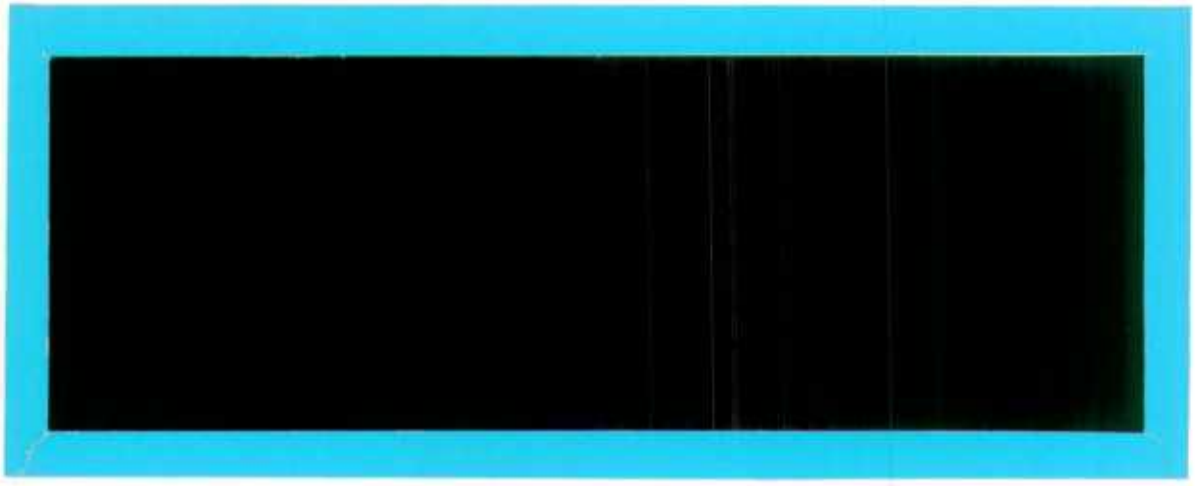
11-613E
no.88-09
c. 2

Statistics Canada Statistique Canada



Methodology Branch

Social Survey
Methods Division

Direction de la méthodologie

Division des méthodes
d'enquêtes sociales

Canada

WORKING PAPER NO. SSMD-88-009 E/F

METHODOLOGY BRANCH

Disclosure Avoidance Techniques in the Canadian Census of Population and Agricuture

SSMD-88-009 E/F

M.J. March and D.A. Norris

STATISTICS CANADA STATISTIQUE CANADA
SEP 8 1987
LIBRARY BIBLIOTHÈQUE

Presented at
The Joint Statistical Meetngs
August 17, 1987
San Francisco

The authors accept sole responsibility for the views expressed in this paper.
They do not necessarily reflect those of Statistics Canada

I. INTRODUCTION

Statistics Canada carries out the Censuses of Population and Agriculture under the authority of the Statistics Act. Under this Act, the Agency can collect and disseminate census data. However, the dissemination must be done in such a way as not to disclose information on an individual respondent.

As a consequence of the need to avoid disclosure, an important aspect of the program to disseminate census data is the development and implementation of procedures which prevent disclosure but at the same time do so in a way that has minimal impact on the amount and quality of information available to census data users.

Disclosure avoidance is an integral part of all statistical programs. However, there are particular problems that arise in the case of census data since they cover all or at least a large sample of the population and provide detailed data for very small geographic areas.

The question of statistical confidentiality has been widely studied. In their 1985 paper, Cox et al provided a comprehensive survey of this issue from the point of view of the U.S. Bureau of the Census and a thorough treatment was also given in Cox (1983) and in the Report on Statistical Disclosure and Disclosure Avoidance Techniques. This issue was considered in the Canadian context by Fellegi (1972) and Fellegi and Phillips (1974).

The purpose of this paper is to describe the disclosure avoidance procedures of the Canadian Censuses of Population and of Agriculture. The two censuses present different disclosure issues, and the approaches adopted are quite different. In the case of the Census of Population, the main emphasis is on disclosures that would result from publication of very small frequencies while for the Census of Agriculture, with tabulations of aggregate magnitude data (e.g. expenditures), it is not only small numbers of respondents that cause a problem but also the amounts contributed to totals by individuals. The main emphasis of this paper is on the procedures to be used for the 1986 Censuses; however, these are put in perspective by a comparison to techniques used in earlier censuses and a description of user reaction to the procedures.

