Special Issue on Longitudinal Methodology, Canadian Studies in Population
Vol. 28(2), 2001, pp. 377-390

After the First Steps: The Evolution of the National Population Health Survey

Douglas Yeo
National Population Health Survey
Household Survey Methods Division
Statistics Canada
Ottawa, Ontario, Canada

Abstract:

Many changes take place over the lifetime of a longitudinal panel survey. Changing priorities, new supplements, and conflicting demands are factors that may be unforeseen. The evolution of the National Population Health Survey (NPHS) since its first cycle in 1994/95 is discussed in this context. Statistics Canada contacts panel members every two years for twenty years, to estimate the health of Canadians and its determinants, health care use, and other characteristics. The NPHS was designed to provide both longitudinal and cross-sectional estimates, and to allow sample and content supplements. This paper describes the NPHS and the changes in focus needed to move the panel forward to cycle 2 and beyond.

Résumé

Plusieurs changements ont lieu pendant la vie d'une enquête longitudinale par panel. Les changements de priorités, de nouveaux suppléments, et des demandes conflictuelles sont des facteurs qui peuvent ne pas avoir été prévus. L'évolution de l'Enquête nationale sur la santé de la population (ENSP) depuis son premier cycle en 1994/95 est discuté dans ce contexte. Statistique Canada contacte les membres du panel tous les deux ans pendant vingt ans, pour estimer la santé des Canadiens et ses déterminants, l'utilisation des soins de santé, et d'autres caractéristiques. L'ENSP a été conçu pour fournir des estimations longitudinales et transversales, et pour permettre l'ajout d'échantillon et de contenu supplémentaires. Ce papier décrit l'ENSP et les mises au point nécessaires pour mener le panel au deuxième cycle et au-delà.

Key Words: Panel surveys, variance estimation, bootstrap, confidentiality, public-use microdata files

Introduction

Longitudinal surveys in general, and Panel surveys in particular, have very different focuses and needs than purely cross-sectional surveys. Sample size and allocation must be considered carefully, to represent the population well over a long period of time. Respondent relations must be stressed to keep attrition of the sampled panel to a minimum. Data management and analytical issues are much more complex. Some issues are a concern only longitudinally, such as tracing of respondents, historical editing, complex non-response definitions, inconsistent data over time, cycle-to-cycle or 'wave' imputation (which can be complicated by changing content from cycle to cycle), and variance estimation for correlated cycle-to-cycle estimates. There are also additional confidentiality concerns with public-use microdata files (PUMFs) for longitudinal surveys, which arise from conspicuous transitions in key variables over time.

In spite of these differences, the first cycle of a panel survey often resembles quite closely a one-time cross-sectional survey. In this paper I shall refer to the NPHS, but many of the points raised will apply to other longitudinal surveys.

'Cycle 1' resembles a cross-sectional survey due to factors both deliberate and random. In the former category, many longitudinal surveys are quite consciously designed to provide both cross-sectional and longitudinal estimates. Both types of estimates are often included in a wide range of products, including PUMFs. Naturally, estimates from cycle 1 will be only cross-sectional in nature. Priority is thus often given in the first year(s) of the panel to a cross-sectional focus, and for good reason, since these cross-sectional estimates may be of critical importance. The decision to produce cross-sectional, as well as longitudinal, estimates may be

the most important decision to be made in the planning stages of a longitudinal survey. Some surveys focus largely or entirely on longitudinal estimates. In other cases, overwhelming need for cross-sectional data leads to the design of a dual-purpose survey. This was certainly the case for the NPHS, since a comprehensive full-scale health survey had not been carried out in the previous fifteen years, and there was a pressing demand for cross-sectional health data.

The challenge arising from this cross-sectional focus early in the life of a panel is to balance it with long-term longitudinal priorities. Production pressures and the demand for data drive the order and even the form of data products released by a longitudinal survey. As well, cycle 1 of a survey has a wide-ranging conditioning effect on the expectations of users, who have seen a release of cross-sectional, often very high-profile, data products. Demand for comparable cross-sectional cycle 2 data naturally increases. At the same time, internal systems have already been put in place to produce these products. Computer systems, documentation, and, especially, analytic tools can be rapidly updated to prepare for the release of cycle 2 cross-sectional data. This conditioning effect may push the cross-sectional component out in front, with the same release schedule as in cycle 1, delaying the longitudinal data release. Finally, on the operational side, the very nature of the cross-sectional component, with its supplementary samples and occasional top-up samples (discussed in the third section, *Switching to a Longitudinal Footing*), will often give it priority.

