Microdata User Guide

NATIONAL LONGITUDINAL SURVEY OF CHILDREN AND YOUTH

Cycle 4

September 2000 to May 2001

Statistics Canada Statistique Canada

Canada

Table of Contents

1.0 Introduction

The National Longitudinal Survey of Children and Youth (NLSCY) Cycle 4 was conducted from September 2000 to May 2001 by Statistics Canada in partnership with Human Resources Development Canada.

This manual has been produced to facilitate the manipulation of the microdata file of the survey results and to document data quality and other analytical issues regarding the NLSCY.

Any questions about the data set or its use should be directed to:

Statistics Canada

Client Services
Special Surveys Division
Telephone: (613) 951-3321 or toll free: 1-800-461-9050
Fax: (613) 951-4527
Email: ssd@statcan.ca

Special Surveys Division, Statistics Canada
2500 - R, Main Building
Tunney's Pasture
Ottawa, Ontario K1A 0T6

Human Resources Development Canada

Child, Youth and Social Development Studies
Applied Research Branch
Human Resources Development Canada
Place du Portage - Phase II, 7th floor
165 Hôtel de Ville Street
Hull, Quebec
K1A OJ2

Telephone: (819) 953-3465
Facsimile: (819) 953-8868

2.0 Background

The National Longitudinal Survey of Children and Youth (NLSCY) is a long-term study of Canadian children that follows their development and well-being from birth to early adulthood. The NLSCY began in 1994 and is jointly conducted by Statistics Canada and Human Resources Development Canada.

The study is designed to collect information about factors influencing a child's social, emotional and behavioural development and to monitor the impact of these factors on the child's development over time.

The survey covers a comprehensive range of topics including the health of children, information on their physical development, learning and behaviour as well as data on their social environment (family, friends, schools and communities).

Information from the NLSCY is being used by a variety of people at all levels of government, in universities, and policy-making organizations.

Survey Population

In Cycle 4, a representative sample of Canadian children aged 0 to 17 years was followed for longitudinal and cross sectional purposes.

Target population
The NLSCY objectives are to produce longitudinal and cross sectional estimates as well. Therefore, several populations are targeted in the Cycle 4 sample.

- Cross-sectionally, the Cycle 4 sample represents all children who were 0 to 17 years old on January 1st, 2001.
- Longitudinally, we have 3 cohorts:
  - The first cohort represents all children who were 0-11 years old in 1994-1995. Those children are now 6-17 years old in Cycle 4.
  - The second cohort represents all children who were 0-1 years old in 1996-1997. Those children are now 4-5 years old in Cycle 4.
  - The third cohort represents all children who were 0-1 years old in 1998-1999. Those children are now 2-3 years old in Cycle 4.

Collection Cycles
Data collection occurs at two-year intervals.

Cycle	Collection Start	Collection End
1	December 1994	April 1995
2	December 1996	April 1997
3	October 1998	June 1999
4	September 2000	May 2001

3.0 Objectives

The objectives of the NLSCY are:

- To determine the prevalence of various risk and protective factors for children and youth.
- To understand how these factors, as well as life events, influence children's development.
- To make this information available for developing policies and programs that will help children and youth.
- To collect information on a wide variety of topics – biological, social, economic.
- To collect information about the environment in which the child is growing up – family, peers, school, community

Information comes from different sources (parent, child and teacher) and from direct measures (PPVT, math/reading tests, etc.)

Data Release Strategy

Along with the release of Cycle 4, revised microdata files are also released for Cycles 1, 2 and 3 since they were re-weighted using revised demographic estimates. Some variables were also updated (MSD Scores,…). A separate microdata file was also released in June 2003 for the North component of the NLSCY.

Release 2 of Cycle 4 will occur in Fall of 2003 and cover the variables related to the teacher's questionnaire and the principal's questionnaire.

4.0 Concepts and Definitions

There are many variables and concepts that are critical to the analysis of NLSCY data. The following is an explanation of the key concepts in the NLSCY. Derived variables are those that are not asked directly to the respondents but are calculated using information they have provided.

