

**National Population Health Survey
Household Component, Cycle 6 (2004/2005)
Dummy File for Remote Access to the Master File**

April 2007

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, six data collection cycles have been completed: NPHS Cycle 1 (1994/1995), NPHS Cycle 2 (1996/1997), NPHS Cycle 3 (1998/1999), NPHS Cycle 4 (2000/2001), NPHS Cycle 5 (2002/2003) and NPHS Cycle 6 (2004/2005). All six cycles contain a household component while only cycles 1 to 5 contain 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. In Cycles 4, 5 and 6, only one file was created for the 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 6 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 6 Dummy File applies to the Cycle 6 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 section of this document describes in more detail the steps leading to the creation of the Dummy File.

2. Creation of the Dummy File

2.1 Subsampling

In an effort to preserve confidentiality, subsampling of the Master file was performed. This involved removing a portion 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 C6 subset were also respected with regards to this subsampling.

2.2 Classes of Records

The records in the file are then 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 6 dummy longitudinal file are based on the age at the time of the interview in 1994/1995, as well as in 1996/1997, 1998/1999, 2000/2001, 2002/2003 and 2004/2005. In order to have a sufficient 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 1994/1995 and Alberta in 1996/1997 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.

2.3 Blocks of Variables

The next step involves the creation of 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 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. Also, several variables are modified in each block. 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.

2.4 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_BYY, AM54_SRC, AM54_LNG, AM54_EMM, AM54_EDD, AM54_EYY, AM64_STA, AM64_BMM, AM64_BDD, AM64_BYY, AM64_SRC, AM64FR, AM64_LNG, AM64_EMM, AM64_EDD, AM64_EYY, AM64_SHA, SP34_STA, AM54_STA, AM64_LNK, AM64_TEL, AM54_PXY, AM64_PXY, AM54_TEL, AM64_ALO, AM64_AFF, GE34DPOP, SP36FOUT, SP36_CPA, AM36_SRC, AM36_LNG, SP36_STA, AM56_STA, AM56_BMM, AM56_BDD, AM56_BYY, AM56_SRC, AM56_LNG, AM56_EMM, AM56_EDD, AM56_EYY, AM66LDUR, AM66_STA, AM66_BMM, AM66_BDD, AM66_BYY, AM66_SRC, AM66FR, AM66_LNG, AM66_EMM, AM66_EDD, AM66_EYY, AM66_SHA, AM66_LNK, AM36_TEL, AM66_TEL, AM56_PXY, AM66_PXY, AM56_TEL, AM66_ALO, AM66_AFF, GE36DPOP, AM68LDUR, SP38FOUT, SP38_CPA, AM38_LP, AM38_SRC, AM38_LNG, SP38_STA, AM58_STA, AM68_STA, AM58_BDD, AM58_BMM, AM58_BYY, AM58_LNG, AM58_EMM, AM58_EDD, AM58_EYY, AM68_BMM, AM68_BDD, AM68_BYY, AM68_SRC, AM68FR, AM68_LNK, AM68_LNG, AM68_EMM, AM68_EDD, AM68_EYY, AM58_SRC, AM38_TEL, AM68_TEL, AM58_PXY, AM68_PXY, AM58_TEL, AM68_ALO, AM68_AFF, GE38DPOP, AM60LDUR, SP30FOUT, SP30_CPA, AM30_PL, SP30_STA, AM30_TEL, AM60_STA, AM60_BMM, AM60_BDD, AM60_BYY, AM60_SRC, AM60FR, AM60_LNK, AM60_SHA, AM60_TEL, AM60_LNG, AM60_PXY, GE30DPOP, AM62LDUR, SP32FOUT, SP32_CPA, AM32_PL, SP32_STA, AM32_TEL, AM62_STA, AM62_BMM, AM62_BDD, AM62_BYY, AM62_SRC, AM62FR, AM62_LNK, AM62_SHA, AM62_TEL, AM62_LNG, AM62_PXY, GE32DPOP, AM6ALDUR, SP3AFOUT, SP3A_CPA, AM3A_PL, SP3A_STA, AM3A_TEL, AM6A_STA, AM6A_BMM, AM6A_BDD, AM6A_BYY, AM6A_SRC, AM6AFR, AM6A_LNK, AM6A_SHA, AM6A_TEL, AM6A_LNG, AM6A_PXY, GE3ADPOP

