

Checklist on Disclosure Potential of Proposed Data Releases

**University of Waterloo
September 2006**

Adapted from

**Interagency Confidentiality and Data Access Group
An Interest Group of the Federal Committee on Statistical
Methodology**

**Statistical Policy Office
Office of Information and Regulatory Affairs
Office of Management and Budget
July 1999**

NOTE: Responses to the questions in this Checklist must be treated as strictly confidential.

Checklist on Disclosure Potential of Proposed Data Releases

Table of Contents

1. Introduction.....	1
1.1 Using the Checklist.....	1
1.2 Keeping Responses to the Checklist Confidential.....	1
2. Cover Sheet.....	2
3. Microdata Files	3
3.1 Geographic Information on the File.....	3
3.2 File Contents Presenting an Unusual Risk of Individual Disclosure.....	5
3.3 Disclosure Risks Associated with Administrative Data and Other External Data	8
3.4 The Addition of Statistical Perturbation (or “Noise”)	13
3.5 Other Issues.....	14
4. Tabular Data from Persons or Households (or “Demographic Tables”)	16
4.1 The Data	16
4.2 Disclosure Risks Associated with Administrative Data and Other External Data	18
4.3 Disclosure Limitation Methods	20
4.4 Coordination of Disclosure Limitation	24
5. Tabular Data from Establishments or Other Types of Organizations	25
Appendix A: Statistical Disclosure Limitation Method by Three Types of Release (Microdata Files, Demographic Tables, and Economic Tables).....	26
Appendix B: Definitions of Selected Statistical Disclosure Limitation Methods.....	27

NOTE: Responses to the questions in this Checklist must be treated as strictly confidential.

Checklist on Disclosure Potential of Proposed Data Releases

Introduction

Researchers storing their data in the Population Health Data Repository have taken great care to protect the identity of their participants and to safeguard the privacy and confidentiality of the information that these participants provide. The following checklist assists ICE staff in limiting the risk of disclosure in data sets released under an assurance of confidentiality. Upon receiving data sets for inclusion in the data repository, an analyst working on the repository will analyze the data to evaluate the risk of disclosure with oversight and review from a **Disclosure Review Board** comprised of ICE investigators.

1.1 Using the Checklist

The checklist will be used in the following manner:

It will be completed by a trained ICE employee who has appropriate statistical knowledge and is familiar with the ICE PHDR policies and procedures. The completed checklist will then be reviewed by a Disclosure Review Board which consists of a subset of ICE investigators as well as an external person with statistical expertise but not related to the ICE project or the PHDR.

Responses to questions in the checklist are not intended to supply all of the information required by a Disclosure Review Board before a microdata file or table is released to the public. Some additional questions may need to be answered and/or given special consideration. Nonetheless, if files and tabular material are reviewed with the aid of the checklist early enough, the need for time-consuming and costly re-programming of the data to be released can be avoided. This allows additional time for coordination with collaborators and/or other potential users.

1.2 Keeping Responses to the Checklist Confidential

It is necessary to walk a fine line between giving data users enough information so that they will have some idea of how the statistical disclosure procedures that were used might affect their analyses and giving them so much information that the data are made more vulnerable to an attacker. In order to release as much information as possible at an acceptable level of disclosure risk, documentation should describe the kind(s) of procedures used to protect the confidentiality of data but not reveal the specific value(s) of the particular disclosure limitation method.

Complete knowledge of a specific statistical disclosure method and its associated parameter value(s) could be used by an attacker to identify one or more individuals on a microdata file or infer the value of table cells that have been suppressed to protect the confidentiality of that information. Because this checklist contains explicit details on disclosure method(s) and parameter value(s), responses to it should be kept confidential.

Cover Sheet

SURVEY TITLE: _____

DATE: _____

What type of data are you releasing?

- Microdata file.** *Please attach the proposed layout and content of the microdata file.*
- Tables.**

When were the data collected? _____

Does (Do) the reference period(s) of the data collection differ from the actual date of collection?

- Yes.** *Please give reference period(s).* _____
- No.**

What is the periodicity of the proposed data release?

- This is a preliminary release.**
- This is a one-time release of a microdata file from a one-time collection of data.**
- This is a release of a special tabulation.**
- This is one in a series of releases (either microdata file or tables) with substantially the same content.** *Please specify the interval at which future products will be released or prior products have been released.* _____
- This is the re-release of an approved product, with the addition of supplemental or previously unreleased data.** *Please give the date the original product was submitted.*

(NOTE. If this is a re-release of a previously approved product, then only complete those Checklist questions for which the answers are now different.)

