



U.S. Agricultural Statistics Current Disclosure Practices and Future Improvements

James M Harris*

United States Department of Agriculture, National Agricultural Statistics Service, Washington DC,
United States – mark.harris@nass.usda.gov

Christy Meyer

United States Department of Agriculture, National Agricultural Statistics Service, Washington DC,
United States – christy.meyer@nass.usda.gov

The United States Department of Agriculture's National Agricultural Statistics Service (NASS) is the statistical organization that conducts survey and census on various agricultural commodities, economics and environmental factors for U.S. agriculture. NASS publishes over 400 releases each year and for each data cell has a legal obligation to ensure that individually identifiable data is protected including preventing identity disclosure, attribute disclosure, and inferential disclosure. This paper will provide information on these types of disclosure and how NASS identifies primary and secondary suppressions to prevent disclosure, different options NASS uses in publications to protect data, the role of Data Labs and special tabulations. This begins the discussion, the paper will conclude with how NASS is looking at future disclosure improvements and challenges.

Keywords: identity disclosure, attribute disclosure, inferential disclosure, primary suppression, complementary suppression



1. Introduction

The mission of the National Agricultural Statistics Service (NASS), an agency of the United States Department of Agriculture (USDA), is “to provide timely, accurate and useful statistics in service to U.S. agriculture.” Towards this goal, NASS conducts hundreds of surveys every year collecting information on virtually every aspect of agricultural activity. NASS also conducts the quinquennial Census of Agriculture and associated follow-on special studies, including Census of Aquaculture, Census of Horticulture Specialties, Farm and Ranch Irrigation Survey, Organic Survey, On-Farm Energy Production and the Tenure, Ownership and Land Transition Study.

The Census of Agriculture is the leading source of facts and figures about American agriculture and provides a detailed picture of U.S. farms and ranches and the people who operate them. It is the only source of uniform, comprehensive agricultural data for every state and county in the United States. Participation is by every farmer and rancher, regardless of the size or type of operation. The census has been conducted since 1840 and serves as the benchmark for change across time.

All NASS releases resulting from the survey or census conducted are subject to a law (Title 7, U.S. Code, and CIPSEA, Public Law 107-347) that guarantees to all respondents that their individual information will remain confidential. NASS uses the information only for statistical purposes and publishes data only in tabulated totals. The report cannot be used for purposes of taxation, investigation or regulation. The privacy of individual Census records is also protected from disclosure through the Freedom of Information Act. The President’s Commission on Federal Statistics (1971) defined confidentiality as: [Confidential should mean that the dissemination] of data in a manner that would allow public identification of the respondent or would in any way be harmful to him is prohibited and that the data are immune from legal process. This paper is designed to discuss how NASS has handled disclosure within publications, special tabulations and access to confidential information through research agreements.

2. Types of Disclosure

NASS must protect against three types of disclosure; identity disclosure, attribute disclosure, and inferential disclosure. Identity disclosure is releasing any data that allows the discovery of one’s identity, such as name, address, phone number, or SSN. This type of disclosure is prohibited and no information defined as personally identifiable information (PII) is released. In fact, PII information is never kept in the same file or database that houses the census or survey responses. Attribute disclosure is revealing information about someone or some organization, such as income, acres, inventory, or demographic characteristics. Inferential disclosure is revealing data that allows others to infer an individual’s or organization’s characteristics more accurately than would have been possible, often using other information sources.

NASS policy is that every number released to the public will have sufficient measure of uncertainty to ensure attribute or inferential disclosure is not possible. NASS does not consider farm counts or number of operations sensitive data and they are not subject to suppression. Any value failing one or more disclosure rules must be suppressed unless an informed consent has been obtain by the data provider.

NASS employs different dominance rules for the agricultural estimates program and the census program which are documented below. Both programs use a threshold rule of at least three operations in the population. As a general practice, statistical agencies do not reveal the numerical parameters of their dominance rules as an added measure of protection.



