

National Population Health Survey (NPHS)
1996-1997 Household Component
Remote Access
Dummy Master File Documentation

1. Background & Overview

The National Population Health Survey Program consists of three component surveys. They are: the household component covering households in the ten provinces, the health care institutions component covering long term residents of health care institutions and the northern territories component covering households in the three northern territories. In the spirit of broad dissemination, public use micro data files are produced for the first two components but not the third. In the transition from a master to a public use file, there is a need to preserve confidentiality and as a consequence the detail in the data is often grouped, capped or suppressed. As a consequence, analysts often discover during their research with the public use product that they have a need for more detail in the data to complete their work. In order to provide clients access to original master files, one of the options available to analysts is the remote access facility of the NPHS.

The remote access facility allows users to submit programs via the Internet by e-mailing them to nphs@statcan.ca. Any user wishing to participate in this option must obtain approval in advance from Health Statistics Division by submitting an abstract describing his or her work using the Internet address above or calling (613) 951-1653 for more information about the process. To allow software testing of custom programs, clients are encouraged to test out their applications using the appropriate dummy master file(s) that are created for distribution. A dummy file is a simulated master file with the same layout and variable names and with the data altered to protect the confidentiality of respondents. When the programs are running correctly, they can be e-mailed and are submitted by NPHS staff to process the internal master file(s) required. The client is responsible for the logical correctness of his or her work. The output will be reviewed for confidentiality concerns and subsequently returned to the client.

In 1996-97, there are four dummy files, the two cross-sectional ones as produced before in 1994-95 representing the General (H35) and Health (H356) master files and two new longitudinal ones. The first longitudinal dummy file (LNGF) represents the longitudinal master file of fully complete responses in both years and the second dummy file (LNGP) represents the longitudinal master file composed of all combinations of both partial and complete responses for the two cycles. For this cycle, 1996-97, due to the size of the original master files, the dummy files are only a sub-sample of the original number of records. The General dummy file is a selection of 5% of the General master file records and the other three dummy files are 10% selections.

Certain administrative variables have been suppressed and coded to "blank" or "9s" on all records since these variables are of no interest from an analytical perspective.

The accompanying data dictionary files distributed on the CD-ROM are those for the corresponding master files and contain weighted and sample counts for each variable value. These files can be viewed through the Adobe Acrobat software installed on the CD-ROM of the public use micro data product. The explanation of what each file contains is detailed below.

It should be noted that total population estimates produced from these files will only yield results that are approximately 5% (General) or 10% (Health, LNGF and LNGP) the size of those using the true master files. Any estimate is for illustrative purposes only to demonstrate that the program is functioning correctly. None should be used for tabular or analytical purposes.

2. Dummy File Creation Strategy

2.1 Class Formation

First, records on each master file were divided into classes based on age, sex and geographic variables. For the longitudinal files, classes also depended on the longitudinal response pattern. For the cross-sectional Health file the classes also made a distinction between RDD (provincial buy-in) and core in Alberta and for children in Manitoba, since many questions were not asked of RDD respondents in Alberta and certain questions were asked only of RDD children in Manitoba and Alberta. One of the objectives was to form classes of records with similar pathways through the questionnaires, so that when random data swapping was applied within classes the resulting artificial data records would be internally coherent.

For the longitudinal dummy files the classes were slightly different. Since the classes for the longitudinal files were based on the age at both the 1994-95 and 1996-97 interviews, the number of classes became very large. In order to maintain a minimum number of master file records in each class, geography was not used in the formation of these classes.

Alberta and Manitoba both had special additional content in 1994-95 and Alberta had special content in 1996-97. To get around this problem a dummy file was first created which did not include this special content, and the special content was then added to the dummy file records from Alberta and Manitoba at a later step. The problems of small classes did not arise in this step since the number of age-based classes for the special content was relatively small.