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

NOTE Household highest level education - Plus haut niveau scolaire menage.pdf Note to read

Dummy File:

DATA\	
Long.txt	Cycle 6 Dummy File, 17,276 records

Record Layouts, Statements:

LAYOUT\	
LONG_fmt.sas	Cycle 6 SAS FORMAT
LONG_i.sas	Cycle 6 SAS INFILE and INPUT
LONG_lbe.sas	Cycle 6 SAS LABEL in English
LONG_lbf.sas	Cycle 6 SAS LABEL in French
LONG_pfe.sas	Cycle 6 SAS PROC FORMAT in English
LONG_pff.sas	Cycle 6 SAS PROC FORMAT in French
READFILE.sas	program to read Cycle 6 data in SAS format
LONG_i.sps	Cycle 6 SPSS DATA LIST
LONGmiss.sps	Cycle 6 SPSS MISSING VALUES
LONGvale.sps	Cycle 6 SPSS VALUE LABELS in English
LONGvalf.sps	Cycle 6 SPSS VALUE LABELS in French
LONGvare.sps	Cycle 6 SPSS VARIABLE LABELS in English
LONGvarf.sps	Cycle 6 SPSS VARIABLE LABELS in French
READFILE.sps	program to read Cycle 6 data in SPSS format

Documentation:

DOC\PDF_E\	
Cycle 6 NPHS Household DV Doc_E 20061102.pdf	English derived variables documentation
Cycle 6 NPHS Household Questionnaire_E 20061018.pdf	English Cycle 6 Questionnaire
Cycle 6 NPHS Household Longdoc_E 20061101.pdf	English Cycle 6 household component documentation

DOC\PDF_E\LNGF	
English Data Dictionary (Freqs).pdf	English data dictionary, Cycle 6 Longitudinal Dummy Full File
English DD Alpha Index.pdf	English alpha index, Cycle 6 Longitudinal Full File
English DD Topical Index.pdf	English topical index, Cycle 6 Longitudinal Full File
English Record Layout.pdf	English record layout, Cycle 6 Longitudinal Full File

DOC\PDF_E\LNGFE	
English Data Dictionary (Freqs).pdf	English data dictionary, Longitudinal Dummy C1 & C6 Full File
English DD Alpha Index.pdf	English alpha index, Longitudinal C1 & C6 Full File
English DD Topical Index.pdf	English topical index, Longitudinal C1 & C6 Full File
English Record Layout.pdf	English record layout, Longitudinal C1 & C6 Full File

DOC\PDF_E\LONG

English Data Dictionary (Freqs).pdf	English data dictionary, Cycle 6 Longitudinal Dummy Square File
English DD Alpha Index.pdf	English alpha index, Cycle 6 Longitudinal Square File
English DD Topical Index.pdf	English topical index, Cycle 6 Longitudinal Square File
English Record Layout.pdf	English record layout, Cycle 6 Longitudinal Square File

DOC\PDF_F\

Cycle 6 NPHS Household DV Doc_F 20061102.pdf	French derived variables documentation
Cycle 6 NPHS Household Questionnaire_F 20061101.pdf	French Cycle 6 Questionnaire
Cycle 6 NPHS Household Longdoc_F 20061101.pdf	French Cycle 6 household component documentation

