Methodology Branch
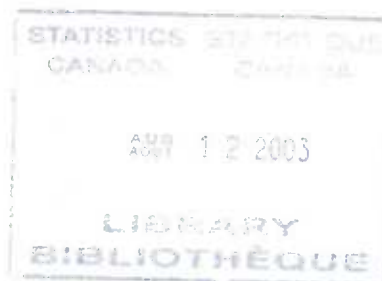
Direction de la méthodologie

Household Survey
Methods Division

Division des méthodes
d'enquêtes des ménages

Canadä

WORKING PAPER
METHODOLOGY BRANCH

NPHS Data Quality: Exploring Non-sampling Errors

HSMD - 2003 - 004E

Sandra Tolusso and François Brisebois

Household Survey Methods Division
Statistics Canada

July 2003

The work presented in this paper is the responsibility of the authors and does not necessarily represent the views or policies of Statistics Canada.

NPHS Data Quality: Exploring Non-sampling Errors

Sandra Tolusso and François Brisebois

ABSTRACT

The National Population Health Survey (NPHS) is a biennial longitudinal survey that collects information on the health of the Canadian population and related socio-demographic information. NPHS also fulfilled cross-sectional needs during its first three cycles, but became purely longitudinal with its recently completed fourth cycle. This paper presents some data quality indicators such as the number of attempted contacts needed to complete a case, the length of interviews, refusal conversion, item refusal and item don't know rates, and edit failure and data inconsistencies, focussing mainly on the longitudinal panel members.

La qualité des données de l'ENSP: une exploration des erreurs non dues à l'échantillonnage

Sandra Tolusso et François Brisebois

RÉSUMÉ

L'Enquête nationale sur la santé de la population (ENSP) est une enquête biennale qui recueille des renseignements sur la santé de la population canadienne ainsi que des renseignements socio-démographiques connexes. L'ENSP a recueilli des renseignements de nature transversale durant les trois premiers cycles mais est devenue uniquement longitudinale avec le quatrième cycle qui a récemment été complété. Ce document présente des indicateurs de qualité comme le nombre d'essais nécessaires pour contacter et compléter un cas, la durée des entrevues, la conversion des refus, les taux de refus et de ne sait pas par item, les taux d'échec au contrôle et l'incohérence des données, en mettant l'accent sur les membres du panel longitudinal.

TABLE OF CONTENTS

1. INTRODUCTION

1.1 Background of the Survey

The National Population Health Survey (NPHS) is designed to collect information on the health of the Canadian population and related socio-demographic information. The first cycle of data collection took place in 1994-1995, and continues every second year thereafter for 20 years. The NPHS fulfilled both cross-sectional and longitudinal needs during its first three cycles, but with Cycle 4 (2000-2001), NPHS became a strictly longitudinal survey. The NPHS is now composed of two components: the survey of households and the survey of health care institutions. This document focuses on the data from the Household component.

The Household component of NPHS includes household residents in all provinces, with the principal exclusion of populations on Indian Reserves and Crown Lands, residents of health care institutions, full-time members of the Canadian Armed Forces, and some remote areas in Ontario and Quebec. The Health Care Institutions component includes long-term residents in health care facilities with four or more beds in all provinces (with an expected stay of longer than six months) with the principal exclusion of the territories and full-time members of the Canadian Armed Forces. If a longitudinal respondent from the Household component moves into a health care institution, the Institutions component will interview them. The Household component of NPHS has released data from the first four cycles: Cycle 1 (1994-1995), Cycle 2 (1996-1997), Cycle 3 (1998-1999) and Cycle 4 (2000-2001).

The Cycle 4 NPHS Household component collected in-depth information on the health of the 17,276 longitudinal respondents (or panel members) who were randomly selected in Cycle 1, as well as demographic information about all members of the longitudinal respondent's household. The health questionnaire includes questions related to health status, use of health services, determinants of health, chronic conditions and activity restrictions. Socio-demographic information is also collected; it includes age, sex, education, ethnicity, labour force status and household income. An entry and demographic questionnaire precedes the health questionnaire, where confirmation of the panel member's residence in that dwelling or the collection of information regarding where they may now reside takes place, as well as the collection of demographic information about all of the household members. An exit questionnaire is administered at the end of the interview, where outcome codes for the interview are assigned, reasons for refusals or non-interviews collected, and contact names recorded for tracing purposes. Data for children under age 12 are collected using a proxy respondent.

NPHS interviews are conducted using Computer Assisted Interviews (CAI). The majority of Cycle 1 interviews were conducted in person, with most panel members choosing to be interviewed by telephone in subsequent cycles. Collection for NPHS takes place every two years, starting in 1994 with Cycle 1, and the task is performed by five Regional Offices (ROs). NPHS has four quarters of collection each cycle, plus a clean-up quarter referred to as Quarter 5. Collection is done over a year, starting in June and continuing in August, November, and February, with Quarter 5 in June. Quarter 5 is used to resend whatever cases were not fully or partially complete during the first four quarters of collection. Nonrespondents in Quarter 1 may also be resent in Quarter 3, and Quarter 2 nonrespondents may be resent in Quarter 4. Cases that refuse are transferred from interviewers to senior interviewers, and to project managers, if necessary.

For more information, please refer to the Cycle 4 Longitudinal Documentation. Questionnaires for all four cycles can be found on Statistics Canada's website at http://www.statcan.ca/english/concepts/nphs/.

1.2 Non-sampling Errors

The survey produces estimates based on information collected from a sample of individuals. Somewhat different estimates might have been obtained if a complete census had been taken using the same questionnaire, interviewers, supervisors, processing methods, etc. as those used in the survey. The difference between the estimates obtained from the sample and those resulting from a complete count taken under similar conditions is called the *sampling error* of the estimate.

Errors that are not related to sampling may occur at almost every phase of a survey operation. Interviewers may misunderstand instructions, respondents may make errors in answering questions, the answers may be incorrectly entered and errors may be introduced in the processing and tabulation of the data. These are all examples of *non-sampling errors*. This document deals strictly with non-sampling errors.

Over a large number of observations, randomly occurring errors will have little effect on estimates derived from the survey. However, errors occurring systematically will contribute to biases in the survey estimates. Considerable time and effort was made to reduce non-sampling errors in the NPHS. Quality assurance measures were implemented at each step of data collection and processing to monitor the quality of the data. These measures included the use of highly skilled interviewers, extensive training with respect to the survey procedures and questionnaire, and the observation of interviewers to detect problems. Testing of the CAI application and field tests were also essential procedures to ensure that data collection errors were minimized.

1.3 Scope of this Document

While reporting measures of data quality is part of our mandate, this document does not cover every aspect of data quality. The Statistics Canada Quality Guidelines (1998) divide data quality studies into two kinds of studies. The first involves *certification* or *validation*, which is the process of reviewing the data prior to the official release to ensure that grossly erroneous data are not released, or to identify data of marginal quality. *Source of error* studies generally provide quantitative information on specific sources of error in the data. This document falls into the second group of studies.

Sections are reported in the order that they occur during the course of a cycle. Before any health information can be obtained, the panel member must be contacted. The first section presented in this document deals with the tracing of panel members, followed by a section on the number of attempted contacts and interview times for panel members. The next section deals with refusals and refusal conversions, which occur once the panel member is contacted. Once panel members (or an acceptable substitute when a proxy interview is necessary) agree to be interviewed, they provide data. Lack of data, in the form of a refusal to answer a particular question or the respondent's inability to recall the information, is presented in the following section. Finally, once collection has ended, edits are performed and data inconsistencies analyzed. These two actions are the subjects of the penultimate section, while the final section mentions forthcoming data quality studies.

Other aspects of data quality such as response rates and attrition rates can be found in the Cycle 4 Longitudinal Documentation.

2. TRACING

Between cycles, panel members may change their phone numbers or move to new residences, cities or provinces. Since this is a longitudinal survey and the goal is to follow the panel members over time, the interviewers must try to find the panel member if they are no longer at the same phone number or residence as when last interviewed. Interviewers are provided with the last known address and telephone number, as well as the name and address of one or two contacts collected in previous cycles. Interviewers are also trained to use resources such as local telephone directories and directory assistance. If unsuccessful, the case is transferred to an experienced interviewer trained specifically in tracing respondents and with access to Canada-wide directories and reverse directories.

NPHS has been incredibly successful at tracing panel members over time. Table 2.1 shows the tracing status of all 17,276 panel members for all four cycles. 96% (16,555) of all records were traced in all four cycles. Of the 296 records that first became untraced in Cycle 2, 35% (103) were traced in the next cycle, 20% (60) were traced two cycles later, and 45% (133) were untraced for three cycles in a row. Of the 169 who first became untraced in Cycle 3, 43% (73) were traced in Cycle 4.

Table 2.1: Distribution of the Tracing Patterns for All 17,276 Panel Members

Cycle 1	Cycle 2	Cycle 3	Cycle 4	Number of panel members
Traced	Untraced	Untraced	Untraced	133
Traced	Untraced	Untraced	Traced	60
Traced	Untraced	Traced	Untraced	19
Traced	Untraced	Traced	Traced	84
Traced	Traced	Untraced	Untraced	96
Traced	Traced	Untraced	Traced	73
Traced	Traced	Traced	Untraced	256
Traced	Traced	Traced	Traced	16,555

Table 2.2 shows the distribution of the panel members by sex and age group (based on their age in 1994) according to whether they were traced in all four cycles or not. The age groups 12–24 and 25–44 are much more likely than other age groups to be untraced at some point. For these two age groups, males in particular make up a large part of those who were untraced.

Table 2.2: Age/sex Distribution of the Panel Members Who Were Traced and Untraced

Age/sex	Traced in all four cycles	Not traced at some point	Proportion untraced at some point
Males 0-11	993	34	3.3%
Males 12-24	1,238	129	9.4%
Males 25-44	2,552	177	6.5%
Males 45-64	1,794	54	2.9%
Males 65+	1,069	5	0.5%
Females 0-11	952	43	4.3%
Females 12-24	1,338	94	6.6%
Females 25-44	2,958	141	4.5%
Females 45-64	2,007	32	1.6%
Females 65+	1,654	12	0.7%

3. NUMBER OF ATTEMPTED CONTACTS AND INTERVIEW LENGTHS IN CYCLE 4

In Cycle 4, cases were managed by the Case Management System, which ensures efficient, reliable and secure movement of data between headquarters in Ottawa and the ROs, as well as between ROs and the interviewers. Case Event files are produced by the Case Management System, consisting of a record for each time that a case was worked on, whether tracing the panel member, attempting unsuccessfully to contact the panel member, or administering the questionnaire. Therefore, the number of contacts per panel member required to complete a case can be computed using the Case Event files. Unfortunately, the Case Management System used for collection in Cycle 4 is different than that used in previous cycles, and information on the number of contacts could only be obtained for Cycle 4.

The H6 file is an intermediate file used in processing that contains health questionnaire data. In Cycle 4, it also contained module start and end times. This file was used to compute the amount of time that it took to complete a module.

3.1 Number of Attempted Contacts in Cycle 4

The number of attempted contacts was computed using the Case Event files from all four quarters of Cycle 4, as well as the clean-up quarter. The Case Event file identifies the date, start time and end time of each contact, and includes case event status variables and the outcome code assigned by the interviewer to each contact.

Not all attempts were used to compute the number of attempted contacts for a panel member. Tracing is considered a separate phase and is thus excluded; including tracing attempts would inflate the actual number of attempted contacts with a panel member needed to obtain a full response. A tracing attempt means that the panel member has not yet been located, and therefore a response to that attempted contact by the panel member is not possible.

Of the 17,276 panel members, 16,051 were used to study the number of attempted contacts. The remaining 1,225 panel members were either institutionalized, confirmed as deceased (through a match to the mortality files) or not sent out due to a previous adamant refusal.

Figure 3.1 – Cumulative Percentage of the Number of Attempted Contacts to Fully Complete an Interview in Cycle 4
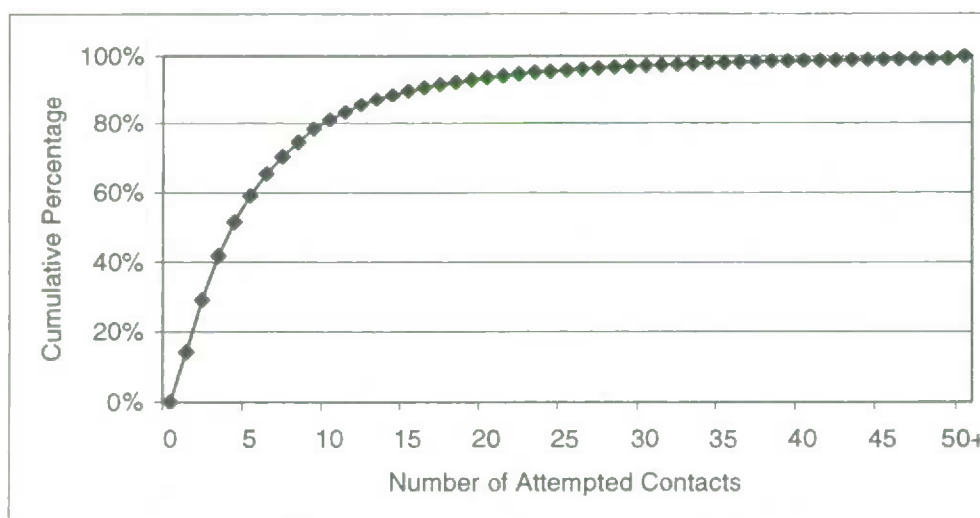
Figure 3.1 shows the cumulative percentage of the number of attempted contacts required to fully complete an interview. About 90% of the respondent interviews were completed in 15 attempted contacts or less. Table 3.1 displays the number of attempted contacts by response type at the RO and national levels. The number of attempted contacts for the respondents is based on the 13,968 panel members who fully completed the interview. This number differs from the 14,650 respondents on the Cycle 4 Master file since panel members who have been confirmed dead or who resided in an institution in Cycle 3 and still reside in an institution are excluded from the Case Event files. Those interviewed by the Institutions component are not on the Case Event files because the questionnaire for the Institutions component is completed on paper. Panel members who are not respondents have been divided into four groups: partial respondents, no one at home, refusals, and other nonresponse.

Table 3.1 - Number of Attempted Contacts by Response Type in Cycle 4

Type	RO	Panel Members	Mean Number of Attempted Contacts	Median Number of Attempted Contacts	25th Percentile	75th Percentile
Fully Complete or Deceased	Canada	13,968	7.2	4	2	9
	Halifax	3,583	7.6	5	2	9
	Montreal	2,483	6.5	4	2	8
	Toronto	3,258	8.0	4	2	9
	Edmonton	3,305	6.6	4	2	8
	Vancouver	1,339	7.3	5	2	9
Partial	Canada	81	11.8	9	5	14
	Halifax	20	11.4	10	4	16
	Montreal	27	8.3	8	4	13
	Toronto	14	20.4	11.5	5	32
	Edmonton	14	11.7	9.5	8	14
	Vancouver	6	9.0	9.5	6	13
No One Home	Canada	131	51.9	39	27	62
	Halifax	17	53.8	52	38	63
	Montreal	19	44.8	40	26	62
	Toronto	50	70.8	52	31	78
	Edmonton	22	32.1	27.5	17	44
	Vancouver	23	34.0	31	16	48
Refusal	Canada	1,015	18.8	14	8	23
	Halifax	261	18.9	13	7	26
	Montreal	170	15.7	13	7	21
	Toronto	284	22.2	16	9	25
	Edmonton	202	16.3	12	7	22
	Vancouver	98	19.2	15	8	24
Other	Canada	856	29.4	23	11	38
	Halifax	109	30.6	25	9	49
	Montreal	135	21.7	20	12	29
	Toronto	283	44.0	34	20	57
	Edmonton	182	20.6	17	10	27
	Vancouver	147	18.2	16	10	24
Total	Canada	16,051	9.5	5	2	11
	Halifax	3,990	9.2	5	2	10
	Montreal	2,834	8.1	5	2	10
	Toronto	3,889	12.5	5	3	13
	Edmonton	3,725	8.0	5	2	10
	Vancouver	1,613	9.4	6	3	12

At the national level, the median number of attempted contacts for respondents was four. The average, median, 25th and 75th percentiles of each RO are similar to those observed at the national level. The number of attempted contacts differs for the four types of nonresponse. The highest number of attempted contacts for all four types are in the Toronto RO, while the lowest can be found in Montreal or Edmonton. Of the nonrespondent groups, the highest number of attempted contacts was for panel members where no one was home, as expected, with a median of 39 contacts at the national level. This is followed by the group "other nonresponse", the refusal group, and partial responses, with medians of 23, 14 and 9 attempted contacts, respectively.

Table 3.2 shows the number of attempted contacts by type of interview, in person or by telephone, using the "type of interview" variable AM60_TEL from the Cycle 4 Master file. They are based on only 13,404 of the 13,968 respondents on the Case Event files, as 564 respondents had no interview type stated, or in a few cases, were interviewed using both methods. Personal visits were made if the respondent did not have a telephone, if the interviewer made a personal visit in the course of tracing a respondent, or upon request by the respondent. The proportion of personal interviews varies from 1.1% in the Toronto RO to 3.2% in the Halifax RO. The median number of attempted contacts at the national level is five for personal interviews, and four for telephone interviews. The median number of attempted contacts for telephone interviews is also four at the RO level, while the medians for personal interviews vary between 3 and 7. The mean and median number of attempted personal contacts is higher than for telephone interviews, with the sole exception being the median in the Montreal RO.

