In the meantime, the agency was able to make minor modifications to an existing system in order to produce and store the burden measurements. The existing system, which is referred to as the Enhanced List Maintenance Operations (ELMO), is a relational database that contains the list frame for the entire United States. By definition, tables in this relational database can be linked to each other by using unique keys.

The burden measurements are calculated using two tables in ELMO. The markedSample table is one of the two tables. As previously mentioned, samples are selected for the major national and most state surveys in the spring, and specific codes are assigned on the list frame to identify the records to be included in each survey. The results of this process are stored in the markedSample table. This table existed prior to the minor modifications referenced above. The unique keys used in linking this table to the other table are *periodId* (numeric code associated with the annual cycle), *stateId* (state identification code, which uses FIPS coding), *surveyId* (survey identification code), and *month* (numeric month in which the survey is conducted). Although not used in the linking process, *poId* (person/operation identification number) is a unique key that identifies a particular record within a state.

Excerpt of markedSample table in ELMO

periodId	*stateId*	*surveyed*	*month*	*poId*
9	17	661	6	987650010
9	17	661	6	987650020
9	17	661	6	987650030
9	17	661	6	987650040

A row in the markedSample table represents a sampled record. Due to the design of this table, the number of sampled records can be determined by counting the number of rows where *periodId*, *stateId*, *surveyId*, and *month* equal values associated with the survey of interest. For example, the sample size in the state of Illinois for the June Quarterly Agricultural Survey is the number of rows with *periodId* = 9, *stateId* =17, *surveyId* = 661 and *month* = 6; the national sample size for the June Quarterly Agricultural is the number of rows with *periodId* = 9, *surveyId* = 661 and *month* = 6. For some surveys, the month in which the survey is conducted may be different than the numeric month specified in the markedSample table. This occurs for surveys conducted every month using the same sample. In this instance, the sampled records appear only once in the markedSample table because the same rows are utilized each month even though the calendar month changes. For example, the sample size in the state of Missouri for the January Milk Production Survey is the number of rows with *periodId* = 8, *stateId* = 29, *surveyId* = 615, and *month* = 1. These same parameters are also used to obtain the sample size in the state of Missouri for the February Milk Production Survey (i.e., *month* = 1 is used rather than *month* = 2).

The surveyMaint table is the other table used in calculating the burden measurements. The parameters used to calculate the *number of surveys*, *number of contacts*, and *total time required* are stored in the surveyMaint table. This table was added to ELMO as part of the minor modifications referenced above. The unique keys in this table are *year* (calendar year associated with the current annual cycle), *periodId*, *stateId*, *surveyId*, and *month,* where the last four variables are used in linking this table to the markedSample table.

A row in the surveyMaint table represents either a survey or a survey contact, where a survey contact is defined as each time an operation is administered a questionnaire for the same survey. The rows in this table must be consistent with the way the survey is stored in the markedSample table. For example, the Quarterly Agricultural Survey is one survey but four survey contacts because it is conducted on a quarterly basis, and the sampled records for this survey appear for every month (June, September, December, and March) in the markedSample table. For this survey, the surveyMaint table contains four rows with the same values for *periodId*, *stateId*, and *surveyId* but different values for *month*, so each row in the table represents a survey contact. Therefore, the surveyMaint table has four rows in the state of Illinois for the Quarterly Agricultural Survey where each row has the values of *periodId* = 9, *stateId* =17, and *surveyId* = 661 but the values for *month* = 6, 9, 12, and 3. In comparison, in some states, the Milk Production Survey is one survey but twelve survey contacts because it is conducted on a monthly basis. As previously mentioned, the sampled records for this survey appear only once in the markedSample table. For this survey, the surveyMaint table contains only one row, so this row represents a survey (rather than a survey contact). Therefore, the surveyMaint table has one row in the state of Missouri for the Milk Production Survey where the row has *periodId* = 8, *stateId* = 29, *surveyId* = 615 and *month* = 1.

