

**National Population Health Survey
Household Component, 2000-2001
Dummy File for Remote Access to the Master File**

July 2002

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 National Population Health Survey (NPHS), which is conducted by the Health Statistics Division at Statistics Canada, collects information on the health and socio-economic characteristics of the Canadian population. To date, four data collection cycles have been completed: NPHS Cycle 1 (1994-1995), NPHS Cycle 2 (1996-1997), NPHS Cycle 3 (1998-1999) and NPHS Cycle 4 (2000-2001). All four cycles contain a household component and a health institutions component. Cycles 1, 2 and 3 also have a North component.

Earlier cycles produced public use microdata files (PUMFs) for the household component (cycles 1, 2 and 3), as well as the health institutions component (cycles 1 and 2) in order to allow broad access to the cross-sectional data. Only one file was created for the cycle 4 household component, and that file is the longitudinal square file containing all 17,276 panel members. 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 confidentiality considerations limit the creation of longitudinal PUMFs, and since the NPHS is now purely longitudinal beginning with cycle 4, a PUMF was not created for the cycle 4 data.

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 (nphs-ensp@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 (for cycles where a PUMF exists),
- 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 NPHS 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 NPHS 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 NPHS team. The cycle 4 Dummy File applies to the cycle 4 longitudinal data file containing all 17,276 members of the longitudinal panel. While certain administrative variables, which are of no analytical interest, have been recoded to “blank” or “9s”, 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.

This CD-ROM also includes SAS and SPSS macro programs for calculating the variance of estimates. Three dummy bootstrap weight files corresponding to the Dummy File are also included to help develop and test the variance calculation programs. These dummy bootstrap weight files simulate the original file, have the same record layout and contain the same variables, 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 Files. 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 Files

2.1 Classes of Records

The records in the Master File are first divided into classes based on age, sex and the longitudinal response pattern. One of the objectives is to create classes of records with similar pathways through the questionnaire, so that when random data swapping is applied within classes, the resulting artificial records are internally coherent.

The classes for the cycle 4 dummy longitudinal file are based on the age at the time of the interview in 1994-95, as well as in 1996-97, 1998-99 and 2000-01. In order to maintain a minimum number of records in each class, the geographic characteristics are not taken into account when creating the classes.

The supplementary content for Alberta and Manitoba in 1994-95 and Alberta in 1996-97 is added back to the Dummy File records in a later step, where fewer problems result from small class sizes because the number of age-based classes for the special content is relatively small. When a problem arises nonetheless, sufficiently large class sizes are applied and the data are adjusted for internal consistency.

2.2 Blocks of Variables

The second step consists of creating blocks of variables. The artificial records for the dummy file are created by randomly swapping blocks of variables among the records within a class. The variables are first grouped into basic blocks corresponding to the various sections of the questionnaire. Where the flow to one basic block depends on the answers to questions in a previous block, the respective basic blocks are integrated into the same block to maintain internal coherence.

The guiding principle is that blocks should be analytically meaningful while also being small enough to conform to data confidentiality requirements. Variables, which when combined could lead to identification of individual respondents, are thus put in different blocks. All modalities of each variable are not necessarily represented on the Dummy File. Users looking for information on all the possible modalities of a variable should consult the Master File documentation. Further details of the block formation are not given as it would provide too much information on the methods used to create the dummy data and compromise confidentiality. While efforts have been made to ensure the coherence of the data between the blocks, this coherence may not always be present.

2.3 Variables Recoded to “Blank” or “9”

For confidentiality purposes, the variables listed below were recoded to “blank” or “9s” during the creation of the dummy file.

DOD, SP34_CPS, SP34_MET, STRATUM, REPLICAT, CYCLE, SUBCYCLE, AM68_SHA, AM34_SRC, AM34_LNG, SP34_CPA, AM54_BMM, AM54_BDD, AM54_BY, AM54_SRC, AM54_LNG, AM54_EMM, AM54_EDD, AM54_EYY, AM64_STA, AM64_BMM, AM64_BDD, AM64_BY, AM64_SRC, AM64FR, AM64_LNG, AM64_EMM, AM64_EDD, AM64_EYY, SP36FOUT, SP36_CPA, AM36_SRC, AM36_LNG, SP36_STA, AM56_STA, AM56_BMM, AM56_BDD, AM56_BY, AM56_SRC, AM56_LNG, AM56_EMM, AM56_EDD, AM56_EYY, AM66LDUR, AM66_STA, AM66_BMM, AM66_BDD, AM66_BY, AM66_SRC, AM66FR, AM66_LNG, AM66_EMM, AM66_EDD, AM66_EYY, AM68LDUR, SP38FOUT, SP38_CPA, AM38_LP, AM38_SRC, AM38_LNG, SP38_STA, AM58_STA, AM68_STA, AM58_BDD, AM58_BMM, AM58_BY, AM58_LNG, AM58_EMM, AM58_EDD, AM58_EYY, AM68_BMM, AM68_BDD, AM68_BY, AM68_SRC, AM68FR, AM68_LNK, AM68_LNG, AM68_EMM, AM68_EDD, AM68_EYY, AM60LDUR, SP30FOUT, SP30_CPA, AM30_PL, SP30_STA, AM30_TEL, AM60_STA, AM60_BMM, AM60_BDD, AM60_BY, AM60_SRC, AM60FR, AM60_LNK, AM60_SHA, AM60_TEL, AM60_LNG

2.4 Subsampling

Finally, in one more effort to preserve confidentiality, subsampling of the file was performed. This involved removing a proportion of the 17,276 records, and duplicating some of the remaining records in order to keep the total number of records equal to 17,276. The number of respondents in the Full subset and the Full C1 and C4 subset were also respected with regards to this subsampling.

