Canadian Community Health Survey (CCHS)

The central objective of the Canadian Community Health Survey (CCHS) is to gather health-related data at the sub-provincial levels of geography (health region or combined health regions).

Detailed information for 2005 (Cycle 3.1)

Data release – December 21, 2005 (this data release covers data collected from January to June 2005)

Survey description

In 1991, the National Task Force on Health Information cited a number of issues and problems with the health information system. These problems were that: data was fragmented; data was incomplete; data could not be easily shared; data was not being analysed to the fullest extent; and the results of research are not consistently reaching Canadians. In responding to the needs, the Canadian Institute for Health Information (CIHI), Statistics Canada and Health Canada have joined forces to create a National Health Information Roadmap.

The plan of action starts by seeking answers to two crucial questions:

1. How healthy is the health care system?
2. How healthy are Canadians?

To answer these questions, the Canadian community Health Survey (CCHS) was created. The primary objectives of the CCHS are to:

- provide timely, reliable, cross-sectional estimates of health determinants, health status and health system utilization across Canada,
- gather data at the sub-provincial levels of geography,
- create a flexible survey instrument that:
  - meets specific health region data gaps,
  - develops focused survey content for key data,
  - deals with emerging health and health care issues as they arise.

As a key component of the Population Health Surveys Program of Statistics Canada, the CCHS helps fulfil broader requirements of health issues in Canada. These are:

- aid in the development of public policy,
- provide data for analytic studies that will assist in understanding the determinants of health,
- collect data on the economic, social, demographic, occupational and environmental correlates of health,
- increase the understanding of the relationship between health status and health care utilization.

In Canada, the primary use of the data is for health surveillance, such as in prevalence of disease and other forms of health research. The data are used extensively by the

research community and other health professionals. The uniqueness of the CCHS arises from the regional nature of both content and survey implementation. These aspects allow for analysis of health data at a regional level, across Canada. Federal and provincial departments of health and human resources, social service agencies, and other types of government agencies use the information collected from the respondents to plan, implement and evaluate programs to improve health and the efficiency of health services. Non-profit health organizations and researchers in the academic fields use the information to make research to improve health. The media uses the results from the surveys to raise awareness about health, an issue of concern to all.

Reference period

Varies according to the question (for example: “over the last 12 months”, “over the last 6 months”, “during the last week”, etc.)

Notes to users about this data release

PLEASE NOTE: This data release covers data collected over the first 6 months (January to June 2005) of the CCHS Cycle 3.1. At that time, the survey had collected information from about 68,000 individuals, aged 12 and older.

Only part of the data collected with the CCHS Cycle 3.1 questionnaire has been processed and finalized for this release. Data covering the entire 12 months collection period (January to December 2005) will be released in summer 2006.

Target population

The CCHS targets persons aged 12 years or older who are living in private dwellings in the ten provinces and the three territories. Persons living on Indian Reserves or Crown lands, residents of institutions, full-time members of the Canadian Armed Forces and residents of certain remote regions are excluded from this survey. The CCHS covers approximately 98% of the Canadian population aged 12 or older.

Instrument design

Each CCHS cycle questionnaire has been conceived in collaboration with specialists from Statistics Canada, other departments and/or academic fields. The CCHS questions were designed for computer-assisted interviewing (CAI), meaning that, as the questions were developed, the associated logical flow into and out of the questions was programmed. This included specifying the type of answer required, the minimum and maximum values, on-line edits associated with the question and what to do in case of item non-response.

With CAI, the interview can be controlled based on answers provided by the respondent. On-screen prompts are shown when an invalid entry is recorded and thus immediate feedback is given to the respondent and/or the interviewer to correct inconsistencies.

Another enhancement is the automatic insertion of reference periods based on current dates. Pre-filling of text or data based on information gathered during the interview allows the interviewer to proceed without having to search back for previous answers. This type of pre-fill includes such things as using the correct name or sex within the questions themselves. Allowable ranges/answers based on data collected during the interview can also be programmed. In other words, the questionnaire can be customized to the respondent based on data collected at that time or during a previous interview.

One field test was conducted for cycle 3.1. The test involved Statistics Canada's Regional Offices. Experienced Labour Force Survey interviewers carried out interviews. The main objectives of the test were to observe respondent reaction to the survey, to obtain estimates of time for the various sections, to study the response rates and to test feedback questions. Field operations and procedures, interviewer training and the data collection computer application were also tested. In addition to the field test, the data collection computer application was extensively tested in-house in order to identify any errors in the program flow and text. The testing of the data collection computer application was an ongoing operation up until the start of the main survey.

Sample design

To provide reliable estimates to the 122 Health Regions (HRs), and given the budget allocated to the CCHS component, a sample of 130,000 respondents was desired. For this 6 month release, the 3 territories have been excluded from the file due to small sample size. A total of approximately 68,000 respondents covering 119 HRs are included in the 6-month file. The sample allocation strategy consisting of three steps, gave relatively equal importance to the HRs and the provinces. In the first two steps, the sample was allocated among the provinces according to their respective populations and the number of HRs they contain. In the third step, each province’s sample was allocated among its HRs proportionally to the square root of the estimated population in each HR.