Other factors affect the evolution of the panel survey. Twenty years (as in the case of the NPHS) is a long time. Along with staff turnover and budget adjustments, data needs and the reactions of respondents will change over time, sometimes with huge impacts on the life of the survey. The NPHS has seen some of these changes in direction already, only three cycles into the ten-cycle lifetime of its first panel.

I shall discuss these issues and conflicts in the section Switching to a Longitudinal Footing, followed by some planned changes and future work envisioned for the NPHS in the final section. First, though, some background on the NPHS, the focus of this paper.

Background

Before 1994, surveys on population health were carried out only occasionally by Statistics Canada. The last Canada Health Survey, in 1978, was truncated because of budgetary pressures. Statistics Canada had also carried out the Health Promotion Survey in 1990, as well as a cycle (1991) of the General Social Survey that focussed on health, but these surveys were limited in scope. In the early 1990s, it was recognised that a major investment was needed in the surveying of health, especially given the cost and high profile of health care in Canada. The National Health Information Council (NHIC) recommended that a national survey of population health be instituted. This recommendation was based on consideration

of the economic and fiscal pressures on the health care system and the commensurate requirements for information to improve the health status of the population in Canada. Existing sources of health data were unable to provide a complete picture of the health status of the population and the myriad factors having an impact on health. For these reasons, Statistics Canada received funding for the development of a longitudinal health survey. The survey was designed to be flexible and to produce valid, reliable, and timely data. Also, it was to be responsive to changing requirements, interests, and policies. Cycle 1 of the NPHS was carried out in 1994, with a longitudinal panel of about 20,000 members to be recontacted every two years for a period of twenty years. Its objectives were the following:

- To aid in the development of public policy:
- by understanding the determinants of health, as well as the economic, social, demographic, occupational, and environmental correlates of health
- by exploring the relationship between health status and health care utilisation
- To follow a panel of people to reflect the dynamic process of health
- To provide means to supplement content or sample
- To allow linkage with administrative data

This panel would provide longitudinal estimates of change over time, as well as cross-sectional 'snapshots' of Canadian health every two years. The panel would comprise one selected person per household. A second questionnaire, covering other household members, would only be used for cross-sectional estimates. Additional sample would be added every second cycle, starting in 1998. These Atop-up@ samples would be used to provide proper cross-sectional representation of the sample over time, and would not be kept longitudinally.

Content

Although the term "questionnaire" is used in this report, data collection was largely done by Computer-Assisted Interviewing (CAI). Further details, and the questionnaires themselves, are available at the Statistics Canada website, *www.statcan.ca*, under "Concepts, definitions and methods." Click on "Questionnaires and data dictionaries", then "NPHS.")

The NPHS was to focus on the health status of Canadians and the determinants of health. Data on all household members would be collected using a brief General questionnaire with sociodemographic and limited health sections, while a randomly selected member would be administered a much more lengthy, in-depth Health questionnaire. Main areas of core content of the NPHS include the following:

General questionnaire

- Two-week Disability
- Health Care Utilization

- Restriction of Activities
- Chronic Conditions
- Sociodemographic Characteristics
- Education
- Labour Force
- Income

Health questionnaire

- Self-perceived Health
- Women=s Health
- Blood Pressure
- Height/Weight
- Health Status
- Physical Activities
- Repetitive Strain (starting in cycle 2)
- Injuries
- Use of Medications
- Smoking
- Alcohol
- Mental Health
- Social Support
- Sense of Coherence (in cycles 1 and 3)
- Alcohol Dependence (in cycle 2)

As well, focus content would be added occasionally to a particular cycle. In cycle 1, questions were added on stress (ongoing problems, recent life events, childhood and adult stressors or traumas, and work stress), self-esteem, and mastery. These modules will be repeated in cycle 4. In cycle 2, the focus was on access to services: blood pressure tests, pap smear tests, mammographies, breast examinations, breastfeeding, physical checkups, flu shots, dental visits, eye examinations, emergency services, and insurance coverage. Self-care, family medical history, and insurance coverage, again, were examined in cycle 3.

