APPLICATIONS OF THE GENERALIZED EDIT AND IMPUTATION SYSTEM AT STATISTICS CANADA

P. Whitridge and J. Kovar, Statistics Canada
P. Whitridge, 11C RHC, Tunney's Pasture, Ottawa, Ontario, K1A 0T6

ABSTRACT

As part of the Generalized Survey Function Development Project at Statistics Canada, the last three years have been spent developing a Generalized Edit and Imputation System (GEIS). This paper gives a brief overview of GEIS and some indication of the early experiences in its use. Emphasis is placed on how to apply the system rather than on a description of its capabilities. In particular, methodological issues such as the development of edits and the imputation strategy are addressed. Recent experiences in several applications are used to illustrate these points.

Key words: survey data editing, data quality, imputation strategy, generalized systems

1. INTRODUCTION

In 1985, Statistics Canada undertook a major project called the Business Survey Redesign Project (BSRP), with the goal of redesigning all of the Bureau's economic surveys. The development of generalized software was an important component of this project that emphasized attempts to conserve resources and eliminate duplication (Outrata and Chinnappa, 1989). The main goal of the Generalized Survey Function Development (GSFD) project was to develop generalized tools that could be adapted easily to the majority of business surveys which would undergo redesign as part of the BSRP. The systems would be based on a limited set of standardized methodological approaches designed to improve timeliness, reduce respondent burden and minimize resources in the production process. In addition, these systems would be flexible enough to permit new modules to be incorporated as they were developed, and would be portable across various system architectures and sites. The systems to be developed were: Generalized Sampling System, Generalized Data Collection and Capture System, Generalized Edit and Imputation System, and Generalized Estimation System.

In developing the generalized systems, the task of edit and imputation has been broken into two stages: preliminary editing, which is done at the data collection and capture stage (Berthelot et al, 1989), followed by automatic edit and imputation. This is different from the current approach which consists of error detection followed by manual correction as the records are received and reviewed. In this approach, several courses of action are available for the correction, including following-up the respondent, overriding the edit, excluding the record or manually supplying ad hoc values. In the GSFD, it was decided that only records that were unresolved after the preliminary edit stage and those with lesser impact would be sent to the Generalized Edit and Imputation System (GEIS), as a last resort. At that stage, all attempts would be made to resolve the remaining cases by automatic imputation.

GEIS, while still undergoing some development, has been in use at Statistics Canada for the past three years. Experience has been gained on how an application should best approach the use of GEIS. This paper concentrates on the applications that have used GEIS, representing both small and large surveys and administrative data.

The paper is divided into five parts. Section 2 provides an overview of GEIS. Section 3 describes the general points that should be addressed by an application considering the use of GEIS. Section 4 outlines the experiences of three different applications: the 1991 Census of Agriculture, Income Tax Data, and the Annual Motor Carrier Freight Survey. The paper concludes with a short summary of the insights gained from these projects.

2. OVERVIEW OF GEIS

GEIS consists of three major components: edit analysis, error localization, and imputation (Kovar, MacMillan and Whitridge, 1988). Each component will be addressed in turn in this section.

The objective of editing is to determine whether a given data record contains invalid, missing, inconsistent or outlying responses by applying certain rules to the data. In other words, editing is the act of error detection. Imputation is the task of replacing an invalid, missing, inconsistent or questionable value with a plausible one. Effectively therefore, imputation is error correction. The two functions are linked through error localization, which is the process of determining which fields to impute.

Historically, editing has been a manual process, with the edits taking the form of "if then else" conditions. Rules of this type effectively perform both edit and imputation: the "if" condition is the edit and the "then" or "else" condition is the imputation. The edits are applied sequentially to the data, with corrections being made for each failing edit. These corrections are often subjective and usually not reproducible. In certain circumstances, corrected fields can be changed subsequently if they fail another edit.

The philosophy behind GEIS is based on the Fellegi-Holt approach to edit and imputation which recommends that the *minimum* amount of respondent data be changed (Fellegi and Holt, 1976). To achieve this, GEIS considers the edits as a set of linear constraints which defines a feasibility region with good records inside and records requiring imputation outside the region. The fields that require imputation are then determined as a function of the edit failures for each individual record. In GEIS, the editing, error localization and imputation are separate steps (Giles and Patrick, 1986). These steps are described in more detail below.

