Survey Methodology, December 1998
Vol. 24, No. 2, pp. 147-156
Statistics Canada

Treatment of Nonresponse in Cycle Two of the National Population Health Survey

JEAN-LOUIS TAMBAY, IOANA ŞCHIOPU-KRATINA, JACQUELINE MAYDA, DIANA STUKEL and SYLVAIN NADON[1]

ABSTRACT

The National Population Health Survey (NPHS) is one of Statistics Canada's three major longitudinal household surveys providing an extensive coverage of the Canadian population. A panel of approximately 17,000 people are being followed up every two years for up to twenty years. The survey data are used for longitudinal analyses, although an important objective is the production of cross-sectional estimates. Each cycle panel respondents provide detailed health information (H) while, to augment the cross-sectional sample, general socio-demographic and health information (G) are collected from all members of their households. This particular collection strategy presents several observable response patterns for Panel Members after two cycles: GH-GH, GH-G*, GH-**, G*-GH, G*-G* and G*-**, where "*" denotes a missing portion of data. The article presents the methodology developed to deal with these types of longitudinal nonresponse as well as with nonresponse from a cross-sectional perspective. The use of weight adjustments for nonresponse and the creation of adjustment cells for weighting using a CHAID algorithm are discussed.

KEY WORDS: Longitudinal surveys; Treatment of nonresponse; CHAID algorithms.

1. INTRODUCTION

In 1996-97, Statistics Canada completed data collection for Cycle 2 of the National Population Health Survey (NPHS). This longitudinal survey was launched in 1994 to provide comprehensive information on the health status of the Canadian population and on the determinants of health. The in-scope population covers residents of households and health institutions throughout the country. In the provinces the household questionnaire has two main components which are administered using computer-assisted interviewing. The General component collects socio-demographic and basic health information for each member of the household. The Health component obtains more detailed health information about the household member selected to participate in the longitudinal panel.

Although the NPHS is a longitudinal survey, its objectives also include the production of periodic cross-sectional estimates (Catlin and Will 1992). The data collection methodology reflects both longitudinal and cross-sectional needs. Panel Members, chosen in Cycle 1, are followed-up every two years for up to 20 years. Persons residing with the Panel Members at those times provide General component information for use in cross-sectional estimation. As the cross-sectional coverage of the sample deteriorates over time, the sample needs to be "topped-up" periodically. The first top-up is planned for Cycle 3, in 1998.

This paper presents the methodology developed in Cycle 2 to deal with nonresponse at the household and person levels (flagging will be used for item nonresponse). The methodology is based on reweighting respondents within sub-populations called weighting cells to account for nonresponse. Reweighting is a common approach for the treatment of item nonresponse. The bias and variance of this approach have been considered by Thomsen (1973), Oh and Scheuren (1983), Kalton and Kasprzyk (1986) and Little (1986), among others. If weighting cells are defined such that nonresponse occurs almost completely at random within each cell then the bias due to nonresponse can become negligible. In a similar vein David, Little, Samuhel and Triest (1983) extended to nonresponse the theory developed by Rosenbaum and Rubin (1983) in the context of propensity score matching in observational studies. Their results imply that reweighting can adjust for nonresponse bias when the weighting cells are formed based on the propensity to respond.

An overview of the NPHS sample design and outputs for the first two Cycles is given in Section 2. Section 3 presents the nonresponse treatment strategies and their results. Concluding remarks are given in Section 4. Note that the methodology presented pertains to the provincial household samples; it does not cover the samples in the territories and in institutions.

2. OVERVIEW OF THE NPHS DESIGN AND OUTPUTS

2.1 Cycle 1 Sample Design

The initial sample of households was selected in 1994 using the sample selection vehicle built for the Canadian Labour Force Survey (LFS), and, in the province of Quebec, using dwellings that had participated in a health survey

[1] Jean-Louis Tambay, Ioana Şchiopu-Kratina, Jacqueline Mayda, Diana Stukel and Sylvain Nadon, Household Survey Methods Division, Statistics Canada, 16th floor, R.H. Coats Building, Tunney's Pasture, Ottawa, Ontario, Canada, K1A 0T6.

conducted by Santé Québec the previous year. In both cases the households or dwellings were selected at random within stratified samples of clusters selected using probability proportional to size. The clusters were organized into replicates and collection period to capture seasonality and for variance estimation purposes. There were two "summer" collection periods (June and August) and two "winter" collection periods (November and March, 1995).

Figure 1 illustrates the Panel selection mechanism applied outside the province of Quebec. Sample households were randomly designated as "Adult" or "Children" households, and as eligible for screening or not, prior to collection. Screening increased the presence in the panel of inhabitants of larger households who would be underrepresented with the selection of only one member per household, particularly children and youths. Households eligible for screening were rejected from the sample if they had no member aged under 25. Screening was not used in Quebec as information from the provincial health survey allowed the application of different sub-sampling rates by household type and size.

