Challenges Facing the Disclosure Review Board at Census

Billy Wisniewski[1], Holly Newman[1]

[1]U.S. Census Bureau, 4600 Silver Hill Rd, Washington, DC 20233

Abstract

Under Title 13 of the U.S. Code, the U.S. Census Bureau is required to release information from its data collection to the public, while also promising to protect the confidentiality of individual respondents. Prior to releasing any data product to the public, we must apply various disclosure limitation procedures and review these in ensure that the data are adequately protected. In most cases, products for public release will also need review and approval from the Census Bureau's Disclosure Review Board (DRB). Many individuals know what the DRB is, and how they operate, but they are not aware of the challenges the DRB faces when approving certain data products. This paper looks to fill that void. We provide detailed examples and discuss many of these challenges. In doing so, we give an inside look at how the DRB overcomes these issues to ensure that the products we approve are safe for public release

Key Words: Disclosure Review Board, Disclosure Avoidance Officer, Confidentiality, Disclosure Protection

1. Introduction

Under Title 13 of the U.S. Code, the U.S. Census Bureau is required to release information from its data collection to the public, while also promising to protect the confidentiality of individual respondents. Prior to releasing any data product to the public, we must apply various disclosure limitation procedures and review these to ensure that the data are adequately protected. In most cases, products for public release will also need review and approval from the Census Bureau's Disclosure Review Board (DRB). This paper will flow as follows. First, we will briefly discuss who the DRB is, and how they operate. Next, we discuss the Disclosure Avoidance Officer (DAO), their role, and why they are a key player in the review and release of data products. Moving along, we will give some in-depth examples of past requests and discuss the challenges that the DRB faces in approving difficult data products. While most requests are standard, some can be very difficult. For example, what occurs when a submitter wants to release tabular data for every county in the United States? What is the process when a submitter wants to include every possible variable on a Public Use Microdata File (PUF)? This section will serve as the bulk of the paper. Finally, we wrap up with a summary of highlights.

2. The Disclosure Review Board

The U.S. Census Bureau's Disclosure Review Board (DRB) has a variety of roles to play in serving their mission. That is, to support the Data Stewardship Executive Policy Committee (DSEP) in its efforts to ensure that the Census Bureau protects all Title 13 and Title 26 respondent confidentiality of publicly released data products. This leads to a variety of roles, which include the following:

To establish and review official Census Bureau disclosure avoidance policies and general rules for the public release of data products.

To review and approve proposed disclosure avoidance procedures for the public release of data products.

To communicate the disclosure avoidance policies to others.

To coordinate research on the disclosure risk of data products, as well as to determine how effective the disclosure avoidance techniques are in protecting such products.

If necessary, revise and update the Census Bureau's disclosure avoidance policies and methodologies.

The DRB consists of nine members. Leading the DRB is the Chair, who is in the Research and Methodology Directorate. Other members represent the Census Bureau's decennial/ACS, demographic and economic programs, as well as the research and policy areas.

The DRB coordinator also plays an important role. This individual is responsible for making sure all DRB members and DAOs are compliant and up to date with the annual Data Stewardship Awareness Training and Title 26 Federal Tax Information (FTI) Mandatory training. The coordinator also makes sure that the meetings are held on a Title 26 secure floor and that all visitors who do not have Title 26 FTI training are escorted to and from the meeting. This includes signing these visitors in and out of a Title 26 log. In addition, the DRB coordinator organizes all DRB requests, making sure that files are properly labeled (e.g., Title 13 or Title 26 markings). At each meeting, the DRB coordinator takes detailed minutes. After the meetings, he/she will send official approval/denial memos via email to the various requesters. The minutes and memos are also made available to view on the CDAR secure shared drive and the CDAR intranet page under the Disclosure Review Board side bar. If the DRB coordinator receives requests containing Title 13 or Title 26 data, the requester will send the information to the DRB Coordinator through Accellion Secure Fire Transfer. Once the DRB coordinator receives the encrypted file(s) over email, he/she will upload them to the DRB directory for the DRB to review.

