

National Population Health Survey Household Component, 1998-1999 Dummy Files for Remote Access to the Master Files

February, 2001

Notice

The Dummy Files 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 Files contain modified data, and must never be used for analytical purposes.

1. Background and Overview

The National Population Health Survey (NPHS), which is conducted by Health Statistics Division at Statistics Canada, collects information on the health and socio-economic characteristics of the Canadian population. To date, three data collection cycles have been completed: NPHS Cycle 1 (1994-1995), NPHS Cycle 2 (1996-1997), and NPHS Cycle 3 (1998-1999), for each of three components: the household component, the health institutions component, and the North component.

In order to allow broad access to the data, public use microdata files (PUMFs) have been produced for the household component (cycles 1, 2 and 3), as well as the health institutions component (cycles 1 and 2). However, 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 well, confidentiality considerations limit the creation of longitudinal PUMFs.

As a rule, most research initiatives can be carried out through the use of the PUMFs. However, some research projects require access to the Master Files, notably the longitudinal files. In order to provide researchers with a means to access these files, 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@statcan.ca), and to receive the results by return e-mail.

To obtain remote access privileges, researchers must necessarily obtain advance approval from 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 title of the research project,
- an abstract of the project,
- the goals of the research,
- the data to which access is required, including why the project requires access to the master data rather than the PUMF(s),
- 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 Colette Koeune, by phone at (613) 951-1653, or by fax at (613) 951-4198.

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 Files that are supplied on this CD-ROM, the researcher can develop and test his/her computer programs before submitting them to the NPHS team. While certain administrative variables, which are of no analytical interest, have been recoded to “blank” or “9s”, the Dummy Files perfectly simulate the Master Files. They contain the same variables and have the same record layout. The data values, however, have been modified in order to protect the confidentiality of respondents.

These Dummy Files pertain to the third cycle (1998-1999) of the NPHS, household component. In the third cycle, approximately 49,000 individuals were surveyed using the General questionnaire, and these data are simulated in the General (“h35”) cross-sectional Dummy File. Approximately 17,000 individuals were surveyed with the detailed Health questionnaire, and their data are simulated in the Health (“h356”) cross-sectional Dummy File. The respondents selected for the longitudinal component of the survey are independently simulated in the Longitudinal, full response (“lngf”) Dummy File. The dummy records for the selected respondents do not have the same values across files due to the random shuffling of blocks.

This CD-ROM also includes a SAS computer program for calculating the variance of estimates. Dummy bootstrap weight files corresponding to each Dummy File are also included to help develop and test the variance calculation programs. These dummy bootstrap weight files simulate the original files, have the same record layout and contain the same variables, but the weight values have been modified. The CD-ROM also contains some examples of SAS programs for variance calculation, 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 Files.

2. Creation of the Dummy Files

2.1 Classes of Records

The records in each Master File are first divided into classes based on age, sex and geographic characteristics. For the longitudinal file, classification is also based on 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 approach for the longitudinal Dummy File is slightly different. Because the classes for the longitudinal file are based on the age at the time of the interview in 1994-95, as well as in 1996-97 and in 1998-99, the resulting number of classes is very large. 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 in 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, and these basic blocks are then grouped into blocks of variables. Where the flow to one section depends on the answers to questions in a previous section, the basic blocks of variables corresponding to the section are integrated into the same block to maintain internal coherence. The only exceptions to this rule are the variables used in the creation of the classes, whose effects on the flow through the questionnaire have already taken into account.

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.

For the creation of the blocks, variables were classified into the following basic blocks (except for some geographic and design variables, individual and household characteristics, and administrative variables):

Basic Blocks

- | | |
|--------------------------------------|--------------------------------|
| 1. Household income | 7. Education |
| 2. Two week disability | 8. Labour force |
| 3. Health care utilisation | 9. Food insecurity (1998 only) |
| 4. Restriction of activities | 10. General Health |
| 5. Chronic conditions | 11. Height/weight |
| 6. Socio-demographic characteristics | 11a. Height/weight (1996 only) |