Sample Unit Type	Household Characteristic	Panel selection restricted to:
"Children" household Eligible for Screening (EFS)	No member under 25	N/A – hhld rejected
	No children, some members under 25	Any member
	Children present	Child members
"Children" household not EFS	No children present	Any member
	Children present	Child members
"Adult" hhld	All	Members over 12

Figure 1. Panel Selection Mechanism Outside Quebec

The classification into "Adult" and "Children" households was done for an operational reason: the Health questionnaire for children, would not be available before the winter collection periods. In "Adult" households, which could be interviewed any time, children under 12 were not eligible for the panel. "Children" households, even those in "summer" clusters, were interviewed in a winter collection period. If children were present in those households then the panel selection was restricted to them. To diminish the seasonal distortions to the data collection workload and the panel representability brought about by these procedures, fewer households were classified as "Children" households in summer clusters, and, with one minor exception, screening was applied only to "Children" households.

Provinces wishing to improve sub-provincial estimates could fund additional sample sizes. In three provinces this was done by augmenting the sample size in targeted regions. In British Columbia an additional sample of about 800 households was selected in a local health unit using Random Digit Dialling (RDD). The expected total sample size in the provinces was approximately 23,000 households after screening.

The above gives a general indication of the 1994 sample design which is sufficient for the needs of this paper. Readers wishing a more precise presentation of the 1994 sample should see Tambay and Catlin (1995), or Statistics Canada (1995).

2.2 Cycle 1 Weighting and Outputs

The major output of the NPHS consists of person-level anonymized Public Use Microdata Files (PUMFs) of survey responses (internal versions of those files that include information suppressed for confidentiality reasons are also created). For 1994 a General PUMF (58,400 records) and a Health PUMF (17,600 records) were released containing the General and Health information collected from every household member and from the selected non-child Panel Members, respectively (Statistics Canada 1995).

The sample weights attached to every record on the PUMF were calculated by applying a series of adjustments to a basic weight representing the household inverse sampling rates (ISR). The ISRs are calculated by multiplying the weights of the original LFS or Santé Québec samples by the inverse of the sub-sampling rates applied by the NPHS. For the sake of brevity we will only describe the main adjustments used outside of Quebec.

Adjustments to the weights for the General PUMF include: (1) a household nonresponse adjustment; (2) an adjustment for the rejective method; (3) an adjustment for person nonresponse [within responding households] and, finally; (4) a simple post-stratification adjustment. Adjustment (2) was applied only to households with no member under 25. It was $1/(1 - r_s)$, where r_s was the sub-sampling rate for the screening applied in the stratum. The post-stratification adjustment was done separately for each province-age group-sex cross-class. Weights resulting from all earlier steps are multiplied by the ratio of known to estimated population sizes within the cross-class. The known population sizes are in fact Census-based projections.

The adjustments for household and person level nonresponse (at 11.3% and 1.4%, respectively) were applied to respondent units as the nonrespondents were excluded from the PUMFs. If w_i is the sample weight of a unit i, the nonresponse-adjusted weight, $w_{\text{adj},i}$ is defined as $w_{\text{adj},i} = w_i (\sum_{\text{all}} w_i)/(\sum_{\text{resp}} w_i)$, where the sums are taken over all sample units and all respondent units, respectively, within nonresponse adjustment weighting cells. Due to a lack of information on nonrespondent households the weighting cells for household level nonresponse were simply cross-classes of NPHS strata and season (*i.e.*, "summer" *vs.* "winter" clusters). Weighting cells for the person level nonresponse, which was very low, were the province-age-sex cross-classes that were used for the post-stratification adjustment.

Adjustments to the weights for the Health PUMF included: (1) a household nonresponse adjustment; (2) an adjustment for the rejective method; (3) an adjustment for the "Adult/Children" household sub-sampling; (4) an adjustment for the longitudinal Panel Member selection; (5) an adjustment for Panel Member nonresponse; and (6) a post-stratification adjustment. The first two adjustments were exactly those for the General PUMF. As the Health PUMF did not include Panel Members who were children, adjustment (3) compensated for those sample households where non-children were ineligible for panel membership. The adjustment thus applied only to households with children and was equal to $1/r$, where r was the proportion of "Adult" households in the sample. Adjustment (4) was the inverse of the probability of having selected the Panel Member. The adjustments for Panel Member nonresponse (at 3.9%) and for post-stratification were similar to those for the General PUMF, and used the same province-age-sex cross-classes. Although child Panel Members were not included in the Health PUMF, for longitudinal purposes their sample weights were obtained as above using $1/(1-r)$ instead of $1/r$ in step (3).

2.3 Cycle 2 Sample Design

In Cycle 2 the focus of the survey was more on longitudinal estimation: no sample "top-up" was planned until the following cycle. The "Core" sample thus consisted of about 17,000 Panel Members and their current households. Panel Members were traced and administered the General and Health questionnaire components, while other members of their household were administered the General component only. No follow-up was done for 1994 nonresponding households. In Alberta, Manitoba and Ontario large (non-Core) additional samples were obtained, using RDD, to allow the production of cross-sectional estimates at sub-provincial levels. In every RDD household one member aged over 12 was selected to complete the Health component. In Alberta and Manitoba, RDD households with children also had a child selected to complete the Health component.