The content for each section of the various questionnaires used in the NLSCY is presented in Chapter 8.

The unit of analysis for the NLSCY is the child or youth. See Chapter 5 for information on the survey design.

4.1 Definitions

Components

The NLSCY is made up of various components; these are generated by the computer application based on the child's age. The main components are: Adult, Child, Youth, Self-completes, PPVT, Direct Measures, Math tests and Cognitive Measure. These components are described in Chapter 6, Data Collection.

Computer Assisted Interviewing (CAI)

There are two types of computer assisted interviewing used in the NLSCY, Computer Assisted Personal Interview (CAPI) and Computer Assisted Telephone Interview (CATI). For these types of interviews, the interviewer will read the questions on the computer and enter the respondent's answers in the computer. CAI allows for complicated flows and edits to be built into the questionnaire. The questions are identical whether the interview is conducted using CAPI or CATI. Depending on the composition of the household and the nature of the required components, the interview will be conducted partly or completely by telephone and/or field visit.

Effective Age (DMMCQ01)

The age of the child determines, in most cases, the questions that will be asked about him or her. Instead of using the child's actual age, the NLSCY uses a calculated age called effective age. This is done to ensure the child stays in the age group to which he/she is assigned regardless of whether collection takes place before or after the child's birthday. For Cycle 4, the effective age is calculated as 2000 minus year of birth. For example a child born in 1998 would have an effective age of 2 years old (2000-1998). Note that the actual age of the child at the time of the interview is sometimes different from the effective age.

4.2 Family Derived variables

Using NLSCY data, a child's family may be described in several different ways. Many of the family variables used to describe the NLSCY children were derived from what is known as the relationship grid. As part of the household questionnaire some basic demographic information was collected for all members of the child's household. As part of this questionnaire, the relationship of everyone in the household to everyone else was asked. Using this information it was possible to create an extensive set of variables to describe the child's family situation.

The following are some of the family derived variables; the names of the derived variable are given in brackets.

Single Parent Families

There are two ways to describe the parental situation of children using NLSCY data.

Using the relationship grid, a child's single-parent status was derived. There were 81.4% of children living with two parents, 18.1% with one parent and 0.6% without a parent (DDMCD04).

A child's parental status can also be defined in terms of the PMK. There were 81.9% of the NLSCY children living in a household where the PMK had a spouse/partner; and for 18.1% of children the PMK did not have a spouse/partner (DDMPD06A).

The two ways of describing the child's family are very similar. The only reason for the small differences is a result of the few cases where the child lived with a parent, but the parent was not selected to be the PMK.

Intact, Step and Blended Families

Children living with two parents are classified as being members of intact, step and/or blended families based on the relationship of these children to the parents.[1]

Intact Family

An intact family consists of a married or common-law couple in which all children are the natural and/or adopted offspring of both members of the couple.

For the NLSCY children, 71.6% were a member of an intact family (DDMCD16), 6.13% were step children themselves (DDMCD03) and 9.8% lived in a step family (DDMCD15).

Step Family

A step family consists of a married or common-law couple, with at least one child living with them who is the biological or adopted child of one parent but not the other. It should be noted that a child who is the biological child of both parents is said to belong to a step family if at least one of these parents has a step child residing in the household.

For the NLSCY children, 6.3% were step children (DDMCD03) and 9.8% lived in a step family (DDMCD15).

Blended Families

A blended family consists of a married or common-law couple living with at least two children, one of whom does not share the same natural and/or adoptive parents as the other child(ren). The following are examples of blended families:

- a couple with biological children of the female partner as well as biological children of the male partner (i.e., hers and his)

[1] *Foster children and children living with only one parent are not included in step, blended or intact families. In the derivation of blended, intact and step families, if a child was the adopted child of one parent and the biological child of the other parent, then this child was treated like a step child, and thus the family labelled as a step family. In other Statistics Canada publications children of this type are treated as if they were biological children of both parents.*

- a couple with biological children of the female partner as well as children out of the new union (i.e., hers and theirs).