The parameters used in calculating the burden measurements are stored as three variables in the surveyMaint table: *surveyLength*, *factor*, and *coordCode*. These parameters are assigned based on the *periodId*, *stateId*, *surveyId*, and *month* and reflect whether the row represents a survey or a survey contact. The *surveyLength* is the amount of time required to complete the survey questionnaire and is specified in minutes. The amount of time is equal to the time stated in the respondent burden statement, which is also equal to the time specified in the OMB clearance package. The *surveyLength* is typically between ten and twenty minutes, but can range from two minutes to ninety minutes for various surveys. The *factor* is the number of times that the survey

Excerpt of surveyMaint table in ELMO

year	*periodId*	*stateId*	*surveyId*	*month*	*surveyLength*	*factor*	*coordCode*
2004	9	17	661	6	20	1	0
2004	9	17	661	9	20	1	0
2004	9	17	661	12	20	1	0
2004	9	17	661	3	20	1	0
2004	8	29	615	1	6	12	1

questionnaire is administered to the sampled records. This parameter is assigned based on how the survey is reflected in the markedSample table and whether the row represents a survey or a survey contact in the surveyMaint table. The *factor* is typically equal to one. The *coordCode* is used to identify surveys or survey contacts where the State Statistical Office typically coordinates data collection activities. When the value is zero for this parameter, the data collection activities for the survey or survey contact are not coordinated with any other survey. When the value is a positive number for this parameter, the data collection activities for the survey or survey contact are coordinated with other rows in the table that have the same *coordCode* value. In this case, multiple questionnaires are administered during the same contact when a record is sampled for more than one of the coordinated surveys.

To give a better understanding of the parameters stored in *surveyLength*, *factor*, and *coordCode*, the parameter values for the examples given earlier are discussed here. In the surveyMaint table, the parameter values for all four rows in the state of Illinois are as follows: the *surveyLength* is twenty, the *factor* is one, and the *coordCode* is equal to zero. The *surveyLength* of twenty means it takes twenty minutes to complete the survey questionnaire. The *factor* is one because each row represents a survey contact where the questionnaire is administered once. The *coordCode* of zero indicates the data collection activities for these survey contacts are not coordinated with any other surveys. In comparison, the parameter values for the one row in the state of Missouri for the Milk Production Survey are as follows: the *surveyLength* is six, the *factor* is twelve, and the *coordCode* is equal to one. The *surveyLength* of six means it takes six minutes to complete the survey questionnaire. The *factor* is twelve because the row represents a survey where the questionnaire is administered every month. The *coordCode* of one indicates the data collection activities for the Milk Production Survey are coordinated with other surveys that also have a *coordCode* value of one.

Calculating the Burden Measurements

The burden measurements are calculated in the spring, after the surveyMaint table and markedSample table are updated. During the annual cycle, personnel in the State Statistical Offices review all rows in the surveyMaint table where the *year* is equal to the calendar year associated with the current annual cycle. Once all steps involved in the annual cycle are complete, personnel in Headquarters store the samples selected during the annual cycle in the markedSample table. The surveyMaint table (a survey-level table) and the markedSample table (a sampled record-level table) are then linked together using the unique keys - *periodId*, *stateId*, *surveyId*, and *month*. All rows in the surveyMaint table where the *year* is equal to the current calendar year should have at least one matching row in the markedSample table. Prior to calculating the burden measurements, the two tables are preliminarily linked together and personnel in Headquarters review the results from this matching process. This linking procedure is extremely important because the rows in the surveyMaint table must be consistent with how the survey is reflected in the markedSample table in order for a survey or survey contact to be accurately reflected in the burden measurements. After the results are reviewed and any necessary updates are made to the surveyMaint table, the two tables are linked together again and a temporary work table is created in ELMO (a table with survey-level parameters at the sampled record-level). A row in the temporary work table represents a sampled record which has the parameter values for the survey or survey contact that are used in calculating the burden measurements for the sampled record.