In order for a meeting to come to session, a quorum of two-thirds of the members must be present. Quite often, other individuals attend the meeting, such as the DAO and guests from various program areas. The DAO and guests are encouraged to attend in case any of the DRB members have questions about the specific request. During a typical meeting, the DRB will go through each request, item by item, and vote on either to approve it as is, approve it with modifications, or deny it. For any decision, six of nine

members must agree. Submitters unhappy with a decision may appeal the decision to the Data Stewardship Executive Policy Committee (DSEP), which consists of a subset of Census Bureau Associate Directors. Although rare, if DSEP reviews a request, their ruling is final. Finally, some requests can be voted on through email. One such example would be for a repeat request where the only difference is year(s) of data used (U.S. Census Bureau, 2013d, 2016).

3. The Disclosure Avoidance Officer

The Disclosure Avoidance Officer (DAO) plays a major role in ensuring the protection of respondent confidentiality for publicly released data products. Internal Census Bureau staff, external researchers who work with internal staff, or researchers with approved projects within the Federal Statistical Research Data Centers (FSRDCs) produce these products. In essence, the DAOs are the individuals who assist the submitters that must go through the DRB review process, as well as perform the final review for data products before sending them to the DRB or into the public domain (after DRB approval). DAOs ensure that the researcher(s) applies specific disclosure rules to their output. Due to the work that a DAO performs, many requests do not need full DRB review, and thus they lessen the burden for the DRB. In this section, I will explain how valuable the DAO is, as well as give examples of some of the supporting documents they use when needing to go to the full DRB. These documents are very important in assisting the DRB to make an informed decision.

Each Census Bureau division/program area that releases data to the public appoints an individual to the position of DAO. A DAO will assist the submitters to go through the full review process. To start, they will organize a submitter's request and contact the DRB chair and DRB coordinator to request a review at the next DRB meeting. Meetings are generally held every Monday. In submitting a request, the DAO will make sure that the following are included (if applicable): a memo to the chair of the DRB explaining the request, a completed DRB checklist, the questionnaire from the survey or census, a list of variables of interest, a record layout for requested microdata, table outlines/shells for requested tabular data, and often, some cross-tabulations of the variables of interest. All of these documents are important in their own right, but the DRB checklist is the most valuable as it asks many detailed questions. For example, the checklist will ask about the data (level of geography), if administrative data was used, if any of the data has been edited, if the data can be matched to external/public data, and if any disclosure avoidance techniques have been, or will be applied. The DRB checklist is only required if the submitters requests release of microdata, demographic tabular data, or establishment level data. Many requests occur in long cycles (up to 10 years) and the DRB checklist can pinpoint what has changed. In short, the DRB checklist greatly assists to ensure the consistency in the DRB's decision-making process (U.S. Census Bureau, 2013e).

For American Community Survey special tabulation requests, DAOs will use a condensed version of the DRB checklist. This 'ACS Special Tab checklist' provides the DRB with some general information about the request, such as the purpose of creating the special tabulation, a summary of the special tabulation, the data sources and geography covered in the request, the type of table that will be produced, which DRB rules are followed, and any other comments the submitter wishes to make (U.S. Census Bureau, 2013a).

The DRB will use all of these documents to make an informed decision for each request. After they make a ruling, the DAO will need to fill out the DAO checklist. This checklist helps the DAO go through the process of ensuring that the disclosure avoidance techniques were not only applied, but also applied correctly. If the DRB requires any modifications to the request, the DAO will also check to make sure the submitter incorporated the modifications. Once the DAO completes this checklist, they will sign off on the release of the request and the submitter can officially release their data product. After this is complete, the DAO must attach their DAO checklist to a copy of the request and the DRB approval memo. Finally, the DAO must ensure that his/her division/program area maintains a copy of this documentation (U.S. Census Bureau, 2013c).

4. How the DRB Handles Challenging Requests

In this section of the paper, I will provide some examples of the common types of requests, and discuss the challenges that the DRB faces in approving such data products. While most requests take great thought, the DRB already has many guidelines in place that can ensure the protection of our respondent's confidentiality. For each example, we lay out some of the disclosure protection methodologies that are suggested, implemented and reviewed before release of a given data product.