2.1 Editing

Editing in GEIS consists of iteratively specifying, analyzing, and applying the edits. The edits must be linear, and the data are assumed to be numeric, non-negative and continuous. If necessary, it is often possible to transform the data or the edits to satisfy these conditions. In addition, GEIS assumes that all follow-ups have already been performed at the collection and capture stage, and that nothing more is to be gained by re-contacting the respondent or by referring to the questionnaire. This allows the system to be fully automatic, requiring no manual intervention in the production

process.

The edit component of GEIS includes several functions that help users analyze their edits. First, a program checks that the edits are consistent, identifies any redundancies, provides a list of bounds for each variable, and specifies the minimal set of edits to be used for further processing. A second program determines the extremal points of the edit set. These are the vertices of the feasible region and describe how "bad" the data is allowed to be and still be accepted by the edits. A third program forms linear combinations of the edits, eliminating one variable at each step and producing a set of edits which shows users the conditions implicit in the edits. This helps users in examining and confirming the edit logic. Once the user is satisfied with the edits, the edits should be fine-tuned as a function of the data. The last function of the edit component applies the edits to a set of data, possibly from a pilot test or a previous survey, and provides counts of failures by edit rule, by fields involved in each edit failure, by number of rules failed etc. These diagnostic measures help the user in deciding whether the edit rules are too rigid and need to be relaxed, or are too relaxed and should be tightened.

2.2 Error Localization

The second major component of GEIS is error localization. This module identifies the minimum set of fields or variables that requires imputation for each record that fails at least one edit. The fields to impute are determined by the data on the record and the edits. Weights may be incorporated to take into account varying levels of reliability of fields. For example, if a respondent is more likely to supply a valid total than the components of the total, then the total would be assigned a higher weight, and hence would be less likely to be imputed. Weights might also be used to preserve certain characteristics that are reported only rarely, since choosing such a field to change in error localization could result in distortion of the reporting frequencies. In GEIS, weights cannot be specified differently for individual records.

2.3 Imputation

Once the fields to be imputed have been determined, what remains is simply the imputation. GEIS provides three main types of imputation. The first is deterministic imputation, in which a value is supplied if there is only one possible way to fix the fields, given the edits and the valid responses on the record. For example, if the components of a total were provided but the total was left blank, it would be deterministically imputed as the sum of the components.

The second type of imputation uses "imputation estimators", which estimate the missing values using previous observations, means, trends, or ratios. One variable is imputed at a time, and the resultant record does not necessarily satisfy the edits.

The last type of imputation is a donor method using the nearest neighbour with respect to a set of matching fields in a pre-specified imputation region. The matching fields are determined for each record in error. They may be system-generated as a function of the edits, specified by the user, or a combination of the two methods. These matching fields are then transformed to a (0,1) range and a search tree is created. The tree is traversed for each record in error and the nearest neighbours are found. Details of the method can be found in Sande (1979). The closest neighbour is used to supply the missing fields, and the imputed record is re-edited. The edit rules for post-imputation re-editing need not be the same set of edit rules that were used in error localization. These "post-imputation" edits might be a relaxed set of the original edits. If these edits are not satisfied, then the next closest neighbour is tried, until either the record is successfully imputed, or the supply of nearest neighbours is exhausted. It is also possible to limit the number of donors that the system tries, if an application prefers to use a different method if a suitable donor could not be found within the first n donors.

Throughout the system, GEIS provides reports to help managers track the process. These performance measures include counts of edit failures by edit, record and field; frequency of imputation by field and by method; identity of donor used; match fields used, etc.

3. GENERAL ISSUES FOR APPLICATIONS USING GEIS

It can be assumed that almost any survey process will require an edit and imputation system. However, what edits and what methods of imputation to use are decisions that must be considered individually by each application. The main objective of GSFD is to provide tools that can be adapted to the majority of surveys. Methodologists can then concentrate their time on developing comprehensive edit and imputation strategies and testing alternatives, rather than writing specifications and testing new systems. The primary issues that must be addressed in the development of any edit and imputation strategy include determining and fine-tuning the edits, choosing imputation methods, and designing the evaluation (see section 3.3). First, a preliminary investigation is needed to decide whether GEIS is feasible for a specific application. GEIS incorporates functions to help the methodologist analyze and fine-tune the edits, but the original edits themselves must be determined outside of the system. There are different imputation methods available, but the applications must decide which methods suit them and specify values for numerous parameters. These include, among others, auxiliary variables to be used with some of the estimators or minimum criteria for calculating means. The question of how to evaluate the imputation is left to the users, though the system makes it easier by providing some summary tables. The next sections address these three issues in order.