A major goal of the NPHS was to allow supplements to the survey. Since health care is a provincial concern in Canada, provincial health ministries have been interested in funding additional content or sample for improved cross-sectional estimates, particularly in cycle 2, when very large sample supplements were funded by Alberta, Manitoba, and Ontario. (More on these supplements in the section *Conflicts*, below.)

Additional content has also been funded externally in every cycle. In cycle 1, Health Canada requested a supplement, separately weighted, covering a variety of topics including nutrition, smoking, injury prevention and safety, breast-feeding, consumption of alcohol and other drugs during pregnancy, sexual health and health care services. In cycle 2, questions from the Health Promotion Survey were integrated into the questionnaire. As well, the cycle 2 NPHS asthma questions were

used as screening questions to create a frame of asthma sufferers. A separate Asthma questionnaire was administered to these persons three to four months after the NPHS data were collected. This survey was funded by Health Canada. Similarly, in cycle 3 Human Resources Development Canada (HRDC) requested a Food Insecurity Survey. Several questions on the NPHS, this time dealing with the quality and quantity of food available to families, were once again used as filter questions. A follow-up interview was done of households answering at least one of these questions positively.

Data Products

NPHS data exist in the form of internal master files, special microdata files shared with Health Canada and the provincial health ministries (only for those respondents who have given permission), PUMFs, and various analytic products. Statistics Canada's Data Liberation Initiative has allowed academia affordable and equitable access to PUMFs and other data. Main results of each cycle are released in the Statistics Canada Daily along with an Overview Report highlighting key findings.

The NPHS is also a principal source of data for *Health Reports*, a quarterly Statistics Canada journal, indexed in Medline and available at *www.statcan.ca* (publication no. 82-003*)*. The Winter 1999 issue (Vol. 11, No. 3) served as the Overview Report for cycle 3 of the NPHS. Two free publications are also available at the website, under "Concepts, definitions and methods" and then "Questionnaires and data dictionaries":

82-F0068-XIE Information about the National Population Health Survey

82-570-XIE Statistical Report on the Health of Canadians

Analysts outside Statistics Canada can gain access to the master files in four ways: through custom tabulations, remote access, direct access at the Statistics Canada offices in Ottawa or in its regional offices, and through the Research Data Centres now being set up in universities across the country, in partnership with the Social Sciences and Humanities Research Council (SSHRC). Referring to the first two means of access, custom tabulations are available for a fee, while remote access is free to those who have purchased a PUMF. In the latter case, dummy survey data files are provided to analysts, to be used to write and debug programs. These programs can then be sent to Statistics Canada to be run against the master files, with the output checked for confidentiality and e-mailed back to the analyst, often with a turnaround time of a day or two. For variance estimation, dummy bootstrap files are also being prepared for the first three cycles. For further details, see *Variance Estimation*, below, and Mantel and Nadon (1999).

The last two methods of direct access are quite similar. Access is limited to researchers with approved projects and who are sworn in as deemed employees of Statistics Canada under the Statistics Act. Researchers must provide a working

paper or other product to Statistics Canada as part of the research agreement, and must attend a one-day training session on confidentiality.

Note also that research funds have been available through the *National Health Research and Development Program*, jointly funded by Health Canada and Statistics Canada, with up to $300,000 annually for NPHS research. As of April 2000, funding is available through the Canadian Institutes for Health Research (CIHR). For further information, contact *hs-ds@statcan.ca*.

Design of the NPHS

The sample design of the survey varies by region and by type of sample unit. The first component of the survey covers households in the ten provinces. It is based largely on the Labour Force Survey frame, like the National Longitudinal Survey of Children and Youth (NLSCY) and the Survey of Labour and Income Dynamics (SLID). However, for the NPHS household component in Quebec, dwellings were selected from those in the 1992 *Enquête sociale et de santé*. Consequently, the target population is identical to that of these 'parent' surveys: households, excluding those on Canadian Forces bases, on Indian reserves, and in some remote areas. Both parent surveys are based on highly complex, heavily stratified, multistage designs. The general LFS design is a stratified, two-stage sample, with six clusters per stratum and dwellings selected within clusters, except in some rural, remote, and apartment strata. For NPHS purposes, strata were created by regrouping LFS strata, keeping at least some of the LFS clusters, but selecting fresh dwellings from them. For more details on the design of the NPHS, see Tambay and Catlin (1995).

