

## HEALTH SURVEYS – CROSS-SECTIONAL SAMPLES

### Aspects That May Explain Differences In The Estimates Obtained From Two Different Survey Occasions

\*\* Work in progress (February 2003)  
STC/HSMD

Since 1994, Health Division has produced, through its surveys, a series of data files from cross-sectional samples. Unlike longitudinal samples, these samples have the characteristic of being uniquely representative of the year in which the data was collected. The available cross-sectional data comes from the National Population Health Survey (NPHS) for the years 1994-95, 1996-97 and 1998-99, and from the Canadian Community Health Survey (CCHS) for 2000-01. In situations where a variable has been collected on several occasions, it is possible for analysts to produce cross-sectional estimates and to therefore examine the trend of that variable over time. Inevitably, differences in these estimates will be observed, and these differences could come from multiple sources. This document reports the various aspects that may explain the differences between estimates obtained from the different NPHS and CCHS cross-sectional files. Note that the NPHS comparisons are made using the Health file (as opposed to the General file), that is the file that contains the data on the selected respondent. This file greatly resembles that of CCHS in terms of content as well as sampling (i.e. they both contain a selected person(s) in from household.)

#### Methodological Aspects

- Target Population:

NPHS (household component) and CCHS cover the same population and have the same exclusions. The only difference comes from the fact that CCHS covers only those persons aged 12 years and over, while NPHS generally covers the entire population. Coverage details for NPHS can be found at a later point in this document. Due to this difference and in order to enable comparisons, the indicators presented in this document refer only to persons aged 12 years and over, whenever possible. In terms of geography, note that both surveys cover the 10 provinces and the territories. However, the territories are covered by an independent component (North component) for NPHS, and have been excluded from this document for that reason.

- Questionnaire:

A difference in how the questions are constructed could have an impact on the estimates. The majority of the concepts measured by NPHS and CCHS use the same question over time; however, one should verify this by checking the questionnaires before interpreting the results. The same holds true for derived variables that may have been constructed differently from occasion to another.

- Collection

	NPHS 1994-95	NPHS 1996-97	NPHS 1998-99	CCHS 2000-01
Collection Period	June 1994 to June 1995	June 1996 to July 1997	June 1998 to June 1999	Sept. 2000 to Oct. 2001
Method (% by telephone; 12+)	27.7%	98.9%	91.1%	53% <sup>2</sup>
Response Rate (household; all ages)	88.7%	82.6%	87.6%	89.9%
Response Rate (person; 12+)	95.8%	95.6%	98.4%	92.6%
Proxy Response Rate (12 + ) <sup>1</sup>	4.2%	2.3%	2.4%	6.3%
Interview Length (approximate)	50 min.	50 min.	50 min.	45 min.

1 Certain modules could not be asked by proxy. Check the questionnaires to see which ones.

2 The operational structure used to conduct interviews by telephone changed in the year 2000. From then on, for NPHS, all telephone interviews were conducted from the interviewer's home. A portion of the CCHS interviews were also made from the interviewer's home, while the others were made from call centres.

- Cross-sectional file composition:

Survey	Composition	Origin (frames used)	Population covered	Number of person selected per household	Geographic representativity	Special characteristics
NPHS 1994-95	Panel members + buy-in sample for 4 provinces (ON, BC, NB, MAN)	Area (panel and buy-in; 84%) + RDD (buy-in; 16%)	0+	1	National + provincial, and regional for ON, BC, NB & MAN	
NPHS 1996-97	Panel members + buy-in sample for 3 provinces (ON, AB, MAN)	Members chosen in 1994-95 are recontacted (panel; 19%) + RDD (buy-in; 81%)	2+, except ON, AB and MAN where it is 0+	<ul style="list-style-type: none"> <li>• Panel = no selection (same person as Cycle 1).</li> <li>• RDD ON = 1 person 12+</li> <li>• RDD AB &amp; MAN = 1 person 12+, and a child (0-11) when possible.</li> </ul>	National + provincial, and regional for ON, AB, MAN	
NPHS 1998-99	Panel members + top-up sample	Members chosen in 1994-95 are recontacted (panel; 87%) + RDD (top-up; 13%)	0+	1 (same person as Cycle 1 for the panel)	National + provincial	Top-up sample is made up of babies (0-1 years) and new immigrants. Drawn from rotation groups exiting the LFS
CCHS 2000-01	Purely cross-sectional sample	Area (82%) + telephone frames (18%) – the percentage varies from one region to another	12+	<ul style="list-style-type: none"> <li>• Area frame = 1 or 2 depending on household composition</li> <li>• Telephone frames = 1</li> </ul>	National + provincial + regional	

Note:

- For NPHS 1996-97 & 1998-99, the part of the sample made up of panel members could be seen as a group of people who are more co-operative since they have already committed to being part of a panel.
- For CCHS, the age group 12-19 was oversampled compared to those 20-64, which will give better variances for estimates of this age group. This oversampling was performed by selecting one or two people by household, depending on the composition of the household.
- The sample was distributed according to the representativity needed on each occasion. For example, the CCHS sample was distributed in order to cover each of the 136 health regions, while the NPHS sample was distributed in order to give good representativity at the provincial level. Therefore, the composition of the sample is much more “rural” for CCHS than for NPHS due to the constraint of covering the entire country.

However, the weighting controls this overrepresentation of the rural area for CCHS (see table below).