Table 3.2 – Number of Attempted Contacts by Type of Interview for Respondent Panel Members in Cycle 4

Type	RO	Panel Members	Mean Number of Attempted Contacts	Median Number of Attempted Contacts	25th Percentile	75th Percentile
Telephone	Canada	13,121	7.1	4	2	8
	Halifax	3,354	7.5	4	2	9
	Montreal	2,318	6.5	4	2	8
	Toronto	3,091	7.8	4	2	9
	Edmonton	3,119	6.5	4	2	8
	Vancouver	1,239	7.3	4	2	9
Personal	Canada	283	10.9	5	2	13
	Halifax	112	10.7	7	3	14
	Montreal	66	7.4	3	2	8
	Toronto	34	17.1	6.5	3	19
	Edmonton	44	12.5	7	3	14.5
	Vancouver	27	9.5	5	2	12

Table 3.3 shows the number of attempted contacts for all respondent panel members by sex and the five age groups (under 12, 12-24, 25-44, 45-64, 65 and older), using the panel member's age in Cycle 4. Only the groups 12-24 and 25-44 have average and median attempted contacts above those of the national level. This result is consistent with the findings of many other studies indicating that younger adults tend to be harder to contact and with the tracing results presented in section 2. The statistics for both sexes are very similar, with males having a slightly higher number of attempted contacts (with a median of 5 and an average of 7.5) than women (with a median of 4 and an average of 7.0). Males 25-44 have the highest number of contacts, with more than double the number of average contacts required for those of age 65 and older. For the under 12 and 12-24 age groups, males require, on average, a lower number of contacts than females, while the reverse is true for the three oldest age groups.

Table 3.3 - Number of Attempted Contacts by Age Group and Sex for Respondent Panel Members in Cycle 4

Age Group	Panel Members	Mean Number of Attempted Contacts	Median Number of Attempted Contacts	25th Percentile	75th Percentile
Total	13,968	7.2	4	2	9
< 12	867	5.6	4	2	7
12-24	2,060	8.2	5	3	9
25-44	4,548	9.1	6	3	11
45-64	3,700	6.9	4	2	8
65+	2,793	4.5	3	1	5
Total Males	6,388	7.5	5	2	9
< 12	437	5.4	3	2	6
12-24	1,025	8.0	5	3	9
25-44	2,063	9.4	6	3	12
45-64	1,772	7.3	4	2	9
65+	1,091	4.7	3	1	5
Total Females	7,580	7.0	4	2	8
< 12	430	5.8	4	2	7
12-24	1,035	8.3	5	2	9
25-44	2,485	8.8	5	3	10
45-64	1,928	6.7	4	2	8
65+	1,702	4.3	3	1	5

Figure 3.2 Cumulative Percentage of the Number of Attempted Contacts to Complete an Interview by Age Group in Cycle 4

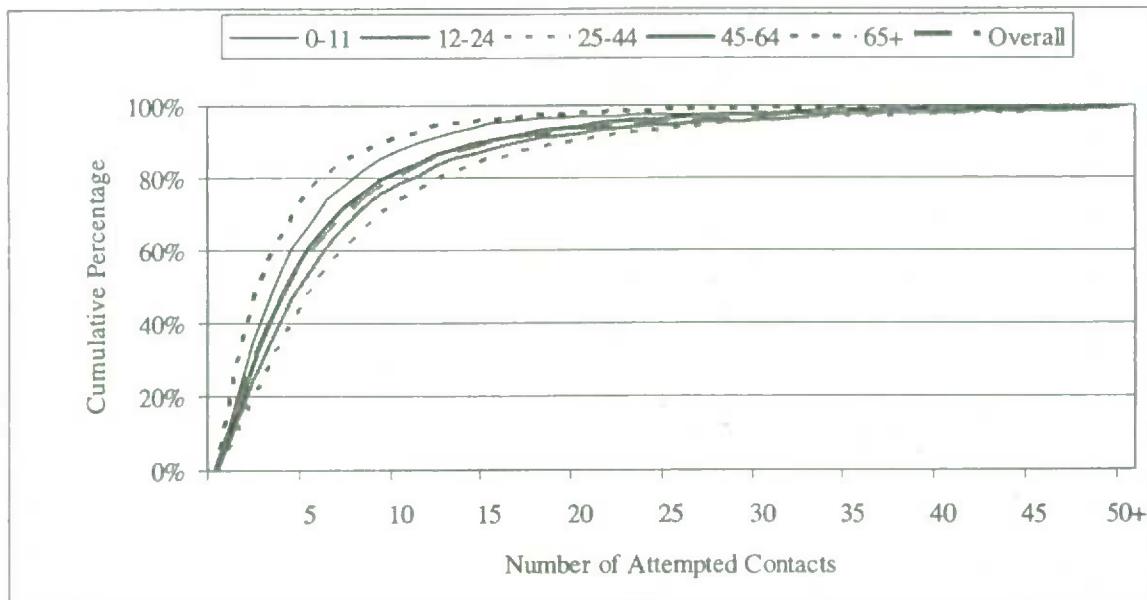Figure 3.2 shows the cumulative percentage of the number of attempted contacts required to fully complete an interview by age group, as well as overall (as seen in Figure 3.1). Like Table 3.3, it shows that the different age groups require a different number of attempted contacts to complete an interview. For example, after 10 attempted contacts, 92% of the interviews with respondents aged 65 and older were

completed, while only 74% of the interviews for the 25-44 age group were completed. Imposing a limit on the number of contacts would cause some age groups to have lower response rates than others.

3.2 Length of Interviews

The interview lengths presented in this section were tabulated using the health questionnaire (H6) only. The entry, demographic, and exit components of the questionnaire are not included in these interview lengths. These components take an average of another four to five minutes to complete. Interviews that were interrupted and consequently completed in more than one session were excluded to ensure that the start and end times of each module were an accurate reflection of how long it took to complete the health questionnaire. 10,086 of the 13,560 fully completed Cycle 4 interviews were identified as having been completed in one session, and were used to calculate the interview lengths.

The total interview lengths obtained from the H6 file indicate that almost 90% of fully completed interviews were completed in 45 minutes or less, and 98% were completed in one hour or less. Figure 3.2 shows the cumulative percentage of the total interview lengths for fully completed interviews, while Table 3.4 displays the median, 25th and 75th percentile total interview lengths by RO.

Table 3.4 – Cycle 4 Health Questionnaire Interview Times by RO

RO	Panel Members	Median Interview Length	25th Percentile	75th Percentile
Canada	10,086	30:53	23:13	38:15
Halifax	2,703	31:19	25:01	37:47
Montreal	1,730	29:31	20:48	37:11
Toronto	2,257	31:05	23:25	39:02
Edmonton	2,395	30:52	21:13	38:53
Vancouver	1,001	31:13	25:01	38:15

Figure 3.3 – Cumulative Percentage of the Lengths of the Health Questionnaire for Fully Completed Interviews in Cycle 4
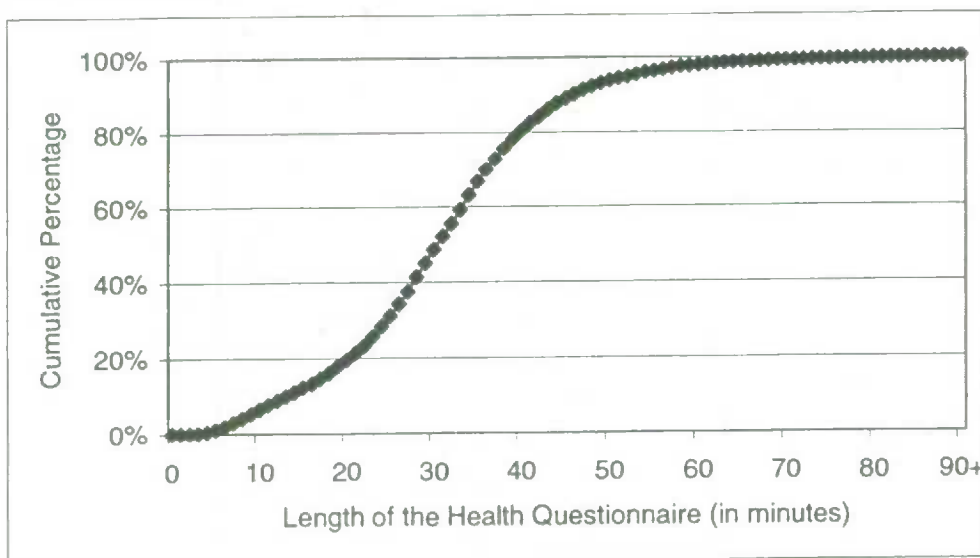

9

Table 3.5 – Cycle 4 Interview Lengths by Module

Module	Maximum Number of Questions	Median Length	25th Percentile	75th Percentile
Tanning and UV Exposure	1	0:11	0:06	0:16
Education	7	0:11	0:07	0:18
Repetitive Strain	3	0:15	0:08	0:23
Injuries	15	0:16	0:09	0:27
Height/Weight	2	0:17	0:13	0:23
Insurance	4	0:25	0:18	0:33
Preventive Health	13	0:29	0:13	0:47
General Health	2	0:32	0:21	0:43
Smoking	23	0:32	0:21	0:50
Socio-Demographic	4	0:33	0:24	0:47
Restriction of Activities	15	0:52	0:31	1:23
Administration	12	0:53	0:34	1:18
Alcohol	15	0:56	0:30	1:22
Income	30	1:15	0:50	1:45
Mental Health	45	1:22	0:52	1:57
Health Status	31	1:23	0:57	1:50
Chronic Conditions	104	1:30	0:55	2:30
Health Care Utilization	28	1:53	1:25	2:31
Drug Use	32	2:04	1:25	3:11
Social Support	20	2:11	1:32	2:48
Labour Force	30	2:12	1:07	3:06
Physical Activities	32	2:58	1:52	4:06
Stress	62	5:49	4:24	7:03

The median interview length at the national level is 30:53 (30 minutes, 53 seconds). All of the median lengths at the RO level are within 30 seconds of the national median, with the exception of the Montreal RO, which is 90 seconds shorter. Interview lengths are very similar for the ROs, although those from the Montreal RO are consistently shorter than those for the other ROs. Table 3.5 displays the interview lengths by module, in ascending order according to the median. Figure 3.3 shows the distribution of the interview lengths, for the entire H6 and by module. The stress module is by far the longest module with a median length of 5:49. Most modules have median interview lengths of under one minute.

Figure 3.4 – Cycle 4 Interview Lengths by Module

Total Health Component Interview Lengths
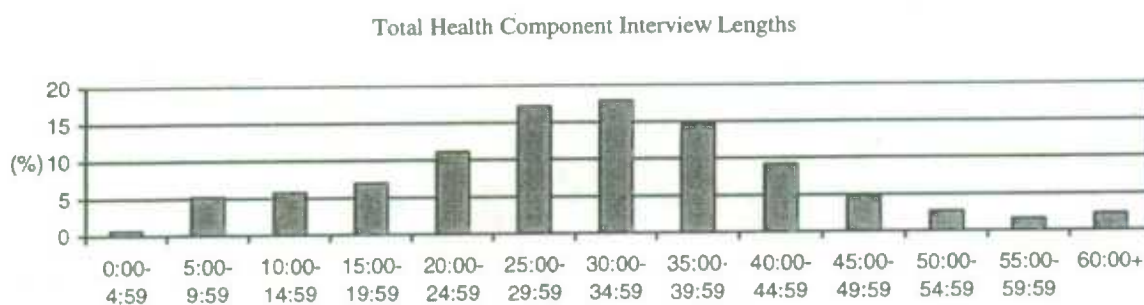
Tanning and UV Exposure
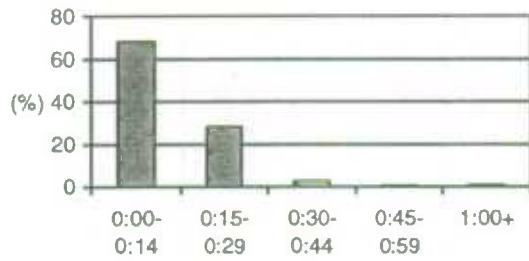
Education
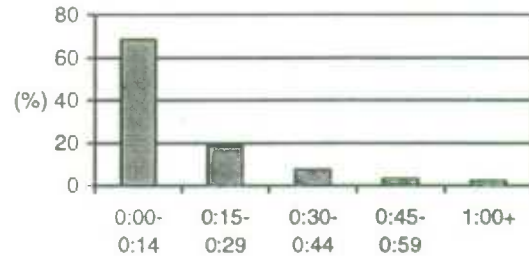
Repetitive Strain
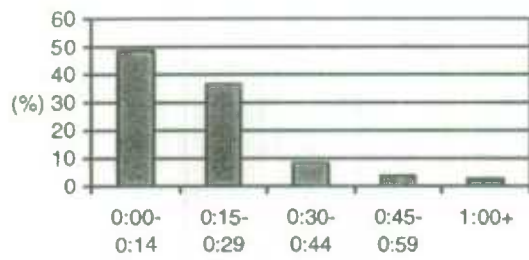
Injuries
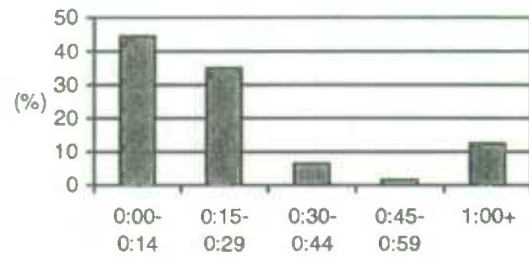
Height/Weight
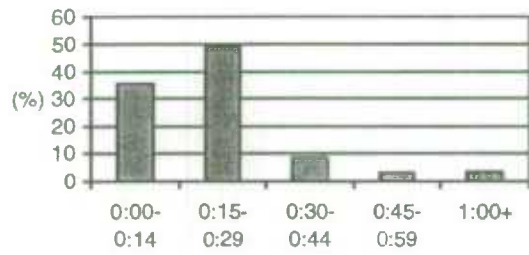
Insurance
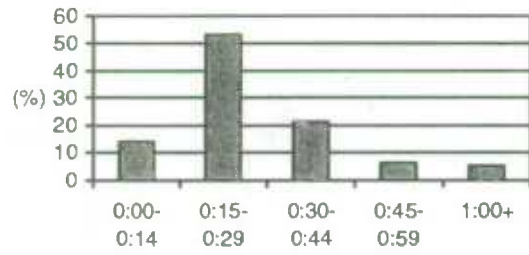
Preventive Health
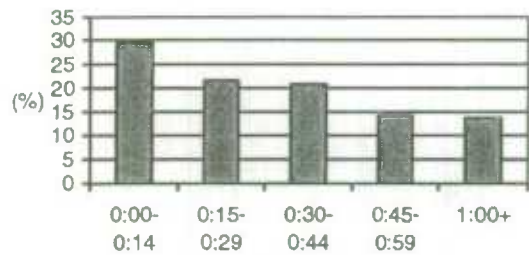
General Health
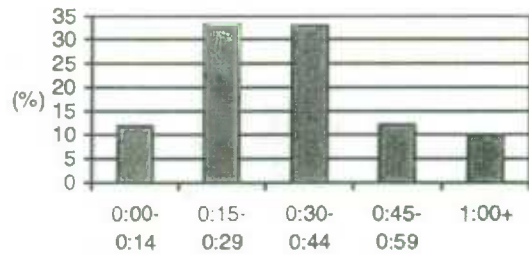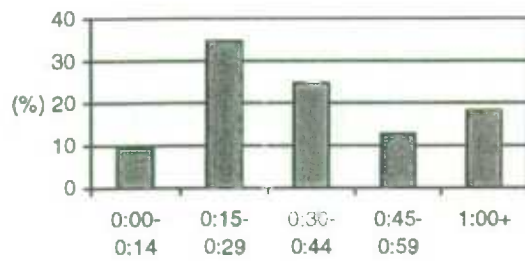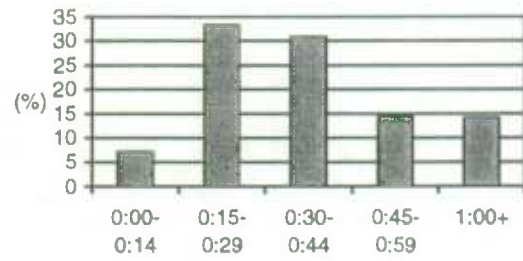
11

Smoking

(%)

40
30
20
10
0

0:00-0:14 0:15-0:29 0:30-0:44 0:45-0:59 1:00+

Socio-Demographic

(%)

35
30
25
20
15
10
5
0

0:00-0:14 0:15-0:29 0:30-0:44 0:45-0:59 1:00+

Restriction of Activities

(%)

40
30
20
10
0

0:00-0:29 0:30-0:59 1:00-1:29 1:30-1:59 2:00+

Administration

(%)

50
40
30
20
10
0

0:00-0:29 0:30-0:59 1:00-1:29 1:30-1:59 2:00+

Alcohol

(%)

35
30
25
20
15
10
5
0

0:00-0:29 0:30-0:59 1:00-1:29 1:30-1:59 2:00+

Income

(%)