The blended family is a sub-set of the step family. For the NLSCY children, 5.3% were members of a blended family (DDMCD14).

Economic family (DDMCD01)

For the NLSCY, an economic family is defined as all family members related by blood, marriage, common-law relationship or adoption. Foster children are considered to be part of the economic family. For example, if a woman lives in a household with her spouse and two children as well as her sister and her sister's child, then all of these individuals would be part of one economic family. If a boarder also resided in the household with her child, then this would constitute a second economic family.

Siblings

Siblings include full, half, step, adopted and foster siblings. Only siblings residing in the household have been included in the calculation of the sibling derived variables. In the case of common-law relationships, if both members have brought their own children into the relationship then these children are considered siblings. All siblings living in the household, including adult siblings, are included in the calculation of the sibling derived variables. The sibling derived variables include total siblings, as well as number of older siblings, younger siblings and siblings of exactly the same date of birth (i.e., twins) (DDMCD08, 09, 10 and 11).

4.3 Person Most Knowledgeable and spouse

In each NLSCY household, for each selected child, a question was asked about who in the household was the person most knowledgeable about this child. This person was labeled as the PMK. The PMK provides the information for all selected children in the household and then gives information about himself/herself and his/her spouse/partner. In some rare cases it might have been appropriate to label two different people in a household as PMKs. For example, in the case of a step family, it may have been appropriate to label the mother as the PMK for one child and the father for another. However, in order to simplify the interview procedures, only one PMK was selected per household.

In some households, there is no PMK. In cases where the selected child is 16 and over and is no longer living with a parent or guardian, there is no PMK selected in the household.

The following is the breakdown of the relationship of the PMK to the NLSCY children for Cycle 4.

For 89.8% of responding children, the PMK was the mother (88.6% the biological mother and 1.2% the step, adoptive or foster mother)

For 8.7% of the children the PMK was the father. For the remaining 1.5% of children the PMK was not a parent.

For the majority of cases of the PMK not being a parent, the child had a parent living in the household, but the parent was not selected as the PMK. For the most part this situation occurred when a child had a very young mother living with her own parents, i.e., the child's grandparents, and the grandmother was selected as the PMK.

If the PMK had a partner residing in the household at the time of the interview, then this person was labeled as the spouse. Spouses included both married and common-law partners. Detailed socio-economic information was collected about the spouse/partner in order to describe the family situation of the child

Change in PMK between cycles

For several reasons, the PMK and his/her spouse could be different people than those designated in the previous cycle. For this reason, analysts should use caution when comparing PMK information from one cycle to the next.

4.4 Respondent

A cross-sectional respondent is a child from whom the Adult, Child or Youth component was completed. These children represent the population as of January 2001.

A longitudinal respondent is a child introduced in a previous cycle for whom the Adult, Child or Youth component was completed. The children introduced in a previous cycle who have died or moved outside of one of the ten Canadian provinces are also longitudinal respondents. They represent children in the reference population who have the same life course (i.e. have died or moved).

A respondent household is a household where an Adult component or a Child or Youth component has been completed.

A respondent child is a child for whom an Adult component or his/her Child or Youth component has been completed. A respondent household without a complete Adult component can have one respondent child and one non-respondent child. There are 19 children in this situation in Cycle 4.

Please see Chapter 10, Survey methodology – Response rates, for more information about the definition of a respondent.

4.5 Socio-ecomonic status

In past cycles of NLSCY, a measure of socio-economic status (SES) was included. This measure will not be available for Cycle 4. The former definition used information about the respondent's employment as classified by the Standard Occupational Classification (SOC 1980). There is now a new coding structure, SOC 1991, and a definition of SES has not been developed using this new classification.

5.0 Survey Methodology - Sample

The NLSCY is a probability survey designed to provide information about children and youth in Canada. To produce reliable estimates that meet the needs stated by clients, a representative sample of children and youth was selected. This chapter describes the sample selection method and the sample size.