We note that, for cross-sectional purposes, the Core sample does not cover very well arrivals in the population such as newborns and recent immigrants. The population administered the General questionnaire consists of residents of households where at least one member was in-scope in Cycle 1; households made up entirely of recent immigrants (and their newborns) are thus missed. The population administered the Health questionnaire consists of persons who were in-scope in Cycle 1: recent immigrants and children under 2 years old are excluded from the Core target population (they are included in the RDD target population). For both the General and the Health questionnaires post-stratification is done using population figures that do not exclude the recent immigrants. The result is that the population of recent immigrants is implicitly being estimated for by the population of non-immigrants because the latter's Core weights are adjusted upwards to account for the former's numbers. This is a limitation that is acknowledged in the PUMF documentation. Alternative methods would have been to post-stratify using only non-immigrant population projections or to somehow adjust only the weights of less recent immigrants (who are covered) to account for the more recent immigrants (who are not). These methods would have been difficult to apply where, for the General questionnaire, a distinction between immigrants in immigrant-only households and immigrants in mixed households would have been required.

2.4 Cycle 2 Weighting and Outputs

Figure 2 summarizes the survey's three major outputs planned for Cycle 2: a Longitudinal PUMF; a Health Cross-Sectional PUMF and a General Cross-Sectional PUMF. The planned Longitudinal PUMF contains General and Health information for both Cycles for the 17,000 Panel respondents [note: confidentiality requirements may prevent the release of a longitudinal PUMF – in which case only an internal microdata file will be produced]. The Health Cross-Sectional PUMF contains 1996 General and Health information for about 70,000 Panel Members and RDD Selected Members. The General Cross-Sectional PUMF contains 1996 General information for about 220,000 members of the Core and RDD samples. The weighting processes involved for each PUMF, presented below for the Core sample, are described in more detail in Stukel, Mohl and Tambay (1997).

Output File	LONGITUDINAL PUMF	HEALTH CROSS-SECTIONAL PUMF	GENERAL CROSS-SECTIONAL PUMF
Contents	General & Health	General & Health	General only
Samples	Core only	Core & RDD (3 provs.)	Core & RDD (3 provs.)
Units	Panel Member (PM)	PM/RDD Sel. Mem.	All Hhld. Members
Size	≈ 17,000 records	≈ 70,000 records	≈ 220,000 records
Weighting Strategy (for Core Sample)	1.Base Year Weight 2.PM Nonresp. Adjustment 3.Post-stratification	1.Base Year Weight 2.PM Nonresp. Adj. 3.Core/RDD integration 4.Post-stratification	1.Base Year Weight 2.Hhld. Nonresp. Adj. 3.Weight Share Adj. 4.Hhld. Mem. NR Adj. 5.Core/RDD integration 6.Post-stratification

Figure 2. Description of Output Files for Cycle 2

Respondent survey weights on the Longitudinal PUMF are obtained by adjusting a base year weight first for 1996 panel nonresponse and then for post-stratification. The base year weight represents the inverse sampling rate for 1994 including all Health PUMF adjustments described in section 2.2 up to adjustment (4) for panel selection (a correction is needed for the "removal" of the 1994 provincial sample additions). The weight adjustment for

nonresponse is the focus of the following section and will be described there. Post-stratification is done to reproduce 1994 provincial population counts by age-sex categories.

For the Health Cross-Sectional PUMF, the weighting process for (Core) Panel Members involves three or four steps. Usually, the base year weight is adjusted for Panel Member nonresponse, as explained in the following section, and for post-stratification (to match 1996 provincial or regional population counts by age-sex categories). In provinces with RDD samples the extra step is the integration with the RDD sample. The integrated estimate is obtained by a somewhat degenerate adaptation of the Skinner-Rao dual frame estimator (Skinner and Rao 1996).

For the General Cross-Sectional PUMF, the weighting process for the core sample involves five or six steps. First, once more, is the calculation of the base year weight. Then comes an adjustment for nonresponse at the household level, discussed in the following section. The next step is the application of the "weight share method". The method was described by Ernst (1989) and developed further by Lavallée (1995). The Panel Member's weight, divided by the number of persons in his/her household who were in-scope in Cycle 1, is assigned to all household members including those who were not in-scope in Cycle 1 (*e.g.*, births, immigrants). The method is unbiased for estimates of totals for the population of households where at least one member was in-scope in Cycle 1. The next step is a household member nonresponse adjustment. In RDD provinces this is followed by integration of the Core sample with the RDD sample (this time for all ages). Post-stratification is done in a similar fashion to that for the Health Cross-Sectional PUMF.

3. CYCLE 2 CORE SAMPLE NONRESPONSE STRATEGY

This section presents the strategy adopted for the treatment of Cycle 2 nonresponse in the Core (non-RDD) sample. Adjusting for nonresponse was done once again using the weighting cell approach except that, this time, Cycle 1 data were available to create weighting cells that are more homogeneous with respect to the propensity to respond, and thus more apt to remove nonresponse bias. Section 3.1 identifies nonrespondents in the NPHS. Section 3.2 discusses two general approaches for the creation of weighting cells, giving the one chosen for the NPHS. The strategy for the adjustment for nonresponse is explained in section 3.3 while section 3.4 describes the creation of the nonresponse weighting cells.

3.1 Definitions of Nonrespondent and Out-of-scope Units

The first step in the treatment of nonresponse consisted of its definition or identification. In Cycle 2, questionnaires were fully completed for 89% of the Core sample and partially completed for another 3%. The rest of the sample consisted of refusals (3.1%), of cases where the Panel Member could not be traced (1.7%), had died (1.7%), had left Canada (0.5%), or was institutionalized (0.4%), and of other types of nonresponse such as temporary absences and special circumstances (0.7%). Within responding households person level nonresponse was very low: 1.8% for the General questionnaire and 1.1% for the Health questionnaire. We first identify cases that are not nonresponses for longitudinal and cross-sectional purposes.