The CCHS used three sampling frames to select the sample of households: 50% of the sample of households came from an area frame, 49% came from a list frame of telephone numbers and the remaining 1% came from a Random Digit Dialling (RDD) sampling frame. For most of the health regions, 50% of the sample was selected from the area frame and 50% from the list frame of telephone numbers. In two health regions (Northern Quebec and Northern Saskatchewan), only the RDD frame was used. In Nunavut, only the area frame was used. In Yukon and Northwest Territories, most of the sample came from the area frame but a small RDD sample was also selected in Whitehorse and Yellowknife.

The CCHS (Cycle 3.1) used the area frame designed for the Canadian Labour Force Survey (LFS) as its primary frame. The sampling plan of the LFS is a multistage stratified cluster design in which the dwelling is the final sampling unit. In the first stage, homogeneous strata were formed and independent samples of clusters were drawn from

each stratum. In the second stage, dwelling lists were prepared for each cluster and dwellings, or households, were selected from the lists.

For the purpose of the plan, each province is divided into three types of regions: major urban centres, cities and rural regions. Geographic or socio-economic strata are created within each major urban centre. Within the strata, dwellings are regrouped to create clusters. Some urban centres have separate strata for apartments or for census enumeration areas (EA) in which the average household income is high. In each stratum, six clusters or residential buildings (sometimes 12 or 18 apartments) are chosen by a random sampling method with a probability proportional to size (PPS), the size of which corresponds to the number of households. The number six was used throughout the sample design to allow a one-sixth rotation of the sample every month for the LFS.

The other cities and rural regions of each province are stratified first on a geographical basis, then according to socio-economic characteristics. In the majority of strata, six clusters (usually census EAs) are selected using the PPS method. Where there is low population density, a three-step plan is used whereby two or three primary sampling units (PSU), which normally correspond to groups of EAs, are selected and divided into clusters, six of which are sampled. The final sample is obtained using a systematic sampling of dwellings.

Data sources

The CCHS questionnaire is administered using computer-assisted interviewing (CAI). Sample units selected from the area frame are interviewed using the Computer-Assisted Personal Interviewing (CAPI) method while units selected from the Random Digit Dialling (RDD) and telephone list frames are interviewed using the Computer-Assisted Telephone Interviewing (CATI) method.

CAI offers a number of data quality advantages over other collection methods. First, question text, including reference periods and pronouns, is customised automatically based on factors such as the age and sex of the respondent, the date of the interview and answers to previous questions. Second, edits to check for inconsistent answers or out-of-range responses are applied automatically and on-screen prompts are shown when an invalid entry is recorded. Immediate feedback is given to the respondent and the interviewer is able to correct any inconsistencies. Third, questions that are not applicable to the respondent are skipped automatically.

CAPI interviewers work independently from their homes using laptop computers and are supervised from a distance by senior interviewers. Completed interviews are transmitted daily to Statistics Canada’s head office using a secure telephone transmission directly from the interviewer’s home. CATI interviewers work in centralised offices and are supervised by a senior interviewer located in the same office. Transmission of cases from each of 5 CATI offices to head office is the responsibility of the regional office project supervisor, senior interviewer and the technical support team.

An automated call scheduler, i.e. a central system to optimise the timing of call-backs and the scheduling of appointments, is used to support CATI collection.

Error detection

Most editing of the data was performed at the time of the interview by the computer-assisted interviewing (CAI) application. It was not possible for interviewers to enter out-of-range values and flow errors were controlled through programmed skip patterns. For example, CAI ensured that questions that did not apply to the respondent were not asked. In response to some types of inconsistent or unusual reporting, warning messages were invoked but no corrective action was taken at the time of the interview. Where appropriate, edits were instead developed to be performed after data collection at Head Office. Inconsistencies were usually corrected by setting one or both of the variables in question to "not stated".

Imputation

There was no imputation done.

Estimation

The principle behind estimation in a probability sample is that each person in the sample "represents", besides himself or herself, several other persons not in the sample. For example, in a simple random 2% sample of the population, each person in the sample represents 50 persons in the population. In the terminology used here, it can be said that each person has a weight of 50. The weighting phase is a step that calculates, for each person, his or her associated sampling weight. This weight appears on the microdata file, and must be used to derive meaningful estimates from the survey. For example, if the number of individuals who had a major depressive episode is to be estimated, it is done by selecting the records referring to those individuals in the sample having that characteristic and summing the weights entered on those records. In order for estimates produced from survey data to be representative of the covered population and not just the sample itself, a user must incorporate the survey weights into their calculations. A survey weight is given to each person included in the final sample, that is, the sample of persons having answered the survey. This weight corresponds to the number of persons represented by the respondent for the entire population.