The sample design of the NLSCY is determined to a large degree by the sample design of the Labour Force Survey (LFS). This applies to the stratification, allocation and selection of the sample as well as the estimation methods.[1]

5.1 LFS Sample Design

The LFS is a monthly survey that collects labour market data from a national sample of about 60,000 dwellings. The current design was implemented at the end of 1994 following a redesign program that included a reassessment of the survey's principal role as a provider of current labour market information as well as a central vehicle for conducting household surveys within Statistics Canada.

5.1.1 LFS Target population

The LFS sample is representative of the civilian, non-institutionalised population 15 years of age or older in Canada's ten provinces. Specifically excluded from the survey's coverage are residents of the Yukon, Nunavut and Northwest Territories, persons living on Indian Reserves, full-time members of the Canadian Armed Forces and inmates of institutions. These groups together represent an exclusion of approximately 2% of the population aged 15 or over.[2]

5.1.2 LFS Stratification

The LFS sample is based upon a stratified, multistage design employing probability sampling at all stages of the design. The design principles of the LFS are the same for each province.

Primary strata

Provinces are divided into economic regions and employment insurance economic regions. Economic Regions (ERs) are geographic areas of more or less homogeneous economic structure formed on the basis of federal provincial agreements. They are relatively stable over time. Employment insurance economic regions (EIER) are also geographic areas, and are roughly the same size and number as ERs, but they do not share the same definitions. Labour force estimates are produced for the EIER regions for the use of Human Resources Development Canada.

[1] For a detailed account of the LFS methodology, see Gambino, J.G., Singh, M.P., Dufour, J., Kennedy, B. and Lindeyer, J. (1998), *Methodology of the Canadian Labour Force Survey*, Statistics Canada, Catalogue no. 71-526.

[2] Since 1992, the LFS has been administered in the Yukon, using an alternative methodology that accommodates some of the operational difficulties inherent to remote locales. To improve reliability due to small sample size, estimates are available on a three-month average basis only. In 2000, the survey was extended to the Northwest Territories and Nunavut. No estimates for any of the territories are included in national totals.

The intersections of the two types of regions form primary strata for the LFS. Note that a third set of regions, Census Metropolitan Areas (CMAs), is also respected by stratification in the current LFS design, since each CMA is also an EIER.

Types of areas

The primary strata (ER/EIER intersections) are classified into three types of areas: rural, urban, and remote areas. Urban and rural areas are loosely based on the Census definitions of urban and rural, with some exceptions. Urban areas include the largest CMAs down to the smallest villages categorised by the 1991 Census as urban (1,000 people or more), while rural areas are made up of areas not designated as urban or remote.

All urban areas are further classified into two types: those using an apartment list frame and an area frame, as well as those using only an area frame.

Approximately 1% of the LFS population is found in remote areas of provinces that are less accessible to LFS interviewers than other areas. For administrative purposes, this portion of the population is sampled separately through the remote area frame. Places with fewer than 10 households or 25 persons and Census Enumeration Areas (EA) with fewer than 25 households are omitted from the design.

Secondary strata

In urban areas with sufficiently large numbers of apartment buildings, the strata are grouped according to those based on apartment frames and those based on area frames. The apartment list frame is a register that is based upon information supplied by Canadian Mortgage Housing Corporation (CMHC) and is maintained in the 18 largest cities across Canada. The purpose of this frame is to ensure better representation of apartment dwellers in the sample as well as to minimise the effect of cluster growth due to construction of new apartment buildings. In the major cities, the apartment strata are further stratified into low-income strata and regular strata.

Where it is possible and/or necessary, the urban area frame is further stratified into regular strata, high-income strata, and low population density strata. Most urban areas fall into the regular urban strata, which, in fact, cover the majority of Canada's population. High-income strata are found in major urban areas, while low-density urban strata consist of small towns that are geographically scattered.

In rural areas, the population density can vary greatly from relatively high population density areas to low population density areas, resulting in the formation of strata that reflect these variations. The different stratification strategies for rural areas were based not only on concentration of population, but also on cost-efficiency and interviewer constraints. Also, within each of the secondary strata in rural areas further stratification is carried out where necessary to reflect differences among a number of socio-economic characteristics within each stratum.