For longitudinal purposes a death is considered a valid survey response. Panel Members who had died before Cycle 2 had their status recorded as such and the 1996 portion of their data coded as "Not Applicable" on the Longitudinal microdata file. Panel Members who moved to an institution or to the Territories were followed-up and their responses were used for longitudinal purposes. Panel Members who left the country were not followed-up but were treated as longitudinal nonrespondents even though it would have been more accurate for some analyses to have considered them as having left the scope of the study. This treatment was chosen because such persons would fall back in-scope should they move back to Canada.

For cross-sectional purposes all the cases presented in the preceding paragraph were treated as out-of-scope situations. This was acceptable because the separate Institutional and Territorial survey vehicles assumed the cross-sectional coverage of these particular in-scope populations. Out-of-scope units were not on the PUMFs but, as they represented other such units, they were treated for weighting purposes like respondents in all the weight adjustment steps except the integration and post-stratification steps.

Refusals and cases where questionnaires were missing for reasons other than those given in the preceding paragraphs were defined as nonresponses. As will be seen, a distinction was later made between "full" and "partial" longitudinal nonrespondents to accommodate different users.

3.2 Approach for Creating Nonresponse Adjustment Weighting Cells

Two statistical approaches for creating response weighting cells involve segmentation modelling and logistic regression. An example of the latter is given in Czajka, Hirabayashi, Little and Rubin (1992). The authors obtained advance taxation estimates from early tax filer returns using adjustment weighting cells that were based on ranges of propensities to file early. Logistic regression was used to estimate tax filers' propensities to file early. The longitudinal Survey of Labour and Income Dynamics (SLID) provides another example involving logistic regression (Grondin 1996). Sample units' response indicators were regressed on known (dichotomous) characteristics. Adjustment cells for nonresponse were generated by cross-classifying the sample units using all the significant covariates. In order to respect minimum cell sizes and response rates some collapsing was done starting with cells sharing all but the least significant covariates.

In the segmentation modelling approach a decision tree structure is generated from the data by successively splitting the data set such that, at each node, the most significant predictor for the response variable is used to define the following split. The splitting continues until one cannot find any significant variable for the split or minimum cell size requirements cannot be respected. An early application of segmentation modelling for nonresponse adjustment was with respect to the Panel Study of Income Dynamics (Institute for Social Research 1979). Because of its advantages, given below, the NPHS adopted the segmentation modelling approach using the CHAID algorithm developed by Kass (1980). The CHAID (Chi-square Automatic Interaction Detection) algorithm uses χ^2 tests to define splits for categorical predictors and retains the most significant split at each stage. The splitting, into two or more categories, is done differently for ordered and unordered predictors. CHAID was applied using the Knowledge Seeker software program (ANGOSS Software 1995). Note that Knowledge Seeker applies CHAID to continuous predictors by first transforming them into ordered discrete variables.

Advantages and disadvantages of the logistic and CHAID approaches are known and documented (for example see Kalton and Kasprzyk 1986). The logistic regression approach is based on theory familiar to many analysts, and can be programmed using a number of standard statistical packages. It also provides individual estimates of response propensity that can be used directly to adjust the weights of respondents. However, to ensure reasonable program execution times the number of variable and interaction terms used must usually be limited. Collapsing cells can also become complicated, as in the case of SLID above. The CHAID algorithm offers the advantages of accepting a large number of covariates and, by its decision tree structure, easily accommodating interactions among them. Moreover, minimum cell size requirements can easily be incorporated as program execution parameters. Its main disadvantages are a less familiar theoretical underpinning (Knowledge Seeker is advertised as a "data mining" tool) and the limited documentation and software available for its implementation. It should also be mentioned that, while some statistical packages such as SUDAAN and PC CARP can incorporate the sample design when fitting logistic models to survey data, this is not the case with CHAID. The NPHS tried to address this limitation by including as predictor variables characteristics that were related to the sample design (see Section 3.4).

Two empirical studies comparing the logistic and CHAID approaches for the treatment of nonresponse obtained different results. Rizzo, Kalton and Brick (1996) did not find much of a difference between the two approaches for the Survey of Income and Program Participation. On the other hand Dufour, Gagnon, Morin, Renaud and Särndal (1998), in a simulation study for SLID, obtained a lower bias after nonresponse adjustment with the CHAID approach.

3.3 Adjusting for Nonresponse in the Core Sample

Nonresponse adjustments had to be developed for each PUMF: Longitudinal, General (Cross-Sectional) and Health (Cross-Sectional). We will deal with the General PUMF first.

As Figure 2 showed, the weighting strategy for the General PUMF required separate adjustments for nonresponse at the household and at the person levels. In creating adjustment cells for household level nonresponse, characteristics of the Panel Member as well as those of the household were considered as nonresponse predictors. This was done for three reasons. Firstly, as the Panel Member was the link to the household in Cycle 2, his or her characteristics may be related to finding the household and obtaining a response (the first contact will often be through him or her). Secondly, a few personal characteristics of the Panel Member, such as race, are in some sense household characteristics. Finally, using Panel Member characteristics was not incompatible with our need to use a variety of information for the construction of weighting cells. If Panel Member characteristics are not significant, then CHAID simply does not retain them.