Using the temporary work table, the three burden measurements - *number of surveys*, *number of contacts*, and *total time required* - are calculated separately for each sampled record. The temporary work table is sorted by *stateId* and *poId*. The burden measurements are calculated for each sampled record by grouping together rows with the same *stateId* and *poId*. The *number of surveys* is calculated by summing the *factor* values for all rows associated with the same *stateId* and *poId*. The *number of contacts* is calculated by summing the *factor* values for all rows associated with the same *stateId* and *poId* where the *coordCode* value is either zero or a unique positive number. The *total time required* is calculated by multiplying the *surveyLength* by the *factor* for each row and summing these values for all rows associated with the same *stateId* and *poId*. After the calculations are performed for each sampled record, the three burden measurements are stored in the controlData table in ELMO, which is a permanent table containing the control data associated with the operations on the list frame.

To give a better understanding of how the burden measurements are calculated, an example is given here. Suppose that the temporary work table contains the following four rows for a sampled record.

Excerpt of temporary work table in ELMO

stateId	*poId*	*year*	*periodId*	*surveyId*	*month*	*surveyLength*	*factor*	*coordCode*
13	987654320	2004	8	615	1	6	1	1
13	987654320	2004	9	665	1	20	1	1
13	987654320	2004	9	665	7	20	1	0
13	987654320	2004	8	760	12	10	12	0

For this sampled record, the *number of surveys* is fifteen, the *number of contacts* is fourteen, and the *total time required* is one hundred sixty-six minutes. These burden measurements are calculated as follows:

number of surveys = 1 + 1 + 1 + 12 = 15
number of contacts = 1 + 1 + 12 = 14
total time required = (6 x 1) + (20 x 1) + (20 x 1) + (10 x 12) = 166

In this example, note that the *number of contacts* would be the same as the *number of surveys* if all of the *coordCode* values were equal to zero. However, since the *coordCode* values of one indicate that the survey in the first row is coordinated with the survey contact in the second row, the *number of contacts* is one less than the *number of surveys*.

Utilizing the Burden Measurements

When possible, the State Statistical Offices coordinate data collection activities for surveys with similar reference dates in an attempt to reduce respondent burden. Survey coordination minimizes the number of times that a sampled record is contacted. However, additional strategies to reduce respondent burden are necessary. Using the burden measurements, highly burdened operations are identified and special handling arrangements can be implemented for these operations. The maximum annual respondent burden is discussed in NASS' policy and standards memorandum on data collection standards. This memorandum states that the maximum annual respondent burden is different for each state due to the diversity in the survey programs at the state level. However, in general, an operation should not be contacted more than twelve times within a year or an operation should not spend more than two hundred forty minutes completing survey questionnaires within a year. If either of these criteria is violated, personnel within the State Statistical Offices are responsible for developing a burden reduction plan for these operations.

Depending on the situation, there are a number of options available for the State Statistical Offices to use. Some special handling arrangements for highly burdened operations are as follows: informing operations of future multiple survey contacts prior to the data collection phase, negotiating special data collection agreements with operations that are hesitant to participate in all surveys, coordinating data collection activities for surveys with different reference dates, utilizing previously reported survey data for questionnaire items that are static over time, assigning an enumerator to collect the data via personal interview for all surveys, or supplying operations with tokens of appreciation.

In addition to using the burden measurements, NASS practices other strategies in an attempt to reduce respondent burden. Although not comprehensive, these strategies include the following: implementing different sampling methodologies to minimize overlap between various surveys, researching modeling techniques to replace data collection for certain questions or certain operations, utilizing alternative data sources to take the place of conducting the survey, reducing the number of questionnaire items while still satisfying the needs of the agency's estimation or reimbursable programs, or limiting the amount of time required to complete a survey questionnaire to sixty minutes when possible.

Conclusion

NASS is extremely concerned about the respondent burden imposed upon all operations, especially highly burdened operations. In an attempt to reduce respondent burden, burden measurements are used to identify highly burdened operations and special arrangements are made for these operations. The agency is optimistic that this strategy, along with other strategies currently being implemented by NASS, will be successful in reducing respondent burden.