This type of clustered design is ideal for controlling costs when personal interviews are needed, as was the case for cycle 1 of the NPHS. Unfortunately, a complex design also causes complexity in the analysis of survey results. Survey weights are quite variable and variance estimation is much more difficult – a point that will be discussed in detail in the Variance Estimation section, below. Approximately 26,000 dwellings yielded a cycle 1 response file of about 20,000 households. One person in each household was selected at random to be administered the Health questionnaire. These persons would comprise the longitudinal panel to be followed over time.

To cover as much of the Canadian population as possible, separate components of the survey were also carried out in the North and in health care institutions. In the North, a simpler stratified design was used. As well, anticipating the creation of Nunavut, separate strata were formed for each of the future territories, the Northwest Territories and Nunavut.

In the third component of the survey, a two-stage design was used to sample residents of long-term, non-correctional, health-related institutions. A list frame of health care facilities was drawn up from the Residential Care Facilities list, which

includes non-hospital institutions, generally with at least four beds, approved, funded or licensed by provincial or territorial departments of health or social service, and the annual Hospital Survey and Hospital Directory. The list was stratified into five regions, by type of institution (institutions for the aged, psychiatric/developmental institutions, and other rehabilitative institutions), and by size. Within each stratum, a subsample of institutions was drawn systematically with probability proportional to the number of beds, and a systematic sample of residents was selected and interviewed within each sampled institution.

Together, these three components of the NPHS constitute a comprehensive, longitudinal national population health survey. Internationally, in the United States, the National Center for Health Statistics (NCHS), part of the Centers for Disease Control and Prevention, carries out several health surveys, including the National Health and Nutrition Examination Survey (NHANES). This survey uses a combination of home interviews and physical health measures done in mobile examination centres. The NCHS also does follow-up studies of respondents for longitudinal analyses. In Great Britain, the Office for National Statistics conducts the Household Panel Survey, an omnibus longitudinal survey of 5,000 households, with a section covering a number of health topics. Other population health surveys include the Russia Longitudinal Monitoring Survey and the Enquête nationale sur la santé et la protection sociale, carried out by the Centre de Recherche, d'Etude et de Documentation en Economie de la Santé in France.

Switching to a Longitudinal Footing

Data from cycle 1 of the NPHS were released in the winter of 1995/96. Two cross-sectional PUMFs were released, one for the General component of the questionnaire for all household members, and one for the Health component, with detailed information for the single selected panel member in each household. These purely cross-sectional data were successfully and quickly disseminated.

At the same time, preparations were already underway on cycle 2, and it was recognised that much work was needed to prepare for a simultaneous cross-sectional and longitudinal data release.

A Double Priority: Longitudinal and Cross-Sectional Estimates

As mentioned earlier, the capacity to add content to the NPHS is an objective of the survey. Focus content, and provincially funded supplementary questions in Alberta and Manitoba, were added to the questionnaire in cycle 2. Because of the huge size of these cross-sectional supplements – bigger than the original cycle 1 national sample – they were treated like a separate survey, using a different frame and data collection method. Computer-assisted telephone interviewing was used, with Random Digit Dialling (RDD) sampling.

Another addition to the CAI questionnaire, taking advantage of the availability of historical data, was the 'feeding back' of data from the previous cycle to the interviewer. This was done in several sections of the questionnaire, such as Chronic Conditions, in an attempt to catch responses that were inconsistent with previously recorded data.