- | | |
|--|---|
| 12. Preventive health practices (1994 and 1998 only) | 33. Alcohol Dependence (1996 only) |
| 13. Access to - blood pressure (1996 only) | 34. Ongoing stress (1994 only) |
| 14. Access to - pap smear (1996 only) | 35. Recent stress (1994 only) |
| 15. Access to - mammography (1996 only) | 36. Trauma (1994 only) |
| 16. Access to - breast exam (1996 only) | 37. Work stress (1994 only) |
| 17. Access to - physical check-up (1996 only) | 38. Self-esteem and mastery (1994 only) |
| 18. Access to - flu shots (1996 only) | 39. Sense of coherence (1994 and 1998 only) |
| 19. Access to - dental visits (1996 only) | 40. Mental health |
| 20. Access to - eye exams (1996 only) | 41. Social support |
| 21. Access to - emergency services (1996 only) | 41a. Social support (Alberta, 1996 only) |
| 22. HIV (1996 only) | 42. Attitude toward parents (Alberta, 1996 only) |
| 23. Health information (Alberta, 1996 only) | 43. Health services (Alberta, 1996 only) |
| 24. Health status | 44. Sexual health (1996 only) |
| 25. Physical activities | 44a. Sexual health (Alberta, 1996 only) |
| 26. Tanning (Alberta, 1996 only) | 45. Road safety (1996 only) |
| 27. Repetitive strain (1996 and 1998 only) | 46. Violence/personal safety (Alberta, 1996 only) |
| 28. Injuries | 47a. Coping (Alberta, 1994) |
| 29. Drug use | 47b. Coping (Alberta, 1996) |
| 30. Smoking | 48. Relate to people (Manitoba, 1994 only) |
| 30a. Smoking (1996 only) | 49. Self-care (1998 only) |
| 31. Tobacco alternatives (1998 only) | 50. Insurance (1998 only) |
| 32. Alcohol | 51. Family medical history (1998 only) |
| 32a. Alcohol (Alberta, 1996 only) | 52. Nutrition (1998 only) |

Note that the basic blocks no. 23, 26, 32a, 41a, 42, 43, 44a, 46, 47a, 47b and 48 contain the special content for Manitoba and Alberta. This content is added at a subsequent step and, as a rule, the classes are based on the province (Manitoba, Alberta and the rest of Canada), the age, the sex and the longitudinal response pattern.

2.3 Recoded Variables

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

General (h35) Dummy File:

AM58_RNO, AM38_LP, AM38_LNG, AM58_LNG, AM58_SHA, AM38_SRC, AM38_CAS, AM58_CAS, AM58_BMM, AM58_BDD, AM58_BYE, AM58_EMM, AM58_EDD, AM58_EYY, SP38_CPS, SP38_CPA, SP38_STA, AM68_STA, SP38_PAR, SP38DPC, CYCLE, SUBCYCLE, MEMCYCLE, STRATUM, REPLICAT.

Health (h356) Dummy File:

All variables listed under the “General file” excluding AM58_RNO, to which the following variables were added:

AM68_RNO, AM68_LNG, AM68_SHA, AM68_SRC, AM68_CAS, AM68_BMM, AM68_BDD, AM68_BYYY, AM68_EMM, AM68_EDD, AM68_EYY, AM68_LNK.

Longitudinal, full response (lngf) Dummy File:

DOD, SP34_CPS, SP34_MET, CYCLE, SUBCYCLE, SP34DPC, 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, SP36FOUT, SP36_CPA, SP36DPC, 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_5, AM66_LNG, AM66_EMM, AM66_EDD, AM66_EYY, AM68LDUR, SP38FOUT, SP38DPC, 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.

3. Content of CD_ROM

Lisez_moi.pdf	Documentation on remote access, creation of the Dummy Files, CD-ROM content (in French)
Read_me.pdf	Documentation on remote access, creation of the Dummy Files, CD-ROM content (in English)

Dummy Files:

DATA\ dumyh35.txt	General (h35) Dummy File, 49,046 records
dumyh356.txt	Health (h356) Dummy File, 17,244 records
dumylngf.txt	Longitudinal, full response (lngf) Dummy File, 14,619 records

Record Layouts, Statements:

LAYOUT\

h35_fmt.sas	h35 SAS FORMAT
h35_i.sas	h35 SAS INFILE and INPUT
h35_lbf.sas	h35 SAS LABEL in French
h35_lbe.sas	h35 SAS LABEL in English
h35_pff.sas	h35 SAS PROC FORMAT in French
h35_pfe.sas	h35 SAS PROC FORMAT in English
h35_i.sps	h35 SPSS FILE HANDLE and DATA LIST
h35miss.sps	h35 SPSS MISSING VALUES
h35valf.sps	h35 SPSS VALUE LABELS in French
h35vale.sps	h35 SPSS VALUE LABELS in English
h35varf.sps	h35 SPSS VARIABLE LABELS in French
h35vare.sps	h35 SPSS VARIABLE LABELS in English

h356_fmt.sas	h356 SAS FORMAT
h356_i.sas	h356 SAS INFILE and INPUT
h356_lbf.sas	h356 SAS LABEL in French
h356_lbe.sas	h356 SAS LABEL in English
h356_pff.sas	h356 SAS PROC FORMAT in French
h356_pfe.sas	h356 SAS PROC FORMAT in English
h356_i.sps	h356 SPSS FILE HANDLE and DATA LIST
h356miss.sps	h356 SPSS MISSING VALUES
h356valf.sps	h356 SPSS VALUE LABELS in French
h356vale.sps	h356 SPSS VALUE LABELS in English
h356varf.sps	h356 SPSS VARIABLE LABELS in French
h356vare.sps	h356 SPSS VARIABLE LABELS in English