In the case of the Census of Population, random rounding to the base 5 has been the principal technique used since its introduction in the 1971 Census; however, to supplement this, various forms of data suppression have been employed and there have been minor changes to these over time. In the case of the Census of Agriculture,most of the development work relating to the confidentiality procedures used at the present time did not take place until the late 1970s. A prototype of a generalized collapse and suppress system was used with some success in 1981. Since results of using this system were not entirely satisfactory, a new custom system was developed for the 1986 Census.

II. CENSUS OF POPULATION DISCLOSURE AVOIDANCE TECHNIQUES

Historically, Census of Population data dissemination consisted of the publication of specified tabulations in a series of census bulletins. With this type of limited dissemination program, the issue of disclosure could generally be handled by manual cell suppressions. As the volume of tables increased and there was no longer the time available to check them manually, other methods were required.

in addition, as data processing technology has evolved so have the demands for more detailed data and for small area data produced on magnetic tape or micro-fiche. To meet these new demands, new census products have been developed. One new type of product has been comprehensive data sets for small geographic areas which can be used as building blocks to further aggregate and manipulate data to user geographic and content specifications. These data sets provide much greater scope for disclosure. Since cell suppression has a major impact on data for such applications, careful consideration must be given to an appropriate choice of disclosure avoidance procedures. In Canada,the census enumeration area (with an average population size of 550) is used as a basic census geographic area. However, for large urban areas that cover about 60% of the population,

the census data are also coded to the block face level that allows for retrieval using the block face as a unit of aggregation. The block face is in most cases also the unit covered by the six digit postal code which, for 1986, will also be available as a unit of dissemination. The ability to tabulate data for user-defined aggregations of block faces may give rise to residual disclosures.

A further development in Census of Population dissemination has been the increasing shift to special tabulations whereby data users can specify both the content and geography of the tabulations. Such tabulations are facilitated by the fact that the census data are retained in a data base format that allows for the retrieval of custom tabulations on a request basis. For the 1981 census, approximately 2000 Census of Population tabulations were produced in pre-planned standard print or machine readable tabulation format while more than 10,000 tables were produced on a custom request basis.

The ability to produce custom tabulations provides a rich source of information for census data users. However, the requirement to process such large volumes of data requires that any disclosure avoidance technique be generalized and operationally easy to implement.

A further development is the increasing requirements for data on specific narrowly defined target groups (e.g. ethnic groups, occupation groups). Tabulations for such groups, especially for user defined small areas, may give rise to residual disclosures. For example this type of situation could occur in development of a special data set for the aboriginal population. A sizable part of Canada's aboriginal population live on Indian Reserves that have been identified as census subdivisions in the standard census geographic hierarchy. Since census subdivisions are basic tabulation units, much data are already available for these same areas. While the great majority of persons living on Indian Reserves are aboriginal persons, there is often a small non- aboriginal population. Release of data for only the aboriginal population could potentially cause a residual disclosure of information for individuals in the non-aboriginal population.

There are two general approaches used to avoid disclosures in the Census of Population. These are:

1. Random rounding;
2. Suppression;

1. RANDOM ROUNDING

Random rounding has been the principal method used in the Canadian Census of Population beginning with the 1971 Census. For each census, all data tabulations, other than total population and dwelling counts, are random rounded to the base 5. With random rounding, each cell frequency is independently rounded up or down to an adjacent multiple of 5 using an un-biased procedure. More specifically, a cell frequency,f, is rounded up or down to a multiple of 5 using r, the remainder when f is divided by 5 as follows:

1. round up with probability r/5 $r = 1, 2, 3, 4$
2. round down with probability $1 - r/5$ $r = 1, 2, 3, 4$
3. do not round if $r = 0$

All numbers in a tabulation, including subtotals and totals are rounded independently. This implies that tabulations are not additive in the sense that the totals will not,in general, exactly equal the sum of the parts.

