

Interpreting Estimates from the Redesigned Canadian Community Health Survey (CCHS)

By Steven Thomas, senior methodologist, CCHS
and Sylvain Tremblay, senior analyst, CCHS



Abstract

In its attempt to better address user needs and to make better use of the interviewer resources, the regional component of the Canadian Community Health Survey, or the .1 survey, was redesigned to include varying types of content and to collect data continually over time. This change in structure allows for the collection and dissemination of various types of information for various time periods for estimation at various geographical and socio-demographic levels. For the user, this implies that several different products will be available for several different time periods. Proper interpretation of the results is now more crucial than ever as the user will have a choice in the product that they use in their analysis. The choice of product will be based on the characteristics they wish to study and the detail required in the estimates. This paper will clarify how the redesign will impact the user and aid in the proper interpretation of the resulting estimates.

1. The CCHS Redesign

After the release of the 2005 regional component of the Canadian Community Health Survey (CCHS cycle 3.1), the CCHS was redesigned to address two main points: to better address user needs and make better use of collection resources¹. The implementation of a continuous collection technique was the key step in addressing these points. At the same time, a flexible content structure was implemented to allow for varying content to be collected over various time periods. These changes affect the dissemination strategy in the types of content that can be released as well as the frequency of releases. With these changes in place, it was decided that it was a good time to implement certain methodological improvements including the implementation of a more time-efficient process.

1.1 Changes in Collection

The change that has the largest impact on users is the change to the data collection approach of the CCHS. In the past, the CCHS regional component collected data from roughly 130,000 respondents over a 12-month period every two years. Starting in January 2007, data are now continually from roughly 65,000 respondents throughout

¹ Béland Y., Dale V., Dufour J., Hamel M. The Canadian Community Health Survey: Building on the Success from the Past. *2005 Proceedings of the American Statistical Association Meeting, Survey Research Methods*. American Statistical Association, 2005.

each year. To ensure that the sample is collected continuously, a new sample of roughly 11,000 respondents is collected every two months where each sample is representative at the health region level for the specific time period. Samples collected in the Territories are representative of the population after 12 months.

1.2 Changes in Content

With the change to a continuous collection approach, it is now possible to collect various types of information (or content) over various time periods. The duration of collection depends on the characteristics of interest and the sample size required. For prevalent characteristics and general domains, the content only needs to be collected for a short time-period before there are enough respondents to produce a quality estimate. For less prevalent characteristics and more detailed domains, the content is collected over an extended time-period in order to obtain an adequate sample of respondents.

The main CCHS content components are still categorized under common and optional content, although the common content is now split into two sub-components: core and theme. While both sub-components are asked of all CCHS respondents, the core content is meant to remain relatively stable over time and the theme content is collected for 12 or 24 months and can rotate back into collection after two, four or six years. The optional content component gives health regions the opportunity to select content that addresses their provincial or regional public health priorities. It can either be collected for one or two years before it is reviewed again.

A new component called Rapid Response is also available which allows the collection of data on emerging health issues from a small sample of respondents over two months of collection (approximately 11,000 respondents). This component, with a maximum duration of 2 minutes, is offered to cost-recovery clients with an immediate need for national-level data.

1.3 Changes in Dissemination

The changes to the collection and content structure of the CCHS have an impact on the dissemination strategy. In the past, information was disseminated every second year after collection of all respondents for the survey. Data files (Master, Share, PUMF) are available for the 2000/2001 (Cycle 1.1), 2003 (Cycle 2.1), and 2005 (Cycle 3.1) reference years. A 6-month file (allowing estimates to be calculated with 65,000 respondents) was produced from the Cycle 3.1 data collected from January 2005 to June 2005.

Beginning in June 2008, with the release of data collected during the 2007 collection period, master and share data files will be released every year. These annual data files will contain about 65,000 respondents, or half the sample size available with previous CCHS data files. These files will include core, theme and optional content collected throughout the year.

In June 2009, two main files will be made available: a main data file based on the 2008 collection period, which will be similar to the main 2007 data file, as well as a main data file based on the 2007-2008 collection period. The 2007-2008 file will be similar in size to files from the previous cycles (approximately 130,000 respondents). It will include core, optional and the theme content collected over the two-year period. One-year themes will not be available on the two-year data file. Also, theme modules collected from sub-sample of respondents will continue to be disseminated in separate files. These files include core content and sub-sample theme modules only. See table 1 for a clarification of what will be available with the 2007 and 2008 releases.