In order to determine the quality of the estimate and to calculate the coefficient of variation (CV), the standard deviation must be calculated. Confidence intervals also require the standard deviation of the estimate. The CCHS uses a multi-stage survey design, which means that there is no simple formula that can be used to calculate variance estimates. Therefore, an approximative method was needed. The bootstrap method is used because the sample design information needs to be taken into account when calculating variance estimates. The bootstrap method does this, and with the use of the Bootvar program, remains a method that is fairly easy for users to use. The bootstrap re-sampling method used in the CCHS involves the selection of simple random samples

known as replicates, and the calculation of the variation between the estimates from replicate to replicate. In each stratum, a simple random sample of (n-1) of the n clusters is selected with replacement to form a replicate. Note that since the selection is with replacement, a cluster may be chosen more than once. In each replicate, the survey weight for each record in the (n-1) selected clusters is recalculated. These weights are then post-stratified according to demographic information in the same way as the sampling design weights in order to obtain the final bootstrap weights. The entire process (selecting simple random samples, recalculating and post-stratifying weights for each stratum) is repeated B times, where B is large. The CCHS typically uses B=500, to produce 500 bootstrap weights. To obtain the bootstrap variance estimator, the point estimate for each of the B samples must be calculated. The standard deviation of these estimates is the bootstrap variance estimator. Statistics Canada has developed a program that can perform all of these calculations for the user: the Bootvar program.

Quality evaluation

Survey design has a profound effect on the objectives of the survey which are listed under "Survey Description". To meet these objectives, a Steering Committee and an Advisory Board comprised of authorities from the provincial and territorial Ministries of Health, the Canadian Institute for Health Information, and Health Canada determined the concepts and focus. Expert Groups were convened to advise on the measures to obtain the results envisioned by the Steering Committee and Advisory Board, and to recommend proven collection vehicles and indices. The resulting data is recognized as valid measures of contemporary concepts such as: depression, activity limitation, weight problems and chronic pain.

The frames chosen to provide the sample, the Labour Force Survey, RDD and list frame of telephone numbers, have been combined with sampling design methodologies which have been tested, used repeatedly, and have been proven to produce accurate results. The large sample in each province/territory helps ensure accurate and meaningful results.

High response rates are essential for quality data. Actions have been taken to reduce non-sampling errors to a minimum. To reduce the number of non-response cases, the interviewers are all extensively trained by Statistics Canada, provided with detailed Interviewer Manuals, and are under the direction of interviewer supervisors. The extent of non-response varies from partial non-response (failure to answer just one or some questions) to total non-response. Partial non-response was basically non-existent. Total non-response occurred because the interviewer was either unable to trace the respondent, no member of the household was able to provide the information or the respondent refused to participate in the survey. Total non-response was handled by adjusting the weight of households that responded to the survey to compensate for those who did not respond. In most cases, partial non-response to the survey occurred when the respondent did not understand or misinterpreted a question, refused to answer a question, could not recall the requested information or could not provide personal or proxy information. Refusals were followed up by senior interviewers, project supervisors or by other

interviewers to encourage respondents to participate in the survey. In addition, to maximize the response rate, non-response cases were also followed up in subsequent collection periods.

Disclosure control

Statistics Canada is prohibited by law from releasing any data which would divulge information obtained under the Statistics Act that relates to any identifiable person, business or organization without the prior knowledge or the consent in writing of that person, business or organization. Various confidentiality rules are applied to all data that are released or published to prevent the publication or disclosure of any information deemed confidential. If necessary, data are suppressed to prevent direct or residual disclosure of identifiable data.

Public Use Microdata Files (PUMFs) based on the entire 12 months data from the CCHS Cycle 3.1 are planned for release in summer 2006. The PUMFs differ in a number of important aspects from the survey 'master' files held by Statistics Canada. These differences are the result of actions taken to protect the anonymity of individual survey respondents. First, only cross-sectional data are available on such files, because longitudinal information can lead to the identification of respondents. Also, some sensible variables are regrouped, capped or completely deleted from the files. Users requiring access to information excluded from the microdata files may purchase custom tabulations, or access the master files through the Research Data Centres program or the Remote Access program. Outputs are vetted for confidentiality before being given to users.

Before releasing and/or publishing any estimate from these files, users should first determine the number of sampled respondents who contribute to the calculation of the estimate. If this number is less than 30, the weighted estimate should not be released regardless of the value of the coefficient of variation for this estimate. For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the rounded estimate and follow the guidelines below.

Estimates in the main body of a statistical table are rounded to the nearest hundred units using the normal rounding technique. If the first or only digit dropped is zero to four, the last digit retained is not changed. If the first or only digit dropped is five to nine, the last digit retained is raised by one. Marginal sub-totals and totals in statistical tables are derived from their corresponding unrounded components and then are rounded themselves to the nearest 100 units using normal rounding methods. Averages, proportions, rates and percentages are computed from unrounded components (for example, numerators and/or denominators) and then are rounded themselves to one decimal using normal rounding. In normal rounding to a single digit, if the final or only digit dropped is zero to four, the last digit retained is not changed. If the first or only digit dropped is five to nine, the last digit retained is increased by one. Sums and differences of aggregates (or ratios) are derived from their corresponding unrounded components and then are rounded themselves to the nearest 100 units (or the nearest one decimal) using

normal rounding. Under no circumstances are unrounded estimates, published or otherwise, released. Unrounded estimates imply greater precision than actually exists.