The sampling weights and the bootstrap weights were then adjusted to reflect the 1994 population totals by province, age group and sex. However, it should be noted that these weights remain artificial.

3. Content of CD_ROM

LisezMoi.pdf	French documentation on remote access, creation of the Dummy Files, CD-ROM content.
ReadMe.pdf	English documentation on remote access, creation of the Dummy Files, CD-ROM content.

Dummy File:

DATA\ Dumylong.txt	Cycle 4 Dummy File, 17,276 records
-----------------------	------------------------------------

Record Layouts, Statements:

LAYOUT\ LONG_fmt.sas LONG_i.sas LONG_lbe.sas LONG_lbf.sas LONG_pfe.sas LONG_pff.sas LONG_i.sps LONGmiss.sps LONGvale.sps LONGvalf.sps LONGvare.sps LONGvarf.sps	Cycle 4 SAS FORMAT Cycle 4 SAS INFILE and INPUT Cycle 4 SAS LABEL in English Cycle 4 SAS LABEL in French Cycle 4 SAS PROC FORMAT in English Cycle 4 SAS PROC FORMAT in French Cycle 4 SPSS DATA LIST Cycle 4 SPSS MISSING VALUES Cycle 4 SPSS VALUE LABELS in English Cycle 4 SPSS VALUE LABELS in French Cycle 4 SPSS VARIABLE LABELS in English Cycle 4 SPSS VARIABLE LABELS in French
---	---

Documentation:

DOC\PDF_E\ Dd_lngf_e.pdf Dd_lngfe_e.pdf Dd_long_e.pdf Dv_e.pdf Index_a_e.pdf Index_t_e.pdf Layout_e.pdf Long00_e.pdf Quest_e.pdf	English data dictionary, Cycle 4 Longitudinal Master Full File English data dictionary, Cycle 4 Longitudinal Master C1 & C4 Full File English data dictionary, Cycle 4 Longitudinal Master File English derived variables documentation English alpha index, Cycle 4 Longitudinal Master File English topical index, Cycle 4 Longitudinal Master File English record layout, Cycle 4 Longitudinal Master File English Cycle 4 household component documentation English Cycle 4 Questionnaire
---	---

DOC\PDF_F\ Dd_lngf_f.pdf Dd_lngfe_f.pdf Dd_long_f.pdf Dv_f.pdf Index_a_f.pdf Index_t_f.pdf Layout_f.pdf Long00_f.pdf Quest_f.pdf	French data dictionary, Cycle 4 Longitudinal Master Full File French data dictionary, Cycle 4 Longitudinal Master C1 & C4 Full File French data dictionary, Cycle 4 Longitudinal Master File French derived variables documentation French alpha index, Cycle 4 Longitudinal Master File French topical index, Cycle 4 Longitudinal Master File French record layout, Cycle 4 Longitudinal Master File French Cycle 4 household component documentation French Cycle 4 Questionnaire
---	--

Files relating to the use of bootstrap weights for variance calculation:

BOOTSTRP\	
Lisez_moi.txt	French documentation of the files related to the use of the Bootstrap variance program
Read_me.txt	English documentation of the files related to the use of the Bootstrap variance program
BOOTSTRP\DATA\	
Bd5long.txt	Square Dummy bootstrap weights file in ASCII format
Bd5lngf.txt	Full Dummy bootstrap weights file in ASCII format
Bd5lngfe.txt	C1 & C4 Full Dummy bootstrap weights file in ASCII format
BOOTSTRP\SAS\DOC\	
SASBootdoc_e.pdf	English documentation of the BOOTVARE_V20.SAS program
SASBootdoc_f.pdf	French documentation of the BOOTVARF_V20.SAS program
BOOTSTRP\SAS\PGM\	
Bootvare_v20.sas	SAS bootstrap variance program with English comments
Bootvarf_v20.sas	SAS bootstrap variance program with French comments
Etape1.sas	Example, with French comments, of the SAS program creating the data analysis file
Macro_e_v20.sas	SAS Macro program with English comments
Macrof_v20.sas	SAS Macro program with French comments
Step1.sas	Example, with English comments, of the SAS program creating the data analysis file
BOOTSTRP\SPSS\DOC\	
SPSSBootdoc_e.pdf	English documentation of the BOOTVARE_V21.SPS program
SPSSBootdoc_f.pdf	French documentation of the BOOTVARE_V21.SPS program
BOOTSTRP\SPSS\PGM\	
Bootvare_v21.sps	SPSS Bootstrap variance calculation program with English comments
Bootvarf_v21.sps	SPSS Bootstrap variance calculation program with French comments
Etape1.sps	Example, with French comments, of the SPSS program creating the data analysis file
Macro_e_v21.sps	SPSS Macro program with English comments
Macrof_v21.sps	SPSS Macro program with French comments
Step1.sps	Example, with English comments, of the SPSS program creating the data analysis file
BOOTSTRP\LAYOOUT\	
B5long_i.sas	SAS Record Layout for the dummy bootstrap weight files BD5LONG.TXT, BD5LNGF.TXT, and BD5LNGFE.TXT
B5long_i.sps	SPSS Record Layout for the dummy bootstrap weight files BD5LONG.TXT, BD5LNGF.TXT, and BD5LNGFE.TXT