Person level nonresponse to the General component occurred when the information was available for some but not all of the household members, perhaps due to members' refusals or temporary absences. Given the low 1.8% nonresponse rate at the person level, it was felt that the creation of weighting cells based on province-age-sex categories (as in Cycle 1) would be sufficient for our needs.

In contrast to the General PUMF, the adjustments for household and person level nonresponse for both the Longitudinal and the Health PUMFs could be combined into a single adjustment as they concerned only one person – the Panel Member. A single set of adjustment cells thus needed to be created.

For the Longitudinal PUMF it was noted that the data items came from both the General and Health components but that response rates for the two components were different. This difference produced data with different Cycle 1-Cycle 2 reporting patterns: GH-GH, GH-G*, G*-GH, G*-G*, not to mention longitudinal nonresponse patterns GH-** and G*-**, where the letters stand for the component reported each Cycle ("*" if not reported). To maximize the utility of the data it was decided to do two adjustments for longitudinal nonresponse. One adjustment would be for the "Full Longitudinal Response" pattern GH-GH. In other words, all other response patterns would be considered as nonresponses. The other adjustment would be for the "Partial Longitudinal Response" pattern which included cases where, at minimum, General information was available for each cycle (patterns GH-GH, GH-G*, G*-GH and G*-G*). The Full Response data set could be used by researchers who would like to analyse a full longitudinal data set covering the entire questionnaire contents. The Partial Response data set could be of use to researchers primarily interested in the types of variables that are on the

General questionnaire. As the counts in Table 1 below show, the Partial Longitudinal Response data set is only about 3% larger than the Full Longitudinal Response data set.

Table 1
Longitudinal Response Patterns

Response Type		Cycle 1-2 Response Pattern	Number of records
Full	Partial		
■	■	GH-GH	15,670
	■	GH-G*	110
	■	G*-GH	366
	■	G*-G*	22
		GH-**	1,014
		G*-**	94
		Total	17,276

Based upon the above, adjustment cells must be built for five types of responses (or nonresponses) in Cycle 2:

- General PUMF – household response
- General PUMF – person response
- Health PUMF – combined response
- Longitudinal PUMF – full response
- Longitudinal PUMF – partial response

Only three sets of adjustment cells were created for those response types. Adjustment cells created for the General PUMF household level responses were also used for the Longitudinal PUMF partial responses because getting a response from a household led almost always to obtaining a partial response for the longitudinal member (there were 53 exceptions). Likewise, adjustment cells generated for full respondents on the Longitudinal PUMF were used for the Health PUMF responses. Although there were 366 more cases of responses of the latter type (pattern G*-GH) it was considered that the same response mechanism was at work in both instances. The third set of adjustment cells was for person level responses on the General PUMF. Province-age-sex categories were used, as was done in Cycle 1.

Note that, although the same adjustment cells would serve for different data sets, the nonresponse weight adjustments would be calculated separately for each data set type. Thus, the 366 records with response pattern G*-GH would be treated as respondents when adjusting weights for the Health PUMF, but as nonrespondents when adjusting weights for full respondents on the Longitudinal PUMF.

3.4 Creation of Weighting Adjustment Cells

Separate sets of weighting adjustment cells were created for each province. The first step consisted of identifying the variables to consider. With CHAID the number of variables that could be considered was not really an issue, and different types were considered. Characteristics of the household, or dwelling, as well as personal characteristics of the Panel Member would of course be considered. In an effort to incorporate the design of the survey into the analysis some characteristics that were related to the design of the survey or to the sampling weight were also considered. These included geographical variables such as the Census Metropolitan Area code or the Urban/Rural indicator, special Cycle 1 design variables such as the flag identifying households for screening and the "Adult/Children" household type, and characteristics related to the application of those design variables, such as the presence in the household of a member aged under 25 or of a child. The household size was used as it was a household characteristic and was also related to the sample weight. From experience, it was also decided to include, in addition to the household income characteristic, a dummy characteristic that identified if household income had been reported in Cycle 1 or not. As a change of address can lead to an unable-to-trace nonresponse situation we would have liked to use a change-of-address identifier. However, in some nonresponse and no contact situations it was difficult to ascertain whether the Panel Member had indeed moved. In the end a "Mover" variable, which identified whether the Panel Member had changed provinces between Cycles, was used in the analysis even though this was far from ideal because the default value would be "no". Personal characteristics from the Health questionnaire component such as Smoker/Drinker status, Health Index Level and Mental Health/Distress Scale were not used because they were not available for almost 500 Panel Members.

The variables used are listed below. The nonresponse indicator, which was the dependent variable, had its values assigned according to the definition of nonresponse being used.