As indicated above, random rounding has been used since 1971. Over time, the technique has been accepted by users who recognize it as an acceptable procedure to guard against disclosure. On occasion, a new census data user may react to the fact that the totals do not agree with the sum of the cells but this has not been a major issue in recent censuses. The 1986 Census will be the fourth census that has used this procedure and users have generally accepted this minor inconvenience. Another concern of some data users is the impact of random rounding on small populations, especially when randomly rounded data are to be aggregated. Again, this has not been a major problem for most users. Randomly rounded data for small geographic areas have been widely used as a basis for aggregation. In addition, where required, customized data tabulations can be

requested where the random rounding is done after geographic aggregation. Nevertheless, even though users have not reacted strongly, more work needs to be done to assess the impact, from the point of view of data quality, of aggregating randomly rounded data.

2. SUPPRESSION

While random rounding is the principal disclosure avoidance technique used in the Census of Population, various forms of data suppression are also employed in addition to random rounding.

Suppression rules take two basic forms (i) so called area suppression whereby no data are shown for very small areas and (ii) cell suppression whereby small cells in a table are suppressed.

In the Census of Population, suppression has been used to supplement random rounding wherever there was concern that even rounded frequencies,especially for characteristics with a number of detailed categories such as income, occupation or industry, may pose a potential for disclosure . In addition, for characteristics collected on a 20% sample basis where the usual weight assigned to an individual is equal to 5, rounded cells of 5 or 10 reflect the characteristics of only one or two individuals and can possibly be disclosures.

The need for some type of suppression is particularly strong if standard data series are being produced and disseminated for very small areas. Table 1 shows the size distribution of enumeration areas (EAs) and census subdivisions (CSDs) (i.e. municipalities) in the 1981 Census, the two standard geographic areas for which small area data are produced. As can be seen, there are areas, albeit a small number of them, where the population is a single individual or a small number of individuals often in a single household. Clearly random rounding is not sufficient to protect against disclosure in such cases.

TABLE 1

NUMBER AND POPULATION BY SIZE OF GEOGRAPHIC AREA

Enumeration Areas	Number	Total Population – All Ages
With Pop 0	2,444	0
With Pop 1-24	1,096	10,423
With Pop 25-39	372	11,682
With Pop 40-49	191	8,379
With Pop 50-99	936	70,636
With Pop 100-249	4,154	741,756
With Pop 250 or more	32,004	23,240,615
Census Subdivisions		
With Pop 0	201	0
With Pop 1-24	124	1,743
With Pop 25-39	90	2,894
With Pop 40-49	43	1,869
With Pop 50-99	197	14,163
With Pop 100-249	487	82,230
With Pop 250-499	885	331,746
With Pop 500-999	1,232	890,494
With Pop 1000-4999	1,801	3,895,742
With Pop 5000-9999	321	2,253,482
With Pop 10000-24999	189	2,821,939
With Pop 25000-49999	66	2,265,458
With Pop 50000-99999	44	2,991,868
With Pop 100000-249999	16	2,170,966
With Pop 250000-499999	8	2,570,129
With Pop 500000-or more	6	3,788,769

In the 1981 Census, rules were developed for suppression of data for very small areas. For example, data for a self-enumeration area were suppressed if its population was less than 50 while for a canvasser area the limit was 25. However, data for those areas were included in all higher level roll-ups. Slightly different rules were applied depending on whether the data were produced as a publication, a summary tabulation or magnetic tape or as a custom tabulation request. This variation in rules caused some confusion among data users since, in a few cases, data suppressed in one media were available on another. For the 1986 Census, area suppression will again be used. However, a consistent rule will be applied so that, in all cases, data will be suppressed for standard geographic areas having a population of less than 40. For non-standard user-defined areas the minimum cut- off point is set at 100 (a higher level as a precaution against residual disclosures).

For both the 1981 and 1986 censuses the population size used to determine the cut-off for a tabulation of data from the 20% sample is the total non-institutional population since sample data were only collected for this sub- universe. For income distributions, no data are shown for areas with a non-institutional population of less than 250.

