

**National Population Health Survey (NPHS)
Household Component
Cycles 1 to 7 (1994/1995 to 2006/2007)**

Dummy File for Remote Access to the Master File

December 2008

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, seven 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), NPHS Cycle 6 (2004-2005) and NPHS Cycle 7 (2006-2007). All seven 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, 6 and 7, 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 7 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.gc.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 7 Dummy File applies to the Cycles 1 to 7 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 provides Statistics Canada's Website links for Bootvar documentation and programs in SAS and SPSS formats for calculating the variance of estimates. Two 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 Classes of Records

The records in the 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 7 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, 2000-01, 2002-03, 2004-05 and 2006-07. 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 1994-95 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 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.

2.3 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_BYYY, AM54_SRC, AM54_LNG, AM54_EMM, AM54_EDD, AM54_EYY, AM64_STA, AM64_BMM, AM64_BDD, AM64_BYYY, 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_BYYY, AM56_SRC, AM56_LNG, AM56_EMM, AM56_EDD, AM56_EYY, AM66LDUR, AM66_STA, AM66_BMM, AM66_BDD, AM66_BYYY, 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_BYYY, AM58_LNG, AM58_EMM, AM58_EDD, AM58_EYY, AM68_BMM, AM68_BDD, AM68_BYYY, 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_BYYY, 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_BYYY, 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_BYYY, AM6A_SRC, AM6AFR, AM6A_LNK, AM6A_SHA, AM6A_TEL, AM6A_LNG, AM6A_PXY, GE3ADPOP, AM6BLDUR, SP3BFOUT, SP3B_CPA, AM3B_PL, SP3B_STA, AM3B_TEL, AM6B_STA, AM6B_BMM, AM6B_BDD, AM6B_BYYY, AM6B_SRC, AM6BFR, AM6B_LNK, AM6B_SHA, AM6B_TEL, AM6B_LNG, AM6B_PXY, GE3BDPOP

2.4 Subsampling

Finally, in one more 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 was 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. CD-ROM: DETAILS OF DIRECTORIES AND FILES

ReadMe.pdf Description of directories and files on this CD-ROM in English
LisezMoi.pdf Description of directories and files on this CD-ROM in French

DUMMY FILE:

DATA

long.txt Cycles 1 to 7 (1994/1995 to 2006/2007) Dummy File
(17,276 records, refers to the longitudinal dummy square subset)

DOCUMENTATION:

DOC\PDF_E\ Documentation in English

Cycle 1-8 NPHS Content Summary List Alpha 20080716_E.pdf	Cycles 1 to 8 summary of content
Cycle 7 NPHS Household Questionnaire 20080716_E.pdf	Cycle 7 Questionnaire
Cycle 7 NPHS DV Doc 20080911_E.pdf	Cycles 1 to 7 Derived variables documentation
Cycle 7 NPHS List of Derived Variables 20080716_E.pdf	Cycles 1 to 7 List of derived variables
Cycle 7 NPHS Household Longdoc 20080716_E.pdf	Cycles 1 to 7 Household component longitudinal documentation

DOC\PDF_E\LONG Documentation in English for the Cycles 1 to 7 longitudinal dummy square subset

Long_English Data Dictionary (Freqs).pdf	Cycles 1 to 7 Data dictionary
Long_English DD Alpha Index.pdf	Cycles 1 to 7 Alpha index
Long_English DD Topical Index.pdf	Cycles 1 to 7 Topical index
Long_English Record Layout.pdf	Cycles 1 to 7 Record layout

DOC\PDF_E\LNFG Documentation in English for the Cycles 1 to 7 longitudinal dummy full subset

Lnfg_English Data Dictionary (Freqs).pdf	Cycles 1 to 7 Data dictionary
Lnfg_English DD Alpha Index.pdf	Cycles 1 to 7 Alpha index
Lnfg_English DD Topical Index.pdf	Cycles 1 to 7 Topical index
Lnfg_English Record Layout.pdf	Cycles 1 to 7 Record layout