2.2 Block Formation

The second step was formation of blocks of variables. The artificial records for the dummy file were created by randomly swapping blocks of variables among the records within a class. The variables were first grouped into basic blocks corresponding to sections of the questionnaire, and the blocks then consisted of collections of these basic blocks. Where the questions asked in one section depended on answers to questions asked in a previous section, the corresponding basic blocks of variables were put into the same block, to maintain internal coherence. The exceptions to this rule

are the variables used in the formation of the classes, whose effects on the pathway through the questionnaires were already taken into account.

A guiding principle was that blocks should be analytically meaningful while also being small enough that they are safe from the point of view of confidentiality. Variables that in combination could identify individuals were put in different blocks.

For the formation of blocks, the variables were classified into the following basic blocks (apart from a number of geographic and design variables, individual and household characteristics, and survey administrative variables):

- | | |
|---|---|
| 1. Household income | 27. Injuries |
| 2. Two week disability | 28. Drug use |
| 3. Health care utilization | 29. Smoking |
| 4. Restriction of activities | 30. Alcohol |
| 5. Chronic conditions | 30a. Alcohol (Alberta, 1996 only) |
| 6. Socio-demographic | 31. Stress (1994 only) |
| 7. Education | 32. Recent life events (1994 only) |
| 8. Labour force | 33. Trauma (1994 only) |
| 9. General Health | 34. Work stress (1994 only) |
| 10. Height/weight | 35. Esteem and mastery (1994 only) |
| 11. Preventive health practices (1994 only) | 36. Sense of coherence (1994 only) |
| 12. Access to - blood pressure (1996 only) | 37. Mental health |
| 13. Access to - pap smear (1996 only) | 38. Social support |
| 14. Access to - mammography (1996 only) | 38a. Social support (Alberta, 1996 only) |
| 15. Access to - breast exam (1996 only) | 39. Att. toward parents (Alberta, 1996 only) |
| 16. Access to - phys. checkup (1996 only) | 40. Health services (Alberta, 1996 only) |
| 17. Access to - flu shots (1996 only) | 41. Sexual health (1996 only) |
| 18. Access to - dental visits (1996 only) | 41a. Sexual health (Alberta, 1996 only) |
| 19. Access to - eye exams (1996 only) | 42. Road safety (1996 only) |
| 20. Access to - emer. services (1996 only) | 43. Violence/per. safety (Alberta, 1996 only) |
| 21. HIV (1996 only) | 44a. Coping (Alberta, 1994) |
| 22. Health information (Alberta, 1996 only) | 44b. Coping (Alberta, 1996) |
| 23. Health status | 45. Relate to people (Manitoba, 1994 only) |
| 24. Physical activity | |
| 25. Tanning (Alberta, 1996 only) | |
| 26. Repetitive strain (1996 only) | |

Note that basic block numbers 22, 25, 39, 40, 43, 44 and 45 contain the special content for Manitoba and Alberta. This content was added at a third step, and in general the classes were based on age, sex and longitudinal response pattern.

2.3 Suppressed Variables

The following variables have been suppressed during the creation of the dummy files.

General Dummy Master File:

AM56_RNO, AM36_PAR, AM36_LP, AM36_LNG, AM56_LNG, AM56_SHA, AM66_SHA, AM36_SRC, AM56_SRC, AM56_CAS, AM56_BMM, AM56_BDD, AM56_BYE, AM56_EMM, AM56_EDD, AM56_EYY, SP36_CPS, SP36_TYP, SP36DPC, CYCLE, SUBCYCLE, MEMCYCLE, WT56_C, WT56_R.

Health Dummy Master File: all the variables listed under the General file excluding AM56_RNO, WT56_C and WT56_R;

AM66_RNO, AM66_LNG, AM66_CAS, AM66_RC, AM66_SRC, AM66_BMM, AM66_BDD, AM66_BYE, AM66_EMM, AM66_EDD, AM66_EYY, AM66_LNK, AM66_HNC, AM66_5, WT66_C, WT66_R, WT66_NR.