3.1 Development of Edits

When an application is considering the use of GEIS, there is a tendency to want to implement the same edits and imputation methods that were in place previously. This is not recommended, since, typically, the old edits are "if then else" rules with the error localization and imputation action implicit in the rules. This type of rule can usually be specified in a linear form, but the particular "then" and "else" actions are lost. Also, these edits identify invalid conditions rather than clean records. Even though the system accepts such "conflict rules" as input, the hazard in this approach is that any unforeseen condition is implicitly considered as acceptable, since that condition was not specified as being incorrect. In the case of continuous data, it is generally better to start by describing a clean record rather than by trying to enumerate all possible invalid conditions.

Several possible sources to help determine the edits must be considered when an application is defining its edit

requirements. The primary source is the subject matter experts.

Subject matter knowledge is highly variable since it depends upon the expert, but it can be extremely useful in identifying possible combinations of variables that should be considered together. The subject matter experts would specify context sensitive relationships (e.g. no farm would have more than 1 bull for every 100 cows (cows < 100*bulls)), sensible bounds (e.g. no cow can produce more than 100 litres of milk per week (milk < 100*cows)), and existency edits, where the presence of one variable indicates the presence of another (e.g. if sales of milk was reported, then dairy cows must be reported).

The second main source of edits involves a logical examination of the questionnaire. Certain relationships are evident, for example, components that sum to a total and variables that are subsets of others (e.g. number of cows milked < = total number of dairy cows). The structure of the questionnaire will identify logical blocks of variables, and pairs of related variables that should be considered together (e.g. # of tractors, $ value of tractors).

Previous or related surveys, administrative data, or pilot studies/surveys are another source that can be useful. Typically the edits, regardless of the form, can be helpful in identifying sets of variables that should be examined at the same time. For example, tax data cover all industries, so the rules used to edit the retail sector of tax data are potentially useful for retail surveys in general. If the underlying concepts of the two sources of data are not the same, then the constants involved in the edits should be recalculated, but the same variables may well be edited together. Edits used for previous surveys should be adjusted based on their failure rates, especially if different definitions are involved. In some cases the constants would require recalibration for inflation.

The last and perhaps the most powerful source for determining edits is data analysis. Principal components analysis techniques can be used to group similar variables together. Correlations can be calculated to determine which variables within a group are related. Graphical analysis or other exploratory data analysis techniques can be used to specify bounds to be used in edits. Statistical edits can also be defined based on the distribution of certain variables or the ratio of certain variables. At present, the tools for such analysis are not included in GEIS.

Due to operational constraints, GEIS cannot edit more than 40 variables at a time. This means that an application must divide the variables into groups of no more than 40. The edits are then formulated for the variables in a group. This grouping of variables can be chosen to correspond to sections of a questionnaire (e.g. livestock, crops, expenses) or to some other logical arrangement of the variables determined by data analysis techniques as explained earlier. Care must be taken to ensure that a variable is not imputed more than once. There are facilities in GEIS to help ensure this.

When the edit rules have been determined, they should be analyzed using the facilities available in GEIS, as discussed in Section 2. If the application includes any variables that could have valid negative values, then they should be transformed by the addition of a constant to make them positive. Any edits involving these transformed variables would need to incorporate the same constant.

Once the edit rules have been analyzed, GEIS can be used to apply the edits to preliminary data (test data, pilot survey data, historical data, or data from other sources) to fine-tune the rules according to the failure rates. This function quickly shows the user if an edit fails in a very high percentage of cases, or never fails at all. These are the two extreme cases when the edit would likely require adjustment of the constants.

The complete set of edits should be shown to the subject matter experts for their acceptance. This is important since the subject matter experts have a great deal of knowledge and experience in editing the data and would have specific concerns which should be addressed. The subject matter experts work with the data at later stages for analysis and validation/certification so they must accept the edits that have been applied at the edit and imputation stage, in part because the final responsibility for the quality of the data produced by the application rests with them.

3.2 Imputation Strategy

The imputation strategy specifies methods of imputation to be used such that the final product is a complete, fully-imputed set of data. Ideally, relationships between variables and higher moments of distributions should be preserved, but this is not always possible. The best imputation strategy depends upon the characteristics of an individual application. The issues to consider include the level of aggregation at which imputation should take place, the choice of imputation methods and certain specific imputation questions.