DESIGN/GEOGRAPHICAL VARIABLES

PROVINCE	The analysis was done at the provincial level
CMA	Census Metropolitan Area (0 if not a CMA)
URBAN	Urban/Rural Indicator
REJECT	Flag if the unit (household) was eligible for screening
ACFLAG	"Adult/Children" design classification for the unit

DWELLING/HOUSEHOLD CHARACTERISTICS

DWELL	Dwelling type (10 categories)
OWNER	Owner/Renter Indicator
FAMTYP	Family Type (unattached individual, single parent hhld., married couple hhld., other)
INC	Household Income Adequacy (5 levels)
INCNR	Nonresponse flag for INC
INCSRC	Main source of income (6 categories)
*HHSIZE	Household size
UND25	Indicator of members under 25 years old
KIDS	Indicator of children under 12 years old

PERSONAL CHARACTERISTICS OF PANEL MEMBER

SEX	Sex
AGE	Age in years
AGE16	Indicator if aged 16 or older
MARIT	Marital Status
FAMID	Family Identifier within household (A, B, C, ...)
RACE	White, Black, Aboriginal or Other
BORN	Place of birth (Canada, USA/Mexico, S. America/Africa, Europe/Australia, Asia)
AGIMM	Age at immigration (for immigrants)
*MOVED	Changed province indicator (see text)
*EDUC	Highest level of education (12 categories)
*STUDNT	Student Indicator
MACT	Main Activity (8 categories)
*NUMJOB	Number of jobs held last year (in Cycle 1)
RESTR	Restriction of Activity Flag
*CAUSE	Main Cause of Restriction (12 categories)
CONSUL	Number of consultations with a Medical Doctor
INHOSP	Overnight Hospital Patient Flag
*CHRONIC	Number of Chronic Conditions

* Indicates the variable was never significant when forming classes.

Figure 3 presents the variables chosen by CHAID to build nonresponse adjustment cells for Household Level Response and for the Full Longitudinal Response in each province. For reasons of confidentiality detail is not given on the individual cell sizes and response rates (some of the variables used are considered sensitive and are not on the PUMFs). However, summary information on the cell construction is given in Tables 2 and 3.

Table 2
Response Adjustment Cell Characteristics
(for Household Level Response)

Prov.	#Units	#NR	Cell Sizes			Cell % NR rates		
			min.	max.	avg.	min.	max.	avg.
Nfld.	1,082	40	354	728	541	1.4	4.8	3.7
P.E.I.	1,037	51	81	478	259	3.0	13.6	4.9
N.S.	1,085	55	46	374	217	0.7	10.9	5.1
N.B.	1,125	59	32	986	281	2.6	34.4	5.2
Que.	3,000	133	123	2363	750	1.8	12.1	4.4
Ont.	4,307	315	44	1,038	308	0.9	25.8	7.3
Man.	1,205	50	1,205	1,205	1,205	4.1	4.1	4.1
Sask.	1,168	59	37	626	167	1.6	35.3	5.1
Alta.	1,544	116	32	837	221	3.9	36.7	7.5
B.C.	1,723	149	82	678	246	5.2	29.0	8.6

The results vary by province. As expected, provinces with larger sample sizes such as Ontario, Quebec, British Columbia and Alberta yield "richer" decision trees. Cell sizes and response rates also vary considerably. In Table 2 on household-level response Manitoba has only one cell, and 88% of New Brunswick's sample is located in one cell. Likewise, in Table 3 almost all of Newfoundland's sample is placed in one of its two cells. Cell nonresponse rates approaching 40% are observed in a few provinces.

Table 3
Response Adjustment Cell Characteristics
(for Full Longitudinal Response)

