1.8 FEBRUARY 1975. DROP-OFF SURVEY ON RETIREMENT AND PRE-RETIREMENT CIRCUMSTANCES.

1. OVERVIEW:

In this section we will describe (a) the survey design (methodology), (b) the response rate, (c) the editing procedures, (d) the link to the Labour Force Survey and screening questions and (e) the weighting procedure.

A. THE SURVEY DESIGN (METHODOLOGY):

If the Labour Force Survey respondent was in the relevant age group and was rotating out of the LFS sample, a separate questionnaire, for self-enumeration, was dropped off at the time of the regular Labour Force Survey interview. The target population was all persons 55 years of age and over who were either still working (pre-retired) or who had retired at the time of the interview (retired). This age group was identified through question 27A on the regular L.F.S. document. Screening questions were also asked to determine whether the respondent should receive a Pre-retired questionnaire or a Retired questionnaire. For a detailed description see the control document used, in Appendix E. The questionnaires were to be picked up by the enumerator during the week following the Labour Force interview. If certain questionnaires were not available at the time of call-back, a self-addresed return envelope was left with the respondent. Also, if the respondent needed some assistance in completing the questionnaire, the interviewer was instructed to offer assistance but only when requested to do so.

B. THE RESPONSE RATES:

Of the 73,736 individual responses to the Labour Force Survey in February 1975, the expected number of responses to the Retirement-Pre-Retirement Survey (that is persons expected to be 55 years of age and over) was approximately 3,500. This estimate was based on the proportion in that age category in the total LFS in other months and on the fact that one-sixt the sample rotate ou each month. The actual number of respondents subsequent to data capture (key-edit) was 2,539.

During the data processing phases some 121 individual responses proved unable to be linked to their LFS record. This left a final response of 2,418, of which 828 were Pre-retired and 1590 were Retired.

C. THE EDITING PROCEDURES:

These 2,418 records served as input to and edit procedure, namely, (1) manual coding and editing; (2) computer editing and update using GEISHA (Generalized Editing and Imputation System Using the Hot Deck Approach).

The Retirement/Pre-Retirement Survey questionnaires were subjected to a manual scrutiny before data capture. Each questionnaire was checked for double entries, clarity of occupational description, and generally an attempt was made to remove any ambiguity in the responses and so to facilitate date capture.

Following the manual coding and editing stage, the Retirement/Pre-Retirement Survey questionnaires were sent to the Central Coding Unit of the Labour Division in Statistics Canada for occupational coding (CCDO). The questionnaire contained five questions which required this coding. All the valid occupational codes were keyed into the computer editing program. The occupational codes as captured were then compared against this list of valid codes.

Following the occupational coding, the questionnaires were then sent to the Data Processing Division for data capture employing the key edit technique. The key editing output was subsequently transferred to magnetic tape which interfaced with the computer edit program as described below. Although a margin of error exists throughout all stages

of survey processing, rough estimates are only possible for the data capture. The data capture could have carried approximately a 2% error rate.

The automated edit stage was composed basically of two main parts:

(1) a pre-edit of the data which did arithmetic and quantitative editing and prepared the file for loading into the GEISHA database, and

(2) GEISHA, a generalized edit and imputation system using the hot-deck approach, based on a theory developed by Fellegi and Holt.

The basic concept of the GEISHA system is to express those values of certain variables which lead to a conflict in the data. Once such a conflict is found, the system will impute values for some fields in order to make the record conflict-free. The fields chosen for imputation will always be the minimum number possible in order to correct the problem. Thus if three fields are tied together in various edit rules, the system will try to change only one field if it is possible to do so and satisfy all edits. The records it searches for the value to use in imputation consist of all records with no conflicts in the same strata. In this survey, strata were:

Pre-retired, urban
Pre-retired, rural
Retired, urban
Retired, rural.

Within the records searched, the system seeks to match the values for as many of the variables involved in the edit rules with the field to be changes as possible. When, a "clean" record (passed all edit rules) is located which matches on as many variables as possible, the value it contains in the field to be imputed is inserted on the originally conflicting record. The record from which the value was taken will not be used for imputation again until all other records which match on the related variables have been used. In this way, the distribution of the orginal clean data is maintained through the imputation process. This is a feature of all hot-deck systems.