Processing was more complex, owing to the new, longitudinal aspect of the data. The definition of response, non-response, and out of scope varied according to the estimate: for example, 'out of country' was considered out of scope cross-sectionally, but in scope (and non-response) for longitudinal weighting and estimation. The description of cycle 2 response rates alone rated an entire chapter in the survey documentation, with a number of cross-sectional and longitudinal rates. The introduction of a second cycle of data yielded inconsistent data over time, in spite of the historical editing done during the interview. Except for key variables such as sex and date of birth, these inconsistencies were not eliminated. Due to the lack of adequate cyclical information with which to impute, it was decided that it would be best to retain the original survey data for access by the analysts. Because of the very high response rate for most variables, no wave imputation was done; this decision will be revisited in future cycles. Missing values were, for the most part, left as missing on the data files. Nonresponse was again treated through reweighting.

Weighting required a long lead-time and careful consideration; eventually *seven* separate survey weights were created. Four were cross-sectional, two each for the household and selected member questionnaires. The second weight in each case was required due to the nature of the cross-sectional supplemental samples. These four cross-sectional weights were created using a dual-frame weighting adjustment to combine the two cross sectional frames: the original cycle 1 frame and the cycle 2 telephone frame. (see Stukel, Mohl and Tambay, 1997)

For longitudinal purposes, another three weights were created. The first corresponded to the 15,670 panel members who provided a full response in both cycles. A second weight covered the 16,168 members with at least partial response in each cycle. Finally, a third, unchanging weight covering the entire 17,276 longitudinal panel members, both respondents and nonrespondents, was also computed.

Nonresponse weighting adjustments were done, based on available historical information, by response propensity group, created using the CHAID (Chi-squared Automatic Interaction Detection) algorithm. (See Tambay et al, 1998) An added complication was the phenomenon of extreme weights, caused by the dual-frame design, as well as by changes such as migration and attrition in the panel over time. Outlier programs detected these weights, some of which were adjusted to reduce disclosure risk.

The demand on resources created by the new longitudinal aspect was especially heavy in terms of data management and dissemination. (See Fobes and Geran,

1999) Changes in coding conventions over time, such as for drug codes, required the recoding of historical data. The original variable naming convention from cycle 1, well suited to that cross-sectional release, needed to be reworked to take into account the longitudinal nature of the questions. The variables were renamed and a naming concordance across cycles was created for use by analysts. The initial analysis of the data, contained in the Overview Report published in the Daily along with the data availability notice, involved much more in-depth longitudinal analyses of the data. (See Swain, Catlin and Beaudet, 1999)

Conflicts

Naturally enough, the need for two different processes (including two partly different samples) and two different sets of data products produced some conflicts. In some ways, cycle 2 of a dual-purpose survey entails twice the work. For the NPHS, separate processing teams were not set up: except for the special RDD data collection of the cross-sectional provincial supplements, each step of the survey function was handled by one common team. Although this is an efficient organisation of work, taking advantage of common aspects and group knowledge and synergy, the occasional resource bottleneck occurred.

The very large cross-sectional supplemental samples had a major impact on the production schedule, including the release of the longitudinal data. Data collection pushed the schedule back a few months, and weighting was a major undertaking, especially the integration of the dual frames. The amount of work needed to process and integrate these large additional samples caused a lengthy delay in the release of the longitudinal (and cross-sectional) estimates.

Another conflict arose in the PUMF release strategy. Because of the huge demand for cross-sectional data in cycle 1, as well as the large additional cross-sectional samples in cycle 2, a decision was made to release cross-sectional PUMFs in both cycles. The cycle 2 cross-sectional PUMF was permitted only on the condition that a successful match could not be made between it and its cycle 1 counterpart. It was acknowledged that the release of these two files would make the future release of a longitudinal PUMF unlikely. This is due to the fact that the cycle 2 longitudinal file is largely just a concatenation of cycle 1 and cycle 2 cross-sectional data. Suppressing variables such as geography on a longitudinal PUMF could not be done successfully, since its other variables could be used as a matching key against the previously released cross-sectional PUMFs, which do contain detailed geography. Although no NPHS longitudinal PUMF has been released to date, this has been considered an acceptable tradeoff, given the richness of the cycle 2 cross-sectional data disseminated in the PUMFs, and the alternate venues for analysis of the longitudinal data, such as remote access and the Research Data Centres.

The final conflict is on the content side. Supplemental content for cross-sectional estimates may be purchased that includes sensitive questions. This type of content may cause increased nonresponse, and, more significantly, long-term longitudinal attrition. Although no studies have been done in this area on Statistics Canada's longitudinal surveys, the danger clearly exists. An example of this type of conflict

was the suggested addition of a module on abuse of the elderly in the Institutions component of the NPHS. The difficulties in implementing this module, as well as the potential damage to the Institutions panel, led to its rejection.