Table 1. Content components included in 2007 and 2008 data files

Files		Core content	2007 Theme ¹	2008 Theme ²	2007-2008 Theme	Optional content ³
2007	Main	Yes	N/A	N/A	Yes	Yes
	Sub-sample	Yes	Yes	N/A	No	No
2008	Main	Yes	N/A	Yes ⁴	Yes	Yes
	Sub-sample	Yes	N/A	Yes ⁵	No	No
2007-2008	Main	Yes	No	No	Yes	Yes

- 1 The 2007 theme was comprised of three modules (Patient satisfaction, Access to health care services and Waiting times) which were all asked to a sub-sample of respondents.
- 2 The 2008 theme is formed of a group of modules related to chronic disease screening and a module on measured height and weight. This last module is asked of a sub-sample of respondents.
- 3 This assumes that optional content remains the same for the two years. If not, it will only be included in the file of the year in which it was collected.
- 4 Chronic disease screening.
- 5 Measured height and weight.

In addition to the regular files, rapid response files will be produced for cost-recovery clients. These files will be available to other users upon request and will contain the rapid response content along with core content for a 2-month period.

Public-use Microdata Files (PUMFs) will be released every second year based on two years of collection. The first PUMF will be released Summer 2009 based on the 2007-2008 collection period. Single year PUMFs will not be available.

1.4 Changes in Survey Methodology

With the changes to the collection, content and dissemination strategies, certain changes were made to the methodology used in calculating survey weights. The redesign meant that weights would be produced more frequently and a methodology consistent with

continuous collection was required. This evolution was also seen as an opportunity to make certain improvements to the weight adjustments that are used in the process².

1.4.1 Period weighting

The weights are controlled, as best possible, to ensure that each collection period is equally represented with the weight and the weighted respondents represent the average population for the extended period of the particular release. Estimates represent the average over the time period.

1.4.2 Changes to integration

The CCHS uses a dual frame methodology where respondents are sampled from a telephone list frame and an area frame. Weights are adjusted / integrated to ensure that the population is represented only once. In the past, the weights on the telephone frame were adjusted for undercoverage (no landline, unlisted numbers, etc.) before integration with the area frame to ensure that the area and telephone list frames covered the same population. This required the assumption that those individuals not on the telephone frame were the same as those who were.

Knowing that the characteristics of telephone respondents can differ from those from those not covered by the telephone frame, the integration method has been updated³. Now, telephone frame respondents are integrated only with those units on the area frame who are also on the telephone frame. Those respondents on the area frame who are not on the telephone frame do not have their weights adjusted. This means that for variables affected by mode of collection, the resulting estimates should be more representative of the actual population.

1.4.3 Changes to calibration

The final step of the weighting procedure is to ensure that the weights sum to known population totals through a process known as calibration. These known totals are usually at the health region by age group by sex level. It is generally accepted that by calibrating weights, estimates for totals are more precise than those not calibrated. However, in order to do a proper calibration adjustment, it is suggested to have at least 20 observations in the domain. This should not be a problem with a 2-year file but with the 1-year file it will not be possible to post-stratify in all domains because of the reduced number of respondents. Users will be provided with a list of post-stata with less than 20 observations and corresponding cells will be suppressed from tabular data produced by Statistics Canada.

² Sarafin C., Simard M., Thomas S. (2007). A Review of the Weighting Strategy for the Canadian Community Health Survey. *2007 Proceedings of the Survey Methods Section, Statistical Society of Canada Annual Meeting*.

³ Skinner, C.J. and Rao, J.N.K. (1996). "Estimation in Dual Frame Surveys with Complex Designs". *Journal of the American Statistical Association*, 91, 349-356.

2. Impact on Users

2.1 More data, more often

Starting with the release of the 2008 and 2007-2008 data in June 2009, users will have the choice of working with one-year or two-year files. Eventually, it will be possible for users to combine these standard files to produce, for example, three-year or four-year files.