2.1 Disclosure Rules Employed by NASS

Agricultural estimates program

The agricultural estimates program uses the (n,k) dominance rule to determine suppressions. Under this rule, a value is suppressed if the n largest operations in the population represent k percent or more of the published total. For example, a (2, 80) dominance rule will identify a suppression when the largest two producers represent 80 percent or more of the total.

The formula is straight forward. If,

(Sum of the n largest operators) / (total to be released)*100 > k

the estimate must be suppressed.

Census program

The census program uses the p-percent rule to determine dominance. Under this rule, the sum of the weighted values of the remaining operations after removing the two largest operations must be more than p-percent of the largest operation. For example, a 20 percent rule will identify a suppression when the remainder is less than or equal to 20 percent of the largest operation.

A remainder is calculated as:

Remainder = Total (weighted) to be released
 - unweighted value of the largest operation
 - unweighted value of the second largest operation

If,

Remainder < unweighted value of the largest operation * (p/100)

the total must be suppressed.

These dominance rules are used to identify primary suppressions as any value to be released that directly fails the disclosure rule. The rules can be easily programmed and primary suppressions are quickly identified.

2.2 Complementary Suppressions

If a data cell is a primary suppression in a table that contains additive subtotals, it is possible to derive (infer) the value of that cell. In order to protect the primary suppression, other cells need to be suppressed and are referred to as complementary suppressions. Sometimes, more than one primary suppression occurs among additive items and they complement each other. However, there will be numerous instances when a value that passes the rules must be sacrificed to protect a primary suppression.

3. Execution of Disclosure Rules

The agricultural estimates program collects data from over 400 surveys each year. For many of these surveys the incidences of primary suppressions is minimal. The commodity statisticians in each of NASS' Regional Field Offices (RFO) monitor the survey and identify primary suppressions. When the individual RFO data is combined in Headquarters (HQ), complementary suppressions are determined. NASS policy is to minimize the number of suppressions without reducing the protection of a suppressed item.

The size and complexity of the census requires a system of computer programs. The system uses network/minimum cost flow methodology to select complements. When the data for the complements are suppressed, no data user should be able to derive the value of any primary suppression. The system must have provisions for statistician intervention to address aberrations in the selections. If the algorithm



selects a crucial estimate for suppression as a complement, an alternate complement may be used in place of the original selection.

4. Publications

Different methods of displaying suppressed data may be used in NASS releases. Each primary and complementary suppression may be reported by explicitly replacing the cell value with a “D”. Cell suppressions may be eliminated by aggregating to a higher level, such as counties to Agricultural Statistics District, or by defining an “all other” group, like all other crops. Regardless of the method chosen, care must be taken to ensure that the suppressed data cannot be reconstructed from public materials.

There are instances of indirect disclosure, when the data represent many operations and no operation is dominant, but the total itself represents a separate individual or business. NASS encounters this with contractors and contractees. A total may be based on numerous contractees and pass all disclosure checks, but, if nearly all of the contractees are producing for the same contractor, release of the total is equivalent to disclosing the contractor’s attribute. For the agricultural estimates program, NASS policy is to suppress these values even though the data were not directly collected from the contractor. For the census of agriculture each farm is considered independently and thus the contractee/contractor relationship is not considered.

Indirect disclosure may also occur when totals are presented in a sorted order such as rankings and top commodities. The values appearing immediately before and after a suppression define a range of values that may be less than the protection required for the suppressed value. When this happens, NASS policy is to protect the original suppression by excluding it from the tables and noting the situation in accompanying text or footnotes. If the range of values exceeds the protection value, the suppression may be left in the proper position.

Data users may find suppressions that are significant obstacles to the usefulness of the data and ask NASS to reverse suppression decisions. In the agricultural estimates program, NASS may release a protected item with the permission of the operation(s) at risk or informed consent. Permission must be in writing and signed by an authorized officer of the operation.

As a matter of policy, informed consent is not allowed for the Census of Agriculture, Volume 1. Table content is too interrelated and complex that exceptions to suppression routines can result in accidental disclosure. However, informed consent may be applied to census follow-on surveys.