Will there be other data release(s) (either microdata files or tables) from this survey?

- Yes.**
- No.**

3. Microdata Files

3.1 Geographic Information

Identify the variables on the file used as geographic identifiers and the minimum population size for each geographic area. The general rule is that all geographic areas identified must at least have 100,000 persons in the sampled area. Smaller geographic areas can be identified and included in a data set if there is no possibility of disclosure.

3.1.1 List all geographic identifiers to be released and the minimum population of each identifier:

Have you chosen to adopt the above rule or another?

Yes, will use the rule of 100,000.

No, will use other rule. Please specify and provide rationale:

3.1.2 Care should be taken before releasing the primary sampling units (PSU's) that are in a sample. In addition to explicit geographic identifiers on the file, the data items, record identifiers, or file structure may provide additional geographic information by inference. Therefore, steps must be taken to avoid inadvertently identifying geographic areas that do not meet the specified minimum population criteria.

How will this problem be avoided on the released file?

Control numbers are deleted or do not contain geographic information.

Control numbers are scrambled. Please describe:

Other. Please describe:

3.1.3 Records in many data bases are sequenced so that the first cases are in the lower numbered PSU or municipality that is first in alphabetic order.

Briefly, describe how the records on this file will be sequenced to avoid such geographic inferences:

3.1.4 Data items that imply specific geography of residence may reveal more than the explicit identifiers displayed.

List all items that will be deleted for this reason:

3.1.5 Identify other geographically-related variables (e.g., urban area, non-urban area) on the file.

List all items that you think might have geographic significance, but could not decide if they should be deleted:

3.1.6 Sampling information may also provide some geographic indicators.

(a) List all sampling information – including that for variance estimation – that will be deleted for confidentiality reasons or subsampling plans to make weights less identifying:

(b) List all other sampling information that you think might have geographic significance, but could not decide if it should be deleted:

3.2 File Contents Presenting an Unusual Risk of Individual Disclosure

The disclosure criteria for microdata require a review of each file to determine if any of the proposed contents present an unusual risk of individual disclosure.

Names, addresses, and other unique numeric identifiers such as Social Insurance Numbers, Health Card numbers *must* be removed from the file.

High income is a visible characteristic of individuals or households and is considered to be a sensitive item of information. Therefore, each income figure on the file, whether for households, persons, or families, including total income and its individual components, should be **topcoded**.

In developing topcode rules, it might be prudent to discuss alternatives with the Disclosure Review Board well in advance of the final submission for approval to release a file.

3.2.1 Please describe the topcoding rule that is used. If you have different rules for different income variables, please give details.

Do all income topcodes satisfy the appropriate rule(s)?

Yes.

No. If not, please specify the percent topcoded and the topcode amount.

3.2.2 In addition to income, certain other characteristics may make an individual more visible than others and some variables may require bottom-coding, such as year of birth before 1914 or large negative value for income.

(a) List all items that will be bottom- or topcoded (or collapsed) and the corresponding codes:

(b) List all other items about which you have questions regarding the need to bottom- or topcode:

Describe any proposed information to be released for the bottom- or topcoded data items (for example, means or medians of the coded values):

3.2.3 Depending on the amount of geographic detail on the file, there are other characteristics that may make a person highly visible. These typically are represented as categorical or non-ordinal variables and, therefore, cannot be topcoded.

(a) List all items that will be collapsed (or deleted) for confidentiality reasons:

(b) List any other items about which you have questions regarding the need to collapse the detail:

3.2.4 Contextual or Ecologic Variables

Contextual or ecologic variables are those that describe some aspect of an area, such as a province, region, census tract, or block group; percent or frequency of the area's population employed, foreign born, receiving public assistance; number of health facilities; etc.

Identify the source(s) of the contextual variables:

Identify any contextual variables and the levels at which they are coded:

List all contextual variables that will be collapsed (or deleted) for confidentiality reasons:

List any other contextual items about which you have questions regarding the need to collapse the detail:

Does the cross tabulation of all contextual variables and geographic variables lead to the identification of one or more geographic areas that are not supposed to be identified on this file?

Yes. If yes, some, variable(s) *must* be removed and/or collapsed. Please describe what will be done.