In addition to the suppression of data for small geographic areas, cell suppression has been used for selected variables: in particular, income, detailed occupation and detailed industry. In publications, cells for these three variables and any calculations derived from these cells are suppressed if the marginal row or column total is less than 250. However, their data have been included in any subtotals or totals that were shown.

A final form of cell suppression relates to the suppression of individual small cells in the case of customized tabulations. For the 1981 Census, all cells of less than 25 were suppressed for tabulations of income, occupation and industry.

In part the suppression limit was set at 25 for data quality reasons and in part for operational reasons since it was not possible to easily implement area suppression for very small areas. The use of this population of 25 suppression rule in custom tabulations was particularly unpopular with users and for 1986 will be replaced by a cell suppression of only the "5s" in tables as discussed below.

The increasing demand for custom tabulations has led to further work on the adequacy of random rounding particularly in custom tabulations for small geographic areas. Studies confirmed that for tabulations of 20% sample data, a very high proportion of the "5s" in tables (after rounding) were in fact based on single individuals. Given the very detailed nature of some custom tabulations,this raised concern about the potential for disclosures even with random rounding. As a result, for the 1986 Census, a decision was made to suppress all "5s", after rounding, in custom tabulations of 20% sample data. While this causes some problems of aggregation, custom tabulations can be specified with a high degree of subtotaling to partially overcome this. Note that the decision to suppress "5s" applied only to custom tabulations and not to planned publications and small area data tabulations in machine-readable form since the latter show only a minimum amount of detail while publications contain detailed tabulations only at the level of Canada, the provinces and census metropolitan areas. The decision to adopt this approach represents a compromise of sorts and this entire issue will be further investigated for the 1991 Census.

III. DISCLOSURE PREVENTION IN THE CENSUS OF AGRICULTURE

In the Census of Agriculture, a large amount of very detailed information is collected. Each resulting data record on the data base contains more than 300 data fields or variables.

Much of the information collected, such as the values of all farm assets, detailed expenses and total sales, is sensitive business data.

Traditionally, aggregations of values collected on the Census of Agriculture questionnaires have been made available to users at five different geographic levels: Canada, the provinces, agricultural regions (which are relatively large areas usually corresponding to crop districts), Census Divisions (usually smaller areas of which there are 266 in Canada) and Census Consolidated Sub-divisions

(which are still smaller but which are at least 25 square kilometers in size). The pre-planned tables of aggregations at these levels cover approximately 20,000 pages.

At Statistics Canada, an aggregate is considered a disclosure of individual information if:

1) It contains only one farm,

2) It is an aggregation of values for more than one farm but so few farms are involved that neighbours can deduce an individual's information without much effort,

3) Although a reasonable number of farms contribute to the total, one or two farms contribute so much to it that to publish it would be to provide a close estimate of their values, or

4) Although a disclosure is not being made directly, a user can derive a value which is a disclosure by combining two or more aggregates which are not themselves disclosures. This last type of disclosure is known as a *residual* disclosure.

This definition has been translated into a set of formal sensitivity rules that are expressed in terms of exact percentages and numbers of farms. The actual values used in the rules are confidential.

Past experiences and study of variable distributions have shown that indiscriminate release of Census of Agriculture aggregated data at the geographic levels mentioned above would result in a significant number of disclosures. It has been observed that:

1) In many Census Consolidated Sub-divisions and even in some Census Divisions and Agricultural regions, there are very few farms because the land is unsuitable for farming, the area is pre-dominantly urban or because farms are being gradually combined together to form a small number of larger operations. A significant proportion of aggregates produced for these types of areas are likely to be disclosures.

2) Some commodities such as fruits or tobacco can only be grown in certain parts of the country and are therefore rare outside these areas. Others are rare because of limited demand or because special conditions or facilities are required to produce them. For these rare commodities, even province level aggregations can be disclosures.

3) At the Census Consolidated Sub-division level, even if the total number of farms is reasonably large, for many of the more than 300 variables, aggregates are disclosures since few variables are evenly distributed, across the country.