The level at which to impute refers to the imputation region or group of records that are considered to be similar and within which donors will be found for records requiring imputation, or which will be used to calculate means and trends. It is important to ensure that geographical and classification structures are preserved. It is also desirable to avoid too many imputation groups, as each imputation group must be processed individually and this increases the number of jobs to be run.

GEIS incorporates three different types of imputation methods, as described in Section 2. For most applications, the donor method is recommended, but the clear exceptions to this include repeated subannual surveys whose characteristics are highly correlated over time, and surveys with very heterogeneous populations in which it is believed that the "nearest neighbour" is the same unit in the previous cycle. Many applications specify the donor method as the primary one, but use estimators as back-ups in case no appropriate donor is available. Surveys with a large number of variables find the imputation estimators cumbersome to use, since a method including any necessary auxiliary information must be specified for each variable requiring imputation. If estimators are to be used, then the order in which they are applied is important, since it is desirable to impute one variable before using it as auxiliary information for imputing another variable. Correlations between variables should be examined when auxiliary information is required.

Certain other imputation questions remain. First, donor imputation requires the specification of post-imputation edits to be used to determine if a donation is successful, as explained earlier. These edits could be the same as the original edits used in error localization, or they could be a relaxed version of the same set. For example, it might be desirable to relax an equality edit into two inequality edits that bound the original equality. Second, if estimators are used, some applications require post-imputation edits to verify the imputation, but this is not easily implemented in GEIS. Re-processing records through error localization is one

possible solution. Third, if donor imputation is used, then, if desired, matching variables can be specified by the user and included with the matching variables determined by the system. This facility can be used to bring variables that are not part of the block of variables being edited into the search for a good donor. For example, in business surveys, key variables such as Gross Business Income or Total Sales can be considered as matching variables for all donor searches. Fourth, minimum imputation criteria can be specified, e.g. the number of records used to calculate a mean or trend, or the minimum donor population size. This is useful in controlling the reliability and stability of the values to be imputed. Finally, it is sometimes necessary to exclude certain records from the donor population, such as zero values, or suspicious or outlying values. This must be considered as part of the imputation strategy.

3.3 Evaluation Strategy

Any application considering the use of a generalized system must first determine whether or not the system is appropriate for that particular application. For example, GEIS imputes continuous numeric variables, so it might not be suitable for an application that involves a large number of categorical or qualitative variables. Usually a feasibility study is carried out to determine whether or not the application should proceed with the implementation of the system. However, the question of how to evaluate the impact of the imputation remains to be addressed. This evaluation often includes examining different imputation strategies that would have been tried in the study.

Evaluation tables that consider counts such as the number of times a reported value was increased, or decreased, or remained unchanged, and the corresponding estimates may be used to evaluate the impact of imputation. Typically, the tables would be produced at several different levels of aggregation, perhaps corresponding to the estimates to be tabulated. They can be used as input to feasibility studies, as well as post-processing documentation.

Evaluating the impact of the imputation is often difficult, since it requires a pre-specified idea of the "true" answers, or of how much imputation is acceptable. These are very subjective measures. Is 5% imputation for one field too much? It depends upon the response rates, the reasons for imputation, and which specific field is being imputed. For certain key variables that are always reported, such as Total Revenues, this could be too high, whereas for other minor variables that are poorly reported, such as certain Other Costs, higher rates may be acceptable. If an application has a high rate of partial non-response for which they are imputing, then zeroes will be replaced by positive values, and the impact will be in a positive direction only. The evaluation should be done co-operatively between the subject matter experts who are responsible for the data being published and the methodologist who is designing the edit and imputation system.

4. SPECIFIC APPLICATIONS

GEIS has been in use at Statistics Canada for the last three years, while it has been undergoing continued development. Experience gained from the early applications has helped determine directions for this development, and has made the product more efficient and user friendly. This section describes the experiences of three applications: the Census of Agriculture, Income Tax Data, and the Annual Motor Carrier Freight Survey.

4.1 Census of Agriculture

In September, 1989 a working group was established to study the feasibility of implementing GEIS for edit and imputation of the 1991 Census of Agriculture. Various processing scenarios were examined and an overall recommendation was made. The primary problems associated with the Census were the large volume of records (300,000 farms) and the large number of variables (320) to be edited and imputed. Of equal concern was the known heterogeneity of the population.