35
30
25
20
15
10
5
0

0:00-0:29 0:30-0:59 1:00-1:29 1:30-1:59 2:00+

Mental Health

(%)

25
20
15
10
5
0

0:00-0:29 0:30-0:59 1:00-1:29 1:30-1:59 2:00-2:29 2:30+

Health Status

(%)

35
30
25
20
15
10
5
0

0:00-0:29 0:30-0:59 1:00-1:29 1:30-1:59 2:00-2:29 2:30+
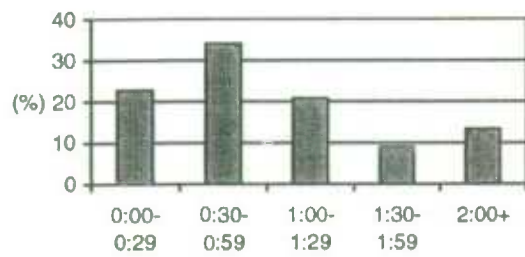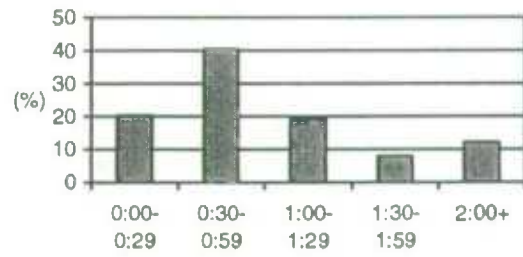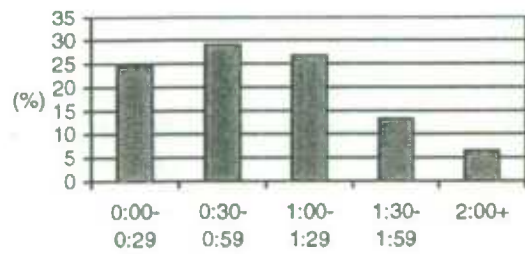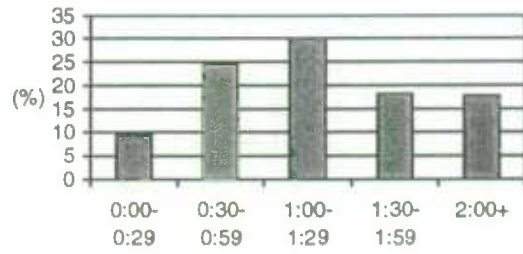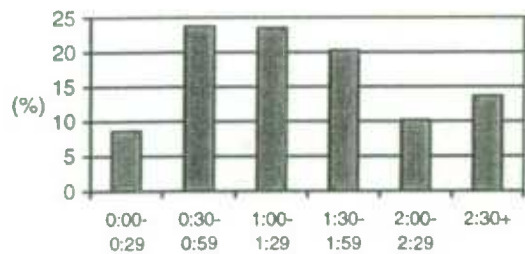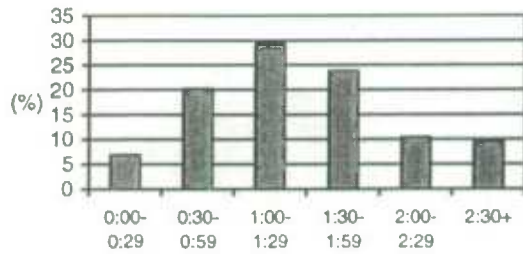
Chronic Conditions
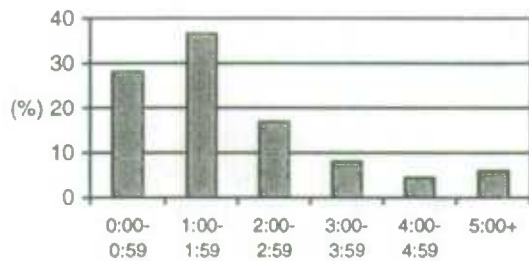
Health Care Utilization
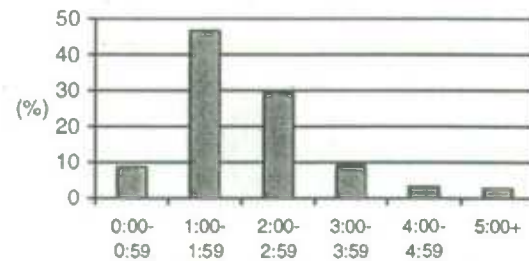
Drug Use
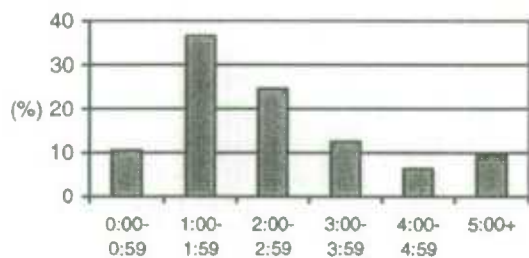
Social Support
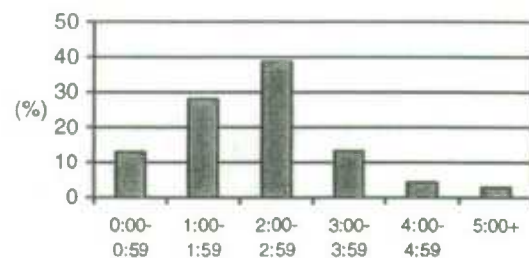
Labour Force
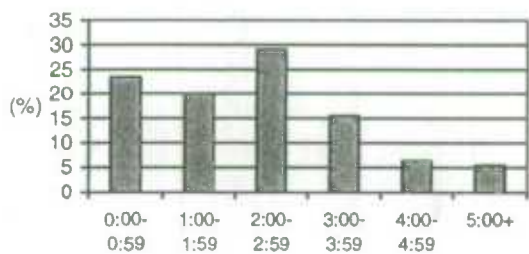
Physical Activities
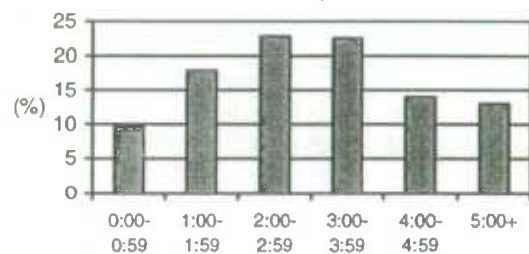
Stress
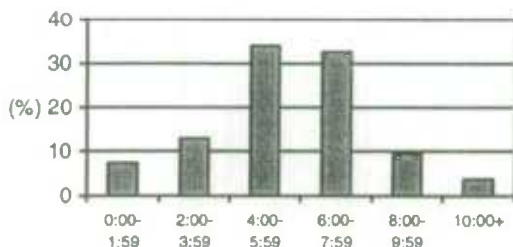
Table 3.6 presents the interview lengths by age group, based on the panel member's age in Cycle 4. As expected, the shortest interview lengths were for the under 12 age group, as several questions and modules are not asked of this age group. The median interview length for the under 12 age group is 8:49. The median time of the 12-24 age group is 28:16, while the medians of the other three age groups vary between 31:55 and 33:38. Some questions, including the entire stress module, which as noted above, is the longest module of the interview, are not asked of panel members under the age of 18, explaining the shorter median interview length for the 12-24 age group.

13

Table 3.6 – Cycle 4 Interview Lengths by Age Group

	Panel Members	Median Interview Length	25th Percentile	75th Percentile
Total	10,086	30:53	23:13	38:15
< 12	722	8:49	6:44	11:03
12-24	1,517	28:16	20:44	35:01
25-44	3,269	31:55	26:06	38:05
45-64	2,775	33:38	26:44	40:38
65+	1,803	32:24	24:50	41:59

Table 3.7 presents the interview lengths by sex. Female respondents tend to have interview lengths that are two to three minutes longer than male respondents. About half of the modules have average lengths that are longer for females, about one quarter have longer lengths for males, and the remaining modules have lengths that are about equal for both sexes. The preventive health and drug use modules have the largest differences between the sexes due to the fact that there are questions in these modules only asked of females.

Table 3.7 – Cycle 4 Interview Lengths by Sex

	Panel Members	Median Interview Length	25th Percentile	75th Percentile
Total	10,086	30:53	23:13	38:15
Males	4,557	29:29	21:46	36:41
Females	5,529	31:59	24:31	39:21

Table 3.8 displays the interview lengths by proxy and non-proxy interviews. Proxy reporting is when someone other than the panel member responds to questions about the panel member, and is allowed only for reasons of illness or incapacity. Proxy interviews are less than half the length of non-proxy interviews on average since many questions or entire modules are skipped in proxy interviews.

Table 3.8 – Cycle 4 Interview Lengths by Proxy Status

	Panel Members	Median Interview Length	25th Percentile	75th Percentile
Total	10,086	30:53	23:13	38:15
Non-Proxy	8,519	32:17	25:52	39:18
Proxy	1,567	14:04	8:52	27:59

Table 3.9 shows the interview lengths by interview type, i.e., telephone or in person. The statistics show that personal interviews are on average three to four minutes longer than telephone interviews, which is consistent with a study of LFS data quality conducted by Brisebois, Dufour and Lévesque (1998).

Table 3.9 – Cycle 4 Interview Lengths by Interview Type

	Panel Members	Median Interview Length	25th Percentile	75th Percentile
Total	10,086*	30:53	23:13	38:15
Telephone	9,907	30:50	23:12	38:12
In Person	177	33:18	23:39	42:12

*2 have an interview type of "both" and are only included in the total only

4. REFUSALS

Refusals are the most substantial source of nonresponse. Refusals make up 49% of Cycle 2 nonresponse, 56% of Cycle 3 nonresponse, and 61% of Cycle 4 nonresponse. Even though the intention is to follow all 17,276 panel members over time, not all records are sent out for collection each cycle, such as the harder refusals. Cases where the panel member has been confirmed dead through a match to the mortality files are considered complete for the rest of the span of the survey, and are no longer sent to the ROs.

Two different refusal rates for each cycle can be calculated, one based only on those records that were sent out, and the other based on all 17,276 records. It can be seen in Table 4.1, which displays both of these rates for Cycles 2, 3 and 4, that both refusal rates increased with each cycle.

Table 4.1 – Refusal Rates by Cycle

	Records that went out	New Refusals	Refusal rate based on records sent out	Refusals that were not sent out	Total number of refusals	Refusal rate based on all 17,276
Cycle 2	17,266	539	3.1%	1	540	3.1%
Cycle 3	16,582	601	3.6%	469	1070	6.2%
Cycle 4	16,186	1017	6.3%	526	1543	8.9%

Refusal conversions can happen within a quarter, between quarters of the same cycle, and between cycles. This section is composed of four subsections, one for each type of conversion and another for reasons for refusals. In this document, a refusal conversion means a refusal that became a respondent who fully completed the questionnaire.

4.1 Refusal Conversion Between Cycles

The refusal conversion rate between cycles is the percentage of records that refused in a particular cycle and provided a full response in a subsequent cycle. Table 4.2 shows the refusal conversion rates by sex and age group, while Table 4.3 shows the rates by RO. These rates are based only on the records that were sent out for collection, since only those records that were sent out could be considered converted. The refusal conversion rate from Cycle 2 to Cycle 3 is based on only 73 records, as 467 of the 540 refusals in Cycle 2 were not sent out for collection in Cycle 3. Two panel members who fully completed the questionnaire in Cycle 2 were not sent out for collection in Cycle 3 and were assigned refusal codes, which brings the number of refusals that were not sent out in Cycle 3 to 469, as shown in Table 4.1. The refusal conversion rate from Cycle 3 to Cycle 4 is based on 539 records, with 531 of the 1070 refusals in Cycle 3 not sent out for collection in Cycle 4. Table 4.1 shows 526 refusals that were not sent out in Cycle 4. This difference of 5 is due to panel members confirmed dead in Cycle 4.

The refusal conversion rate has remained stable at around 30% between cycles. Of the 22 individuals who refused in Cycle 2 and were converted in Cycle 3, 14 were also fully complete in Cycle 4. Two individuals refused in Cycle 4, while the remaining six were either partially complete or another form of nonresponse in Cycle 4. Of the 26 panel members who refused in both Cycles 2 and 3, eight were converted in Cycle 4. Females have slightly higher conversion rates than males for all cycles, except for those who refuse in both Cycles 2 and 3.

As for the other comparisons, the age used for the refusal rates by age group is the panel member's age in Cycle 4. When looking at the rates by age group for conversions from Cycle 2 to Cycle 3 and conversions from Cycle 2 to Cycle 4, the number of records in each age group is small, making it difficult to compare rates. The number of records in each age group for the conversions from Cycle 3 to Cycle 4 are much larger. The 25-44 age group has the lowest refusal conversion rate at 24%, while the two highest refusal conversion rates come from the oldest and youngest age groups with rates of 43% and 41%, respectively.

Table 4.2 – Refusal Conversion Rates by Sex and Age Group

	Number of Refusals	Number Converted in Cycle 3	Number Converted in Cycle 4	Number Not Converted
Cycle 2 Total	73	22 (30%)	16 (22%)	35 (48%)
Males	33	9 (27%)	7 (21%)	17 (52%)
Females	40	13 (33%)	9 (23%)	18 (45%)
Under 12	6	3 (50%)	1 (17%)	2 (33%)
12-24	15	5 (33%)	6 (40%)	4 (27%)
25-44	26	8 (31%)	5 (19%)	13 (50%)
45-64	18	5 (28%)	2 (11%)	11 (61%)
65+	8	1 (13%)	2 (25%)	5 (63%)
Cycle 3 Total	539	-	164 (30%)	375 (70%)
Males	240	-	72 (30%)	168 (70%)
Females	299	-	92 (31%)	207 (69%)
Under 12	37	-	15 (41%)	22 (59%)
12-24	78	-	22 (28%)	56 (72%)
25-44	234	-	57 (24%)	177 (76%)
45-64	127	-	43 (34%)	84 (66%)
65+	63	-	27 (43%)	36 (57%)
Refusals to both Cycles 2 and 3 – Total	26	-	8 (31%)	18 (69%)
Males	12	-	5 (48%)	7 (58%)
Females	14	-	3 (21%)	11 (79%)
Under 12	1	-	0 (0%)	1 (100%)
12-24	7	-	4 (57%)	3 (43%)
25-44	8	-	2 (25%)	6 (75%)
45-64	6	-	1 (17%)	5 (83%)
65+	4	-	1 (25%)	3 (75%)

Table 4.3 – Refusal Rates by RO

	Number of Refusals	Number Converted in Cycle 3	Number Converted in Cycle 4	Number Not Converted
Cycle 2 Total	73	22 (30%)	16 (22%)	35 (48%)
Halifax	11	2 (18%)	4 (36%)	5 (46%)
Montreal	12	5 (42%)	1 (8%)	6 (50%)
Toronto	37	12 (32%)	5 (14%)	20 (54%)
Edmonton	7	1 (14%)	2 (29%)	4 (57%)
Vancouver	6	2 (33%)	4 (67%)	0 (0%)
Cycle 3 Total	539	-	164 (30%)	375 (70%)
Halifax	113	-	30 (27%)	83 (73%)
Montreal	127	-	44 (35%)	83 (65%)
Toronto	100	-	21 (21%)	79 (79%)
Edmonton	138	-	52 (38%)	86 (62%)
Vancouver	61	-	17 (28%)	44 (72%)
Refusals to both Cycles 2 and 3 – Total	26	-	8 (31%)	18 (69%)
Halifax	6	-	3 (50%)	3 (50%)
Montreal	4	-	0 (0%)	4 (100%)
Toronto	10	-	2 (20%)	8 (80%)
Edmonton	5	-	2 (40%)	3 (60%)
Vancouver	1	-	1 (100%)	0 (0%)

As with the age groups, it is difficult to compare the refusal conversion rates from Cycle 2 to Cycle 3 and from Cycle 2 to Cycle 4 by RO due to the small number of records (see Table 4.3). For the Cycle 3 refusals, the Toronto RO has the lowest conversion rate at 21.0%, and the Edmonton RO has the highest at 37.7%.

4.2 Refusal Conversion Within Cycle 4

In Cycle 4, a status code was assigned during the processing of each quarter to determine if the panel member was a fully complete, a refusal, or neither. Panel members not sent out for collection or with data collected by the Institutions component are excluded, leaving 16,071 records for use in the study of Cycle 4 refusal reasons (section 4.4) and refusal conversion rates within Cycle 4.

Most panel members who refuse in Quarter 1 are resent for collection in Quarter 3, while most panel members who refuse in Quarter 2 are resent for collection in Quarter 4. The majority of those who refuse in Quarters 3 or 4 are resent in Quarter 5, as well as some of the stronger refusals from Quarters 1 and 2 who were not resent in Quarters 3 and 4. Table 4.5 shows the refusal rates by quarter.

Table 4.5: Cycle 4 Refusal Rates by Quarter

Quarter	Records that went out	Refusals	Refusal rate
1	3,434	238	7%
2	3,610	208	6%
3	4,734	333	7%
4	4,640	330	7%
5	2,935	911	31%

The refusal rate is highest in Quarter 5, with a rate of 31%, not surprising since Quarter 5 contains only two kinds of panel members: nonrespondents from the previous four quarters of the cycle, or adamant refusals to a previous cycle that are only sent out for collection in Quarter 5.