4) Distributions of amounts produced of certain commodities are sometimes highly skewed. For example, there a few very big turkey producers as compared to a large number of small operators and a small number of farms produce vegetables for the canning and frozen food industries while at the same time many market gardeners produce the same vegetables for direct sale. For these types of commodities, values reported by the bigger producers are large enough to dominate almost any aggregated value.

For some time, production of the large number of detailed Census of Agriculture tables of aggregated data has been automated. Prior to 1981, various procedures were used to avoid disclosures. One procedure used (prior to and in 1981) was the "Rule of Ten Farms" which did not allow release of data for a geographic area of less than ten farms. Another procedure consisted of designation of certain variables that would not be released below a given geographic level. A third procedure, used on frequency distribution data, required that categories containing less than three farms be collapsed with adjacent categories. Some cells dominated by one or two larger operators were suppressed prior to 1981 but this was not done systematically. Random rounding (i.e. random rounding of the number of farms reporting to a multiple of five and changing the aggregated value reported to preserve the original average) was used for user-requested tables and tables produced from the Agriculture Population Linkage data base.

A short time before the release of the 1981 Census data, a new Generalized Confidentiality system became available. This system was capable of removing every type of disclosure that affects the Census of Agriculture making use of sensitivity rules and a Collapse and Suppress approach as described by Sande (1977) and Cox and Sande (1979). With this approach, values in complementary cells are suppressed along with values in sensitive cells in order to avoid residual disclosures. The result is in effect a collapsing of cells. This system, where it was applied, was very effective. Unfortunately, it was not possible to apply the system to all of the tables produced, since determination of the cells to be suppressed in the types of detailed tables produced by the Census of Agriculture turned out to be a very large problem that consumed a large amount of computer and human resources.

For 1986, the system used in 1981 was not considered a viable alternative because it was:

1. not being actively maintained and therefore extremely difficult to incorporate into and use in a large production system.

2. a "prototype" system not designed to be used directly on a large retrieval data base. Interface programs would have to be developed that would be used to create system input files.

3. not designed to interface directly with print -ready tables and to make required changes. Therefore, additional interface programs would be needed to use results of the Confidentiality system.

4. Costly to run.

Random rounding was also rejected as an alternative because of a user preference for suppression rather than distortion.

It was concluded that there were not sufficient time and resources available to be able to fully develop a new methodology and system and there was no suitable ready-made system available. Therefore, development of a permanent (re-usable) system would have to wait until 1991 and that for 1986 a simplified custom system would be developed that could handle most of the worst disclosure problems with the help of some manual intervention by subject-matter analysts. Such a system was developed and it has been used to produce tables released in June of this year.

To briefly describe the methodology of the system that was used:

The first phase of the system consisted of the entry of a number of subject-matter analyst specified input files required by the system. An important file among them contained geographic codes corresponding to areas that were to be collapsed to avoid publishing for any area where most of the data would have to be suppressed anyway because of a small number of farms. Areas to be collapsed had been determined after analysis of 1981 tables and using results of an early test run of the 1986 system.

The next phase of the system was the analysis phase. The analysis program was run once and using it, aggregates of the each of the 309 Census of Agriculture variables were obtained for every geographic level to be published. While the values were being summed for each variable, the largest values contributing to each aggregate were determined and once the sum was available, calculations were carried out and comparisons were made to determine if any sensitivity rule was being violated. Information pertaining to all such cases including the name of the variable, the size of complement needed (in the one-dimensional case) and the type of rule violated was written on a computer file.

The second phase of the system, known as the suppress program, was applied to the final tables produced from the retrieval data base. Each table had been described on one of the input files entered earlier through a computer terminal using an inter-active program. The description had included codes corresponding to variables in each column, the type of geographic breakdown being

used and an identification of columns that were being summed to obtain sub-totals. For each table, the system checked the analysis file to determine if there were any sensitivity rule violations recorded for the variables in the table at the geographic levels included. If not, the table did not require suppressions. If there were records on the analysis file corresponding to the table, the cells in question were automatically suppressed and, if necessary, other cells were chosen to be suppressed in order to avoid complementary disclosures. Complementary cells were chosen using a prescribed geographic complements input file that linked adjoining areas together since it was a user preference that if collapsing was necessary, adjacent areas should be collapsed. The adjacent geographic complement was not used if another cell in the area had also been suppressed because of violation of a sensitivity rule and the two automatically suppressed cells could serve as complements for each other. If the prescribed geographic complement contained a zero value, a system algorithm chose some other geographic area as a complement – preferably one where other related variables were also being suppressed. Additional complements were chosen using a set of specified rules if a variable was being summed in the table with other related variables (e.g. milk cows, beef cows and heifers).