D. THE LINK:

As described earlier the Labour Force Survey data and Retirement Survey data employed different techniques of data collection and data capture. As a result, the two data tapes required merging before editing was done. Each data record from the Retirement Survey was linked to its associated LFS data record via a unique fourteen (14) digit identifier which theoretically is the same on both records. However, problems complicating this link were: (1) the possibility of error on the coding of the LFS identification by the interviewer; (2) the possibility of error at the time of coding the Retirement Survey identification; and (3) the possibility of error while key editing the Retirement Survey questionnaires.

Due to the limitations pointed out above, a certain margin of error can be expected in the linking process. The objectives of the link are to match the Retirement Survey data record to its associate LFS data record and to minimize data lost through non-matches. The link is a three-stage process.

STAGE 1:

The match was performed on Identification, Age and Sex and Marital status, the full identification information carried on the drop-off.

STAGE 2:

The non-match records from Stage 1 were then subjected to a manual scan which employed a combination of the best information. Age and sex were compared and minor adjustments were made to the Retirement Survey identification to make it compatible with the LFS identification. to make it compatible with the LFS identification. This in most cases involved one or two blank columns in the Retirement Survey identification. The corrections were made to the Retirement Survey non-matched tape prior to the second matching pass.

STAGE 3:

Apart from the problems caused by Retirement Survey records with blanks in the identification fields, other non-matches result from blank responses in the age or sex field of the Retirement questionnaire. At this stage, the match was done on identification and sex. The missing information for age was inputed from the Labour Force record. No match was made for records with sex missing due to the different patterns of work and retirement for men and women. The number of records matched in total was 2,418.

E. WEIGHTING PROCEDURE:

As discussed earlier, the Retirement Survey and LFS employed different techniques for the collection of data. Quite naturally, they have different response rates. Each LFS record contains a final universal weight as derived from the Labour Force Survey weighting stage. As the Retirement Survey represents a distinguishable sub-population of the Labour Force Survey population, all persons not responding to the Retirement Survey were considered purely as non-respondent. The final universal weights for the Retirement Survey were derived by adjusting the LFS weights for this non-response and for the procedure of taking only a sub-population of the LFS sample.

The procedures followed for this adjustment were developed by the Labour Force Survey methodologist of the Household Survey Development Staff. As pointed out in Section 1.4, the final LFS weight is created at the Province, Age, Sex level. The adjustment for non-response is made by creating a ratio of expected responses (the aggregate weight to the Labour Force Survey within a particular Age, Sex, Province category) to the actual response (the aggregate weight to the Retirement Survey within a particular Age, Sex, Province category). This ratio is then applied to the

weights of each individual record within that grouping.

In order to achieve this readjustment, at least one response to the Retirement Survey must be found for each age, sex, area group for which responses exist for the LFS. If at least one response for each age, sex, area group cannot be found for the Retirement Survey, collapsing of characteristics must take place.

Total Response to the February 1975 Labour Force Survey (55 + yrs.)	17,687
Target Retirement/Pre-Retirement Responses (1 rotation group)	3,011
Response to Retirement and Pre-Retirement After Collection and Capture	2,539
Invalid Records Due Either to Totally Blank Identification or to Totally Blank Responses to the Retirement and Pre-Retirement Questionnaires	0
Responses as Input to Edit and Matching Programs	2,539
Unmatched Responses	121
Matched Responses	2,418
Responses Declared Invalid by Subject Matter Update Procedures	0
Responses Contained on the Final Tape	2,418

This leaves a Response Rate: (1,590 Retired) (828 Pre-Retired)

Usuable Response	2,418
Expected Response	3,011
Response Rate	80.3%

These population figures should be consistent with those published by the Labour Force Survey for February, 1975. Since the adjustment

was made on the basis of demographic (age, sex) and geographic province characteristics only, one would not expect to be able to derive comparable estimates for any other characteristics (Labour Force Status, etc.). Therefore, in release of any of these characteristics it should be pointed out they are not official estimates and vary from published figures as a result of the different response rates to the two surveys.