DOC\PDF_F**Documentation in French**

Cycle 1-8 NPHS Content Summary List Alpha 20080716_F.pdf	Cycles 1 to 8 summary of content
Cycle 7 NPHS Household Questionnaire 20080716_F.pdf	Cycle 7 Questionnaire
Cycle 7 NPHS DV Doc 20080911_F.pdf	Cycles 1 to 7 Derived variables documentation
Cycle 7 NPHS List of Derived Variables 20080716_F.pdf	Cycles 1 to 7 List of derived variables
Cycle 7 NPHS Household Longdoc 20080716_F.pdf	Cycles 1 to 7 Household component longitudinal documentation

DOC\PDF_FLONG**Documentation in French for Cycles 1 to 7 longitudinal dummy square subset**

Long_French Data Dictionary (Freqs).pdf	Cycles 1 to 7 Data dictionary
Long_French DD Alpha Index.pdf	Cycles 1 to 7 Alpha index
Long_French DD Topical Index.pdf	Cycles 1 to 7 Topical index
Long_French Record Layout.pdf	Cycles 1 to 7 Record layout

DOC\PDF_FLNGF**Documentation in French for the Cycle 1 to 7 longitudinal dummy full subset of respondents**

Lngf_French Data Dictionary (Freqs).pdf	Cycle 1 to 7 Data dictionary
Lngf_French DD Alpha Index.pdf	Cycle 1 to 7 Alpha index
Lngf_French DD Topical Index.pdf	Cycle 1 to 7 Topical index
Lngf_French Record Layout.pdf	Cycle 1 to 7 Record layout

RECORD LAYOUTS, STATEMENTS:**LAYOUT**

LONG_fmt.sas	Cycles 1 to 7 SAS FORMAT
LONG_i.sas	Cycles 1 to 7 SAS INFILE and INPUT
LONG_lbe.sas	Cycles 1 to 7 SAS LABEL in English
LONG_lbf.sas	Cycles 1 to 7 SAS LABEL in French
LONG_pfe.sas	Cycles 1 to 7 SAS PROC FORMAT in English
LONG_pff.sas	Cycles 1 to 7 SAS PROC FORMAT in French
READFILE.sas	Program to read Cycles 1 to 7 data in SAS format
LONG_i.sps	Cycles 1 to 7 SPSS DATA LIST
LONGmiss.sps	Cycles 1 to 7 SPSS MISSING VALUES
LONGvale.sps	Cycles 1 to 7 SPSS VALUE LABELS in English
LONGvalf.sps	Cycles 1 to 7 SPSS VALUE LABELS in French
LONGvare.sps	Cycles 1 to 7 SPSS VARIABLE LABELS in English
LONGvarf.sps	Cycles 1 to 7 SPSS VARIABLE LABELS in French
READFILE.sps	Program to read Cycles 1 to 7 data in SPSS format

BOOTSTRAP WEIGHTS AND PROGRAMS FOR VARIANCE CALCULATION:

BOOTSTRP\DATA

B5long.txt	Bootstrap weights file in ASCII format for the Cycles 1 to 7 longitudinal dummy square subset
B5lngf.txt	Bootstrap weights file in ASCII format for the Cycles 1 to 7 longitudinal dummy full subset

BOOTSTRP\LAYOUT

B5LONG_i.SAS	SAS Record Layout for the bootstrap weights file B5LONG.TXT
B5LNGF_i.SAS	SAS Record Layout for the bootstrap weights file B5LNGF.TXT
B5LONG_i.SPS	SPSS Record Layout for the bootstrap weights file B5LONG.TXT
B5LNGF_i.SPS	SPSS Record Layout for the bootstrap weights file B5LNGF.TXT

BOOTSTRP\BOOTVAR

Bootvar_E.pdf	Document that provides Statistics Canada's Website links for Bootvar documentation and programs in SAS and SPSS formats in English.
Bootvar_F.pdf	Document that provides Statistics Canada's Website links for Bootvar documentation and programs in SAS and SPSS formats in French.