No.

Don't know/did not check. Please explain.

3.3 Disclosure Risks Associated with Administrative Data and Other External Data

3.3.1 External Files Matchable to Proposed File

Were any of the sample cases contained in the microdata file selected from a list provided by an outside source?

Yes. Please identify the source, and describe how and by whom sample cases were selected from the list:

No.
 Don't know.

Were any administrative data or data from another external source used to “expand” the content of this microdata file (e.g., merging administrative data with survey data)?

Yes. Please identify the source, and describe the variables obtained from the external source(s):

No.
 Don't know.

Are you aware of administrative records, research files, or a mailing list that contains data also included in this proposed microdata file? Such external data could be used to “compromise” your data, that is, they pose a risk because an “attacker” would use these external data sources to deliberately try to identify individual respondents and gain access to their confidential responses.

Yes. Please identify:

No.
 Don't know/did not check.

Based on available information, will any data item on the microdata file identify residence in a particular type of institution of which there may be only one in an identified area; or for which a system of records could be obtained?

Yes. Please identify the type of institution:

No.

Don't know.

3.3.2 Matching

When an external file related to the proposed file to be released exists, several steps may be taken to reduce the possibility of matching survey data to this file; for example, selected items may be deleted or recoded, or "noise" (i.e., small amounts of random variation) may be introduced into these items.

...if any data item, or combination of items, isolates a small and readily identifiable population subgroup or class. The inclusion of codes that identify very small population segments should be avoided; for example, First Nations tribes or detailed occupation groups in combination with the release of highly specific geographic identifiers. Normally, one has to consider more than one variable at a time if a group of variables is likely to appear together on a file or list. For example, age and sex, together with country of birth and occupation, could permit the disclosure of individual identity.

(a) List all data item(s) proposed for inclusion on the file that, in various combinations, may isolate a small, readily identifiable population:

(b) List all data item(s) that will be altered (i.e., deleted, recoded, noise added) for this reason:

...if the file includes essentially every member of a population (say $p > 0.5$). Examples include: establishments/institutions with large numbers of employees, high-income individuals, doctors, scientists of a specified type, or residents of certain types of institutions. In such instances, prior to data release, it may be appropriate to do additional subsampling.

Identify these populations, if any are on the file, and how they will be subsampled:

...if the file contains any information obtained from records or other sources where that information could serve as a link to an external file which has individual identifiers or detailed geographic information. Examples include: fuel consumption or cost records from a utility company; neighborhood, tract, or summary characteristics from a decennial census; welfare or health-related data from a Government agency; arrest record from a police department; benefits provided to employees, such as pensions and health insurance.

(a) List all data item(s) proposed for release on the file that were not obtained from an interview with the respondent:

(b) List all data item(s) altered or deleted for this reason:

...if the file includes data items frequently used for matching, such as exact date of birth, sex, and race, or if it includes other items that should be identical on both files, such as exact income amount, real estate taxes or other taxes, or date of entry or termination from a Government-sponsored program.

(a) List these data items, if any:

(b) List all data item(s) altered or deleted for this reason:

...if longitudinal data are being collected; i.e., if the data for the same respondents/units will be collected for several different reference periods. Primary concern relates to time series of data items potentially matchable to outside records; e.g., income tax or employment records.

If data are collected from the same respondents more than once, indicate the frequency of interview, length of time that any one unit may be in the sample, and factors affecting the likelihood of matching a sample unit from one time period to the next:

...if highly specific geography is included on the file, such as province or State, Metropolitan statistical area, etc.

List all geographic identifiers below the level of region:

...if data collected from multiple persons in a household are linked on the released file. Disclosure risks associated with linking of household members are well-known. For example, households can be identified because of significant difference in spouses' ages, atypical number and ages of children, a "unique" multi-racial composition of the household, etc. -- not to mention the fact that one household member, by self-identification, could look up other members' reported information.

(a) Are data collected from multiple persons in a household?

- Yes.**
- No.** *If no, skip to 3.3.3.*

(b) If yes, describe the strategy for releasing these data, and indicate whether or not the data from these will be linked:

Describe any considerations not previously mentioned that reduce the ability to match this file to external data; e.g., unreliability or natural noise in the data:

3.3.3 Cross-Tabulations to Identify Unique Sets of Characteristics

Were any cross-tabulations performed to identify sets of unique characteristics?