lngf_fmt.sas	lngf SAS FORMAT
lngf_i.sas	lngf SAS INFILE and INPUT
lngf_lbf.sas	lngf SAS LABEL in French
lngf_lbe.sas	lngf SAS LABEL in English
lngf_pff.sas	lngf SAS PROC FORMAT in French
lngf_pfe.sas	lngf SAS PROC FORMAT in English
lngf_i.sps	lngf SPSS FILE HANDLE and DATA LIST
lngfmiss.sps	lngf SPSS MISSING VALUES
lngfvalf.sps	lngf SPSS VALUE LABELS in French
lngfvale.sps	lngf SPSS VALUE LABELS in English
lngfvarf.sps	lngf SPSS VARIABLE LABELS in French
lngfvare.sps	lngf SPSS VARIABLE LABELS in English

Documentation:

DOC\
PDF_E\
h35_dd.pdf

h35_ind.pdf

h35_lay.pdf

h35_top.pdf

h356_dd.pdf

h356_ind.pdf

h356_lay.pdf

h356_top.pdf

lngf_dd.pdf

lngf_ind.pdf

lngf_lay.pdf

lngf_top.pdf

Data dictionary, General Master File (h35) in English

Alpha index, General Master File (h35) in English

Record layout, General Master File (h35) in English

Topical index, General Master File (h35) in English

Data dictionary, Health Master File (h356) in English

Alpha index, Health Master File (h356) in English

Record layout, Health Master File (h356) in English

Topical index, Health Master File (h356) in English

Data dictionary, Longitudinal Master File, full response (lngf) in English

Alpha index, Longitudinal Master File, full response (lngf) in English

Record layout, Longitudinal Master File, full response (lngf) in English

Topical index, Longitudinal Master File, full response (lngf) in English

DOC\
PDF_F\
h35_dd.pdf

h35_ind.pdf

h35_lay.pdf

h35_top.pdf

h356_dd.pdf

h356_ind.pdf

h356_lay.pdf

h356_top.pdf

lngf_dd.pdf

lngf_ind.pdf

lngf_lay.pdf

lngf_top.pdf

Data dictionary, General Master File (h35) in French

Alpha index, General Master File (h35) in French

Record layout, General Master File (h35) in French

Topical index, General Master File (h35) in French

Data dictionary, Health Master File(h356) in French

Alpha index, Health Master File(h356) in French

Record layout, Health Master File(h356) in French

Topical index, Health Master File (h356) in French

Data dictionary, Longitudinal Master File, full response (lngf) in French

Alpha index, Longitudinal Master File, full response (lngf) in French

Record layout, Longitudinal Master File, full response (lngf) in French

Topical index, Longitudinal Master File, full response (lngf) in French

BOOTSTRAP Variance Calculation:

Bootstrap Weights:

BOOTSTRP\ bd5h35.txt	Dummy bootstrap weights for General Dummy File (h35)
bd5h356.txt	Dummy bootstrap weights for Health Dummy File (h356)
bd5lngf.txt	Dummy bootstrap weights for Longitudinal Dummy File, full response (lngf)

Documentation:

BOOTSTRP\ DOC\ btdoceng.pdf	Documentation for bootvar.sas program in English
btdocfr.pdf	Documentation bootvarf.sas program in French
lisezmoi.txt	Description of files for bootstrap variance calculation (in French)
readme.txt	Description of files for bootstrap variance calculation (in English)

Variance program and example:

BOOTSTRP\ PGM_E\ bootvar.sas	Bootstrap variance calculation program with comments in English
ex1a_eng.sas	Example 1a (Comments in English)
ex1b_eng.sas	Example 1b (Comments in English)
ex2a_eng.sas	Example 2a (Comments in English)
ex2b_eng.sas	Example 2b (Comments in English)

BOOTSTRP\ PGM_F\ bootvarf.sas	Bootstrap variance calculation program with comments in French
ex1a_fr.sas	Example 1a (Comments in French)
ex1b_fr.sas	Example 1b (Comments in French)
ex2a_fr.sas	Example 2a (Comments in French)
ex2b_fr.sas	Example 2b (Comments in French)

Record layouts:

BOOTSTRP\
LAYOUT\
 b35_i.sas

 b356_i.sas
 blngf_i.sas

bd5h35 SAS INFILE and INPUT
bd5h356 SAS INFILE and INPUT
bd5lngf SAS INFILE and INPUT