Variance Estimation

Standard statistical packages such as SAS and SPSS are not able to deal with complex, clustered designs such as that of the NPHS, often seriously underestimating standard errors. As well, design information cannot be released with the PUMF because of confidentiality concerns, prohibiting users from correctly incorporating this information directly into their variance estimation procedures, whether a Taylor approximation or those output from statistical packages such as SUDAAN or STATA. For this reason, cycle 1 of the NPHS relied for variance estimation on a stand-alone jackknife system, run in-house by the Methodology function, and on the dissemination of approximate coefficient of variance (CV) look-up tables to PUMF users. These CV tables were useful only for giving approximate variance estimates for specific domains and simple estimates, such as totals and proportions. In cycle 2, the original plan was to produce a special "collapsed jackknife" program for PUMF users, which would have allowed them to calculate their own variance estimates for regression models and other complex analyses, as well as more accurate variance estimates for simple statistics, without providing detailed design information.

However, since CV tables would still have been required for some users, this plan would have resulted in an awkward, three-part variance estimation system, difficult to implement and maintain. What was needed was a single system to stand the test of time, properly handling interprovincial movers, different levels of geography, nonsmooth statistics such as medians, and subproducts like CV tables, an integrated system covering longitudinal and cross-sectional estimation, internal and external use. For this reason, and because of problems with each part of the suggested three-part system above, and the likelihood that the original jackknife program would not be able to handle the much-larger survey files in cycle 2, a complete overhaul of the variance estimation system was imperative.

In the end, the NPHS successfully moved to the comprehensive use of bootstrap weights, and was the first Statistics Canada survey to provide these weights directly to analysts for variance estimation. The methodological research, simulation studies, and programming required were another task resulting from the switch-over from a cross-sectional survey to a multipurpose, cross-sectional/longitudinal survey.

In each stratum, a simple random sample, with replacement, of n_h-1 clusters from the n_h sampled clusters was taken. These samples across all strata formed the first bootstrap replicate and were used to create the first set of bootstrap weights obtained by adjusting each unit's original survey weight according to the number of times its cluster was resampled and by expanding by n_h / (n_h-1) to account for the

subsampling. This set of bootstrap weights was then post-stratified to population control totals in the same way as the survey weights. This process was repeated many times to create B replicates and sets of bootstrap weights. For any survey estimate, an analogous bootstrap estimate can be calculated using a set of bootstrap weights. The bootstrap variance estimator is just the sum of squares of the deviation of each bootstrap estimate from the average of all B bootstrap estimates, divided by B.

For the NPHS, simulation studies showed that taking B=500 sets of bootstrap weights was a reasonable compromise between precision of variance estimation and use-ability of the bootstrap weight files. Bootstrap weights were created for every survey weight for every type of file, the only exception being the PUMFs, again

because of confidentiality restrictions. (Several possible adjustments to the bootstrap algorithm are being examined to overcome this problem in future cycles.) These weights are calculated only once and then provided to the researcher to be used for all analyses. In the case of remote access, dummy bootstrap weights are being provided, as mentioned above in Data Products. There is no longer any need to provide analysts with design variables, since the design effect is implicit in the structure of the bootstrap weights. These weights were also used to calculate directly the design effects needed to produce the CV tables, simplifying that process as well. For more details on the implementation of the use of bootstrap weights for NPHS variance estimation, see Yeo, Mantel, and Liu (1999).

Sample Maintenance

Several other new functions were required with the advent of the longitudinal component. To deal with respondents over time, respondent relations and tracing functions were created. The tracing function in the NPHS has been highly successful to date, with only 1.7% of the sample untraced in cycle 2 and a cumulative 3% after cycle 3. Slippage, the gradually increasing separation between weighted sample totals and population projections, needed to be analysed. Deaths discovered in data collection were matched to the mortality database for validation. A match is also planned between the nonresponse/unable-to-trace portion of the sample and this database, to check for deaths and to add cause of death. On the cross-sectional side, a top-up sample was picked in cycle 3 to cover the new entrants to the population (i.e., new immigrants and infants) and to counter attrition.