2.2 Period estimation

Whether a multi-year, two-year or one-year file is being used, users are encouraged to think of CCHS data as involving *period estimation*, in which the interviews corresponding to a period of time are combined and an updated sampling weight calculated. An annual estimate of a given characteristic is reflective of the average characteristics of the average population for the time period. In the case of the 2007 file, estimates are reflective of the average from January to December 2007. The result is a *period estimate* which is different from the snapshot idea that is often presented with most cross-sectional surveys. Technically, this is true only of the Census, where estimates represent a point in time.

The idea of period estimation is simply an extension of the methods used for previous cycles of CCHS, in which a set of interviews conducted over a 12-month period were combined. Similarly, the techniques involved in combined standard one-year or two-year data sets to create customized period estimates will be very similar to those used in combining cycles 1.1, 2.1 and 3.1 of the survey⁴.

Decisions about which period to use in a given analysis should be guided by the level of detail and the quality required. With a one year file, estimates will not always be available because of the quality associated with the limited sample size. The CCHS recommends having a Coefficient of Variation of less than 33% and having at least 10 respondents in the domain with the characteristic before publishing an estimate. This will not be possible for rare characteristics and detailed domains with a one-year file. Instead, users will have to rely on two-year files or multi-year accumulations.

Where the use of either a one-year or two-year file is viable, the user should consider the trade-off between accuracy and currency. If it is important to reflect the current characteristics of a population as closely as possible, the one-year file would be preferable. With two-year files, year-to-year trends will be masked, just as the seasonal trends are masked in a one-year file. However, with the increased sample size, more detailed estimates and analyses can be carried out.

⁴ Thomas S. Combining Cycles of the Canadian Community Health Survey. *Proceedings of Statistics Canada Symposium* (Statistics Canada, Catalogue no. 11-522-XIE), 2006.

2.3 Impact on variable naming convention

The variable naming convention has been changed slightly to reflect the fact that the same variable is being collected each year. In the past a letter designating the cycle was included in the variable name. For example, the 'e' in 'ccce_101' meant that it was the information collected from cycle 3.1. From now on the variable will be labeled 'ccc_101'. To help users wanting to combine two data files or more, a new variable showing the reference period "REFPER" was added. This variable uses the following format YYYYMM-YYYYMM (collection start year and month - collection end year and month).

2.4 Differences in Estimates Compared to the Past

Users should be aware that changes to sampling and the production of sampling weights introduced in 2007 might partially explain differences from previous cycles. In terms of sampling, the sample is controlled to have roughly the same number of respondents collected throughout the year and controlled to ensure that half the sample is from each of the two frames. This is not a dramatic change from the previous releases where the sample was divided into monthly collection periods. In terms of the production of weights, changes made to the process of integrating telephone and area frame samples could have the effect of influencing characteristics which are strongly correlated with having a listed phone number⁵. Further studies of this possibility are planned.

⁵ St-Pierre M, Béland Y. Mode effects in the Canadian Community Health Survey: a comparison of CAPI and CATI. *2004 Proceedings of the American Statistical Association Meeting, Survey Research Methods*. Toronto: American Statistical Association, 2004.

Highlights

- Beginning with the June 18, 2008 release, master and share data files will be released every year. These annual files will contain about 65,000 respondents or half the sample size of previous data files. Data files based on two years of data will continue to be produced and will be similar in size to files from the previous cycles (~130,000 respondents).
- Theme content was introduced with the CCHS redesign. This content is asked of all CCHS respondents and collected for one or two years only.
- Annual sample files will include core content, annual theme content and the 2-year theme and optional content collected that year. The two year files, will include the core content, the 2-year theme and all optional content collected for two years.
- Beginning in June 2009, users will have a choice between using one-year or two-year files.
- With single-year estimates, year-to-year trends can be calculated. Given the idea of continuous collection, each annual estimate is reflective of the average characteristics of the average population for the time period.
- To estimate rarer characteristics in more detailed domains, the use of two-year files, or even multi-year accumulations, will be necessary to ensure good data quality (33% CV with minimum of 10 respondents having the characteristics).
- The CCHS variable naming convention has been changed slightly to reflect the fact that the same variable is being collected. The letter designating the cycle (e.g., “e” for cycle 3.1) was dropped from the variable name.