- Yes.
- No. *If no, skip to 3.4.*

If yes, what were the results?

Will any additional steps be taken to reduce disclosure risk based on these results?

3.4 The Addition of Statistical Perturbation (or “Noise”)

The addition of statistical perturbation, called “noise,” is another statistical disclosure limitation technique. **This is not a recommended technique for the ICE PHDR**, though it may occur in very rare cases.

Was any noise added to the data?

- Yes.**
- No.** *If no, skip to 3.5.*

What procedure(s) was (were) used to add noise to the data? Please give specifics for that procedure (i.e., percent of records affected, distribution of noise, etc.).

Was any attempt made to match back the noise-added data to the original file?

- Yes.**
- No.** *If no, skip to 3.5.*

If yes, how was it done, and what was the rate of success in matching?

3.5 Other Issues

3.5.1 Files that include every sample case or cases in strata that are sampled at high rates ($p > 0.5$) are more likely to lead to disclosure than files containing only a subsample of cases.

Does this file contain every case?

- Yes, it includes every case.
- No, it includes a subsample of cases. If so, specify the range of sampling rates:

3.5.2 Project managers should be aware that confidentiality problems may arise if special tabulations are made from an internal version of the file, which includes detail omitted from the public-use file. For example, tabulation might provide specific geography, not included on the public-use file that is cross-tabulated by multiple data items on the file. Consult with the Disclosure Review Board if you are planning to release tabulations that make use of detail not available on the public-use file.

Do you plan to release tables that make use of detail not available on the public-use file?

- Yes.
- No.

3.5.3 Describe the Sample Design

Briefly describe the sample design by answering these questions:

Describe the sample design, including stratification, clustering, and stages. Be certain to identify the units that were sampled at any stage with probability > 0.5 :

Provide a brief paragraph that compares and contrasts the proposed sampling units, units of enumeration, and units of analysis in the study:

The following subset of questions pertains to how much information is (will be) publicly available about the sample design, i.e., sampling plan and estimators (including the identity of PSU's):

(a) What information is already publicly available?

(b) What information will be made public with this release?

(c) What information will be withheld?

Describe how users will estimate sampling variances, potentially identifying any proposed "nesting variables" on the proposed file layout or the design of any weights used for replication approaches:

3.5.4 Supplements

Was this information gathered as a supplement to another survey?

- Yes.**
- No.** *If no, you are finished with this section of the Checklist.*

Can this microdata file be linked to the file produced from the main survey?

- Yes.**
- No.**
- Don't know.**

If yes, what geographic information is on the main file?

4. Tabular Data from Persons or Households (or “Demographic Tables”)

4.1 The Data

4.1.1 What data will be released and in what formats (i.e., table dimensions; variables and their detail)? One way to describe this is to use a table like the following:

Tabulations	Cell Contents	Comments
VbleA x VbleB x VbleC	ListA	
VbleA x VbleD	ListB	
VbleB x VbleD	ListB	
...		

Description of the variables:

VbleA = county

VbleB = income ranges (0; 1-9,999; 10,000-49,999;...)

VbleX = income

etc...

Description of the contents of each list (magnitude data):

ListA = unit (person/household) counts, VbleW, VbleX, VbleY, and VbleZ

ListB = unit (person/household) counts, VbleX, VbleY, and VbleZ

etc...

Use this space for additional descriptive information for item 4.1.1

4.1.2 What is the highest dimension of a table that you are releasing? _____

4.1.3 What level(s) of geography is (are) released?

4.1.4 Do you collect written waivers from individuals or households that would enable you to publish information that might otherwise have to be suppressed or masked?

Yes.

No.

If yes, please describe:

4.1.5 One method of protecting the confidentiality of data is to conduct a **sample** survey rather than a census. In reporting results of large-scale sample surveys, estimates are made by multiplying a respondent's data by a sampling weight before they are aggregated.

Are the data from a sample or a census (complete count)?

- sample.**
- census.** *If census, please skip to Section 4.2.*

If data are from a sample,

(a) Are some groups of individuals selected with certainty?

- Yes.**
- No.**

(b) Briefly describe the sample design, including overall sampling rate:

(c) Are weights common knowledge (or could easily be inferred) so that a cell with a value of 10, for example, could be linked to one person/household?

- Yes.**
- No.**

If yes, please explain:

4.1.7 Coordinating preliminary release(s) with final release(s).

Is this a preliminary release or a final release?