It was decided that the best way to proceed for this application would be to develop an actual prototype and to pass some data through the complete edit and imputation system in order to test the preliminary edits and the imputation strategy. The section of the questionnaire dealing with livestock was selected and seven imputation regions from across Canada were chosen. The livestock section had 29 variables, so it was ideal for one edit group. It was felt that livestock was a good choice since there would be a reasonable number of relationship edits that could be derived and since editing and imputing this section was of at least average difficulty. The seven imputation regions involved 35,000 records, so they would adequately test the volume. Since the regions were from across Canada, they would represent different types of farms: beef cattle in Alberta, pork producers in Quebec, and dairy farms in Ontario.

The questionnaire for the Census has a number of yes/no boxes to indicate whether or not a farm has certain characteristics such as fruit trees, field crops, or cattle. If the box is ticked yes but no data are provided, then the entire section of the questionnaire requires imputation. Such response patterns should be detected by the capture system. However, the Census data capture system does not do so, therefore a pre-processor program was written to help resolve such cases.

Edits were determined based on subject matter edits that had been used in the previous Census, current subject matter knowledge, intense data analysis and examination of the questionnaire itself. The data analysis yielded ratio edits between pairs of variables. The bounds for the ratios were adjusted with respect to the slopes and the intercepts with the axes of the linear relationship between the variables until the units that were rejected because they fell outside the bounds were considered to be those that needed to be imputed. The analysis was based upon a combination of data from the 1986 Census and data from a pilot test that took place in 1989. The edits were presented to the subject matter experts and explained graphically. It was felt that these edits might be better than the edits used in 1986 for the identification of the units to be imputed and that they would reflect the heterogeneity of the population correctly. A mechanism was put into place to fine-tune the edits based on a sample of the incoming 1991 data.

The prototype simulated different processing scenarios that might be used in production. All the regions were processed together to simulate one entire province. Data from many sections of the questionnaire and some very preliminary edits were run together to test the system under great volume. Different combinations of edit and data groups were then run to determine the optimal scenario.

The impact of the imputation was evaluated using tables

that broke down the imputation by reason for imputation: whether due to non-response or edit failure. The number of times a field was changed for each reason, as well as the magnitude of this change was examined at an aggregate level. Based on these tables and the comparison of the resulting estimates with those of 1986, it was felt that the overall quality of the data after imputation would be as good if not better than that of the 1986 Census of Agriculture, which was considered "excellent".

The overall conclusions of the feasibility study were positive (Statistics Canada Working Group, 1990). The working group recommended that GEIS be used for the 1991 Census of Agriculture and this was accepted by senior management.

4.2 Income Tax Data

At Statistics Canada, the list of Income Tax filers is used as a sampling frame for various business surveys. The data on this list has not been imputed in the past. The situation is changing as economic surveys are encouraged to take advantage of administrative sources of data especially for small businesses. Studies have recently taken place to determine if it would be feasible to edit and impute Income Tax data using GEIS.

Based on the results of these studies, it was decided to use GEIS to perform automatic edit and imputation on Income Tax Data from Tax Year 1988. The issues to be addressed included differentiating between zeroes and missing values and developing an edit and imputation system that would satisfy all potential users of the tax data, both those who are interested in one industry such as Construction, and those who cover all industries such as Small Business Statistics. The system had to be the best for the set of all industries, which meant that it might not be the best for some specific industries.

The problem was also a question of volume: 200,000 records were involved with 24 variables each. The file was divided into imputation groups according to industry; only one edit group was required.

Businesses supply their financial statements to Revenue Canada Taxation (RCT) in whatever format they choose; if the data is complete then RCT is satisfied, since they use the data for compliance audits only. However, when Statistics Canada transcribes data from the financial statements, there are often items that the business grouped together or specified in a form that is incompatible with the transcription form. These items are then captured as zeroes, when they should really be identified as missing values. The main objective of the edit and imputation function for tax data is to identify these missing values and replace them with valid values. This issue of missing values is really one of non-response that should be identified at the capture stage. The edits to perform this task are largely conditional rules that are not suited for implementation in GEIS. To solve this problem a program was written to pre-process the data and supply GEIS with flags indicating which values were missing and should be imputed.

Edits were developed by subject matter experts for their specific areas of interest: some were for one specific industry, and others covered all industries. The edits were assembled into one group for each industry and then analyzed. Any inconsistencies or redundancies were resolved co-operatively between the appropriate experts. The final set of edits was approved by all subject matter experts involved.

Transcription and capture of the tax records for Tax Year 1988 was completed in March 1990. The edit and imputation was run early in the summer of 1990. The preliminary results were positive, and a more thorough evaluation of the impact of imputation is expected (Block, 1990). It has been decided to use GEIS for the edit and imputation of tax data for Tax Year 1989 as well.