Planned Changes and Future Work

The long-range plan for the NPHS was always for an increasing concentration on the longitudinal aspects of the survey, such as more complex analyses and wave imputation. A second panel will be needed somewhere around 2003. The issue of sample size must be carefully examined: a much larger sample than the first panel's

20,000 may be needed to deal with demands for more precise provincial and subprovincial estimates, especially given the effects on sample size of cumulative attrition in later cycles. Confidentiality and variance estimation issues will continue to arise over time.

What does the future hold? More surprises. With the institution of a new, solely cross-sectional survey, the Canadian Community Health Survey (CCHS), many of the long-range plans for the NPHS have been set aside. This new survey, which started collection in the fall of 2000, will obviate the need for a cross-sectional component to the NPHS. The 'disentanglement' of these two processes to create two separate surveys, one purely longitudinal and the other cross-sectional, will not be trivial. Supplemental samples and cross-sectional top-up samples will no longer be needed in the NPHS; the CCHS will assume this function. Only the single longitudinal member in each NPHS household will be interviewed, allowing the Health and General questionnaires to be collapsed into one large CAI module, reducing collection and processing costs significantly. The Territories and Institutions components will also need to be integrated, and a decision must be taken on how long these subpanels should continue, given high attrition rates.

Most longitudinal surveys make this transition, from a largely cross-sectional focus to a more heavy concentration on the longitudinal. For the NPHS, this has been a complete transformation: from cross-sectional in cycle 1 to longitudinal only in cycle 4.

Acknowledgements:

I would like to thank Jean-Louis Tambay and François Brisebois for their helpful comments during the writing of this paper.

References:

Fobes, P. and L. Geran. 1999. Cycle 2 and beyond: Preparing and storing longitudinal data of the National Population Health Survey. *Proceedings of Symposium 98: Longitudinal Analysis for Complex Surveys*, Statistics Canada.

Mantel, H. and S. Nadon. 1999. Dummy File Creation for the Remote Access Program of the National Population Health Survey. *Proceedings of the Survey Methods Section*, Statistical Society of Canada, 181B186.

Statistics Canada. 1995. National Population Health Survey, 1994/95: Public use microdata files. Statistics Canada, Catalogue 82F0001.

Statistics Canada. 1998. National Population Health Survey: Cycle 2, 1996/97. *The Daily*, May 29, 1998, Statistics Canada, Catalogue 11-001, 3B5.

Statistics Canada. 1998. National Population Health Survey Overview, 1996/97. Statistics Canada, Catalogue 82-567.

Statistics Canada. 1998. National Population Health Survey: Public-use CD-ROM, 1996/97. Statistics Canada, Catalogue 82M0009XCB.

Statistics Canada. 2000. How Healthy are Canadians? A Special Issue. *Health Reports*, Statistics Canada, Catalogue 82-003-XPB, 11 (3).

Statistics Canada. 2000. National Population Health Survey, household component public-use microdata files, 1998B1999. Statistics Canada, Catalogue 82M0009XCB.

Stukel, D. M., C.A. Mohl and J.-L. Tambay. 1997. Weighting for cycle two of Statistics Canada's National Population Health Survey. *Proceedings of the Survey Methods Section*, Statistical Society of Canada, 111B116.

Swain, L., G. Catlin and M. Beaudet. 1999. The National Population Health Survey: Its Longitudinal Nature. *Health Reports*, Statistics Canada, Catalogue 82-003-XIE, 10 (4): 69B82.

Tambay, J. L. and G. Catlin. 1995. Sample design of the National Population Health Survey. *Health Reports*, Statistics Canada, Catalogue 82-003, 7 (1): 29B38.

Tambay, J-L., I. Schiopu-Kratina, J. Mayda, D. Stukel and S. Nadon. 1998. "Treatment of Nonresponse in Cycle Two of the National Population Health Survey," *Survey Methodology* 24: 147B156.

Yeo, D., H. Mantel and T.P. Liu. 1999. Bootstrap Variance Estimation for the National Population Health Survey, *1999 Proceedings of the Survey Research Methods Section*, American Statistical Association, to appear.