- Final release.**
- Preliminary release.** *(NOTE: Identical suppression patterns must be used in both the preliminary and final releases. If this is a preliminary release, consult with the Disclosure Review Board. Do **not** lock yourself into a disclosure method/pattern too early in the process because the disclosure limitation method(s) used for the preliminary release may restrict subsequent release(s), including the final release.)*

4.2 Disclosure Risks Associated with Administrative Data and Other External Data

The disclosure risk of a table is increased if it contains administrative data or any type of data from an outside/external source.

4.2.1 Were any administrative data or data from an external source used as a sampling frame for this survey?

- Yes.
- No. *If no, please skip to 4.2.4.*

If yes, please describe:

4.2.2 Were any administrative data or data from an external source used to “expand” the content of this census or survey (i.e., by merging administrative data with survey data)?

- Yes.
- No. *If no, please skip to 4.2.4.*

If yes, please describe:

4.2.3 Does the agency providing these administrative/external data have different confidentiality rules than your office/program?

- Yes.
- No.

If yes, please describe rules used by the other agency and how your office/program met both sets of rules:

4.2.4 Are you aware of any external data sources (i.e., State data files, files released by another Federal agency) that could be used to “compromise” your data? When such files exist, there is a risk that an “attacker” would use these external data sources to deliberately try to identify individual respondents and gain access to their

confidential responses.

Yes. If yes, please describe:

No.

Don't know/did not check. Please explain.

4.3 Disclosure Limitation Methods

4.3.1 Cell Suppression

One historical method of protecting sensitive cells in tables is cell suppression.

Did you use cell suppression as a disclosure limitation method?

- Yes.
- No. *If cell suppression was not used, please skip to Section 4.3.2.*

(a) What rule(s) was (were) used to determine primary suppressions? Please check all that apply and give parameter(s).

- the (n,k) rule. Give parameters: _____
 - the pq rule. Give parameters: _____
 - the p-percent rule. Give parameters: _____
 - other rule. Please describe this rule: _____
-
-

(b) Please describe how you applied this (these) rule(s). For example, if you used the (n,k) rule, describe the parameters used for each variable:

Was (were) a key item(s) chosen in performing cell suppression?

- Yes.
- No.

If yes, what was this key item, and why was it selected?

In determining secondary suppressions, what was your goal? (*used for magnitude data.*)

- minimize the number of cells that are suppressed.
 - minimize the amount of information that is suppressed.
 - other – please describe:
-

(a) Did you suppress values that were based on less than a specified threshold number?

- Yes. Please give threshold number: _____
- No. If you did not use a threshold rule, please explain why not

If yes, was it applied to weighted or unweighted data?

- Weighted data.
- Unweighted data.

Do any cells contain all members of a domain?

- Yes.
- No.
- Don't know.

If yes, please describe: _____

(a) Were secondary suppressions selected by hand or suppression software? (Please check all that apply, and describe what you did.)

- by hand.
- by suppression software developed in this agency.
- by suppression software developed by an outside vendor. Please provide name of vendor and product:

(b) Please describe your method:

4.3.2 Auditing

Were the suppression patterns in the tables audited? That is, were all tables audited?

- Yes. If yes, please explain how:
-

No. If no, explain why not:

Were any suppressions removed by hand?

- Yes.**
- No.**

If yes, please explain why:

If yes, how were suppressions chosen to replace them?

Will any additional information be released for values that were suppressed (i.e. ranges, medians, estimates, rounded values, values with noise, etc.)?

- Yes.**
- No.**

If yes, please give details:

4.3.3 Addition of Noise

Did you add noise to the underlying microdata before creating the tables?

- Yes.**
- No.** *If no, please skip to Section 4.3.4.*

Which items received noise?

How was noise added to the data?

How much noise was added to the data?

4.3.4 Additional Methods For Frequency Count Data

Did you use additional disclosure methods, other than cell suppression, for your frequency count data?

- Yes.**
- No.** *If no, please skip to Section 4.4.*

If yes, please describe:

4.4 Coordination of Disclosure Limitation

Have these data been used for research purposes (for example, by an external researcher) or released as in a special tabulation?

- Yes.
- No.

If yes, please give a brief description of this use or what was released:

Has the same (or very similar) data also been released by your unit or by another division, program area, or branch?