There were 1,311 panel members who were classified as a refusal at the end of at least one quarterly collection period in Cycle 4. Of those, 235 fully completed the questionnaire in a later quarter, for a refusal conversion rate of 18%. Tables 4.6, 4.7 and 4.8 show the refusal conversion rates by sex, age group, and RO.

Table 4.6 – Cycle 4 Refusal Conversion Rates by Sex

	Total	Conversions	Non-Conversions	Conversion rate
Canada	1,311	235	1,076	17.9%
Males	650	117	533	18.0%
Females	661	118	543	17.9%

Table 4.7 – Cycle 4 Refusal Conversion Rates by Age Group

	Total	Conversions	Non-Conversions	Conversion rate
Canada	1,311	235	1,076	17.9%
Under 12	31	9	22	29.0%
12-24	152	27	125	17.8%
25-44	539	88	451	16.3%
45-64	383	62	321	16.2%
65+	206	49	157	23.8%

Table 4.8 – Cycle 4 Refusal Conversion Rates by RO

RO	Total	Conversions	Non-Conversions	Conversion rate
Canada	1,311	235	1,076	17.9%
Halifax	339	68	271	20.0%
Montreal	225	43	182	19.1%
Toronto	356	54	302	15.2%
Edmonton	253	38	215	15.0%
Vancouver	138	32	106	23.2%

The refusal conversion rates by sex are almost identical at roughly 18%. As with the between-cycle conversion rates, the oldest and youngest age groups have the highest within-Cycle 4 conversion rates, with rates of 24% and 29%, respectively. The Edmonton and Toronto ROs have the lowest conversion rates at 15%, while the Vancouver RO has the highest rate at 23%.

Table 4.9 shows the quarters in which the 1,311 panel members who refused went out for collection. There were 216 records who were converted in the next quarter, and 19 were converted two quarters later. Of these 235 conversions, 26 were converted in Quarter 3, 12 in Quarter 4 and 197 in Quarter 5. Those records that went out in two quarters had a refusal conversion rate of 32%, while those that went out in three quarters had a rate of 16%. These rates show that it is worth it to resend refusals for collection.

Table 4.9 – Quarters of Attempted Collection for Cycle 4 Refusal Cases

Quarters sent out	Total	Converted	Not converted	Conversion rate
Q1 only	40	n/a	n/a	n/a
Q1, Q3	31	26	5	83.9%
Q1, Q5	97	23	74	23.7%
Q1, Q3, Q5	70	14	56	20.0%
Q2 only	14	n/a	n/a	n/a
Q2, Q4	21	12	9	57.1%
Q2, Q5	125	33	92	26.4%
Q2, Q4, Q5	48	5	43	10.4%
Q3 only	20	n/a	n/a	n/a
Q3, Q5	258	67	191	26.0%
Q4 only	16	n/a	n/a	n/a
Q4, Q5	267	55	212	20.6%
Q5 only	304	n/a	n/a	n/a
Total	1,311	235	1,076	17.9%

4.3 Refusal Conversion Within a Quarter

The interviewer assigns an outcome code for each contact within a quarter. Some panel members refused in an early contact(s) of a quarter, but became fully complete during a contact later in the same quarter. Table 4.13 shows the number of panel members who refused at some point during a given quarter, the number of refusals that converted to fully completed within that quarter, and the rate.

Table 4.13 – Cycle 4 Conversion Rates Within Quarters

Quarter	Number that refuse at some point within a quarter	Number that convert in that quarter	Conversion Rate
1	304	45	15%
2	279	43	15%
3	413	39	9%
4	459	75	16%
5	920	38	4%

4.4 Reasons for Refusal

When an interview is coded as a refusal, the interviewer assigns a code identifying the reason for the refusal. Not all of those records considered as refusals have a reason for their refusal. In some cases, the outcome code of "refusal" was not assigned to the record by the interviewer, but was changed to a refusal during processing. Table 4.10 shows the reasons for refusal given by quarter. "Not interested / doesn't want to participate" is the reason that is given most often, with "Same household, refusal maintained", "Doesn't want to continue the survey" and "Adamant refusal" following.

Table 4.11 shows the refusal reasons given upon the panel member's first, second and third refusals (if applicable). The "Not interested / doesn't want to participate" reason is the reason given the most often for a respondent's first refusal, making up 24% of the first refusal reasons. This percentage drops to near 15% for the second and third refusals. "Same household, refusal maintained" makes up 16% of the first refusal reasons, making it the third most common. However, not surprisingly, it becomes the top reason given for both second and third refusals, making up 25% and 31% of the refusal reasons, respectively. Of the 635 panel members that refused at least twice, 153 (24%) gave the same reason twice, while only 4 of the 74 (5%) that refused three times gave the same reason three times.

Table 4.10 – Cycle 4 Refusal Reasons by Quarter

Reason for refusal	Q1	Q2	Q3	Q4	Q5	Total
Not interested / doesn't want to participate	57	55	72	93	149	426 (21%)
Same household, refusal maintained (for follow-ups only)	41	8	38	48	254	389 (19%)
Doesn't want to continue the survey (no more follow-ups)	35	49	76	46	86	292 (15%)
Adamant refusal (no specific reason, include shuts the door or hangs up the phone)	36	43	36	48	104	267 (13%)
Doesn't have the time	14	18	39	25	71	167 (8%)
Doesn't want to give out personal information	6	8	6	2	46	68 (3%)
Doesn't believe in or want to hear about statistics	0	6	20	4	27	57 (3%)
Against the government or Statistics Canada	5	3	4	6	20	38 (2%)
Doesn't want to be disturbed	3	2	7	12	12	36 (2%)
Doesn't believe in or want to hear about surveys	6	4	5	2	8	25 (1%)
Dangerous / rude attitude	0	1	2	7	14	24 (1%)
Doesn't believe the info is secure (confidentiality)	1	0	2	3	11	17 (1%)
Says not obligated / wants legal proof	7	2	2	2	3	16 (1%)
Won't answer the door	0	0	1	1	3	5 (<1%)
Recently completed a survey (doesn't want to again)	1	0	1	1	0	3 (<1%)
Why me? Tells you to choose someone else	0	1	0	0	0	1 (<1%)
Can get info somewhere else (e.g. Revenue Canada)	0	0	0	0	1	1 (<1%)
Other – Specify	13	3	5	12	11	44 (2%)
Not stated (were not initially classified as refusals, so no reason assigned)	13	5	17	18	91	144 (7%)
Total	238	208	333	330	911	2,020

Table 4.11 – Cycle 4 Refusal Reasons

Reason for refusal	First refusal	Second refusal	Third refusal
Not interested / doesn't want to participate	312 (24%)	103 (16%)	11 (15%)
Same household, refusal maintained (for follow-ups only)	205 (16%)	161 (25%)	23 (31%)
Doesn't want to continue the survey (no more follow-ups)	217 (17%)	72 (11%)	3 (4%)
Adamant refusal (no specific reason, include shuts the door or hangs up the phone)	178 (14%)	82 (13%)	7 (10%)
Doesn't have the time	118 (9%)	43 (7%)	6 (8%)
Doesn't want to give out personal information	34 (3%)	31 (5%)	3 (4%)
Doesn't believe in or want to hear about statistics	27 (2%)	26 (4%)	4 (5%)
Against the government or Statistics Canada	24 (2%)	13 (2%)	1 (1%)
Doesn't want to be disturbed	28 (2%)	6 (1%)	2 (3%)
Doesn't believe in or want to hear about surveys	17 (1%)	8 (1%)	0 (0%)
Dangerous / rude attitude	12 (1%)	10 (2%)	2 (3%)
Doesn't believe the info is secure (confidentiality)	7 (1%)	8 (1%)	2 (3%)
Says not obligated / wants legal proof	11 (1%)	5 (1%)	0 (0%)
Won't answer the door	4 (<1%)	1 (<1%)	0 (0%)
Recently completed a survey (doesn't want to again)	3 (<1%)	0 (0%)	0 (0%)
Why me? Tells you to choose someone else	1 (<1%)	0 (0%)	0 (0%)
Can get info somewhere else (e.g. Revenue Canada)	0 (0%)	0 (0%)	1 (1%)
Other – Specify	33 (3%)	10 (2%)	1 (1%)
Not stated (were not initially classified as refusals, so no reason assigned)	80 (6%)	56 (9%)	8 (11%)

Table 4.12 – Cycle 4 Reasons of Those Converted

Reason for refusal	Converted next quarter – first refusal	Converted two quarters later – first refusal	Converted two quarters later – second refusal*	Percentage who gave this reason and were converted
Not interested / doesn't want to participate	62	7	2	17%
Same household, refusal maintained (for follow-ups only)	22	2	1	6%
Doesn't want to continue the survey (no more follow-ups)	25	3	3	11%
Adamant refusal (no specific reason, include shuts the door or hangs up the phone)	42	0	3	17%
Doesn't have the time	27	3	1	19%
Doesn't want to give out personal information	2	0	0	3%
Doesn't believe in or want to hear about statistics	5	0	1	11%
Against the government or Statistics Canada	2	1	1	11%
Doesn't want to be disturbed	5	2	0	19%
Doesn't believe in or want to hear about surveys	2	0	0	8%
Other – Specify	8	0	0	18%
Not stated (were not initially classified as refusals, so no reason assigned)	14	1	1	11%
Total	216	19	13	12%

*Sums to only 13 since some were nonrespondent but not a refusal during the second quarter that they were contacted and therefore do not have a refusal reason

Table 4.12 shows the refusal reasons of those panel members who initially refused and then became fully completed in a subsequent quarter. It is interesting to note that 17% of individuals who were an adamant refusal at some point in Cycle 4 fully completed the questionnaire in a subsequent quarter. Panel members who say that they do not want to give out personal information are the least likely to fully complete the questionnaire in a subsequent quarter, with only 3% doing so.

5. ITEM REFUSAL AND DON'T KNOW RATES

A major source of non-sampling errors in surveys is the effect of nonresponse on the survey results. The extent of nonresponse varies from partial nonresponse (failure to answer just one or several questions) to total nonresponse. Partial nonresponse to NPHS was minimal: once the questionnaire was started, it tended to be completed with very little nonresponse. In most cases, partial nonresponse to the survey occurred when the respondent did not understand or misinterpreted a question, refused to answer a question, could not recall the requested information, or could not provide personal or proxy information. Total nonresponse mainly occurred because it was impossible to trace the respondent, no member of the household was present or able to provide the information, or the respondent refused to participate in the survey. Total nonresponse was handled by adjusting the weight of persons who responded to the survey to compensate for those who did not respond. Note that this adjustment was done by nonresponse class, using many of the following variables to group similar units together, in order to minimize the bias associated with nonresponse. For more details on this adjustment, please refer to the Cycle 4 Longitudinal Documentation.

This section examines two types of partial non-response: item *refusals* and *don't knows*. Item refusal rates and don't know rates were calculated for each Cycle 4 variable by age group, sex and overall. Don't know rates were also computed and compared for those whose information was collected directly from the respondent and for those collected through an intermediary person (proxy). The variables were grouped into modules (and submodules when possible), and refusal and don't know rates by module were calculated. The module and submodule rates were computed by dividing the total number of refusals obtained for all questions in the module (or submodule), by the total number of questions asked to all panel members specifically in that module (or submodule). A rate based on the entire questionnaire was also computed using the same method. Note that within a given module or submodule, not all questions were necessarily asked to the same group of panel members. The calculation of the refusal and don't know rates are based on the 12,575 panel members who fully completed the questionnaire in all four cycles, which excludes panel members who died or became institutionalized during the course of the survey.

5.1 Item Refusal Rates

The refusal rate based on the entire questionnaire is 0.09%. Table 5.1 displays the refusal rates by module and some submodules. The 24 chronic condition submodules and the 9 health status submodules all have refusal rates of 0.03% or lower, and have been excluded from the table. The income module is by far the module with the highest refusal rate, with a rate of 0.62%. The labour force module is made up of eight submodules, and four of these submodules have the highest refusal rates after the income module. These four submodules are job description (0.36%), looking for work (0.25%), weeks worked (0.16%) and absence/hours (0.14%).

Table 5.2 shows the ten variables with the highest refusal rates. The top three rates are all variables from the income module. The only other variable with a refusal rate above 1% at the national level is the variable CCC0_G2M, which represents the month that the respondent's migraines disappeared for those who said that they were diagnosed with migraines in a previous cycle but no longer have them. This variable is based only on 171 respondents, whereas the three income variables are all based on over 10,000 respondents.

The stress module is made up of 6 submodules, and 3 of them (self-esteem, mastery, recent life events) have refusal rates among the top 10 highest. The first question in each of the self-esteem and mastery submodules of the stress module are among the highest refusal rates for individual variables, with rates of 0.58% and 0.64% respectively. The refusal rates of the other variables in the submodules are all below 0.05%, since if a respondent refuses to answer the first question, the rest of the module is skipped.

5.1.1 Sex

Table 5.3 shows the ten modules and submodules with the highest refusal rates by sex. Refusal rates based on the entire questionnaire were fairly similar, with males having a slightly higher refusal rate of 0.10% compared to 0.08% for females. Nine of the ten modules and submodules with the highest refusal rates are

the same for both sexes, although their order within the top 10 and their actual rates differ. The majority of the modules and submodules have higher refusal rates for males than females, and in the cases where the rate for females is higher, it is never by more than 0.03 percentage points.

Table 5.1 – Refusal Rates by Module and Submodule

Module	Submodule	Refusal Rate (in %)	Module	Submodule	Refusal Rate (in %)
Overall		0.09	Drug Use (DGC)		0.02
Household Record Variables (DHC)		0.01	Smoking (SMC)		0.03
General Health (GHC, STS)		0.01	Alcohol (ALC)		0.03
Height/Weight (HWC)		0.00	Mental Health (MHC)		0.07
Preventive Health (BPC,WHC,GHC)		0.05	Social Support (SSC)		0.07
	BPC – Blood Pressure	0.05	Socio-Demographic (SDC)		0.07
	WHC – Women's Health	0.06	Education (EDC)		0.01
	GHC – Births	0.00	Labour Force (LSC, LFC)		0.12
Health Care Utilization (HCC)		0.00		Job Attachment	0.04
Restriction Of Activities (RAC)		0.01		Job Search - Last 4 Wks	0.01
Chronic Conditions (CCC, CC_)		0.01		Past Job Attachment	0.06
Insurance (ISC)		0.04		Job Description	0.36
Health Status (HSC)		0.01		Absence/Hours	0.14
Physical Activities (PAC)		0.12		Other Job	0.08
UV Exposure (TUC)		0.04		Weeks Worked	0.16
Repetitive Strain (RPC)		0.04		Looking for Work	0.25
Injuries (IJC)		0.02	Income (INC)		0.62
Stress (ST_, PY_)		0.07	Cycle 1 Variables (Birth Weight, Age and Year of Immigration, Country of Birth, Day, Month, Year of Birth)		0.00
	SY_ - Stress	0.06			
	PY_ - Self-Esteem and Mastery	0.10			
	Ongoing Problems	0.09			
	Recent Life Events	0.10			
	Childhood/Adult Stressors	0.05			
	Work Stress	0.06			
	Self-Esteem	0.13			
	Mastery	0.11			

Table 5.2 - Variables with the Ten Highest Refusal Rates

Variable Name	Variable Concept	Refusal Rate (in %)	Sample Size
INC0_3	Total hhld inc. - best estimate	3.77	12,226
INC0_3A	Total hhld inc. - <$20,000 or >=$20,000	1.87	12,225
INC0_4	Total pers. inc. - best estimate	1.49	10,416
CCC0_G2M	Migraines - month cond. disappeared	1.17	171
CCC0_C2M	Asthma - month cond. disappeared	0.85	118
LSC0F32	Flag for name of own business	0.76	1,191
INC0_1A - INC0_1N	Source of income	0.66	12,475
PY_0_M1A	Stress/mastery - lack of control	0.64	11,267
LSC0F33	Flag for name of employer	0.60	6,385
PY_0_E1A	Stress/self-esteem – have good qualities	0.58	11,269
ST_0_C1	Stress - trying take on too many things	0.56	10,523

Table 5.3 – Modules with the Highest Refusal Rates by Sex

Males		Females	
Module or submodule	Refusal rate (in %)	Module or submodule	Refusal rate (in %)
Overall	0.10	Overall	0.08
Income (INC)	0.63	Income (INC)	0.62
Job description	0.46	Job description	0.27
Looking for work	0.35	Looking for work	0.17
Weeks worked	0.21	Physical Activity (PAC)	0.12
Self-esteem	0.19	Weeks worked	0.10
Absence/Hours	0.18	Absence/Hours	0.10
Labour Force (LSC, LFC)	0.16	Recent Life Events	0.09
PY_ - Self-esteem and Mastery	0.14	Self-esteem	0.09
Mastery	0.14	Mastery	0.09
Physical Activity (PAC)	0.13	Labour Force (LSC, LFC)	0.08

Table 5.4 – Variables with the Highest Refusal Rates by Sex

Males				Females			
Variable	Concept	Sample size	Refusal Rate (in %)	Variable	Concept	Sample size	Refusal Rate (in %)
INC0_3	Total hhld inc. - best estimate	5,500	3.95	INC0_3	Total hhld inc. – best estimate	6,726	3.63
CCC0_G2M	Migraines – month cond. Disappeared	47	2.13	INC0_3A	Total hhld inc. - <$20,000 or >=$20,000	6,725	1.78
INC0_3A	Total hhld inc. - <$20,000 or >=$20,000	5,500	1.98	INC0_4	Total pers. inc. – best estimate	5,791	1.54
INC0_4	Total pers. inc. – best estimate	4,625	1.43	CCC0_C2M	Asthma - month cond. disappeared	67	1.49
PY_0_M1A	Stress/mastery – lack of control	4,946	0.89	CCC0_G2M	Migraines – month cond. disappeared	124	0.81
LSC0F32	Flag for name of own business	733	0.82	INC0_1A – INC0_1N	Source of income	6,861	0.66
LSC0F33	Flag for name of employer	3,011	0.80	LSC0F32	Flag for name of own business	458	0.66
PY_0_E1A	Stress/self-esteem - have good qualities	4,948	0.79	MHC0_14	Depr. - no. of weeks in past 12 months	405	0.49
ST_0_C1	Stress – trying take on too many things	4,578	0.74	ST_0_T1	Stress/trauma – 2 weeks/more in hospital	437	0.46
CCC0_N3	Ulcers – age first diagnosed	144	0.69	INC0_4A	Total pers. inc. - <$20,000 or >=$20,000	5,791	0.45

Table 5.4 shows the variables with the ten highest refusal rates for each sex. Of the ten highest refusal rates by variable, five are common to both sexes, three of them being income variables. The first variable from the mastery and self-esteem submodules, PY_0_M1A and PY_0_E1A, respectively, are both among the top 10 refusal rates for males, at approximately double the female rates.