The remote area frame is stratified only by province.

5.1.3 LFS Cluster delineation and selection

Households in final strata are not selected directly. Instead, each stratum is divided into clusters, and then a sample of clusters is selected within the stratum. Dwellings are then sampled from selected clusters. Different methods are used to define the clusters, depending on the type of stratum.

Within each urban stratum in the urban area frame, a number of geographically contiguous groups of dwellings, or clusters, are formed based upon 1991 Census counts. These clusters are generally a set of one or more city blocks or block faces. The selection of a sample of clusters (always six or a multiple of six clusters) from each of these secondary strata represents the first stage of sampling in most urban areas. In some other urban areas, Census EAs are used as clusters. In the low-density urban strata, a three-stage design is followed. Under this design, two towns within a stratum are sampled and then a multiple of six clusters within each town are sampled.

For urban apartment strata, instead of defining clusters, the apartment building is the primary sampling unit. Apartment buildings are sampled from the list frame with probability proportional to the number of units in each building.

Other procedures are applied in rural and remote areas. Within each rural stratum, six EAs or two or three groups of EAs are sampled as clusters, whereas remote settlements within each province are sampled proportional to the number of dwellings in the settlement.

5.1.4 LFS Dwelling selection

In all three types of areas (urban, rural and remote areas) enumerators in the field first visit selected clusters and a listing of all private dwellings in the cluster is prepared. From the listing, a sample of dwellings is then selected. The sample yield depends on the type of stratum. For example, in the urban area frame, sample yields in regular strata within major urban areas are either six or eight dwellings, depending on the size of the city. In the urban apartment frame, each cluster yields five dwellings, while in the rural areas and EA parts of cities, each cluster yields ten dwellings. In all clusters, dwellings are sampled systematically. This represents the final stage of sampling.

5.1.5 LFS Sample rotation

The LFS employs a panel design whereby the entire monthly sample of dwellings consists of six panels or rotation groups, of approximately equal size. Each of these panels can be considered to be representative of the entire LFS population. Dwellings are in the LFS for six consecutive months. Each month a new panel of dwellings selected from the same or similar clusters replaces the sample dwellings in one of the rotation groups.

This rotation pattern has the statistical advantage of providing a common sample base for month-to-month comparisons of LFS characteristics. It also ensures that the sample of dwellings constantly reflects changes in the current housing stock and helps to minimise the respondent burden and non-response that could result if households were to remain in the sample longer than six months. Surveys that use the LFS frame or sample can take advantage of the rotation group feature to use larger or smaller sample sizes than that of the LFS.

5.1.6 Household members eligible for the LFS

The first month a dwelling is in the LFS a roster containing information on the household composition is completed. Demographic information including the name, sex, date of birth

and education level is obtained for all persons for whom the selected dwelling is the usual place of residence. LFS information is obtained for all civilian household members 15 years of age or older.

When the dwelling is contacted in subsequent months the roster is updated to reflect changes in household membership from the previous month. Response burden is minimised for the elderly (70 years of age or older) by carrying forward their responses for the initial interview to the subsequent five months in the survey.

5.2 Birth Registry

When a significantly larger number of 1 year old or 5 year old children were needed, the birth registry was used.

This registry is created by the Health Statistics Division which is maintained by updates provided by the provincial registries to Statistics Canada. Some of the information given on the birth certificate is available on an electronic file. This file contains personal information such as the age of the mother, the birth date, the weight of the child at birth, the postal code and the code for the Census Subdivision (CSD). The remainder of the information is available on microfilm which consists of copies of the birth certificates.

The advantages and the disadvantages of this survey frame are summarized in the following table:

Advantages	Disadvantages
Good coverage. File is already available at STC. The reverse record check of the Census of Population uses a similar approach. This allows us to take advantage of experiences gained from that project.	Does not include immigrants. This situation is significant for the 5 year old children. The address provided is the one at the time of birth. In constructing the frame of 1 year old children, there can be administrative delays as some provinces are late in providing STC with their registries. A large amount of time and resources are needed to create a sampling frame (stratification, formation of PSU's, capturing the information from the microfilm, merging information …) The name of the child is often incomplete.