DOC\PDF_F\LNGF

French Data Dictionary (Freqs).pdf	French data dictionary, Cycle 6 Longitudinal Dummy Full File
French DD Alpha Index.pdf	French alpha index, Cycle 6 Longitudinal Full File
French DD Topical Index.pdf	French topical index, Cycle 6 Longitudinal Full File
French Record Layout.pdf	French record layout, Cycle 6 Longitudinal Full File

DOC\PDF_F\LNGFE

French Data Dictionary (Freqs).pdf	French data dictionary, Longitudinal Dummy C1 & C6 Full File
French DD Alpha Index.pdf	French alpha index, Longitudinal C1 & C6 Full File
French DD Topical Index.pdf	French topical index, Longitudinal C1 & C6 Full File
French Record Layout.pdf	French record layout, Longitudinal C1 & C6 Full File

DOC\PDF_F\LONG

French Data Dictionary (Freqs).pdf	French data dictionary, Cycle 6 Longitudinal Dummy Square File
French DD Alpha Index.pdf	French alpha index, Cycle 6 Longitudinal Square File
French DD Topical Index.pdf	French topical index, Cycle 6 Longitudinal Square File
French Record Layout.pdf	French record layout, Cycle 6 Longitudinal Square File

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

BOOTSTRP\DATA\

Bd5long.txt	Bootstrap weights file in ASCII format for the dummy square file
Bd5lngf.txt	Bootstrap weights file in ASCII format for the dummy full file
Bd5lngfe.txt	Bootstrap weights file in ASCII format for the dummy C1 & C6 full file

BOOTSTRP\LAYOUT\

B5LONG_i.SAS	SAS Record Layout for the bootstrap weights file BD5LONG.TXT
B5LNGF_i.SAS	SAS Record Layout for the bootstrap weights file BD5LNGF.TXT
B5LNGFE_i.SAS	SAS Record Layout for the bootstrap weights file BD5LNGFE.TXT
B5LONG_i.SPS	SPSS Record Layout for the bootstrap weights file BD5LONG.TXT
B5LNGF_i.SPS	SPSS Record Layout for the bootstrap weights file BD5LNGF.TXT
B5LNGFE_i.SPS	SPSS Record Layout for the bootstrap weights file BD5LNGFE.TXT

BOOTSTRP\SAS\DOC\

SASBootdoc_eng.pdf	English documentation of the Bootvare_v31.sas program
SASBootdoc_fr.pdf	French documentation of the Bootvarf_v31.sas program
AppendixC.pdf	Appendix for the SASBootdoc_eng.pdf documentation
AnnexeC.pdf	Appendix for the SASBootdoc_fr.pdf documentation

BOOTSTRP\SAS\PGM\	
Bootvare_v31.sas	SAS bootstrap variance program with English comments
Bootvarf_v31.sas	SAS bootstrap variance program with French comments
Step1.sas	Example, with English comments, of the SAS program creating the data analysis file
Etape1.sas	Example, with French comments, of the SAS program creating the data analysis file
Macroe_v31.sas	SAS Macro program with English comments
Macrof_v31.sas	SAS Macro program with French comments
BOOTSTRP\SPSS\DOC\	
SPSSBootdoc_eng.pdf	English documentation of the Bootvare_v30.sps program
SPSSBootdoc_fr.pdf	French documentation of the Bootvarf_v30.sps program
AppendixC.pdf	Appendix for the SPSSBootdoc_eng.pdf documentation
AnnexeC.pdf	Appendix for the SPSSBootdoc_fr.pdf documentation
BOOTSTRP\SPSS\PGM\	
Bootvare_v30.sps	SPSS Bootstrap variance calculation program with English comments
Bootvarf_v30.sps	SPSS Bootstrap variance calculation program with French comments
Step1.sps	Example, with English comments, of the SPSS program creating the data analysis file
Etape1.sps	Example, with French comments, of the SPSS program creating the data analysis file
Macroe_v30.sps	SPSS Macro program with English comments
Macrof_v30.sps	SPSS Macro program with French comments