

**Canadian Community Health Survey
Cycle 1.2, 2002
Dummy File for Remote Access to the Master File**

January 2005

Notice

The Dummy File should not be used for purposes other than to develop and test the computer programs that are to be submitted by remote access. The Dummy File contains modified data, and must never be used for analytical purposes.

1. Background and Overview

The Canadian Community Health Survey (CCHS) is a cross-sectional survey that collects information related to health status, health care utilization and health determinants for the Canadian population. The CCHS operates on a two-year collection cycle. The first year of the survey cycle “.1” is a large sample, general population health survey, designed to provide reliable estimates at the health region level. The second year of the survey cycle “.2” is a smaller survey designed to provide provincial level results on specific focused health topics.

In order to provide extensive access to the data, a public use microdata file (PUMF) was produced for Cycle 1.2. The creation of a PUMF involves the application of rigorous procedures to ensure data confidentiality. Consequently, survey variables may have to be grouped, capped, or simply suppressed.

As an alternative, in cases where the PUMF does not meet their needs, researchers can perform their analyses on the CCHS Master File. In order to provide researchers with a means to access the Master File, a remote access facility has been implemented. Remote access provides researchers with the possibility to submit computer programs via e-mail to a dedicated address (cchs-esc@statcan.ca), and to receive the results by return e-mail. To obtain remote access privileges, researchers must necessarily obtain advance approval from the Health Statistics Division. Requests must be submitted to the aforementioned e-mail address and must provide the following, clearly itemised information:

- the researcher’s affiliation,
- the name of all researchers involved in the project,
- the title of the research project,
- an abstract of the project,
- the goals of the research,
- the data to which access is required (survey, cycle),
- why the project requires the access to the master data rather than the PUMF,
- why the Remote Access service is chosen rather than the on-site access in a Research Data Centre (RDC),
- the expected results, and
- the project’s expected completion date.

Further information is available by contacting the CCHS team at the above e-mail address or by phone at (613) 951-1653. Once the request for remote access has been approved, the researcher can submit his/her computer programs to the CCHS team for processing on the Master File(s). The computer output is reviewed by the team for confidentiality concerns and returned to the researcher. However, the correctness and accuracy of each program submission remains at all times the sole responsibility of the researcher.

With the Dummy File supplied on this CD-ROM, the researcher can develop and test his/her computer programs before submitting them to the CCHS team. The Cycle 1.2 Dummy File applies to the data file containing 36,984 respondents. While certain administrative variables, which are of no analytical interest, have been recoded to “blank” or “9”, the Dummy File simulates the Master File perfectly. It contains the same variables and has the same record layout. The data values, however, have been modified in order to protect the confidentiality of respondents. The CCHS Cycle 1.2 Dummy File contains 36,984 records.

This CD-ROM also includes SAS and SPSS macro programs for calculating the variance of estimates. A dummy bootstrap weight file corresponding to the Master File weights present on the Dummy File, are also included to help develop and test the variance calculation programs. This dummy bootstrap weight file simulates the original file, has the same record layout and contains the same variables as those in the original file, but the weight values have been modified. The CD-ROM also contains peripheral SAS and SPSS layout statement files, as well as the technical and methodological documentation usually accompanying the Master File. Please refer to Section 3 of the present document for the complete list of files on this CD-ROM.

The following sections of this document describe in more detail the steps leading to the creation of the Dummy File.

2. Creation of the Dummy File

2.1 Classes of Variables

The master file records are first classified by age, sex. The purpose for this is to create classes of records having the same questionnaire flow so that, after random permutation of the data within the classes, each dummy record contain data respecting a certain internal consistency.

2.2 Sampling and Restrictions of Variable Values

As an additional measure of preserving data confidentiality in the Master File, the Dummy File is created by sub-sampling 50% of the Master File records. However, this sub-sample of 50% is duplicated before it is perturbed in order for the dummy file to be exactly the same size as the master file. In addition, quantitative variables are top-coded and bottom-coded. Variables considered as sensitive or without analytical value are recoded to “blank” or to “9” (see section 2.4). All these modifications mean that we do not necessarily find, on the Dummy File, all the possible categories for each variable. To obtain complete information on a specific variable, users are asked to consult the Master File documentation. Finally, all the weight variables have been repoststratified according to the same methods used to poststratify the Master File weights. However, it should be noted that the weights remain “dummy” weights. Other details related to the creation of this file are not available because they would disclose too much information on the method used for creating the dummy data, hence putting confidentiality at risk.

2.3 Block of variables

The third step consists of creating blocks of variables. Dummy records are produced by randomly swapping blocks of variables for the records in the same class. At first, elementary blocks of variables are grouped following the various modules of the questionnaire. If the content of the elementary block depends on the answers obtained in a previous block, the respective elementary block is amalgamated with the other one to preserve the internal consistency of the records.