In order to reduce collection costs, it was decided to implement a two phase sampling plan. In the first phase, a geographical area was selected and then children born in this area. Many different geographic areas were considered to form the primary sampling units (PSU). None seemed to be appropriate. The CSD had too small a number of births and the Census Division had too many and covered too large of an area.

The PSU's were formed by regrouping CSD's. In order to improve the efficiency of the sampling plan, the PSU's were selected with probability proportional to size (number of births). With this scenario, the Census Metropolitan Areas (CMA) would have had a large probability of selection. As the CMA's are relatively dense and 2/3 of children are born in CMA's, it was decided to classify these children into distinct strata.

To summarize, each province was sub-divided into two strata, the stratum of children born in a CMA and the stratum of children born outside of a CMA. For efficiency sake, simple random sampling (SRS) was used in the CMA stratum. For the stratum of children born outside of a CMA, a PSU was selected using probability proportional to size and then a SRS of children living in the selected PSU's was chosen. Given the small number of births in Prince Edward Island, only an SRS of children was chosen for this province.

5.3 NLSCY Sample design

Because the NLSCY has both longitudinal and cross-sectional estimation goals, it has several different target populations.

5.3.1 Longitudinal target populations

Longitudinally, the following populations are represented:

Children aged 0-11 in 1994-95:

In Cycle 1, in 1994-95, a sample of children aged 0 to 11 was selected. In Cycle 4, those children were between the ages of 6 and 17. Sample reductions were made in the sample in Cycle 2. As a result, only part of the sample is being followed longitudinally. Children dropped between Cycles 1 and 2 can be regarded as Cycle 1 cross-sectional children. It is important to note that, longitudinally, this cohort still represents children aged 0 to 11 in 1994-95, who were aged 6 to 17 in 2000-2001. This cohort will be followed until the children reach the age of 25.

Children aged 0 and 1 in 1996-97:

In Cycle 2, in 1996-97, a longitudinal sample of children aged 0 and 1 was selected. About 2,000 children aged 0 and 2,000 children aged 1 were selected. In Cycle 4, those children were 4 and 5 years old. Hence this cohort represented children aged 0 and 1 in 1996-97. This cohort has been followed for only three cycles (2 through 4).

Children aged 0 and 1 in 1998-99:

In Cycle 3, in 1998-99, a longitudinal sample of children aged 0 and 1 was selected. About 2,000 children aged 0 and 8,000 children aged 1 were selected. In Cycle 4, those children were 2 and 3 years old and represent children aged 0 and 1 in 1998-99. They will be surveyed for the last time in Cycle 5.

Children aged 0 and 1 in 2000-01

In Cycle 4, in 2000-01, a longitudinal sample of children aged 0 and 1 was also selected. About 2,000 children aged 0 and 2,000 children aged 1 were selected.

5.3.2 Cross-sectional target population

The NLSCY cross-sectional estimate, in Cycle 4, covers children aged 0 to 17 on January 1st, 2001. The user should note, however, that this cross-sectional sample is made up of various components:

(1) The children between 6 and 17 come from the longitudinal sample selected for Cycle 1.
(2) The four-year-olds and some five-year-olds are from the sample of 0-1 year-olds selected in Cycle 2.
(3) The rest of the five-year-olds are from an additional sample of five-year-olds that was selected in Cycle 4 in order to produce more precise estimates for that age.
(4) The two- and three-year-olds are from the sample of 0-1 year-olds selected in Cycle 3.
(5) The sample of children aged 0 and 1 were newly selected in Cycle 4.

It should be noted that Cycle 4 is the last cycle for which data for all ages are available with no interruption. Starting In Cycle 5 there will be no six- and seven-year-olds on the sample file, since the children aged 0 and 1 introduced in Cycle 2 will no longer be part of the survey, which represents a gap in the population of children covered by the survey.