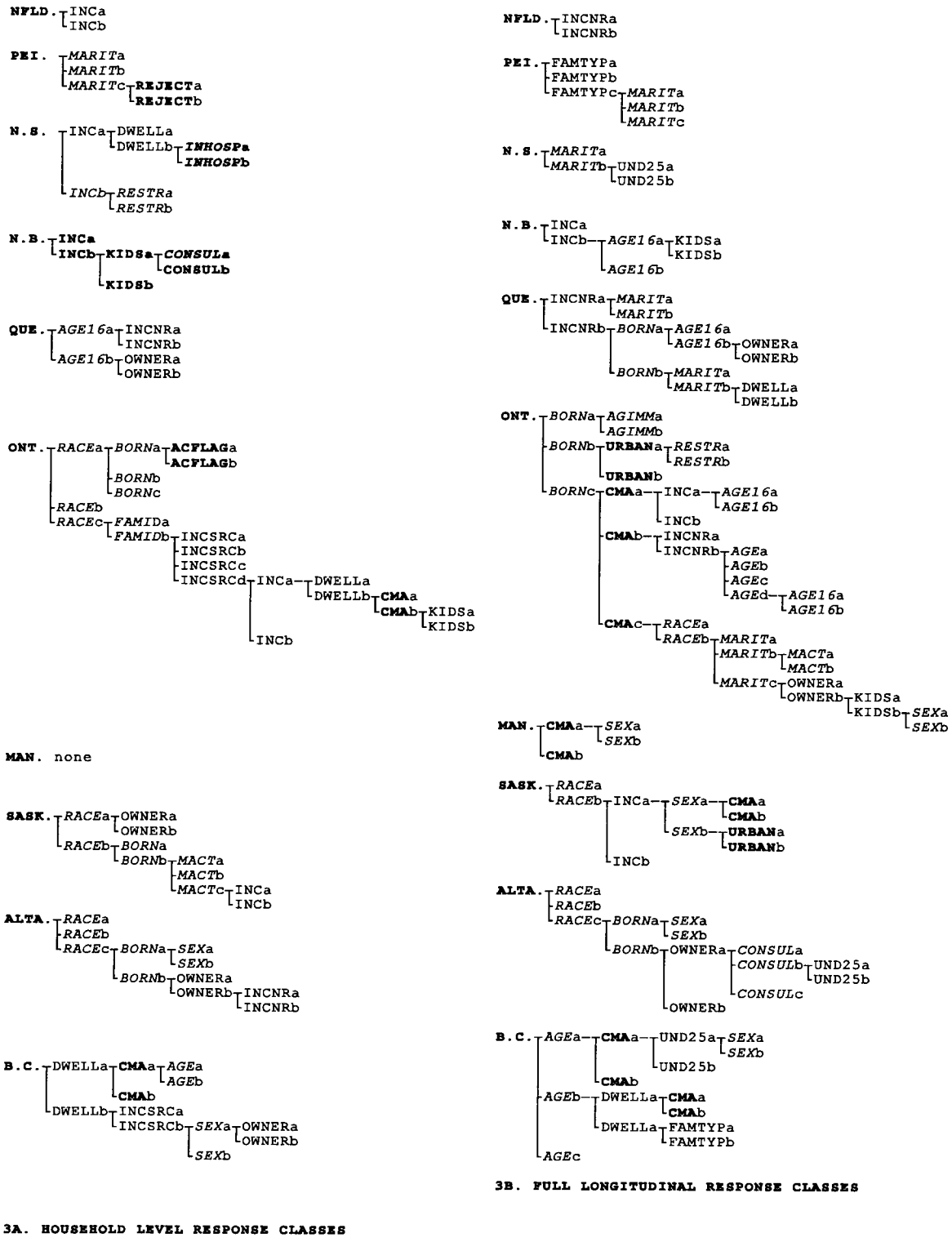
Prov.	#Units	#NR	Cell Sizes			Cell % NR rates		
			min.	max.	avg.	min.	max.	avg.
Nfld.	1,082	73	35	1,047	541	6.2	22.9	6.7
P.E.I.	1,037	80	41	453	207	4.1	26.8	7.7
N.S.	1,085	96	236	555	362	6.5	14.3	8.8
N.B.	1,125	86	59	819	281	4.8	16.8	7.6
Que.	3,000	211	42	2,202	375	2.5	37.8	7.0
Ont.	4,307	470	34	619	196	0.0	38.0	10.9
Man.	1,205	91	186	763	402	5.6	15.1	7.6
Sask.	1,168	83	90	339	195	0.0	28.9	7.1
Alta.	1,544	148	41	866	172	1.1	39.0	9.6
B.C.	1,723	192	33	408	191	4.5	37.3	11.1

Figure 3 shows a variety in the characteristics of weighting classes both between provinces and between the two types of nonresponse within provinces. In all provinces except Alberta the CHAID algorithm uses different characteristics for the two nonresponse types as early as at the first or second level of branching. A few characteristics figure prominently in the early stages of branching in many of the trees for both types of nonresponse. They are: household income adequacy level (INCNR), income nonresponse flag (INCNR), Race (RACE) and Place of Birth (BORN).

In Figure 3a household income and its related variables (INCNR and INCSRC), Owner/Renter status (OWNER), Race, Place of Birth and Dwelling Type (DWELL) all were used three or more times in forming weighting classes for Household Level nonresponse. It is also remarked that in five out of nine provinces a personal characteristic of the Panel Member was selected at the first stage of branching by CHAID. This supports the decision to consider personal characteristics when adjusting for household level nonresponse.

In Figure 3b for Full Longitudinal nonresponse Census Metropolitan Area (CMA), Marital Status (MARIT) and SEX, although not as important at first as Income, Race and Place of Birth, were used the most often (5 times each).

As mentioned earlier, design variables such as the rejection flag (REJECT) and the "Adult/Children" flag (ACFLAG) were considered in an attempt to incorporate the sample design in the CHAID analyses. Although these variables were selected only once each, household characteristics used by the design, such as the presence of children (KIDS) and under 25 year-olds (UND25) did get selected occasionally. Household size was not used but



(TYPES OF CHARACTERISTICS: **DESIGN/GEOGRAPHICAL**, DWELLING/HOUSEHOLD, *PERSONAL*)

Figure 3. Provincial Response Classes Obtained for Cycle 2 Nonresponse

Family Type (FAMTYP), which is related to the household size, did get selected twice.

The adjustment cells produced by CHAID were reviewed but only in rare cases were they manually altered. Within each cell, the weights of responding units were prorated to add up to the total weight for responding and nonresponding units. The magnitude of the nonresponse weight adjustments never exceeded 1.83.

4. CONCLUSION

This paper presented the strategy developed for the treatment of both longitudinal and cross-sectional nonresponse to Cycle 2 of the NPHS. The approach adopted took into account practical considerations such as the need for an easy-to-use, yet statistically valid, way of defining weight adjustment cells and the need to provide a more useful data set (by having separate adjustments for "Full" and "Partial" Longitudinal Responses) while keeping the additional effort required at a reasonable level (*e.g.*, by using weight adjustment cells for more than one purpose). Having chosen the CHAID algorithm approach rather than logistic regression allowed us more freedom in the number and choice of variables to consider: many design variables and personal variables could thus easily be considered – and some were retained. This did seem to offer some promise about the usefulness of those characteristics in the treatment of nonresponse.