Edits were applied to the outgoing tables to ensure that no residual disclosures remained and that complements selected appeared large enough to avoid a disclosure. Edit failures were listed in a system report provided to subject matter analysts who reviewed it and a draft copy of the table and who could make use of a system feature that allowed on-line access to tables to suppress additional cells and to restore cells no longer needed to be used as complements.

Once the subject-matter analyst was satisfied with a table, a print-ready version was created.

The analysis program also handled planned cross-classified tables by calculating totals of all cells that would appear in these tables and at the same time applying the sensitivity rules. However, the suppression program did not handle these types of tables. Instead, subject matter analysts used a report from the analysis program along with a printed version of the table to manually select sensitive cells and their complements to be suppressed. The on-line access component of the system was used to make changes to the tables.

The major strength of this system was that it could be developed and tested in time for production of the outputs. Furthermore, costs of its development and of production runs have been reasonable and outputs, even if intervention of subject matter staff has been required, have been timely.

A second strength of the system is that it is well understood by subject matter staff and data users since they were heavily involved in its development.

The major weakness of the system is its limited flexibility. Only table formats used in the planned products can be handled by it. All variables, derived variables, variables to be cross- tabulated and geographic breakdowns had to be defined beforehand. The analysis program is designed to be run once or twice at the beginning of production and then not again. Also, major revisions to it would be necessary if table formats and variables were to be changed. The suppress program also has a limitation in that it has been designed specifically for tables produced using Statistics Canada's Statpak output system.

Because of its lack of flexibility, a second weakness of the system is that it does not handle ad hoc user requested tables. Approximately 1000 such requests were received for 1981 data and at least that many requests are expected after this census. Analysts must continue to handle these requests using the tools that are available to them (i.e. random rounding, suppression of cells with only a few farms reporting and visual analysis). There are plans to use reports produced by the Confidentiality system (one of them containing records for each of the sensitive cells detected by the analysis program and the other an evaluation file with information regarding every suppression) as an additional helpful reference tool.

A possible third weakness of the system is in the selection of complements. The complement selection algorithm is designed to select a complement according to an arbitrary rule that does not consider its size. Subsequent edits identify complements that are too small. However, there is no

further edit to ensure that the analyst responsible for the table has made an appropriate alternative choice. Validation of changes made by the analyst must be a manual operation. As such, it is subject to error.

A full evaluation of the 1986 Census of Agriculture confidentiality system is planned. Issues to be addressed include:

a) Costs
b) Quality of outputs from users' point of view
c) Effectiveness of the system in removing disclosures
d) Manual intervention required and consistency of subject matter decisions

At the time of writing this paper, users have only begun to receive tabulations produced using the new Confidentiality system. Although it is too early to assess user reactions, they are not expected to be negative since collapsing strategies used were influenced by expressed user preferences. Analysts involved in the preparation of the tables have expressed concern that some Census Consolidated Sub-division tables contain a significant number of suppressions and geographic collapsing and that this may frustrate users. Feedback from users will be monitored to determine if this is the case.

IV. CONCLUSIONS

This paper has described the procedures being used, for disclosure protection in the Canadian Censuses of Population and Agriculture.

In the case of the Census of Population the principal technique used is random rounding to the base 5. This technique has been used since the 1971 Census and in general has come to be accepted by census data users. The technique has the limitation of not being additive in the sense that totals are independently rounded and therefore do not equal the row or column sums. The use of random rounding has allowed for the development of a flexible data dissemination program that includes the production of large numbers of tabulations where data users specify the detailed content and geography. The geographic areas can be built up from block face level data.