4.1 Public Use Files

What occurs if a submitter wants to release all possible variables as part of a Public Use Microdata File (PUF)? A PUF is a microdata file where each record represents the values and/or characteristics of a particular respondent. For example, variables may include demographic characteristics such as age, race, sex and income. Most PUF files are demographic in nature (i.e. American Community Survey PUMS), but recently, there have been some PUF files from economic surveys (i.e. Survey of Business Owners PUF). In theory, a submitter may be able to report all possible variables within a PUMS request. However, in practice, this may not be easily achievable and poses a challenge. Imagine if the Census Bureau released a PUF file without any restrictions. First, they would violate their pledge to confidentiality, as respondents could be easily identifiable. Second, those releasing the unrestricted file would be violating Title 13 laws. Third, the Census Bureau would lose the public's trust as a PUF file would be a very easy way to find out confidential information on a respondent. This would cause a number of significant long-term consequences, especially to response rates. To make sure none of these occur, the Census Bureau currently uses many disclosure avoidance techniques for microdata files, regardless of how many variables are part of the request. All of these techniques are applied to the variables within the PUF file before it is released into the public domain. When reviewing a PUF request, the DRB will need to agree on which techniques will be utilized for the specific set of variables for each request. In most cases, the submitter has already applied these techniques and the DAO has checked to make sure they were applied correctly prior to obtaining DRB approval. If the DRB rules to make any modifications, then the submitter and DAO must work together to ensure that these are applied. Below are most of the techniques used for PUF files.

4.1.1 Direct Identifiers

First, staff will remove all direct identifiers. A respondent's name or address would be the easiest way for re-identification within a PUF file.

4.1.2 Geographic Thresholds

Staff will ensure that all identified geographic areas contain a population of at least 100,000. Additional geographic thresholds are sometimes required depending on other factors, such as the level of detail of the variables requested, whether the survey used is longitudinal, or depending on similar publicly available data.

4.1.3 Rounding Content

Certain variables, such as dollar amounts, are rounded. Staff will apply rounding prior to all summaries and ratio calculations. There are some variables that are already publicly available, such as Property Taxes. In these instances, the Census Bureau will use larger categories than the ones used in the traditional scheme, listed below.

The Census Bureau uses the following traditional scheme:

$0 remains $0
$1-7 rounds to $4
$8-$999 rounds to the nearest $10
$1,000-$49,999 rounds to the nearest $100
$50,000+ rounds to the nearest $1,000

More recently, the DRB has also encouraged rounding to three significant digits, where the traditional scheme is not applicable.

4.1.4 Topcoding and Bottomcoding

Topcoding and bottomcoding are used to mask some of the outliers for a given set of continuous variables. For a given geographic area, an outlier is an easy way to identify a respondent. The Census Bureau uses the half-percent/three-percent rule for topcoding. Here, the topcode includes at least half-percent of all cases within a given universe. For example, if a submitter wants to include the variable age, any age that is within the top one-half percent will receive the topcode. For variables that are included in subpopulations (such as farm income), the topcode must include three percent of non-zero cases, or one-half percent of all cases, whichever is larger. Some variables, such as year born, are bottomcoded. Other information about the coded variables, such as mean and median, are released.

4.1.5 Categorical Thresholds

If the PUF file contains any categorical variables, each published category must contain 10,000 people nationwide. Any categories not meeting this threshold will be recoded into larger categories.

4.1.6 Data Swapping

Some PUF files also receive data swapping, the disclosure avoidance methodology that swaps a small number of households between different geographic areas. Here, risky households are targeted and must match on a few key characteristics in order to be part of the swapping procedure.

4.1.7 Noise Infusion

In some cases, variables will also receive a small amount of noise to protect uncommon observations. For example, noise is added to age where a household contains 10 or more people.

4.2 Tabular Data

What occurs when a submitter wants to release tabular data for every county in the United States? The Census Bureau produces a multitude of count, or frequency tables. Many times, submitters wish to release count tables at the county-level or at smaller levels of geography. The DRB must review all tables where the level of geography is below the state level. These types of tables are certainly allowable (e.g., large counties in CA), but there are other times where it is much more difficult (Loving County, TX – 112, 2015 Population Estimates). Regardless, the DRB and DAO must ensure that the submitters apply certain disclosure avoidance techniques to consider releasing these types of counts (U.S. Census Bureau, 2013b).

There are two types of these requests: standard tables, and special tabulations. In both cases, almost all requests will use American Community Survey or Decennial data. For these data sources, data swapping is the main disclosure technique used to protect the data. Data swapping is used to swap risky households in different geographic areas. These households will match on a few key characteristics. Even if a specific county has a small number of respondents, there is a probability that a reported household may have been swapped with another from a different geographic area. Sometimes, synthetic data are used as well (e.g., for Group Quarters).