5.1.2 Age groups

The panel members were grouped into five age groups based on their age in Cycle 4: under 12, 12-24, 25-44, 45-64, 65+. The refusal rates based on the entire questionnaire vary little, between 0.06% for the 12-24 age group and 0.14% for the 65+ age group. Table 5.5 shows the modules with the ten highest refusal rates for each age group. Income is the module with the highest refusal rate for all five age groups, with the 65+ group having the highest rate at 1.05%. The questionnaire for those under 12 is completed by proxy, and

24

only two modules, income and socio-demographic, have refusal rates above 0%. All other modules either have a refusal rate of 0% or were not asked. Due to the fact that the under 12 age group is so different from the other four age groups, it is excluded from the discussion that follows.

Table 5.5 – Modules with the Highest Rates by Age Group

0-11		12-24		25-44	
Module or submodule	Refusal rate (in %)	Module or submodule	Refusal rate (in %)	Module or submodule	Refusal rate (in %)
Overall	0.10	Overall	0.06	Overall	0.09
Income (INC)	0.50	Income (INC)	0.60	Income (INC)	0.43
Socio-demographic (SDC)	0.10	Job description	0.32	Job description	0.37
All other modules 0% or not asked		Blood pressure (BPC)	0.18	Looking for work	0.33
		Absence/Hours	0.13	Past job attachment	0.33
		Insurance (ISC)	0.11	Physical Activity (PAC)	0.21
		Work stress	0.10	Women's health WHC)	0.21
		Labour force (LSC, LFC)	0.09	Other job	0.16
		Preventive health (BPC, WHC, GHC)	0.08	Labour force (LSC, LFC)	0.14
		Weeks worked	0.08	Weeks worked	0.13
		Self-esteem	0.07	Socio-demographic (SDC)	0.13

45-64		65+	
Module or submodule	Refusal rate (in %)	Module or submodule	Refusal rate (in %)
Overall	0.07	Overall	0.14
Income (INC)	0.62	Income (INC)	1.05
Looking for work	0.38	Job description	0.43
Job description	0.38	Self-esteem	0.29
Weeks worked	0.24	Mastery	0.28
Absence/Hours	0.18	Recent life events	0.25
Labour force (LSC, LFC)	0.14	PY_ - Self -esteem and mastery	0.25
Self-esteem	0.10	Ongoing problems	0.23
Recent life events	0.09	Stress	0.18
Mastery	0.08	Work stress	0.15
PY_ - Self-esteem and mastery	0.08	Social support (SSC)	0.15

When looking at the module and submodule level, most of the ten highest refusal rates for each age group are found in submodules of the labour force and stress modules. Income and job description are the only module or submodules found among the ten highest refusal rates for all four age groups.

The blood pressure submodule of the preventive health module has a much higher refusal rate in the 12-24 age group than the other groups, with a rate of 0.18%, while the highest rate in the other three age groups is 0.04%. This is due mostly to the variable BPC0_10, which asks if the respondent has ever had their blood pressure taken, with a refusal rate of 0.23% for the 12-24 group, compared to rates of 0% for the three other groups.

The self-esteem and mastery submodules have higher refusal rates in the 65+ age group. Two other submodules from the stress section (recent life events and ongoing problems) also have refusal rates that are about three times higher than those of the other age groups.

The social support module is among the top ten refusal rates by module only for the 65+ group. Most of the variables that make up this module have refusal rates that increase by the age of the group.

Table 5.6 displays the ten variables with the highest refusal rates for each age group. Not surprisingly, an income variable has the highest refusal rate for each of the age groups, ranging from 2.68% for the 25-44 group to 5.72% for the 65+ group. Many of the variables that are among the ten highest refusal rates for a particular age group but are not found among the highest rates of the other three groups are based on a small number of respondents. Therefore, when looking at refusal rates, the number of respondents should be taken into account.

Table 5.6 – Variables with the Highest Refusal Rates by Age Group

0-11				12-24			
Variable	Concept	Sample Size	Refusal Rate (in %)	Variable	Concept	Sample Size	Refusal Rate (in %)
INC0_3	Total hhld inc. – best estimate	825	3.27	INC0_3	Total hhld inc. – best estimate	1,776	3.72
INC0_3A	Total hhld inc. - <$20,000 or >=$20,000	825	1.33	CCC0_C2M	Asthma – month cond. disappeared	32	3.13
INC0_3F	Total hhld inc. - <$30,000 or >=$30,000	174	0.57	INC0_3A	Total hhld inc. - <$20,000 or >=$20,000	1,776	1.18
INC0_1A – INC0_1N	Source of income	833	0.48	INC0_3D	Total hhld inc. - <$15,000 or >=$15,000	139	0.72
INC0_3G	Total hhld inc. - >= $50,000	526	0.38	INC0_1A – INC0_1N	Source of income	1,915	0.68
INC0_3E	Total hhld inc. - <$40,000 or >=$40,000	704	0.14	INC0_4	Total pers. inc. – best estimate	1,186	0.59
SDC0_4A – SDC0_4S	Ethnic Origin	833	0.12	RAC0CIC & RAC0F3	Main health prob. & Flag for prob. causing limitation	173	0.58
SDC0_5A – SDC0_5S	Can converse in	833	0.12	LSC0F33	Flag for name of employer	1,183	0.51
SDC0_6A – SDC0_6S	First language learned and still understand	833	0.12	INC0_3B	Total hhld inc. - <$10,000 or >=$10,000	210	0.48
NO OTHER REFUSALS				INC0_4F	Total pers. inc. - <$30,000 or >=$30,000	243	0.41
25-44				45-64			
Variable	Concept	Sample Size	Refusal Rate (in %)	Variable	Concept	Sample Size	Refusal Rate (in %)
INC0_3	Total hhld inc. – best estimate	4,061	2.68	INC0_3	Total hhld inc. – best estimate	3,345	3.95
DGC0_1T1	Hormone therapy – type	48	2.08	CCC0_G2M	Migraines - month cond. disappeared	55	3.64
INC0_3A	Total hhld inc. - <$20,000 or >=$20,000	4,060	1.53	INC0_3A	Total hhld inc. - <$20,000 or >=$20,000	3,345	2.12
INC0_4	Total pers. inc. – best estimate	3,937	1.07	INC0_4	Total pers. inc. - best estimate	3,227	1.64
LSC0F32	Flag for name of own business	549	0.91	PAC0_2I	No. times partic. - downhill skiing	66	1.52
WHC0_32	Last time mammogram was done	466	0.86	HSC0_7	Hearing - in group with hearing aid	114	0.88
LSC0F33	Flag for name of employer	3,171	0.66	LSC0_71	No. of weeks looked for work past year	582	0.86
LSC0_53	Usual no. hour at work - other job/bus.	307	0.65	LSC0F32	Flag for name of own business	522	0.77
WHC0_33A – WHC0_33H	Reason had mammogram	466	0.64	CCC0_N3	Ulcers – age first diagnosed	148	0.68
LSC0_22	Looked for work in the past 12 months	314	0.64	INC0_1A – INC0_1N	Source of income	3,376	0.62

65+			
Variable	Concept	Sample Size	Refusal Rate (in %)
INC0_3	Total hhld inc. – best estimate	2,219	5.72
INC0_3A	Total hhld inc. - <$20,000 or >=$20,000	2,219	2.88
INC0_4	Total pers. inc. – best estimate	2,066	2.57
MHC0_3	Depr. - usual duration	97	2.06
MHC0_14	Depr. - no. of weeks in past 12 months	53	1.89
MHC0_12	Depr. - felt down on self	56	1.79
PY_0_M1A	Stress/mastery – lack of control	2,122	1.32
LSC0F33	Flag for name of employer	76	1.32
PY_0_E1A	Stress/self-esteem - have good qualities	2,124	1.18
INC0_1A – INC0_1N	Source of income	2,263	1.15

The majority of the variables among the ten highest refusal rates for the 12-24 group are from the income module. It is interesting to note that the variables with the ten highest rates for the 25-44 age group fall into three categories: income, labour force, and women's health. In the 45-64 age group, the majority of the variables with the ten highest rates were income and labour force variables. Income, mastery, self-esteem and depression represent the majority of the variables with the ten highest rates for the 65+ group.

5.1.3 Other cycles

Of the variables that are common to Cycle 4 and one or more previous cycles, those with high refusal rates in Cycle 4 also tend to have high refusal rates in previous cycles. Income variables, for example, have high refusal rates in all cycles. Many of the same patterns seen with the Cycle 4 variables are also seen in previous cycles, such as an increase in refusal rates to mastery and self-esteem variables as age increases.

Some variables that were only part of one cycle have high refusal rates compared to the other variables of that cycle. Many of the Alberta buy-in variables and the Manitoba buy-in variables in Cycle 1, which were about coping with day-to-day demands, have high refusal rates. The sexual health variables in Cycle 2 have refusal rates that are quite high, as well as the sexual health variables that were part of the Alberta buy-in questionnaire only.

5.2 Don't Know Rates

The don't know rate based on the entire questionnaire is 0.28% for all respondents. Table 5.7 shows the don't know rates by module and submodule.

At the module level, income (1.82%) and insurance (1.88%) clearly have higher rates than the other modules. At the submodule level, 4 of the 22 chronic conditions submodules have substantially higher rates: heart disease (1.23%), focus content (1.02%), arthritis (0.89%), and high blood pressure (0.82%).

Five of the 17 variables in the heart disease submodule are responsible for its high don't know rate: the age of the respondent at the time of the most recent heart attack (5.79%), age at the time of first heart attack (3.36%), the number of heart attacks (2.24%), whether the respondent has congestive heart failure (1.56%), and has been referred to a cardiac rehabilitation program (1.12%).

Table 5.7 - Don't Know Rates by Module and Submodule

Module	Submodule	Don't Know Rate (in %)	Module	Submodule	Don't Know Rate (in %)
Overall		0.28	Physical Activities (PAC)		0.12
Household Record Variables (DHC)		0.01	UV Exposure (TUC)		0.04
General Health (GHC, STC)		0.05	Repetitive Strain (RPC)		0.36
Height/Weight (HWC)		0.01	Injuries (IJC)		0.08
Preventive Health (BPC,WHC,GHC)		0.29	Stress (ST_, PY_)		0.22
	BPC – Blood Pressure	0.22		ST_ - Stress	0.19
	WHC – Women's Health	0.40		PY_ - Self-Esteem and Mastery	0.34
	GHC – Births	0.01		Ongoing Problems	0.38
Health Care Utilization (HCC)		0.06		Recent Life Events	0.31
Restriction of Activities (RAC)		0.04		Childhood/Adult Stressors	0.24
Chronic Conditions (CCC, CC_)		0.25		Work Stress	0.12
	CCC – Core Content	0.18		Self-Esteem	0.30
	CC_- Focus Content	1.02		Mastery	0.46
	Food or Digestive Allergy	0.14	Drug Use (DGC)		0.18
	Other Allergies	0.06	Smoking (SMC)		0.27
	Asthma	0.39	Alcohol (ALC)		0.13
	Fibromyalgia	0.05	Mental Health (MHC)		0.08
	Arthritis or Rheumatism Excluding Fibromyalgia	0.89	Social Support (SSC)		0.16
	Back Problems	0.23	Socio-Demographic (SDC)		0.23
	High Blood Pressure	0.82	Education (EDC)		0.01
	Migraine Headaches	0.30	Labour Force (LSC, LFC)		0.09
	Chronic Bronchitis or Emphysema	0.07		Job Attachment	0.01
	Diabetes	0.15		Job Search - Last 4 Wks	0.02
	Epilepsy	0.03		Past Job Attachment	0.01
	Heart Disease	1.23		Job Description	0.13
	Cancer	0.08		Absence/Hours	0.20
	Stomach or Intestinal Ulcers	0.42		Other Job	0.81
	Effects of a Stroke	0.05		Weeks Worked	0.45
	Urinary Incontinence	0.11		Looking for Work	0.43
	Bowel Disorder	0.08	Income (INC)		1.82
	Alzheimer's Disease or Dementia	0.06	Cycle 1 Variables (Birth Weight, Age and Year of Immigration, Country of Birth, Day, Month, Year of Birth)		0.02
	Cataracts	0.15			
	Glaucoma	0.08			
	Thyroid Condition	0.13			
	Other Long-Term Condition	0.02			
Insurance (ISC)		1.88			
Health Status (HSC)		0.05			
	Vision	0.13			
	Hearing	0.39			
	Speech	0.02			
	Getting Around	0.00			
	Hands and Fingers	0.01			
	Feelings	0.08			
	Memory	0.05			
	Thinking	0.05			
	Pain And Discomfort	0.07			

For the arthritis submodule, the high don't know rate is largely due to the fact that 11.42% of respondents who have arthritis responded don't know when asked what kind of arthritis they have. The high blood pressure submodule's high rate is due primarily to three variables: the respondent's age when first diagnosed (2.51%), and the year (8.96%) and month (16.42%) that the condition disappeared. It is worth

noting that many of the other variables dealing with the month or year that a particular condition disappeared have don't know rates between 5% and 30%, but are based on very few panel members. The focus content submodule is made up of the focus content variables from several chronic condition modules, including the heart disease, arthritis and high blood pressure module, so it is not surprising that its don't know rate is one of the highest rates.

5.2.1 Proxy vs. non-proxy

The don't know rate based on the entire questionnaire is 63% higher for the proxy respondents than the non-proxy respondents. The proxy group is made up of 1,257 respondents, while the non-proxy group has 11,318 respondents. Table 5.8 displays the modules with the ten highest don't know rates by interview type. Five modules or submodules are among the ten highest don't know rates for both types: insurance, income, weeks worked, and two chronic conditions submodules (heart disease, arthritis/rheumatism).

Most of the modules and submodules have higher don't know rates for proxy respondents than for non-proxy respondents. The weeks worked submodule shows the biggest gap at with a difference of 2.02 percentage points.

Table 5.9 lists the variables with the ten highest don't know rates by proxy and non-proxy interview. On the proxy side, two variables from the income module have the highest don't know rates. For non-proxy interviews, all top 10 variables but one come from the chronic conditions module.

Table 5.8 – Modules With The Highest Don't Know Rates By Interview Type

Proxy		Non-Proxy	
Module or submodule	Don't Know Rate (in %)	Module or submodule	Don't Know Rate (in %)
Overall	0.44	Overall	0.27
Weeks worked	2.44	Insurance (ISC)	1.92
Absence/Hours	2.06	Income (INC)	1.85
Looking for work	1.69	Heart disease	1.21
Heart disease	1.57	CC_- chronic conditions focus content	1.02
Insurance (ISC)	1.56	Arthritis/Rheumatism	0.87
Income (INC)	1.53	Other job	0.82
Job description	1.46	High blood pressure	0.81
Arthritis/Rheumatism	1.46	Mastery*	0.46
Smoking (SMC)	1.31	Stomach ulcers	0.42
Cataracts	1.25	Weeks worked	0.42

*Not asked in proxy interviews

When looking more specifically at the variables with these high rates, eight of them involve the month or year that a particular condition disappeared, which is understandable. Note that this type of question is not asked of proxy respondents. Finally, the variable CC_0_D1 (kind of arthritis/rheumatism) is one of only two variables among the ten highest rates for both proxy and non-proxy interviews, the other being total household income.