Longitudinal Dummy Master Files: all the variables listed under the Health file excluding AM66_RNO, WT66_C, WT66_R, WT66_NR;

AM34_LP, AM34_SRC, AM34_LNG, SP34_CPS, SP34_TYP, SP34_MET, AM64_LNK, AM54_BMM, AM54_BDD, AM54_BYE, AM54_SRC, AM54_EMM, AM54_EDD, AM54_EYY, AM64_BMM, AM64_BDD, AM64_BYE, AM54_LNG, AM64_STA, AM64_SRC, AM64_RC, AM64_LNG, AM64_EMM, AM64_EDD, AM64_EYY, DHC6ADOD, SP34_DPC, SP36FOUT, SP36_STA, AM56_STA, AM66_STA, AM66LDUR, AM66_HN.

2.4 Files Contained on the CD_R

ASCII files:

DUMYH35.TXT	General (H35) dummy file
DUMYH356.TXT	Health (H356) dummy file
DUMYLNFG.TXT	Full response longitudinal dummy file
DUMYLNFP.TXT	Partial response longitudinal dummy file

Data Dictionary Reports:

General - H35:

DICGEN_E.PDF	English data dictionary listing
DICGEN_F.PDF	French data dictionary listing
INDGEN_E.PDF	English alphabetical index of variables
INDGEN_F.PDF	French alphabetical index of variables
LAYGEN_E.PDF	English record layout
LAYGEN_F.PDF	French record layout
TOPGEN_E.PDF	English topical index of variables
TOPGEN_F.PDF	French topical index of variables

Health - H356:

DICHLT_E.PDF	English data dictionary listing
DICHLT_F.PDF	French data dictionary listing
INDHLT_E.PDF	English alphabetical index of variables
INDHLT_F.PDF	French alphabetical index of variables
LAYHLT_E.PDF	English record layout
LAYHLT_F.PDF	French record layout
TOPHLT_E.PDF	English topical index of variables
TOPHLT_F.PDF	French topical index of variables

Longitudinal Full - LNFG

DICLNF_E.PDF	English data dictionary listing
DICLNF_F.PDF	French data dictionary listing
INDLNF_E.PDF	English alphabetical index of variables
INDLNF_F.PDF	French alphabetical index of variables
LAYLNF_E.PDF	English record layout
LAYLNF_F.PDF	French record layout
TOPLNF_E.PDF	English topical index of variables
TOPLNF_F.PDF	French topical index of variables

Longitudinal Partial - LNGP

DICLNP_E.PDF	English data dictionary listing
DICLNP_F.PDF	French data dictionary listing
INDLNP_E.PDF	English alphabetical index of variables
INDLNP_F.PDF	French alphabetical index of variables
LAYLNP_E.PDF	English record layout
LAYLNP_F.PDF	French record layout
TOPLNP_E.PDF	English topical index of variables
TOPLNP_F.PDF	French topical index of variables

SAS Peripheral Files:

H35_IN.SAS	H35 input record layout
H35_FMT.SAS	H35 format statement
H35_LBE.SAS	H35 English variable labels (name + English concept (49 char.))
H35_LBF.SAS	H35 French variable labels (name + French concept (49 char.))
H35_OUT.SAS	H35 output record layout
H35_PFE.SAS	H35 proc format in English
H35_PFF.SAS	H35 proc format in French
H356_IN.SAS	H356 input record layout
H356_FMT.SAS	H356 format statement
H356_LBE.SAS	H356 English variable labels (name + English concept (49 char.))
H356_LBF.SAS	H356 French variable labels (name + French concept (49 char.))
H356_OUT.SAS	H356 output record layout
H356_PFE.SAS	H356 proc format in English
H356_PFF.SAS	H356 proc format in French
LNGF_IN.SAS	LNGF input record layout
LNGF_FMT.SAS	LNGF format statement
LNGF_LBE.SAS	LNGF English variable labels (name + English concept (49 char.))
LNGF_LBF.SAS	LNGF French variable labels (name + French concept (49 char.))
LNGF_OUT.SAS	LNGF output record layout
LNGF_PFE.SAS	LNGF proc format in English
LNGF_PFF.SAS	LNGF proc format in French
LNGP_IN.SAS	LNGP input record layout
LNGP_FMT.SAS	LNGP format statement
LNGP_LBE.SAS	LNGP English variable labels (name + English concept (49 char.))
LNGP_LBF.SAS	LNGP French variable labels (name + French concept (49 char.))
LNGP_OUT.SAS	LNGP output record layout