In addition, the coverage of the cross-sectional sample is deteriorating over time. As mentioned earlier, children aged 6 to 17 in the sample were selected in 1994, and no updated sample was added to reflect changes in the population during that period. The coverage of the sample for cross-sectional estimates is analyzed in the chapter on data quality (Chapter 13).

5.4 Sample selection

As described in the section on target population, the Cycle 4 sample consists of several distinct parts. To help the user understand the current composition of the Cycle 4 sample, we should explain briefly how the sample was selected and how it has changed since Cycle 1. Following is a brief history of the sample selection for Cycles 1 to 4.

5.4.1 Cycle 1

Cohort 1

In Cycle 1, in 1994, the initial strategy involved selecting children aged 0 to 11 in each of Canada's 10 provinces. The objective was to be able to produce reliable provincial estimates by age group. The children were then to be followed until they reached the age of 25. Several frames were used to select the initial sample. Households with children in the target population (ages 0-11) were selected from the old-design Labour Force Survey (LFS), from the new-design LFS, from the National Population Health Survey (NPHS) outside Quebec and from the NPHS in Quebec. A total of 22,831 responding children made up our longitudinal sample. A breakdown of this total is provided in the Cycle 1 User Guide.

5.4.2 Cycle 2

Cohort 1

Sample reductions were made between Cycles 1 and 2 on the longitudinal cohort. First, the children from the NPHS in and outside Quebec were dropped. Then, to reduce the response burden on households with several eligible children, the number of children selected was limited to two per household. Some children were dropped from the sample. 16,903 children remained in our longitudinal sample. These children, all Cycle 1 respondents, were selected for Cycle 2 of the survey.

Cohort 2

A new initiative was introduced to the main survey in Cycle 2: Understanding the Early Years. This new initiative focussed on a sample of children aged 0 and 1 to be followed until the age of 5. Two sources were used for the sample: first, we selected children from the LFS; then we added the siblings of Cycle 1 longitudinal children who were already in our sample. A total of 4,153 children were included in our second longitudinal cohort (our first UEY cohort).

Sample Buy-in

Following a request for additional sample, we also added a sample specifically for New Brunswick. The LFS was used for this sample as well. In all, 549 households were selected for a total of 480 responding children. These children were selected for cross-sectional estimation purposes and were not intended to be followed in subsequent cycles.

5.4.3 Cycle 3

Cohort 1

In Cycle 3, we had the opportunity to convert Cycle 2 non-respondents back into the fold. The same initial sample as in Cycle 2 was used, except for deceased children (12), duplicate cases (3), children who were the wrong age for the survey (3), households that were not traceable in Cycle 2 (2), households that had moved permanently out of the country (52), children on Indian reserves (1) and households that were adamant refusals (112) as recorded in Cycle 2. In all, we excluded 185 Cohort 1 children from Cycle 3, for a longitudinal sample of 16,718.

Cohort 2

Responding children from Cycle 2 as part of the UEY initiative should all have been in the Cycle 3 sample. Unfortunately, an error in assigning cases to interviewers reduced the sample by 164 children. In all, only 2,506 of the 2,670 children selected from the LFS were in the Cycle 3 sample. Similarly only responding siblings of Cycle 1 longitudinal children were contacted. A total of 1,483 children made up the second part of the Cohort 2 longitudinal sample.

Cohort 3

The UEY initiative continued in Cycle 3 with a new sample of children aged 0 and 1. The initiative was also interested in "the readiness to learn" aspect of children entering the school system. It was determined that a large sample of 5 year-olds was required to meet these analytical goals. Simultaneously, it was decided that a larger sample of 1 year old children selected in Cycle 3 would meet those objectives once they were 5 years-old in Cycle 6. A sample of about 10,000 would be selected. The LFS is not large enough to provide a sample of 10,000 children as it would require too many rotation groups. Typically, 12 LFS rotation groups ? a year's worth of data ? generates about 2,000 children aged 0 and 2,000 one-year-olds. Consequently, an additional sampling frame was used to procure a large sample of 1 year-olds. A sample of approximately 2,000 children aged 0 to 11 months was selected using the LFS. A sample of about 8,000 children aged 1 year old was selected using the Birth Register. Stratification for this second sample is done differently than in the LFS. Because of the overlap between the two frames, certain eligible children were dropped from selection as their household had already been selected for children in the other cohorts. After excluding the children in households already in the survey, we had a sample of 7,944.