While random rounding is the main disclosure avoidance technique used, it is in some cases supplemented by various rules that suppress data for very small areas or small cells in tabulations of certain variables.

In addition, for the 1986 Census,all cells of size 5 (after rounding) will be suppressed in custom tabulations based on sample data since it has been shown that in most cases these data are for single individuals. The problem of very small cells (particularly 5s) might be addressed by the introduction of what might be termed "double random rounding" for small cells. This would involve eliminating the 5s in tables by a second random rounding to either 15, 10 or 0. This procedure would be preferable to the current suppression since it would be un-biased; however for the 1986 Census it was not possible to fully assess the implications of using such a technique or to adapt it to the retrieval software currently in use.

While much further work can be done in the area of disclosure avoidance in the Census of Population, it is doubtful that any one technique will be developed to meet all concerns. The decision on which rules to apply will continue to represent a trade-off between minimizing the chances of disclosure and maintaining the usefulness of the data for research and policy purposes.

As planning begins for the next round of censuses in 1991, disclosure protection has been identified as one of the issues to be subjected to a comprehensive review. While the non- additivity of the current random rounding technique has not been a major concern of data users, careful consideration will be given to possible implementation of controlled random rounding. The experience with controlled rounding in the U.S. 1990 census will be watched carefully.

Attention will also be given to alternatives to suppression for dealing with small cells particularly in very small geographic areas.

Before deciding whether to develop a new Census of Agriculture Confidentiality system for 1991, the present system will have to be fully evaluated from the points of view of cost, practicality, data user satisfaction and quality given changing data distributions. Possibly, data releases at the lowest geographic level, the Census Consolidated Sub-division, may have to be eliminated for data quality reasons. The likelihood must be assessed that a generalized Confidentiality system would become available that could meet Census of Agriculture needs, and be easily adapted to be used with Output systems that will be employed in the future. The problem of user requests and the flexibility that they require is also an important consideration.

REFERENCES

1. COX, LAWRENCE H. (1983): "Some Mathematical Problems Arising from Confidentiality Concerns", Statistical Review, 21, 5, Statisitcs Sweden, Stockholm, 179-189.

2. COX, L.H., JOHNSON, B., MCDONALD, S.-K., NELSON, D. and VASQUEZ, V. (1985): "Confidentiality Issues at the Census Bureau", Proceedings of First Annual Census Bureau Research Conference, Reston, VA.

3. COX, L. and Ernst, L. (1982): "Controlled Rounding". INFOR, 20, 4, pp. 423-432.

4. COX, L. and Sande, G. (1979): "Techniques for Preserving Statistical Confidentiality.", International Statistical Institute, Vol. 3, Proceedings of the 42nd Session, Manila.

5. FELLEGI, Ivan P. (1972): "On the Question of Statistical Confidentiality", J.A.S.A., March, 1979, Vol. 67, No. 337, pp. 7-18.

6. FELLEGI, Ivan P. (1975): "Controlled Random Rounding", Survey Methodology, 1, Statistics Canada, 123-135.

7. FELLEGI, I. and PHILLIPS, J. (1976): "Statistical Confidentiality: Some Theory and Applications to Data Dissemination". Annals of Economic and Social Measurement. pp. 399-409.

8. NARGUNDKAR, M.S. and SAVELAND, . (1972): "Random Rounding: A Means of Preventing Disclosure about Individual Respondents in Aggregate Data". American Statistical Association, Proceedings of the Social Statisitics Section, Washington, D.C. pp. 382-385.

9. SANDE, G. (1977): "Towards Automated Disclosure Analysis for Enterprise-Based Statistics", Statistics Canada, unpublished.

10. U.S. DEPARTMENT OF COMMERCE (1978): "Statistical Policy Working Paper 2: Report on Statistical Disclosure and Disclosure Avoidance Techniques". U.S. Government Printing Office, Washington, D.C.

Ca 008

STATISTICS CANADA LIBRARY
BIBLIOTHEQUE STATISTIQUE CANADA

1010248885

E. 2