LNGP_PFE.SAS	LNGP proc format in English
LNGP_PFF.SAS	LNGP proc format in French

SPSS Peripheral Files:

H35_IN.SPS	H35 input record layout
H35VALE.SPS	H35 value labels in English
H35VALF.SPS	H35 value labels in French
H35VARE.SPS	H35 variable labels (name + English concept (49 char.))
H35VARF.SPS	H35 variable labels (name + French concept (49 char.))
H356_IN.SPS	H356 input record layout
H356VALE.SPS	H356 value labels in English
H356VALF.SPS	H356 value labels in French
H356VARE.SPS	H356 variable labels (name + English concept (49 char.))
H356VARF.SPS	H356 variable labels (name + French concept (49 char.))
LNGF_IN.SPS	LNGF input record layout
LNGFVALE.SPS	LNGF value labels in English
LNGFVALF.SPS	LNGF value labels in French
LNGFVARE.SPS	LNGF variable labels (name + English concept (49 char.))
LNGFVARF.SPS	LNGF variable labels (name + French concept (49 char.))
LNGP_IN.SPS	LNGP input record layout
LNGPVALE.SPS	LNGP value labels in English
LNGPVALF.SPS	LNGP value labels in French
LNGPVARE.SPS	LNGP variable labels (name + English concept (49 char.))
LNGPVARF.SPS	LNGP variable labels (name + French concept (49 char.))

Bootstrap Variance Program

Bootstrap weights used by the program BOOTVAR.SAS:

BD5H35.TXT	Bootstrap weights for the file dummyH35.TXT
BD5H356.TXT	Bootstrap weights for the file dummyH356.TXT
BD5H356A.TXT	Bootstrap weights for the file dummyH356.TXT Health Promotion Survey questions
BD5H356C.TXT	Bootstrap weights for the file dummyH356.TXT Child Health Services questions
BD5LNGF.TXT	Bootstrap weights for the file dummyLNGF.TXT
BD5LNGP.TXT	Bootstrap weights for the file dummyLNGP.TXT

Record layout:

BOOT_IN.SAS Record layout for the files containing the bootstrap weights

Documentation:

BTDOCENG.PDF	English documentation of the BOOTVAR.SAS program
BTDOCFR.PDF	French documentation of the BOOTVAR.SAS program
BOOTVAR.SAS	Bootstrap variance program with English comments
BOOTVARF.SAS	Bootstrap variance program with French comments
BOOTSAS.SAS	English Documentation of the BOOTSAS.SAS program
BOOTSASF.SAS	French Documentation of the BOOTSASF.SAS program
EX1A_ENG.SAS	Example 1A (English comments)
EX1A_FR.SAS	Example 1A (French comments)
EX1B_ENG.SAS	Example 1B (English comments)
EX1B_FR.SAS	Example 1B (French comments)
EX1C_ENG.SAS	Example 1C (English comments)
EX1C_FR.SAS	Example 1C (French comments)
EX2A_ENG.SAS	Example 2A (English comments)
EX2A_FR.SAS	Example 2A (French comments)
EX2B_ENG.SAS	Example 2B (English comments)
EX2B_FR.SAS	Example 2B (French comments)
EX2C_ENG.SAS	Example 2C (English comments)
EX2C_FR.SAS	Example 2C (French comments)
EX2BTSTE.SAS	Example 4 (English comments)
EX2BTSTF.SAS	Example 4 (French comments)