Table 5.9 – Variables with the Highest Don't Know Rates by Interview Type

Proxy				Non-Proxy			
Variable	Concept	Sample Size	Don't Know Rate (in %)	Variable	Concept	Sample size	Don't Know Rate (in %)
INC0_4	Total pers. inc. - best estimate	288	15.63	CCC0_O2M	Stroke - month cond. Disappeared	17	29.41
INC0_3	Total hhld inc. - best estimate	1,233	15.57	CCC0_J2M	Diabetes - month cond. Disappeared	7	28.57
SMC0_5C	No. days smoked >=1 cig. - occ. smoker	7	14.29	CCC0_C2M	Asthma - month cond. Disappeared	96	20.83
CC_0_D1	Arthr./rheum. - kind	77	11.69	CCC0_N2M	Ulcers - month cond. Disappeared	103	19.42
LSC0_43	Preferred no. hours - main job/bus.	68	10.29	INC0_3	Total hhld inc. - best estimate	10,993	17.89
HSC0_7	Hearing - in group with hearing aid	84	9.52	CCC0_F2M	High B.P. - month cond. Disappeared	134	16.42
CCC0_D3	Arthr./rheum. - age first diagnosed	77	9.09	CCC0_G2M	Migraines - month cond. Disappeared	171	16.37
CCC0_N3	Ulcers - age first diagnosed	12	8.33	CCC0_O2Y	Stroke - year cond. Disappeared	17	11.76
CCC0_S3	Cataracts – age first diagnosed	27	7.41	CCC0_N2Y	Ulcers - year cond. Disappeared	103	11.65
SMC0_6	Age started smoking - former	101	6.93	CC_0_D1	Arthr./rheum. – kind	1998	11.41

5.2.1 Sex

Table 5.10 displays the modules with the ten highest don't know rates by sex. The differences between the sexes are more pronounced when analyzing the don't know rates than the refusal rates, despite the fact that the don't know rate for the entire questionnaire is identical for both sexes at 0.28%. Of the modules and submodules with the ten highest rates for each sex, seven are the same for both sexes, although not in the same order. The heart disease submodule shows the biggest difference in don't know rates between the sexes, with females having triple the rate of males.

Table 5.10 – Modules with the Highest Don't Know Rate by Sex

Males		Females	
Module or submodule	Don't know rate (in %)	Module or submodule	Don't know rate (in %)
Overall	0.28	Overall	0.28
Insurance (ISC)	2.21	Income (INC)	1.94
Income (INC)	1.67	Heart Disease	1.89
CC_ - chronic conditions focus content	1.17	Insurance (ISC)	1.61
Arthritis/Rheumatism	0.98	CC_ - chronic conditions focus content	0.94
High blood pressure	0.87	Arthritis/Rheumatism	0.84
Other job	0.84	High blood pressure	0.79
Heart Disease	0.62	Other job	0.78
Looking for work	0.52	Weeks worked	0.50
Asthma	0.49	Mastery	0.47
Stomach Ulcers	0.47	Women's health (WHC)	0.40

Table 5.11 shows the modules with the ten highest don't know rates specifically for proxy interviews of both sexes. Here again, the overall rates are similar between both sexes. Five modules or submodules (insurance, smoking, arthritis/rheumatism, absence/hours, and looking for work) are among the ten highest rates for both sexes. Note that the sex reported in the table is the sex of the panel member, and not that of the proxy respondent.

Table 5.11 – Modules with the Highest Don't Know Rates by Sex for Proxy Interviews

Proxy interviews only			
Males		Females	
Module or submodule	Don't Know Rate (in %)	Module or submodule	Don't Know Rate (in %)
Overall	0.45	Overall	0.42
Weeks worked	3.19	Absence/Hours	4.07
Alzh./Dementia	1.88	Heart disease	3.88
Income (INC)	1.68	Looking for work	2.22
Insurance (ISC)	1.60	Cataracts	1.77
Arthritis/Rheumatism	1.50	Smoking (SMC)	1.63
Job description	1.49	Insurance (ISC)	1.52
Absence/Hours	1.38	Alcohol (ALC)	1.42
Looking for work	1.37	Arthritis/Rheumatism	1.42
Back problems	1.31	High blood pressure	1.42
Smoking (SMC)	1.16	CC_ - chronic conditions focus content	1.40

Table 5.12 shows more results, but at the variable level. The table presents results for both proxy and non-proxy interviews by sex. The only variable found among the top 10 highest rates for all four proxy/sex groups is INC0_3 (total household income) with rates ranging between 14.00% and 19.93%. Although the rates are high, most of the other variables in the top 10 are based on a relatively small number of respondents. For both male and female non-proxy groups, all variables but one (total hhld income) are from the chronic conditions module, and the majority of them are related to the month or year when the condition disappeared.

Aside from INC0_3, the only common variable for both male groups is CC_0_D1 (kind of arthritis/rheumatism), with similar don't know rates near 14%. The female don't know rates for this variable are lower, but still quite high, near 10%. Finally, LSC0_43 (preferred number of hours) has a much higher proxy rate with a combined don't know rate of 10.29% for both sexes, compared to 0.52% for the non-proxy respondents.

5.2.2 Age groups

The don't know rates for the entire questionnaire vary from 0.16% for the 25-44 age group to 0.59% for the 12-24 age group. There were 819 under-12-year-old proxy interviews conducted, while the number of proxy interviews for the other three groups varies between 51 (25-44) and 172 (12-24). Because of these small numbers, a single response of Don't Know in the proxy group influences the percentage of don't knows much more than in non-proxy groups.

Table 5.13 shows the modules with the highest don't know rates by age group and type of interview. Note that there are no non-proxy results reported for the younger than 12 age group. Interviews for this group are supposed to be done by proxy, however 14 interviews were non-proxy, and have not been included in this analysis.

The number of chronic condition submodules among the ten highest don't know rates for an age/proxy group increases with the older age groups, while the number of labour force submodules among the ten highest rates decreases with the older age groups.

Table 5.12 – Variables with the Highest Don't Know Rates by Sex and Interview Type

Proxy							
Males				Females			
Variable	Concept	Sample Size	Don't Know Rate (in %)	Variable	Concept	Sample Size	Don't Know Rate (in %)
INC0_3	Total hhld inc. - best estimate	683	16.84	LSC0_43	Preferred no. hours - main job/bus.	20	20.00
INC0_3C	Total hhld inc. - <$5,000 or >=$5,000	6	16.67	SMC0_6	Age started smoking daily – former	22	18.18
SMC0_5C	No. days smoked >=1 cig. - occ. Smoker	6	16.67	SMC0_4	No. cig. smoked each day - daily smoker	12	16.67
INC0_4	Total pers. inc. - best estimate	188	15.43	INC0_4	Total pers. inc. - best estimate	100	16.00
CC_0_D1	Arthr./rheum. - kind	35	14.29	CCC0_H3	Chronic bronchitis - age first diagnosed	7	14.29
CCC0_D3	Arthr./rheum. - age first diagnosed	35	8.57	INC0_3	Total hhld inc. - best estimate	550	14.00
CCC0_E3	Back problems - age first diagnosed	25	8.00	SMC0_7	No. of cig. daily - former daily smoker	22	13.64
HSC0_7	Hearing - in group with hearing aid	53	7.55	CCC0_S3	Cataracts - age first diagnosed	15	13.33
LSC0_43	Preferred no. hours - main job/bus.	48	6.25	HSC0_7	Hearing – in group with hearing aid	31	12.90
CCC0_R3	Alzh./other dem. - age first diagnosed	17	5.88	CCC0_N3	Ulcers – age first diagnosed	8	12.50
Non-proxy							
Males				Females			
Variable	Concept	Sample Size	Don't Know Rate	Variable	Concept	Sample Size	Don't Know Rate
CCC0_N2M	Ulcers - month cond. Disappeared	51	25.49	CCC0_J2M	Diabetes - month cond. disappeared	5	40.00
CCC0_O2M	Stroke - month cond. Disappeared	4	25.00	CCC0_O2M	Stroke – month cond. disappeared	13	30.77
CCC0_C2M	Asthma - month cond. Disappeared	40	20.00	CCC0_C2M	Asthma – month cond. disappeared	56	21.43
CCC0_F2M	High B.P. - month cond. Disappeared	60	18.33	INC0_3	Total hhld inc. - best estimate	6,176	19.93
INC0_3	Total hhld inc. - best estimate	4,817	15.28	CCC0_G2M	Migraines - month cond. disappeared	124	17.74
CCC0_N2Y	Ulcers - year cond. Disappeared	51	13.73	CCC0_O2Y	Stroke – year cond. disappeared	13	15.38
CC_0_D1	Arthr./rheum. - kind	655	13.44	CCC0_F2M	High B.P. - month cond. disappeared	74	14.86
CCC0_G2M	Migraines - month cond. Disappeared	47	12.77	CCC0_N2M	Ulcers – month cond. disappeared	52	13.46
CCC0_O1	Stroke - condition disappeared	8	12.50	CC_0_L1D	Heart disease – age of most recent attack	48	12.50
CCC0_F2Y	High B.P. - year cond. Disappeared	60	11.67	CC_0_D1	Arthr./rheum - kind	1,343	10.42

The income module is among the ten highest rates for all 9 age/proxy groups. The don't know rate for the income module for the proxy group is higher than that of the non-proxy group for the age groups 25-44, 45-64 and 65+. The non-proxy rate is higher for the 12-24 group, due mostly to the fact that the variable INC0_3 (total household income) has a rate of 41.33%. The insurance module is also among the highest rates for all age/proxy groups, with the 12-24 non-proxy group having the highest rate at 7.03%. In fact,

the four variables that make up the insurance module have rates that vary from 5.59% to 13.59% for the 12-24 non-proxy group. The insurance and income modules have the highest rates for the under 12 group at 1.08% and 0.96%, which are clearly higher than all other modules that have rates below 0.25% for this group.

The self-esteem and mastery submodules, which are only asked in non-proxy interviews, have higher don't know rates in the 65+ age group. Two other submodules from the stress section (recent life events and ongoing problems – not shown in table) also have rates that are about three times higher than those of the other age groups.

Several labour force submodules are among the ten highest rates for the various age/proxy groups. With a rate of 1.26%, the entire labour force module has a much higher rate for the 12-24 non-proxy group, compared to the next highest rate for this module among the other groups at 0.19%.

For most chronic condition submodules, the older the age group, the higher the don't know rate. The chronic conditions focus content is among the ten highest rates of all four non-proxy groups, as well as the proxy group for the 45-64 group. The rate for the Alzheimer's/dementia submodule is strongly correlated to age, at 0% for all age groups with the exception of the 65+ group, which has a proxy don't know rate of 2.13%, putting it among the ten highest. The heart disease submodule shows a similar pattern with low rates for all groups except 65+, which has a proxy don't know rate of 1.75% and a non-proxy don't know rate of 1.83%.

Finally, it is interesting to note that the alcohol module has don't know rates for the 12-24 and 25-44 proxy groups that are twenty times higher than their non-proxy counterparts. The smoking module has higher proxy don't know rates that are three to six times higher than non-proxy don't know rates, varying from 0.79% to 1.83%, compared to 0.22% to 0.55% for the non-proxy interviews. For all of the questions in the drug module that ask if a respondent took a particular medication in the past month (DGC0_1A to DGC0_1V), the 65+ proxy group has a much higher rate than the other groups.

Table 5.13 – Modules with the Highest Don't Know Rates by Age Group and Interview Type

Under 12	
Proxy	
Module or submodule	Don't Know Rate (in %)
Overall	0.23
Insurance (ISC)	1.08
Income (INC)	0.96
Other Allergies	0.24
Vision	0.23
Socio-demographic (SDC)	0.20
Chronic Bronchitis	0.12
Food Allergies	0.12
Cycle 1 Variables (birth weight, immigration variables, date of birth)	0.11
CCC – chronic conditions core content	0.05
CCC2 – all chronic conditions content	0.05

12-24			
Proxy		**Non-Proxy**	
Module or submodule	Don't Know Rate (in %)	Module or submodule	Don't Know Rate (in %)
Overall	0.39	Overall	0.60
Weeks worked	4.76	Insurance (ISC)	7.03
Looking for work	3.57	Income (INC)	6.53
Job description	2.86	Blood pressure (BPC)*	1.04
Absence/Hours	2.29	Weeks worked	0.58
Insurance (ISC)	2.00	Preventive health (BPC, WHC, GHC)*	0.53
Income (INC)	1.68	Mastery*	0.47
Alcohol (ALC)	1.65	Socio-demographic (SDC)	0.44
Past job attachment	1.56	Looking for work	0.41
Labour Force (LSC, LFC)	1.26	Asthma	0.34
Asthma	1.17	CC_ - chronic conditions focus content	0.34

*Not asked in proxy interviews

25-44			
Proxy		**Non-Proxy**	
Module or submodule	Don't Know Rate (in %)	Module or submodule	Don't Know Rate (in %)
Overall	0.59	Overall	0.16
Absence/Hours	3.23	CC_ - chronic conditions focus content	1.30
Insurance (ISC)	3.19	Other job	1.15
Alcohol (ALC)	2.31	Insurance (ISC)	1.10
Income (INC)	2.30	High blood pressure	0.86
Memory	1.96	Income (INC)	0.80
Thinking	1.96	Arthritis/Rheumatism	0.68
Food Allergies	1.72	Asthma	0.50
Hearing	1.67	Weeks worked	0.41
Job description	1.50	Looking for work	0.39
Smoking (SMC)	1.01	Repetitive Strain (RPC)	0.33

45-64			
Proxy		**Non-Proxy**	
Module or submodule	Don't Know Rate (in %)	Module or submodule	Don't Know Rate (in %)
Overall	0.30	Overall	0.17
Weeks worked	2.78	Income (INC)	0.98
Income (INC)	2.17	Other job	0.85
Back problems	1.49	CC_ - chronic conditions focus content	0.81
CC_ - chronic conditions focus content	1.27	Arthritis/Rheumatism	0.80
Insurance (ISC)	1.10	Insurance (ISC)	0.79
Smoking (SMC)	1.06	Stomach Ulcers	0.48
Arthritis/Rheumatism	0.78	Looking for work	0.47
Absence/Hours	0.73	High blood pressure	0.41
Drug Use (DGC)	0.29	Migraine	0.36
Labour Force (LSC, LFC)	0.19	Weeks worked	0.33

65+				
Proxy		**Non-Proxy**		
Module or submodule	Don't Know Rate (in %)	Module or submodule	Don't Know Rate (in %)
Overall	1.14	Overall	0.41
Income (INC)	3.54	Heart Disease	1.83
Insurance (ISC)	3.20	Income (INC)	1.78
Hearing	2.54	Mastery*	1.31
Cataracts	2.16	Arthritis/Rheumatism	1.18
Alzheimer's/Dementia	2.13	CC_ - chronic conditions focus content	1.13
Arthritis/Rheumatism	1.96	Insurance (ISC)	1.11
Drug Use (DGC)	1.94	High blood pressure	1.11
Vision	1.93	Stomach Ulcers	1.05
Feelings	1.90	Self-esteem and mastery*	1.00
Memory	1.90	Self-esteem*	0.96

*Not asked in proxy interviews

Table 5.14 shows the variables with the highest don't know rates by age group and interview type. It is no surprise that five income variables and two insurance variables are among the ten variables with the highest don't know rates for the under 12 group as these two modules have by far the highest rates for this group. For the 12-24 age group, alcohol and labour force related variables make up most of the top 10 for the proxy interviews. As for the ten highest 12-24 non-proxy don't know rates, four are income variables and two are insurance variables.

Three income variables, three alcohol variables and two labour force variables have don't know rates among the ten highest for the 25-44 proxy interviews. For the 25-44 non-proxy, the list is made up mostly of variables from the chronic condition submodules, and again often relate to the month or year that a condition disappeared. The top four are based on only a handful of respondents and are included only for completeness.