In most cases, submitters cannot obtain the data they want or need via a standard tabulation. Thus, they must request and pay for a special tabulation. These carry additional disclosure techniques. Let us look at some of the procedures that are in place when a submitter requests a special tabulation for population counts for all counties in the United States (U.S. Census Bureau, 2008, 2011).

4.2.1 Rounding Count

Whether using the Decennial Census or American Community Survey, all cells must be rounded. The rounding scheme for all tables is:

0 remains 0
1-7 rounds to 4
8 or greater rounds to the nearest multiple of 5

In addition, any totals or subtotals should be constructed before rounding. In doing so, we are assured that the universes remain consistent between tables.

4.2.2 Thresholds

In many cases, all cells must meet a cell threshold requirement. Regardless of the level of geography, each cell must contain at least three unweighted respondents and/or the universe of the table must included 50 unweighted cases to avoid reporting data for very small geographic areas or for small population groups. For survey data, usually only weighted estimates are published. Generally, tables may not have more than three or four dimensions as well. The tables' level of detail must also be taken into consideration. Sometimes, submitters wish to produce county level counts with a variety of respondent characteristics. If this occurs, the DRB may need to apply certain thresholds based on the requested universe. If a cell or universe count fails the thresholds, it will not be released.

4.2.3 Other Techniques

Of course, there are times when a submitter requests additional summary information in addition to population counts. In the case of distributional estimates, medians or other quantiles must be interpolated, or calculated as a point quantile. For point quantiles, the quantile must be rounded to two significant digits, and contain at least five non-overlapping cases on either side of the quantile. If percentages or rates are required, these should be calculated after rounding.

4.3 Magnitude Tables

What occurs if a submitter is interested in releasing magnitude tables using economic data? Another type of commonly requested data comes in the form of economic magnitude tables. Magnitude tables usually contain establishment frequency counts, as well as an aggregate of some quantity of interest (e.g. sales) over all units in each cell. The frequency counts in the tables are not considered sensitive because a lot of information about establishments is already publicly available. However, the magnitude values are considered sensitive and must be protected. Protection is applied to the firm level, and determined by how closely one can estimate an individual firm's data via a sensitivity rule. The level of disclosure risk depends on the specific request. Generally, as more detail is requested (e.g., smaller geographies, more defined NAICS categories), it becomes more likely that cells will need suppression, and thus, the request becomes more challenging. When a submitter comes to the DRB, in most cases, the researcher applies the cell suppression program to the table(s) and makes the appropriate suppressions. This technique is explained next.

For most of its magnitude tables, the Census Bureau uses a disclosure avoidance technique called cell suppression. Here, they suppress any table cell value that is determined to be sensitive. That is, cells that could allow users to estimate a responding firm's value too closely. These sensitive cells (called primary suppressions), are suppressed from the table and replaced with a "D" for disclosure. These are identified using the p% rule, which is designed to ensure that a user cannot estimate a respondent's value to within p% of that value (Federal Committee on Statistical Methodology, 2005). However, the disclosure procedure does not end there. Since many magnitude tables contain marginal totals, other cells, called complementary suppressions must also be suppressed. This is necessary so that the primary suppressions cannot be estimated too closely via addition or subtraction of published values. For complex, and multi-dimensional tables, the Census Bureau uses advanced cell suppression software, based on linear programming (Steel, et al, 2013). For small tables, it is possible to perform cell suppression by hand. However, this will require extensive review to ensure that none of the primary or complementary suppressions can be backed out.

If a submitter does not want to go through the cell suppression avenue, they have other options as well in producing magnitude tables. These would involve using noise infused or synthetic data to create the magnitude tables. In these cases, the DRB would generally grant product approval in a quick manner. This is especially true if the noise infusion or synthetic data methodology has previously been approved and applied.

