

Canadian Community Health Survey, 2005 (3.1)

Dummy Files for Remote Access to the Master Files

Mars 2007

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 Canadian Community Health Survey (CCHS) is a cross-sectional survey that collects information related to health status, health care utilization and health determinants for the Canadian population. The CCHS operated on a two-year collection cycle for 2005, 2003 and 2000/01. The first year of the survey cycle “.1” was a large sample, general population health regional survey, designed to provide reliable estimates at the health region level. The second year of the survey cycle “.2” is a smaller survey designed to provide provincial level results on specific focused health topics. The regional survey now collects information on a continuous basis since January 2007.

In order to provide extensive access to the data, a public use microdata file (PUMF) was produced for 2005 data. 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 an alternative, in cases where the PUMF does not meet their needs, researchers can perform their analyses on the CCHS master file(s). In order to provide researchers with a means to access the master file(s), 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 (cchs-esc@statcan.ca), and to receive the results by return e-mail. To obtain remote access privileges, it is necessary that researchers obtain advance approval from the Health Statistics Division. Requests must be submitted to the aforementioned e-mail address and must provide the following, clearly itemized 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 access to the master data rather than the PUMF,
- why 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 CCHS 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 CCHS 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 supplied on this CD-ROM, the researcher can develop and test his/her computer programs before submitting them to the CCHS team. For CCHS 3.1, the five following dummy data were created:

- **Full sample file:** File with all respondents, and all variables from common and optional modules; 44,796 records.
- **Subsample 1 file:** File with only respondents who were administered subsample 1 modules. Variables included are the ones specifically collected for subsample 1, as well as the ones from the common modules; 11,569 records.
- **Subsample 2 file:** Similar to subsample 1 file but specific to subsample 2; 3,253 records.
- **Subsample 3 file:** Similar to subsample 1 file but specific to subsample 3; 12,528 records.

While certain administrative variables (which are of no analytical interest) have been recoded to “blank” or “9”, the dummy files simulate the master files perfectly. 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. However, note that the number of records has been reduced by more than 60% compared to what is available in master versions of these files.

Corresponding dummy bootstrap weights files were also created for each of these files (except for the Group file which consists of the same sample as the full sample file, therefore, bootstrap weights for that later file must be used when dealing with the Group file). The CD-ROM includes a link to Bootvar (a program used for calculating the variance of estimates using the bootstrap method). SAS and SPSS versions of the program are available for free download from Statistics Canada's Website.

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.

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 Variables

All records are first classified by age, sex. The purpose for this is to create classes of records having the same questionnaire flow so that, after random permutation of the data within the classes, each dummy record contains data respecting a certain internal consistency.

2.2 Sampling and Restrictions of Variable Values

As an additional measure of preserving data confidentiality in the master files, the dummy files are created by sub-sampling the master files records. Also, quantitative variables are top-coded and bottom-coded. Variables considered as sensitive or without analytical value are recoded to “blank” or to “9” (see section 2.4). All of these modifications mean that we do not necessarily find, on the dummy files, all the possible categories for each variable. To obtain complete information on a specific variable, users are asked to consult the master files documentation. Finally, all the weight variables have been repoststratified according to the same methods used to poststratify the master files weights. However, it should be noted that the weights remain “dummy” weights. Other details related to the creation of this file are not available because they would disclose too much information on the method used for creating the dummy data, hence putting confidentiality at risk.

2.3 Block of variables

The third step consists of creating blocks of variables. Dummy records are produced by randomly swapping blocks of variables for the records in the same class. At first, elementary blocks of variables are grouped following the various modules of the questionnaire. Blocks do not always correspond to modules of the questionnaire, however, as some modules are split differently. If the content of the elementary block depends on the answers obtained in a previous block, the respective elementary block is amalgamated with the other one to preserve the internal consistency of the records.

Recode Variables to “blank” and to “9”

The following variables are recoded to “blank” or to “9” during the creation of the dummy files:

PERSONID SAME_TYP SAME_CP SAMEDSHR SAMEDLNK ADME_STA ADME_PRX DME_YOI ADME_MOI
ADME_DOI ADME_N09 ADME_N10 ADME_N11 ADME_LHH ADME_N12 DHHE_YOB DHHE_MOB DHHE_DOB