5. Data Labs and Special Tabulations

NASS operates data labs to address custom needs of data users whose requirements are not met by an existing data series. A full-time Headquarters Datalab is operated by the Summary, Estimation, and Disclosure Methodology Branch, Methodology Division. The HQ Datalab services special tabulation requests and operates workstations for customers with approved access to micro-data. A data lab may be opened in any NASS Regional Field Office to accommodate local customers with approved access. NASS also enters into a Memorandum of Understanding with other Federal agencies to establish remote data labs. Under U.S. Code, Title 44 §3501 (CIPSEA) NASS has the authority to establish these data labs. U.S. Office of Management and Budget (OMB) has encouraged U.S. federal agencies to share information and data to the fullest extent of confidentiality restrictions. Remote labs are operated at the cooperator’s location under the same physical and cyber security rules as NASS data labs. The cooperator must provide an approved supervisor to administer access authorities and disclosure requirements as if it were in a NASS facility.



NASS participates in a federally-sanctioned, privately operated data enclave. Federal agencies may make selected micro datasets available to individuals who have obtained agency approval. Customers access the data at their location through a web-based interface. NASS retains the services of researchers to work on special projects or research topics of interest to the Agency. Access to unpublished NASS data is often required to fulfill both parties research needs.

Special tabulations from census records is the primary mechanism for filling special data requests. All special tabulations are reviewed for disclosure following NASS' established disclosure rules. The same disclosure programs used in the original census disclosure checking are used for the special tabulations to ensure consistent application of disclosure suppressions. These special tabulations often request data to be tabulated at a smaller geographical region than previously published or to provide more information about a particular sector of agriculture. The challenge with these types of requests is to compare the data being requested to the data that has already been published. We do not want to inadvertently allow a suppressed cell in the original publication to be discerned by examining the data from the special tabulation.

6. Future Improvements and Conclusions

The disclosure checks for census programs has been automated and allows for a degree of adjustments to ensure complementary suppressions minimize the suppression of key values. The establishment of these checks for each publication and special tabulation requires an investment of time and training. The art of working with the complementary suppressions is acquired over many years of experience. NASS is currently working towards expanding the number of staff that are familiar with the programs and increase their experience in establishing the checks for publications.

For survey programs NASS is examining current systems and discussing options for potentially automating the primary disclosure checks for more surveys. NASS will need to take into consideration the limited number of complementary suppressions to determine if automated or manual methods of determining complementary suppressions is the best practice for the agency.

On occasion, NASS has projects that span the survey and census programs. There are usually two or more different publications that are released at different times. These pose a challenge for NASS to determine which disclosure rule to use and how to coordinate the disclosure suppressions. Currently the coordination is handled on an individual project basis. In the future, NASS would like to provide a more standard system for addressing these challenges.

The structure of U.S. farms has been changing from what has historically been observed. More farms are considered large farms specializing in a limited number of commodities. As this trend continues, the risks of disclosing an individual operation's information becomes greater. NASS will continue to examine and improve our methods to mitigate this risk and protect each operation's information.



References

NASS (2012). Policy and Standards Memorandum No. PSM-CS-2 Access to Unpublished Data. April 2, 2010.

NASS (2012). Policy and Standards Memorandum No. PSM-CS-4 Standard for Suppressing Data Due to Confidentiality. April 2, 2010.

OMB (2002).

https://www.whitehouse.gov/sites/default/files/omb/assets/omb/inforeg/cipsea/cipsea_statute.pdf.

Office of Management and Budget website, Accessed on April 24, 2015.

OMB (2006).

https://www.whitehouse.gov/sites/default/files/omb/assets/omb/inforeg/proposed_cispea_guidance.pdf

. Office of Management and Budget website, Accessed on April 24, 2015.

White House (2012). https://www.whitehouse.gov/sites/default/files/docs/2012sharingstrategy_1.pdf.

White House/President of the United States, Accessed on April 24, 2015.