2.4 Recode Variables to “blank” and to “9”

The following variables are recoded to “blank” or to “9” during the creation of the Dummy file:

PERSONID SAMB_ACP SAMB_OCP ADMB_ENT ADMB_DAT ADMB_STA ADMBDN09 ADMB_N10
ADMB_LHH ADMB_N12 DHHB_YOB DHHB_MOB DHHB_DOB FILLR001 FILLR002 FILLR004 FILLROO1

3. Content of CD-ROM

LISEZ_MOI.PDF	French documentation about remote-access, creation of the Dummy File and contents of the CD-ROM
READ_ME.PDF	English documentation about remote-access, creation of the Dummy File and contents of the CD-ROM

Dummy File :

DATA\
HSI.TXT Dummy file of Cycle 1.2, 36,984 records

Record Layouts, Statements :

LAYOUT\
 HSI_FMT.SAS SAS Format statement.
 HSI_I.SAS SAS Infile and Input statements.
 HSI_LBF.SAS SAS Label statement in French.
 HSI_LBE.SAS SAS Label statement in English.
 HSI_PFF.SAS SAS Proc Format statement in French.
 HSI_PFE.SAS SAS Proc Format statement in English.

 HSI_I.SPS SPSS Infile statement.
 HSIMISS.SPS SPSS Missing Values statement.
 HSIVALF.SPS SPSS Value Labels statement in French.
 HSIVALE.SPS SPSS Value Labels statement in English.
 HSIVARF.SPS SPSS Variable Labels statement in French.
 HSIVARE.SPS SPSS Variable Labels statement in English.

 READFILE.SAS SAS Statements to read, label and format ASCII data.
 READFILE.SPS SPSS Statements to read, label and format ASCII data.

Documentation:

DOC\PDF_E\
 DD_E.PDF Data Dictionary in English.
 INDEX_A_E.PDF English alpha index.
 INDEX_T_E.PDF English topical index.
 LAYOUT_E.PDF English record layout.
 QUEST_E.PDF English Questionnaire.
 QUEST_HHLD_E.PDF English Household Questionnaire.
 DVDOC_E.PDF English derived variables documentation.
 GUIDE_E.PDF English Userguide.

DOC\PDF_F\
 DD_F.PDF Data Dictionary in French.
 INDEX_A_F.PDF French alpha index.
 INDEX_T_F.PDF French topical index.
 LAYOUT_F.PDF French record layout.
 QUEST_F.PDF French Questionnaire.
 QUEST_HHLD_F.PDF French Household Questionnaire.
 DVDOC_F.PDF French Derived variables documentation.
 GUIDE_F.PDF French Userguide.

Files related to bootstrap weights for variance calculation:

BOOTSTRP\DATA\
 B5.TXT Bootstrap weights file in ASCII format.

BOOTSTRP\SAS\DOC\
 SASBOOTDOC_F.PDF French documentation of the BOOTVARF_V30.SAS program.
 SASBOOTDOC_E.PDF English documentation of the BOOTVARE_V30.SAS program.
 APPENDIXC_HEALTH.PDF Variable and file names to use with the BOOTVAR program in English.
 ANNEXEC_SANTE.PDF Variable and file names to use with the BOOTVAR program in French.

BOOTSTRP\SAS\PGM\	
BOOTVARF_V30.SAS	SAS bootstrap variance program with French comments.
BOOTVARE_V30.SAS	SAS bootstrap variance program with English comments.
MACROF_V30.SAS	SAS Macro program with French comments.
MACROE_V30.SAS	SAS Macro program with English comments.
ETAPE1.SAS	Example, with French comments, of the SAS program creating the data analysis file.
STEP1.SAS	Example, with English comments, of the SAS program creating the data analysis file.

BOOTSTRP\SPSS\DOC\	
SPSSBOOTDOC_F.PDF	French documentation of the BOOTVARF_V30.SPS program.
SPSSBOOTDOC_E.PDF	English documentation of the BOOTVARE_V30.SPS program.
APPENDIXC_HEALTH.PDF	Variable and file names to use with the BOOTVAR program in English.
ANNEXEC_SANTE.PDF	Variable and file names to use with the BOOTVAR program in French.

BOOTSTRP\SPSS\PGM\	
BOOTVARF_V30.SPS	SPSS bootstrap variance program with French comments.
BOOTVARE_V30.SPS	SPSS bootstrap variance program with English comments
MACROF_V30.SPS	SPSS Macro program with French comments.
MACROE_V30.SPS	SPSS Macro program with English comments.
ETAPE1.SPS	Example, with French comments, of the SPSS program creating the data analysis file.
STEP1.SPS	Example, with English comments, of the SPSS program creating the data analysis file.

BOOTSTRP\LAYOUT\	
B5_I.SAS	SAS Record Layout for the bootstrap weight file B5.TXT
B5_I.SPS	SPSS Record Layout for the bootstrap weight file B5.TXT