- Yes.
- No. *If no, you have completed the Checklist.*

If yes, please describe:

Were disclosure limitation techniques (such as cell suppression patterns) coordinated with those previously used?

- Yes. **If yes, please explain:**

- No. **If no, explain why you did not coordinate:**

5. Tabular Data from Establishments or Other Types of Organizations

In contrast to demographic data, it is presumed that data collected from establishments or other organizations (such as hospitals, schools) often pose a higher risk for several reasons: the data are typically very skewed, the size of the universe may be small, and there are many high visibility variables. For these reasons, establishment data collected under a pledge of confidentiality are rarely released as public-use microdata files and are typically released in tables.

A checklist for establishment surveys will not be included here. The principles are similar to those given for tabular data in the previous section.

Appendix A: Statistical Disclosure Limitation Method by Three Types of Release (Microdata Files, Demographic Tables, and Economic Tables)

Statistical Disclosure Limitation Method	Type of Release		
	Microdata File	Demographic Tables	Economic Tables
Record swapping	Yes	Yes	
Blanking and imputing	Yes	Yes	
Rank swapping	Yes		
Traditional rounding		Yes	
Controlled rounding		Yes	
Random rounding		Yes	
Noise	Yes		Yes
Cell suppression		Yes	Yes
Local suppression	Yes		
Recoding into broader categories (includes top-coding, bottom-coding, and geographic restrictions)	Yes	Yes	Yes
Blurring	Yes		
Microaggregation	Yes		
Multiple imputation	Yes		
Data modification		Yes	

Appendix B: Definitions of Selected Statistical Disclosure Limitation Methods¹

Blanking and imputation: Blanking and imputation involve selecting a sample of respondent records from internal files and blanking a subset of the values on those records and then using imputation techniques to fill in the blanked values.

Data modification (data blurring): Data modification (or data blurring) involves adding uncertainty to all cell values in a table or to cell values that are less than a prescribed threshold (e.g., 10). Modifications to cell values (e.g., -1, 0, or +1) are added with prescribed probabilities. For example, with these modifications, a published value of 4 could represent an actual value of 3, 4, or 5. To the extent that tables have actual cell values below a prescribed threshold, there will be rounding error and inconsistencies within tables and among tables.

Data swapping: Data swapping involves selecting a sample of respondent records from internal files and interchanging data for these respondents with other respondents that have identical characteristics on a set of key variables.

Rounding: Rounding involves adjusting the values in all cells in a table to a specified base so as to create uncertainty about the real value for any cell while adding a small but acceptable amount of distortion to the data. There are several ways in which the rounding technique can be implemented:

Controlled rounding: In controlled rounding, an attempt is made to round the values in cells of a table so as to preserve summation to subtotals and table totals. With controlled rounding, there may be inconsistencies among tables. For example, the total population in a table showing data on age may be different from the total population in a table showing data on race.

Fixed (or traditional) rounding: In fixed (or traditional) rounding, each cell value is rounded independently in a prescribed manner, which produces nonadditivity within tables but consistency among tables. An example of fixed rounding is to round values ending in 8, 9, 0, 1, or 2 to a value ending in 0 (e.g., a value of 18, 19, 20, 21, or 22 would be rounded to 20). In the example of tables on age and on race in the preceding paragraph, the total population with fixed rounding will be the same in the two tables, but the sum of cell values by age or by race may not add to the total populations in the respective tables.

Random rounding: In random rounding, each cell value is rounded independently in a random manner. For example, values of 6, 7, 8, or 9 could be rounded to 5 or 10 based on assigned probabilities. With random rounding, there may be inconsistency in data within tables and among tables.

Note # 1. With controlled rounding, fixed rounding, and random rounding, there is a choice of the base for rounding, the most common choices being 3 and 5. All rounded values

¹ For more information consult the Federal Committee on Statistical Methodology's *Report on Statistical Disclosure Limitation Methodology* (May 1994).

(other than zeros) are multiples of 3 or 5, respectively. Base 3 rounding introduces less distortion than base 5 rounding, but since multiples of 3 can end in any digit from 0 to 9, the use of base 3 rounding is less obvious to data users than the use of base 5 rounding, which produces values ending only in 0 or 5.

Note # 2. The techniques of rounding and data modification are similar in that they introduce uncertainty into the published data. The primary difference is that the rounding technique uses a specified base (usually 3 or 5) for rounding values in table cells whereas data modification does not use a specific base.