4.4 Federal Statistical Research Data Center Output

Occasionally, the DRB will need to review requests originating from the Federal Statistical Research Data Centers (FSRDCs). These occur when the level of geography is below the state level, or when an approved external researcher asks for a modification of

a prior guideline. Most FSRDC requests do not make it to the DRB because the DAOs review and approve the output on behalf of the DRB. They are responsible for ensuring that the rules, as listed in the RDC Researcher's Handbook are met (U.S. Census Bureau, 2010). The output must not reveal the confidential information of any particular individual, household, firm or establishment. In the RDCs, output must abide to count rules (both for demographic and economic data), and concentration ratios (for economic data). Concentration ratios are a calculated value that determines if the quantity in a cell (or for a set of firms) is too concentrated. If so, it is determined that there exists a possibility that a user can closely approximate the true value of a respondent (in that given cell, or group of firms). There is also a wide variety of other rules that the DAOs must ensure are accounted for before approving any output for public release. Some examples include rounding the number of observations, interpolating quantiles and keeping track of samples and implicit samples within a given project as well as between projects. If samples or cells fail any of the thresholds, then suppression is required.

5. Conclusion

When it comes to protecting the confidentiality of respondents in Census Bureau data, the DRB and DAOs play a critical role. The Census Bureau produces thousands of products each year, and with the disclosure avoidance techniques in place, protection of our respondents is ensured through the work of the DRB and DAOs. As explained in this paper, there are many types of disclosure avoidance techniques that the DRB and DAOs look to and apply when approving requests. Most requests are standard, but from time to time, the DRB will obtain difficult requests. However, due to these techniques, the DRB is confident that even the most difficult requests can be approved or modified. A compromise on content can usually achieve a useful, safe data product. In doing so, the DRB can help the Census Bureau fulfill its mission. That is, to serve as the leading source of quality data about the nation's people and economy, while honoring privacy and protecting confidentiality.

Acknowledgements

We would like to acknowledge Laura McKenna, Aref Dajani and Phil Steel for their helpful contributions and feedback.

References

Federal Committee on Statistical Methodology 2005. *Statistical Policy Working Paper 22: Report on Statistical Disclosure Limitation Methodology* Version 2, Washington, DC: U.S. Office of Management and Budget. Available at http://fcsm.sites.usa.gov/files/2014/04/spwp22.pdf (accessed April 2016).

Steel, P., J. Fagan, P. Massell, R. Moore Jr., J. Slanta, B. Wang. "Re-development of the Cell Suppression Methodology at the US Census Bureau." Joint United Nations Economic Commission for Europe (UNECE)/Statistical Office of the European Union (EUROSTAT) work session on statistical data confidentiality, October 2013.

U.S. Census Bureau (2008), "Disclosure Rules for American Community Survey Special tabulations based on Group Quarters data." Suitland, MD: U.S. Census Bureau. Available upon request (accessed May 2016).

U.S. Census Bureau (2010), "The Researcher Handbook: U.S. Bureau of the Census Center for Economic Studies Research Data Center's Handbook for Researchers." Suitland, MD: U.S. Census Bureau. Available upon request. (accessed June 2016).

U.S. Census Bureau (2011), "2010 Census Special Tabulation Program: Disclosure Review Board Rules/Requirements." Suitland, MD: U.S. Census Bureau. Available at: http://www.census.gov/population/www/cen2010/spec-tab/stp_DRBrules.html (accessed May 2016).

U.S. Census Bureau (2013a), "ACS Special Tabulation Requirement Template." Suitland, MD: U.S. Census Bureau. Available upon request. (accessed June 2016).

U.S. Census Bureau (2013b), "American Community Survey: Data Suppression." Suitland, MD: U.S. Census Bureau. Available at: http://www2.census.gov/programs-surveys/acs/tech_docs/data_suppression/ACSO_Data_Suppression.pdf (accessed April 2016).

U.S. Census Bureau (2013c), "Checklist for Disclosure Avoidance Officers." Suitland, MD: U.S. Census Bureau. Available upon request. (accessed May 2016).

U.S. Census Bureau (2013d), "Disclosure Review Procedures." Suitland, MD: U.S. Census Bureau. Available upon request. (accessed May 2016).

U.S. Census Bureau (2013e), "Supporting Document Checklist on Disclosure Potential of Data Disclosure Review." Suitland, MD: U.S. Census Bureau. Available at http://www.census.gov/srd/sdc/drbchecklist51313.docx (accessed May 2016).

U.S. Census Bureau (2016), "Disclosure Review Board Charter." Suitland, MD: U.S. Census Bureau. Available at https://collab.ecm.census.gov/div/pco/PDSIntranet/Documents/DRB%20Charter%207-15-16.pdf (accessed June 2016).