## Percentage of the sample and population living in a rural area (12+ & provinces only)

	<b>NPHS 1994-95</b>	<b>NPHS 1996-97</b>	<b>NPHS 1998-99</b>	<b>CCHS 2000-01</b>
Sample (% rural)	23.4%	21.2%	23.2%	26.4%
Population (% weighted rural)	16.8%	17.5%	18.5%	18.3%

- Sample size (respondents aged 12 and over from the Master files)

<b>Province</b>	<b>NPHS 1994-95</b>	<b>NPHS 1996-97</b>	<b>NPHS 1998-99</b>	<b>CCHS 2000-01</b>
CANADA (excluding the territories)	17,626	73,402	15,249	129,018
Newfoundland	918	868	875	3,870
Prince Edward Island	899	829	844	3,651
Nova Scotia	911	882	943	5,319
New Brunswick	1,111	929	948	4,996
Quebec	2,581	2,521	2,593	22,667
Ontario	5,187	39,010	4,148	39,278
Manitoba	1,420	11,816	1,021	8,470
Saskatchewan	1,005	942	980	8,009
Alberta	1,310	14,203	1,384	14,456
British Columbia	2,284	1,402	1,513	18,302

NOTE: The difference in sample sizes will obviously be reflected in the precision of the estimates produced with the various data files.

- Weighting:

- Seasonality:

For NPHS, the weighting never included specific adjustments to control seasonality. However, collection was conducted in equal time periods (quarters) to more or less cover the four seasons. For CCHS, collection was also planned to evenly distribute the sample over the four seasons, however, operational problems during collection caused the sample to be unbalanced. To remedy this situation, an adjustment controlling for seasonality was incorporated in the weighting.

- Post-stratification:

The goal of post-stratification is to restore the sums of the weights so that they correspond exactly to the estimated population. Post-stratification is done independently within each region/province for a number of age-sex groups. These groups, as presented below, were defined differently during the various survey occasions.

Age groups used for post-stratification:

- NPHS 1994-95: 12-24, 25-44, 45-64, 65+ (no children in Cycle 1)
- NPHS 1996-97: 2-11, 12-24, 25-44, 45-64, 65+ (except for provinces with a buy-in sample where the group 0-1 was added)
- NPHS 1998-99: 0-11, 12-24, 25-44, 45-64, 65+ (a pre-poststratification step was applied to the 0-3 & 4-11 groups, at the Canada\*sex level)
- CCHS 2000-01: 12-19, 20-29, 30-44, 45-64, 65+

Note: For estimates of the total number of people per age group, the closer the age group is to the interval used for the post-stratum, the smaller the variance will be (for example, with equal sample sizes, an estimate for the number of 12-17 will have a much smaller CV for CCHS than NPHS since this age group is almost the same as one of the post-strata, i.e. 12-19).

- Imputation:

NPHS did not use imputation for any of the first three cycles that are discussed in this document. Any missing value is coded as such, without being replaced by another value in the data file. As for CCHS, some variables had to be imputed due to a proxy response rate that was too high. In the case of proxy responses, many questions were not asked due to their private or personal nature. Consequently, a high nonresponse rate to these questions was observed. Imputation was therefore used to obtain data for these questions that were unanswered due to a proxy interview. An article by St-Pierre and Béland (2002) explains the situation, as well as the method used.

Reference: St-Pierre, M. & Béland, Y. (2002). Imputation of Proxy Respondents in the Canadian Community Health Survey. *Proceedings of the Survey Methods Section*. Statistical Society of Canada.

- Method to calculate to variance (bootstrap):

The bootstrap is used for all survey occasions, however certain technical details differ from one occasion to another.

- NPHS 1994-95: incorporates post-stratification only
- NPHS 1996-97: incorporates post-stratification only
- NPHS 1998-99: incorporates nonresponse (household and person) and post-stratification
- CCHS 2000-01: incorporates all of the adjustments, from the household nonresponse adjustment onwards, in the bootstrap

- Sample variability

The fact that information is collected from a sample, and not from the entire population, means that the results obtained will all be subject to sample variability. The variability related to each estimate produced may, in some cases, explain the difference between the estimates obtained at different survey occasions. To find out if a difference really is significant and not due only to the variability of the estimates, statistical tests must be performed. For example, a Student test will check if two aggregate values differ significantly from one another.

## Contextual Aspects

- Changes in health standards

Some variables are derived according to a particular standard. For example, depending on the value, the body mass index determines that a person is obese if their index is above a certain standard. Similarly, certain clinical standards are used to determine if a person suffers from a particular illness or chronic health problem. These standards sometimes change over time according to advances in the field of health.

For example, the criteria used to determine if a person is diabetic was modified in the 1990s. According to the standards set out by the World Health Organization (WHO) in 1985, diabetes was defined as: *a fasting glucose level equal to or exceeding 7.8 mmol/L or a 2-hour post-challenge glucose level equal to or exceeding 11.1 mmol/L, or both*. In 1997, the *American Diabetes Association* adopted fasting glucose levels as the primary standard and reduced its level from 7.8 to 7.0 mmol/L. The *1998 Canadian Clinical practice guidelines for the management of diabetes* then adopted this change. This change could in theory have an effect on the incidence (and prevalence) of diabetes in Canada. It is therefore important to keep up-to-date on the changes adopted for the diagnosis of an illness or chronic health condition by clinical organizations.

- True change in the population

After examining all of the methodological aspects, it remains that the difference observed between two survey occasions could in fact be real. Health is a very dynamic field and is constantly evolving; different health indicators are therefore subject to fluctuations.