Table 5.14 – Variables with the Highest Don't Know Rates by Age Group and Interview Type

Under 12			
Proxy			
Variable	Concept	Sample Size	Don't Know Rate (in %)
INC0_3	Total hhld inc. – best estimate	812	13.79
CCC0_H3	Chronic bronchitis - age first diagnosed	9	11.11
ISC0_4	Insurance – hospital charges	819	3.79
INC0_3D	Total hhld inc. - <$15,000 or >=$15,000	79	2.53
INC0_3A	Total hhld inc. - <$20,000 or >=$20,000	812	2.09
ISC0_3	Insurance – eye glasses/contact lenses	819	1.34
INC0_3G	Total hhld inc. - >= $50,000	517	1.16
INC0_3B	Total hhld inc. - <$10,000 or >=$10,000	89	1.12
IJC0_9	Most serious injury - act. when injured	90	1.11
HWB	Birth weight	819	0.85

12-24							
Proxy				Non-proxy			
Variable	Concept	Sample Size	Don't Know Rate (in %)	Variable	Concept	Sample Size	Don't Know Rate (in %)
ALC0_5A7	Number of drinks – Saturday	6	16.67	PAC0_2G	No. times partic. - ice hockey	3	100
INC0_3	Total hhld inc. - best estimate	167	11.38	INC0_3	Total hhld inc. - best estimate	1,609	41.33
INC0_4	Total pers. inc. - best estimate	55	10.91	INC0_3A	Total hhld inc. - <$20,000 or >=$20,000	1,609	15.72
ALC0_5	Drank alcohol in past week	39	10.26	HSC0_7	Hearing – in group with hearing aid	13	15.38
SMC0_4	No. cig. smoked each day – daily smoker	11	9.09	ISC0_4	Insurance - hospital charges	1,751	13.59
ALC0_6	Regularly drank > 12 drinks a week	12	8.33	CCC0_G2M	Migraines - month cond. disappeared	22	9.09
LSC0_43	Preferred no. hours - main job/bus.	25	8.00	ISC0_3	Insurance – eye glasses/contact lenses	1,751	8.91
ALC0_3	Freq. of having 5 or more drinks	39	7.69	INC0_4	Total pers. inc. - best estimate	1,131	7.60
LSC0_71	No. of weeks looked for work past year	28	7.14	INC0_1A-N	Source of income	1,743	7.06
LSC0F33	Flag for name of employer	40	5.00	MHC0_27	Loss interest - no. weeks past 12 m.	29	6.90

25-44							
Proxy				Non-proxy			
Variable	Concept	Sample Size	Don't Know Rate (in %)	Variable	Concept	Sample Size	Don't Know Rate (in %)
INC0_3	Total hhld inc. - best estimate	48	18.75	PAC0_3Y	Time spent – in-line skate/rollerblade	1	100.00
LSC0_43	Preferred no. hours - main job/bus.	25	16.00	PAC0_2Y	No. times partic. - in-line skate./roll.	1	100.00
CCC0_A3	Food allergies - age first diagnosed	7	14.29	CCC0_O2M	Stroke – month cond. disappeared	4	50.00
ALC0_5A6	Number of drinks – Friday	16	12.50	CCC0_O2Y	Stroke – year cond. disappeared	4	25.00
INC0_4	Total pers. inc. - best estimate	42	9.52	CCC0_C2M	Asthma – month cond. disappeared	34	20.59
LSC0F32	Flag for name of own business	11	9.09	CCC0_F2M	High B.P. – month cond. disappeared	28	17.86
INC0_3A	Total hhld inc. - <$20,000 or >=$20,000	48	8.33	CCC0_G2M	Migraines – month cond. disappeared	75	14.67
ALC0_5B	Ever had a drink	14	7.14	CC_0_D1	Arthr./rheum. – kind	276	14.13
ALC0_5A1, A2, A3, A4, A5, A7	Number of drinks – Sun, Mon, Tue, Wed, Thu, Sat	16	6.25	CCC0_N2M	Ulcers – month cond. disappeared	36	13.89
ISC0_3	Insurance – eyeglasses/contact lenses	51	5.88	CCC0_C2Y	Asthma – year cond. disappeared	34	11.76

| 45-64 | | | | | | | |
| Proxy | | | | Non-proxy | | | |
Variable	Concept	Sample Size	Don't Know Rate (in %)	Variable	Concept	Sample Size	Don't Know Rate (in %)
INC0_3	Total hhld inc. – best estimate	55	14.55	PAC0_2D	No. times partic. – bicycling	5	80.00
CC_0_D1	Arthr./rheum. – kind	7	14.29	CCC0_C2M	Asthma – month cond. disappeared	21	23.81
INC0_4	Total pers. inc. – best estimate	53	13.21	CCC0_G2M	Migraines – month cond. disappeared	55	20.00
INC0_3D	Total hhld inc. - <$15,000 or >=$15,000	9	11.11	CCC0_N2M	Ulcers – month cond. disappeared	34	17.65
SMC0_7	No. of cig. daily - former daily smoker	18	11.11	CCC0_F2M	High B.P. – month cond. disappeared	44	13.64
CCC0_E3	Back problems - age first diagnosed	11	9.09	INC0_3	Total hhld inc. - best estimate	3,290	13.19
LSC0_43	Preferred no. hours – main job/bus.	17	5.88	CCC0_N2Y	Ulcers – year cond. disappeared	34	11.76
SMC0_6	Age started smoking daily – former	18	5.56	HSC0_7	Hearing – in group with hearing aid	110	10.00
SMC0_8	Age stopped smoking daily – former	18	5.56	CC_0_D1	Arthr./rheum. – kind	826	9.56
ALC0_5A5	Number of drinks – Thursday	19	5.26	CCC0_G2Y	Migraines – year cond. disappeared	55	9.09

| 65+ | | | | | | | |
| Proxy | | | | Non-proxy | | | |
Variable	Concept	Sample Size	Don't Know Rate (in %)	Variable	Concept	Sample Size	Don't Know Rate (in %)
CCC0_R3	Alzh./other dem. - age first diagnosed	1	100.00	CCC0_C2M	Asthma – month cond. disappeared	9	44.44
SMC0_5C	No. days smoked >=1 cig. - occ. smoker	3	33.33	CCC0_N2M	Ulcers – month cond. disappeared	30	30.00
INC0_3	Total hhld inc. - best estimate	151	29.14	CCC0_J2M	Diabetes - month cond. disappeared	4	25.00
INC0_3C	Total hhld inc. - <$5,000 or >=$5,000	4	25.00	CCC0_O2M	Stroke – month cond. disappeared	12	25.00
INC0_4	Total pers. inc. - best estimate	138	20.29	CCC0_C2Y	Asthma – year cond. disappeared	9	22.22
INC0_4G	Total pers. inc. - >=$50,000	5	20.00	CCC0_G2M	Migraines – month cond. disappeared	19	21.05
CCC0_N3	Ulcers - age first diagnosed	8	12.50	CCC0_N2Y	Ulcers – year cond. disappeared	30	20.00
SMC0_3	Age started smoking daily - daily smoker	16	12.50	INC0_3	Total hhld inc. – best estimate	2,068	19.39
CC_0_D1	Arthr./rheum. – kind	66	12.12	SMC0_2_5	Considers quitt. smoking - next 6 months	26	19.23
HSC0_7	Hearing – in group with hearing aid	69	11.59	CCC0_F2M	High b.p. – month condition disappeared	61	18.03

For the 45-64 proxy group, three income variables and three smoking variables are among the ten highest. The non-proxy group, like the 25-44 group, contains several chronic condition variables among the ten highest. PAC0_2D (number of times participated in bicycling) has a proxy don't know rate of 80%, but is based on only five respondents.

Finally, for the 65+ proxy group, four income, three chronic conditions and two smoking variables are among the ten highest don't know rates. As with the 25-44 and 45-64 age groups, chronic condition variables make up the majority of the ten highest non-proxy rates.

5.2.3 Other cycles

As with the refusals, variables with high don't know rates in Cycle 4 that were asked in previous cycles also have high don't know rates. For example, the majority of the variables representing the month that a particular chronic condition disappeared also have high don't know rates in Cycles 2 and 3.

Many of the Manitoba buy-in and Alberta buy-in variables (day-to-day demands) that were only part of Cycle 1 have high don't know rates. Of the content unique to Cycle 2, the sexual health variables in Alberta had high don't know rates, as well as two of the three variables in the emergency services module. Virtually all of the variables in the family medical history module, which is only part of Cycle 3, have high don't know rates.

5.3 Respondent Level Rates

Refusal and don't know rates were also calculated at the person level to determine the percentage of questions an individual refuses to answer or answers Don't Know. Table 5.15 shows the proportion of panel members who did not refuse any questions, who refused less than 1% of the questions asked, and who refused less than 3% of the questions asked. Don't know rates, which show much more variation than refusal rates when separated into sex, age groups and interview type, are also shown.

Of the 12,575 fully completes, 11,676 have refusal rates equal to 0%, meaning that 92.9% of the respondents did not refuse to answer any of the questions. Over 99% have refusal rates less than 3%. As for don't know rates, 8471 (67.4%) have don't know rates of 0%. Over 99% have don't know rates less than 5%, and 95% have don't know rates less than 1%. This shows that almost everyone who refuses or responds Don't Know does so for only a few questions.

Only 48.4% of the 12-24 group did not answer Don't Know to any questions, which is not surprising. As reported earlier, a large proportion responded Don't Know to the household income question. The under 12 and 65+ age groups tend to answer Don't Know more often than the other three age groups.

Table 5.15 – Refusal and Don't Know Rates at the Respondent Level

	Refusals to 0% of questions	Refusals to less than 1% of questions	Refusals to less than 3% of questions	Don't know to 0% of questions	Don't know to less than 1% of questions	Don't know to less than 5% of questions
Overall	92.9%	98.4%	99.1%	67.4%	94.8%	99.1%
Males	92.3%	98.2%	98.9%	68.9%	94.7%	99.1%
Females	93.3%	98.6%	99.2%	66.1%	94.8%	99.1%
Under 12	96.0%	98.2%	99.5%	80.0%	94.8%	98.8%
12-24	93.8%	98.9%	99.2%	48.4%	87.7%	97.8%
25-44	94.3%	98.7%	99.2%	75.9%	97.5%	99.7%
45-64	92.3%	98.6%	99.3%	71.3%	97.4%	99.6%
65+	89.2%	97.3%	98.5%	57.6%	91.8%	98.5%
Proxy	94.7%	98.1%	99.2%	74.6%	92.5%	98.1%
Non-Proxy	92.6%	98.5%	99.1%	66.6%	95.0%	99.2%

6. EDIT FAILURES AND DATA INCONSISTENCIES

6.1 Cycle 4 Edit Failures

Editing is the identification of inconsistencies between data items, with some sort of corrective action usually implemented at the same time. Two sets of edits are applied to the NPHS data. First, Computer Assisted Interviewing (CAI) edits are programmed into the application and are performed in the field during collection. Next, head office (HO) edits are performed at Statistics Canada after the data is received.

There are two types of CAI edits, hard edits and soft edits. Hard edits check that programming restrictions are met, such as whether or not the value falls within the minimum/maximum range or that the selected category is valid. If a hard edit fails, a message appears on the screen during collection. The interviewer must enter a valid value or category to proceed to the next question. Soft edits check to see if a response is less than the maximum allowed for that variable, but still high enough to raise suspicion. When a soft edit fails, a message appears on the screen. To continue to the next question, the interviewer must then confirm the response with the respondent and either change the value or leave it as a valid response.

HO edits are performed after the receipt of data. If a HO edit fails, the inconsistency is left as is, because data from previous cycles are never changed, and, unlike with CAI edits, the respondent is not available to verify if this cycle's data are correct. The exception to this are the relationship edits, in which inconsistencies go through a manual correction process.

CAI and HO edits can be classified into three categories: validity edits, consistency edits, and longitudinal edits. Validity edits check that only valid or allowable responses are given for each question, and that range restrictions are respected. Consistency edits check that expected or known relationships between variables are not contradicted. These edits involve two or more variables for the same respondent or more than one respondent. Longitudinal edits check that expected or known relationships between variables across cycles are not contradictory. These edits involve comparing one variable between at least two cycles to ensure that the data is consistent.

In the case of Cycle 4 data, edits were created for four different files that make up the interview: household (H3HH), members (H3MM), relationships (REL) and health (H6). The H3HH is made up of household info, such as address, phone number and dwelling type, and contains information from all 17,276 panel members. The H6 file contains information from the health component for the 13,774 panel members whose interviews were either fully or partially completed, or who were institutionalized. The H3MM contains information on each household member, such as date of birth, age, sex and marital status, for a total of 38,178 individuals. The relationship file contains the relationship of each person in the household with each of the other members in the household, for a total of 100,081 records.

Table 6.1 shows the number of edit rules for each of the four files, as well as the number of times that an edit rule failed. Table 6.2 shows the H6 counts separated by module. Soft edit failures are much more common than hard edit failures for both CAI validity edits and CAI consistency edits. Table 6.3 lists the nine HO longitudinal consistency edits, as well as the number of failures for each. These tables present only Cycle 4 data since the processing systems changed for Cycle 4, making comparisons to previous cycles difficult.

Table 6.4 shows the ten edits with the highest number of failures. The edit dealing with the consistency across cycles of the respondent's ethnic or cultural groups has by far the highest number of failures. This is due to the fact that if the respondent does not list exactly the same group(s) in each cycle, the edit fails. For example, if a respondent lists four different groups in one cycle and only three of the four in another cycle, the edit fails.

A detailed document on the editing process entitled "NPHS Cycle 4: Production Edit Specifications and Documentation" covers all of the Cycle 4 edits in depth. Readers who wish to obtain more information on edits should consult this document.

Table 6.1 – Cycle 4 Edit Rules and Failures for Each File

| | CAI Validity Edits | | | | CAI Consistency Edits | | | |
| | Hard | | Soft | | Hard | | Soft | |
	Number of edit rules	Number of times an edit rule failed	Number of edit rules	Number of times an edit rule failed	Number of edit rules	Number of times an edit rule failed	Number of edit rules	Number of times an edit rule failed
H3HH	2	0	2	8	0	-	0	-
H3MM	3	0	0	-	3	0	0	-
Rel	24	72	0	-	0	-	5	384
H6	96	1	58	4,206	62	1	22	304
Total	125	73	60	4,214	65	1	27	688

| | HO Validity Edits | | HO Consistency Edits | | HO Longitudinal Edits | | Total | |
	Number of edit rules	Number of times an edit rule failed	Number of edit rules	Number of times an edit rule failed	Number of edit rules	Number of times an edit rule failed	Number of edit rules	Number of times an edit rule failed
H3HH	1	2	0	-	0	-	5	10
H3MM	0	-	0	-	0	-	6	0
Rel	0	-	33	275	0	-	62	731
H6	27	0	17	268	9	12,061	291	16,841
Total	28	2	50	543	9	12,061	364	17,582

Table 6.2 – Cycle 4 Edit Rules and Failures in the Health Questionnaire by Module

Health Component Module	CAI Validity Edits				CAI Consistency Edits			
	Hard		Soft		Hard		Soft	
	Number of edit rules	Number of times an edit rule failed	Number of edit rules	Number of times an edit rule failed	Number of edit rules	Number of times an edit rule failed	Number of edit rules	Number of times an edit rule failed
Height/Weight	0	-	0	-	33	0	1	33
Preventive Health	1	0	0	-	0	-	0	-
Health Care Utilization	11	0	11	1,721	0	-	1	0
Restriction of Activities	0	-	0	-	2	0	0	-
Chronic Conditions	30	0	1	5	7	0	2	11
Physical Activities	23	0	23	1,512	1	0	0	-
Injuries	2	0	2	25	0	-	2	6
Stress	0	-	0	-	0	-	0	-
Drug Use	2	0	1	55	0	-	3	9
Smoking	7	1	3	46	1	1	3	0
Alcohol	7	0	7	89	7	0	0	-
Mental Health	5	0	5	83	0	-	5	47
Social Support	1	0	1	261	0	-	0	-
Socio-Demographic	0	-	0	-	0	-	0	-
Education	0	-	0	-	1	0	0	-
Labour Force	5	0	2	93	8	0	3	169
Income	2	0	2	316	2	0	2	29
Total	96	1	58	4,206	62	1	22	304

Table 6.2 — continued

	HO Validity Edits		HO Consistency Edits		HO Longitudinal Edits		Total	
	Number of edit rules	Number of times an edit rule failed	Number of edit rules	Number of times an edit rule failed	Number of edit rules	Number of times an edit rule failed	Number of edit rules	Number of times an edit rule failed
Height/Weight	0	-	0	-	0	-	34	33
Preventive Health	0	-	0	-	0	-	1	0
Health Care Utilization	0	-	0	-	0	-	23	1,721
Restriction of Activities	0	-	0	-	0	-	2	0
Chronic Conditions	0	-	7	0	0	-	47	16
Physical Activities	3	0	0	-	0	-	50	1,512
Injuries	0	-	4	0	0	-	10	31
Stress	0	-	1	5	0	-	1	5
Drug Use	24	0	1	0	0	-	31	64
Smoking	0	-	0	1	1	8	15	57
Alcohol	0	-	1	43	2	556	24	688
Mental Health	0	-	0	-	0	-	15	130
Social Support	0	-	0	-	0	-	2	261
Socio-Demographic	0	-	0	-	2	10,890	2	10,890
Education	0	-	0	-	4	607	5	607
Labour Force	0	-	0	-	0	-	18	262
Income	0	-	3	219	0	-	11	564
Total	27	0	17	268	9	12,061	291	16,841

Table 6.3 – Cycle 4 HO Longitudinal Edits

Edit rule	Number of failures
Data on whether the respondent ever smoked is not consistent across all cycles.	8
Data on whether the respondent ever had a drink is not consistent across all cycles.	455
Data on whether the respondent ever drank regularly is not consistent across all cycles.	101
Data on the ethnic or cultural groups to which the respondent's ancestors belonged is not consistent across all cycles.	10,304
Data on the respondent's race or colour is not consistent across all cycles.	586
Number of years of schooling is less than what was reported in a previous cycle.	85
Respondent has not graduated from high school but he/she graduated in a previous cycle.	3
Respondent has never attended a post-secondary school but he/she attended in a previous cycle.	10
Highest level of education ever attained is less than reported in a previous cycle.	509

Table 6.4 – Cycle 4 Edits With the Ten Highest Number of Failures

Type of edit	Edit rule	Number of failures
HO Longitudinal	Data on the ethnic or cultural groups to which the respondent's ancestors belonged is not consistent across all cycles.	10,304
HO Longitudinal	Data on the respondent's race or colour is not consistent across all cycles.	586
CAI Validity Soft	Number of times the respondent did home exercises (PA_Q2F) is greater than 60.	562
HO Longitudinal	Highest level of education ever attained is less than what was reported in a previous cycle.	509
HO Longitudinal	Data on whether the respondent ever had a drink is not consistent across all cycles.	455
CAI Validity Soft	Number of times the respondent has seen or talked to a dentist or orthodontist in the past 12 months (HC_Q02E) is greater than 4.	434
CAI Validity Soft	Number of times the respondent has seen or talked to a family doctor or general practitioner in the past 12 months (HC_Q02A) is greater than 12.	407
CAI Validity Soft	Number of times the respondent gardened or did yard work (PA_Q2B) is greater than 60.	276
CAI Validity Soft	Number of times the respondent has seen or talked to any other medical doctor in the past 12 months (HC_Q02C) is greater than 15.	269
CAI Validity Soft	Number of close friends and close relatives (SS_Q01) is greater than 20.	261

6.2 Data Inconsistencies

While some variables have edit rules to flag inconsistencies, other variables do not. This section discusses inconsistencies between cycles, some of which are identified by questions that flag inconsistencies to the respondent during the interview, others by comparing variables between cycles once the data is back at Head Office.