On the other hand, a tight production schedule meant that some analysis that we wished to have carried out was not performed. It would have been interesting to pursue the possibilities offered by the CHAID algorithm, for example, as CHAID allows the use of a categorical response variable we could have classified sample units into three groups: live respondents, dead or out-of-scope units, and nonrespondents. We would have liked to do our own comparison of CHAID with a logistic regression approach. We could also have attempted to use Health questionnaire variables such as the Health Index or Smoker/Drinker status in defining weight adjustment cells, although their usefulness would have been reduced by the fact that they were not present for all units (they are missing in response patterns G*-GH, G*-G* and G*-**). Decisions to use the same weight adjustment cells for different types of nonresponse should be revisited. For example, could the adjustment cells built for household level response have been more suitable for the Health cross-sectional nonresponse? An attempt to compare the efficiency of various nonresponse adjustment strategies would involve evaluating their impact on the variance of estimators. We could also evaluate the impact of our Cycle 2 nonresponse adjustment on the nonresponse bias by using the Cycle 1 data available for all panel members. Estimates using the full sample would be compared to nonresponse-adjusted estimates generated from the responding units.

Cycle 3 itself will present new problems. A global sample "top-up" is planned in that year, which will have an impact on our cross-sectional estimation strategy and therefore on the treatment of nonresponse. As longitudinal nonresponse is increasing we will have to consider side effects of the weighting adjustment such as the possible creation of outlier weights. Providing sets of weights for different types of longitudinal analyses will become cumbersome as the number of "partial" response patterns will increase. How many patterns can reasonably be treated, and which ones? The choice of additional information, such as Mover status, for the treatment of nonresponse should be reconsidered. Some imputation for nonresponse will likely be used in Cycle 3: the question is how to reconcile imputation with the weight adjustment approach to nonresponse. As can be seen, a lot of work remains to be done for the NPHS. One hopes that we will have time to investigate many of those issues before Cycle 3 processing is finished.

ACKNOWLEDGEMENTS

The authors would like to thank Douglas Yeo and the referees for their helpful comments.

REFERENCES

ANGOSS SOFTWARE (1995). *Knowledge SEEKER IV for Windows – User's Guide.* ANGOSS Software International Limited.

CATLIN, G., and WILL, P. (1992). The National Population Health Survey: Highlights of initial developments. *Health Reports*, Statistics Canada, 4, 313-319.

CZAJKA, J.L., HIRABAYASHI, S.M., LITTLE, R.J.A., and RUBIN, B.R. (1992). Projecting from advance data using propensity modelling: an application to income and tax statistics. *Journal of Business & Economic Statistics*, 10, 117-131.

DAVID, M.H., LITTLE, R., SAMUHEL, M., and TRIEST, R. (1983). Imputation models based on the propensity to respond. *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, 168-173.

DUFOUR, J., GAGNON, F., MORIN, Y., RENAUD, M., and SÄRNDAL, C.-E. (1998). Measuring the impact of alternative weighting schemes for longitudinal data. *Proceedings of the Section on Survey Research Methods, American Statistical Association.* To appear.

ERNST, L. (1989). Weighting issues for household and family estimates. In *Panel Surveys*, (Eds. D. Kasprzyk, G. Duncan, G. Kalton and M.P. Singh). New York: John Wiley and Sons, 135-159.

GRONDIN, C. (1996). Pondération longitudinale – Première vague de l'EDTR. Internal note, Social Survey Methods Division, Statistics Canada.

INSTITUTE FOR SOCIAL RESEARCH (1979). A Panel Study of Income Dynamics: Procedures and Tape Codes – 1978 Interviewing Year – Wave XI – A Supplement. The University of Michigan.

KALTON, G., and KASPRZYK, D. (1986). The treatment of missing survey data. *Survey Methodology*, 12, 1-16.

KASS, G.V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics*, 29, 119-127.

LAVALLÉE, P. (1995). Cross-sectional weighting of longitudinal surveys of individuals and households using the weight share method. *Survey Methodology*, 21, 25-32.

LITTLE, R.J.A. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review*, 54, 137-157.

OH, H.L., and SCHEUREN, F. (1983). Weighting adjustments for unit nonresponse. In *Incomplete Data in Sample Surveys, Volume 2: Theory and Bibliographies*, (Eds. W.G. Madow, I. Olkin and D. Rubin). New York: Academic Press, 143-184.

RIZZO, L., KALTON, G., and BRICK, J.M. (1996). A comparison of some weighting adjustment methods for panel nonresponse. *Survey Methodology*, 22, 43-53.

ROSENBAUM, P.R., and RUBIN, D.B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70, 41-55.

SKINNER, C.J., and RAO, J.N.K. (1996). Estimation in dual frame surveys with complex designs. *Journal of the American Statistical Association*, 91, 433, 349-356.

STATISTICS CANADA (1995). National Population Health Survey 1994-95 public use microdata file. Catalogue No. 82F0001XCB.

STUKEL, D.M., MOHL, C.A., and TAMBAY, J.L. (1997). Weighting for cycle two of Statistics Canada's National Population Health Survey. *Proceedings of the Survey Methods Section, Statistical Society of Canada*, 111-116.

TAMBAY, J.L., and CATLIN, G. (1995). Sample design of the National Population Health Survey. *Health Reports*, Statistics Canada, 7, 29-38.

THOMSEN, I. (1973). A note on the efficiency of weighting subclass means to reduce the effects of nonresponse when analyzing survey data. *Statistik Tidskrift*, 11, 278-283.