4.3 Annual Motor Carrier Freight Survey

The Annual Motor Carrier Freight Survey is currently undergoing a redesign and edit and imputation was identified as one area which should be addressed. The system in place relies on clerical staff who manually perform detailed edits, follow-ups and imputation. For this reason, a study was undertaken in February, 1990 to investigate the feasibility of using GEIS for this application. Various imputation methods were examined in order to gain information that would help determine an optimal imputation strategy.

The feasibility study concentrated on one section of the questionnaire for one class of motor carriers. Data from the previous year was used for the population of 1300 units. The twenty-nine variables on the Balance Sheet were considered. A large number of edits were specified, mostly defined by the accounting structure of the Balance Sheet. The edits were complex and highly recursive, some dealing with a subsequent subtotal. There were many equalities. A pre-processor program was required to correctly error localize the cases where a total was provided but the components were all missing, so that the entire set of components would be identified to be imputed. Specific problems that had to be addressed concerned the large amount of non-response and the complex pattern of edits required by the survey. Some problems were encountered in the error localization module during the study due to the large number of variables and edits.

The results of the feasibility study were evaluated by comparing the estimates obtained from the imputed file to those published by the survey. The different imputation strategies were examined individually, but the overall conclusions with respect to the optimal imputation strategy were inconclusive. As a result of the feasibility study, a recommendation was made to use GEIS for the Annual Motor Carrier Freight Survey, if some minor enhancements could be made to the error localization module of GEIS (Gossen, 1990). Further study was recommended for establishing the best imputation strategy.

5. SUMMARY

The overall experience to date with GEIS has been positive. The system allows methodologists to spend more time developing edits and an imputation strategy instead of writing specifications and testing new systems. The applications that have studied GEIS have provided valuable input to the development team working on GEIS concerning additional functions that need to be included in GEIS and how to make the system more user-friendly. Many enhancements currently being programmed were designed as a direct result of requests from the users.

A common concern about the system is the need for a pre-processor program to treat the data before it can be processed by GEIS. This was the case for all three applications described here. It is felt that such a program will

not be necessary once an intelligent data capture system, such as the Generalized Data Collection and Capture System, is in place. For example, this system would correctly identify that all missing components of a total required imputation, unlike GEIS which would identify only one field amongst the components if the missing items were recorded as zeroes.

Currently, surveys that are considering the use of GEIS tend to prefer to simply implement existing edit rules using GEIS instead of taking the opportunity to re-examine the way the edit and imputation has been done in the past. Such a review would show that the edit rules should be re-developed for use in GEIS to ensure better quality of final data. The need for a pre-processor to handle conditional edits, as in the tax application, reflects this problem. It is suggested that applications should develop a set of linear edits that take advantage of the way GEIS functions. Similar edit requirements to those currently in place can be used in GEIS, but the actual edit specifications would look very different. This problem should resolve itself with time, as applications become more accustomed to working with this generalized system.

REFERENCES

Berthelot, J.M., Danis, J., Latouche, M., Lindeyer, J., Horricks, T., Ross, E. and Ruscher, P. (1989). Methodological and Operational Concepts in the Collection and Capture Module. Statistics Canada Technical Report.

Block, C. (1990). Report on the Edit and Imputation of the NIP Tax Sample for Tax Year 1988. Statistics Canada Technical Report.

Fellegi, I.P. and Holt, D. (1976). A Systematic Approach to Automatic Edit and Imputation. Journal of the American Statistical Association 71, 17-35.

Giles, P. and Patrick, C. (1986). Imputation Options in Generalized Edit and Imputation Systems. Survey Methodology 12, 49-60.

Gossen, M. (1990). GEIS Feasibility Study for the Annual Motor Carrier Freight Survey. Statistics Canada Technical Report.

Kovar, J.G., MacMillan, J.H. and Whitridge, P. (1988). Overview and Strategy for the Generalized Edit and Imputation System. Statistics Canada Methodology Branch Working Paper No. BSMD 88-007.

Outrata, E. and Chinnappa, N. (1989). General Survey Functions Design at Statistics Canada. Bulleting of the International Statistical Institute, Proceedings of the 47th Session, Book 2, 219-238.

Sande, G. (1979). Numerical Edit and Imputation. Presented at the 42nd Session of the International Statistical Institute, Manila, Philippines.

Statistics Canada Working Group on the Use of GEIS for the 1991 Census of Agriculture, (1990). Statistics Canada Technical Report.