A study in an issue of Health Reports (Shields) described inconsistencies that occurred when people who reported having a particular condition (having not reported having it in the previous cycle) were asked when the condition was first diagnosed. Some people reported diagnosis dates from before the previous interview, leading to many inconsistencies. In this section, the reverse is examined - that is, when a particular condition was present in an earlier cycle and is no longer present in the current cycle.

Eight of the chronic conditions have questions programmed into the application that point out inconsistencies to the respondent. If the respondent says that they do not have the condition, but reported in a previous cycle that they did, they are reminded that they reported the presence of the condition during the previous interview and are asked if it has disappeared since then. This question can only be asked of non-proxy respondents, except when the panel member is a child. This is done to ensure the confidentiality of the panel member's data provided in a previous cycle; asking somebody who is providing a proxy response if a condition has disappeared would reveal that the panel member reported the condition in the previous cycle. Several respondents reported that they had never had the condition. Most respondents who report having a condition in one cycle and report that they never had it in a subsequent cycle have given these inconsistent responses personally; at least 70% were non-proxy in the previous cycle. Table 6.5 shows the number of cases where the respondent reported never having the condition, as well as the proportion of these cases that were proxy interviews in the previous cycle.

Some of the chronic conditions do not have checks programmed into the application to see if a reported condition has disappeared. For these conditions, frequencies of those that said they did not have a particular condition but did have it in a previous cycle were computed by proxy type for both cycles. Table 6.6 displays these frequencies. The number of panel members asked each question is slightly lower than the maximum number of interviews shown in the table as some partial respondents stopped answering partway through. Also, some questions are only asked if the respondent is over a certain age, or are not asked if the respondent is part of the institutions component. It should be noted that these inconsistencies could very well be valid (i.e. the condition has really disappeared) but it is also likely that some of them are true inconsistencies, as with the chronic conditions examined earlier.

Three smoking questions identify inconsistent responses when it comes to identifying a respondent as a smoker or a non-smoker. Respondents who were smokers in a previous cycle and report that they do not smoke in the current cycle are asked why they quit (SMCn_9). A few respondents report that they have never smoked, even though it was reported in a previous cycle that they did. Respondents who reported being non-smokers in the previous cycle and now report being smokers are asked why they started to smoke (SMCn_10). Several respondents asked this question reported that they smoked at the time of the last interview. Smokers who reported smoking daily in a previous cycle and now report smoking occasionally are asked why they cut down (SMCn_11). One person who reported being a daily smoker in a previous cycle and an occasional smoker in the current cycle responded that they did not smoke at the time of the last interview. Table 6.7 shows the number of inconsistent responses for these three questions.

Table 6.5 – Number of Respondents Reporting That They Never Had the Condition

"During our last interview in (month and year), it was reported that (person) had (condition), but this time it was not. Has the condition disappeared since then?" (Possible responses: Yes, No, Never had, DK, Ref)					
Condition		Number of respondents asked the question	Respondents who said that they never had the condition.		
			Total	Proportion with proxy interview in previous cycle	Proportion with non-proxy interview in previous cycle
Asthma	C2	91	30 (33%)	20%	80%
	C3	100	36 (36%)	6%	94%
	C4	174	49 (28%)	20%	80%
Arthritis/rheumatism	C2	222	122 (55%)	18%	82%
	C3	236	123 (52%)	12%	88%
	C4	not asked	not asked	not asked	not asked
High blood pressure	C2	156	27 (17%)	30%	70%
	C3	170	25 (15%)	16%	84%
	C4	172	31 (18%)	13%	87%
Migraines	C2	255	46 (18%)	26%	74%
	C3	214	41 (19%)	12%	88%
	C4	228	39 (17%)	3%	97%
Diabetes	C2	31	11 (36%)	27%	73%
	C3	24	9 (38%)	0%	100%
	C4	26	18 (69%)	6%	94%
Epilepsy	C2	15	5 (33%)	20%	80%
	C3	10	7 (70%)	30%	70%
	C4	8	4 (50%)	25%	75%
Stomach/intestinal ulcers	C2	185	59 (32%)	27%	73%
	C3	151	29 (19%)	10%	90%
	C4	152	37 (24%)	0%	100%
Effects of a Stroke	C2	31	16 (52%)	6%	94%
	C3	31	17 (55%)	12%	88%
	C4	28	10 (36%)	0%	100%

Table 6.6 – Chronic Condition Inconsistencies

		Previous = proxy Current = proxy	Previous = non-proxy Current = proxy	Previous = proxy Current = non-proxy	Previous = non-proxy Current = non-proxy	Total
Number of people (0+) interviewed	C2	3,319	824	2,871	8,867	15,881
	C3	2,307	903	1,554	9,773	14,537
	C4	1,165	248	1,775	10,132	13,320
Chronic condition inconsistencies						
Food allergy (0+)	C2	34	20	76	232	362
	C3	48	27	39	303	417
	C4	23	3	50	324	400
Other allergy (0+)	C2	75	69	143	491	778
	C3	127	85	91	791	1094
	C4	63	21	127	841	1052
Back problems (12+)	C2	79	53	165	672	969
	C3	46	77	94	732	949
	C4	14	19	67	744	844
Chronic bronchitis or Emphysema (0+)	C2	55	11	38	191	295
	C3	41	16	17	165	239
	C4	24	3	19	152	198
Sinusitis (12+)	C2	10	28	37	346	421
	C3	25	8	32	151	216
	C4	not asked in C4	not asked in C4	not asked in C4	not asked in C4	-
Heart Disease (0+)	C2	25	8	32	151	216
	C3	22	17	16	152	207
	C4	6	8	18	143	175
Cancer (12+)	C2	9	5	19	105	138
	C3	1	6	9	88	104
	C4	6	9	7	95	117
Urinary Incontinence (12+)	C2	4	4	9	99	116
	C3	4	14	11	119	148
	C4	5	7	7	158	125
Bowel disorder (12+)	C2	not asked in C1	not asked in C1	not asked in C1	not asked in C1	-
	C3	4	8	5	109	126
	C4	6	1	9	109	20
Alzheimer's or other dementia (18+)	C2	2	0	1	3	6
	C3	10	3	4	15	32
	C4	3	2	2	13	244
Cataracts (18+)	C2	6	12	17	163	198
	C3	12	15	16	195	238
	C4	14	18	10	202	69
Glaucoma (18+)	C2	3	2	7	42	54
	C3	4	2	4	50	60
	C4	4	6	4	55	116
Thyroid condition (12+)	C2	not asked in C1	not asked in C1	not asked in C1	not asked in C1	-
	C3	4	9	6	106	125
	C4	2	3	4	107	641

Table 6.7 – Inconsistencies for Smoking Questions

	Proxy in Previous Cycle	Non-Proxy in Previous Cycle
Compared to our interview in (month and year) you are reporting that you no longer smoke. Why did you quit?		
	Respondent says never smoked	Respondent says never smoked
Cycle 2	0	3
Cycle 3	0	2
Cycle 4	0	6
Compared to our interview in (month and year) you are reporting that you currently smoke. Why did you start smoking?		
	Respondent says smoked at last interview	Respondent says smoked at last interview
Cycle 2	43	59
Cycle 3	16	79
Cycle 4	6	87
Compared to our interview in (month and year), you are reporting that you smoke less. Why did you cut down?		
	Respondent says never smoked	Respondent says never smoked
Cycle 2	0	1
Cycle 3	0	0
Cycle 4	0	0

Respondents who say that they presently do not smoke cigarettes at all are asked if they have ever smoked cigarettes. In Cycles 3 and 4, if they say that they have never smoked but the data shows that they have reported smoking in a previous cycle, they are reminded that they stated in a previous interview that they had previously smoked, and asked again if they have ever smoked. This probing only occurs for non-proxy interviews. In Cycle 3, 37.6% of these respondents then reply that yes, they have previously smoked, as did 34.0% of the Cycle 4 respondents. This shows that probing the respondent does have an effect on the consistency of the data. This probe question did not exist in Cycle 2, however a comparison between the Cycle 1 and 2 questions asking if the respondent ever smoked shows that there are a few inconsistencies. Table 6.8 shows the frequencies for the Cycle 3 and 4 "ever smoked" probe questions, while Table 6.9 shows the number of respondents who in Cycle 2 said that they had never smoked, but said that they had in Cycle 1.

There is a longitudinal edit, mentioned in the previous subsection, that identifies if the data on whether or not the respondent ever smoked are not consistent across all cycles. This edit fails if a respondent answers no to both the "ever smoke" and "ever smoke" (probe) questions, but in a previous cycle was a daily smoker who smoked at least one cigarette a day. This different definition is why the longitudinal edit has only 8 failures for Cycle 4, compared to the 200 inconsistencies for the "ever smoke" (probe) question.

Women above a certain age are asked if they have ever had a PAP smear or a mammogram, and both sexes are asked if they have ever had their blood pressure taken. These three items have probe questions in Cycle 3 similar to the "ever smoke" probe questions, that is, those that say no but said yes in a previous cycle are reminded of that fact and asked again. As with the smoking question, many respondents change their answer to yes, especially for the blood pressure question where 60.7% of the respondents change their answer from no to yes. The frequencies can be seen in Table 6.8. Since there are no probe questions for these variables in Cycle 2, a comparison between the Cycle 1 and 2 responses similar to that done for the "ever smoked" question in Cycle 2 can be seen in Table 6.9. There are no probe questions for these variables in Cycle 4, since if they said yes in a previous cycle, yes was pre-filled by the application.

Just as there is a question asking if the respondent has ever smoked, there is one asking if the respondent has ever smoked daily. As there are no probe questions for smoking daily, the number of people who say no in one cycle and yes in the previous cycle was calculated. These individuals were separated into four groups based on their proxy status from two cycles. Almost all respondents who said that they have not smoked daily but said yes in the previous cycle were non-proxy interviews in both cycles. Table 6.9 shows

the number of respondents who said no in a cycle despite saying yes in the previous cycle. It should be noted that Cycle 4 is treated differently than Cycles 2 and 3 in that if a respondent said yes in a previous cycle, yes is filled in by the application. The three people who said yes in Cycle 3 and no in Cycle 4 are cases who were interviewed by the Institution component, where the yes response was not pre-filled as it is in the household component.

Table 6.8 – Responses to the Probe Questions

	Yes			No			Total
	Proxy previous cycle	Non-proxy previous cycle	Total	Proxy previous cycle	Non-proxy previous cycle	Total	
Cycle 3 - During our last interview in (month and year) we recorded that you had previously smoked BUT this time we did not. In fact, have you EVER smoked cigarettes?	1	81	82 (37.6%)	0	136	136 (62.4%)	218
Cycle 4 - During our last interview in (month and year) we recorded that you had previously smoked BUT this time we did not. In fact, have you EVER smoked cigarettes?	1	102	103 (34.0%)	2	198	200 (66.0%)	303
Cycle 3 – During our last interview in (month and year), we recorded that you had previously had your blood pressure taken BUT this time we did not. In fact, have you EVER had your blood pressure taken?	0	82	82 (60.7%)	0	53	53 (39.3%)	135
Cycle 3 - During our last interview in (month and year), we recorded that you had previously had a PAP smear BUT this time we did not. In fact, have you EVER had a PAP smear?	0	38	38 (38.8%)	0	60	60 (61.2%)	98
Cycle 3 - During our last interview in (month and year), we recorded that you had previously had a mammogram BUT this time we did not. In fact, have you EVER had a mammogram?	0	20	20 (27.8%)	0	52	52 (72.2%)	72

Table 6.9 – Inconsistent Data by Proxy Status

		Previous = proxy Current = proxy	Previous = non-proxy Current = proxy	Previous = proxy Current = non-proxy	Previous = non-proxy Current = non-proxy	Total
Ever smoked - C2		1	6	10	523	540
Ever smoked daily	C2	2	4	7	161	174
	C3	0	2	2	199	203
	C4	0	2	0	1	3
Ever had blood pressure taken - C2		0	0	0	407	407
Ever had PAP smear - C2		n/a	n/a	n/a	240	240
Ever had mammogram – C2		n/a	n/a	n/a	194	194

Table 6.10 – Restriction of Activities Inconsistencies

		Previous = proxy Current = proxy	Previous = non-proxy Current = proxy	Previous = proxy Current = non-proxy	Previous = non-proxy Current = non-proxy	Total
Restriction/dis-ability in previous cycle, but now report not having one	C2	6	7	32	79	124
	C3	5	7	11	57	80
	C4	0	0	11	34	45
Restriction/dis-ability in previous cycle, but now report not having one	C2	2	1	7	25	35
	C3	0	1	2	29	32
	C4	3	0	5	34	42
No restr./disab. in previous cycle, but now report having one and that it is the same as before	C2	6	8	20	70	104
	C3	8	6	15	83	112
	C4	11	4	17	108	143

If a respondent has no activity restriction or disability, but had one in the previous cycle, they are asked to what this change is due (RACn_2B). Some respondents said that they never had the restriction, or that it was not present at the time of the last interview. As with the other questions compared for inconsistencies, most respondents were non-proxy interviews in both cycles. Respondents who did not have a restriction or disability in the previous cycle but have one now are asked if it is due to a new condition, a worsening of a condition, or the same as before (RACn_2A). Several respondents say that it is the same restriction or disability as before, even though a restriction had not been reported in the previous cycle. Table 6.10 shows the distribution of both of the restriction of activities inconsistencies by proxy type.

7. SUMMARY OF FINDINGS AND RECOMMENDATIONS

The various data quality indicators that were presented in this document help us to understand what is necessary to successfully complete an interview for the NPHS. They show areas in which we are successful, as well as areas where there is room for improvement. Here are some of the highlights and recommendations for each topic covered.

Tracing
- Untraced individuals are most likely to be in the 12-44 age group. Males are slightly more likely to be untraced than females.
- It is good to note that even for records untraced in both Cycles 2 and 3, interviewers successfully traced 31% of them in Cycle 4.
⇒ Tracing attempts are worthwhile and should continue.

Contacts and Length of Interviews
- 90% of fully completed interviews are completed in 45 minutes or less, the target time for NPHS.
- The median number of attempted contacts for fully completed interviews is 4.
⇒ These observed results are acceptable.

Refusal Conversion
- The refusal conversion rate between cycles is roughly 30%.
⇒ Refusal conversion attempts should continue, even after a series of refusals.

Item Refusal and Don't Knows
- The income module had a high refusal rate, while the refusal rates for other modules are negligible.
- The income and insurance modules had substantially higher don't know rates than the other modules.
- As expected, interviews that are done by proxy tend to have more don't knows than non-proxy interviews.
- The questionnaire appears to be quite clear for respondents since there is a low amount of don't knows.
- Respondents aged 12-24 have exceptionally high don't know rates for income (7.03%) and insurance (6.53%) modules.
⇒ Existing procedures should be improved in order to reduce the high don't know rates for the income and insurance modules, particularly for the 12-24 age group.

Edits and Inconsistencies
- Almost two-thirds of all cases fail the edit that identifies if the respondent's ancestors' ethnic or cultural group(s) are not consistent across all cycles.
- For variables without prompts, it is impossible to verify if there has been an actual change between cycles, or if an incorrect response was provided (either in the current cycle or a previous cycle)
⇒ More prompts confirming the respondent's responses in the current cycle compared to their responses from previous cycles would eliminate some discrepancies and edit failures. The downside is that the programming of the application more complicated. The literature also shows that this could decrease the amount of change observed as some respondents may choose to stick with the response given in a previous cycle, even when there is actual change.

8. FUTURE STUDIES

Other aspects of data quality involving the profiles and characteristics of certain panel members are also of interest. Some internal studies involving data quality have been conducted. These include the profiles of nonrespondents, refusals, individuals who are unable to be traced, proxy respondents, those who are unwilling to share their data and those unwilling to have their data linked to other files. While some of these profiles, such as the profiles of nonrespondents at the time of weight creation in order to adjust for nonresponse have already been investigated, more formal studies will be completed in the future.

9. ACKNOWLDEGEMENTS

10. REFERENCES

Brisebois, F., Dufour J., Lévesque, I. (1998). *Nouveaux indicateurs de qualité à l'enquête sur la population active*, Statistics Canada, Methodology Branch Working Paper, HSMD-98-004 F.

Cycle 4 (2000-2001) Household Component Longitudinal Documentation. Health Division, Statistics Canada, unpublished, May 2002.

NPHS Cycle 4: Production Edit Specifications and Documentation. Health Divisions, Statistics Canada, unpublished, June2002.

Shields, M. "Enquête nationale sur la santé de la population – Déclaration par proxy." *Rapports sur la santé, vol. 12, n°1.* Statistics Canada, October 2002.

Statistics Canada Quality Guidelines (Third Edition). Statistics Canada, October 1998. http://dissemination.statcan.ca/english/freepub/12-539-XIE/free.htm