The number of five-year-olds in Cohort 1 was not large enough to meet the analytical goals set out for studying the "readiness to learn" aspect of children entering the school system. To meet these goals, we also used the Birth Register to select about 7,000 five-year-olds. These children were part of the UEY initiative and were sampled to meet analytical goals for the cross-sectional estimates in Cycle 3 only. After identifying and removing the duplicates from Cohort 1, an extra sample of 7,052 5 year-olds was added.

5.4.4 Cycle 4

Cohort 1

From the initial longitudinal sample of children introduced in Cycle 1, some had been responding since the beginning of the survey; others had not responded at some point during the next two subsequent cycles. Some that did not respond in Cycle 2 were also non respondents in Cycle 3, while others were converted to respondents in Cycle 3. Some responded in the first two cycles, but not in the third. The NLSCY strategy was to try to re-interview as many of the initial cohort as possible. However, it became inefficient to contact households that were unlikely to cooperate or untraceable. It was therefore decided to exclude households after two consecutive cycles of non-response. There were 518 such households excluded in Cycle 4. There were also hard refusals (473), deceased children (7), children who had moved away permanently out of the country (79) and children who had not responded in Cycle 2 and had moved temporarily in Cycle 3 (8). Those children were also considered to be non-respondents for two consecutive cycles. In all, 1,086 children were dropped from the longitudinal Cohort sample from Cycle 1, leaving a total of 15,632 selected children.

Cohort 2

The longitudinal children introduced in Cycle 2 are now 4 and 5 years old. Cycle 4 is the last contact cycle for these children. First, to correct the error that occurred in Cycle 3, we added the 164 omitted children to the sample. We also attempted to convert non-respondents from the previous cycle. Only 38 children from the LFS were dropped from the Cycle 4 sample. The reasons for dropping them were age (2), death (2), permanent move (13) and hard refusal (21). A total of 2,632 children from the LFS were included in the Cycle 4 sample.

For the siblings of children introduced in Cycle 1, the situation was more complicated. We had decided in Cycle 2 to limit the number of children surveyed per household to two. However, the addition of siblings contravened that rule for many cases. Consequently, for households in which two children were already being surveyed, we excluded their younger sibling. This reduced the sample by 484 children. In addition we excluded seven cases that had permanently moved out of the country, 13 hard refusals and one death. A total of 978 siblings of Cycle 1 longitudinal children remained in the sample for Cycle 4.

Cohort 3

For children introduced in Cycle 3, only respondents were contacted for this cycle. A total of 1,735 children from the LFS (one hard refusal was excluded) were contacted again in Cycle 4, along with 6,383 children selected using the Birth Register (seven hard refusals were excluded).

Cohort 4

The sample of children aged 0 and 1 from the LFS for this cycle was scaled back to the original UEY goals. In all, 5,031 unique households were chosen.

As in the previous cycle, there were not enough five-year-olds (introduced in Cycle 2) to meet the analytical goals of the "readiness to learn" aspect of 5 year-olds entering the school system. The Birth Register was used once again to select a supplemental sample of 4,399 children.

5.4.5 Sample sizes in Cycle 4

The number of responding children in Cycle 4 is shown by age and province in the following tables. Note that some children are purely cross-sectional and others are purely longitudinal. Among the children who were cross-sectional only, there was the top-up of five-year-olds introduced in Cycle 4. Examples of children who were longitudinal only are children introduced in a previous cycle that had died or moved outside the country. These children were no longer in the target population for the cross-sectional sample, but longitudinally, they still represented the children of the year in which they were selected. For more details on these concepts, please see the chapter on weighting (Chapter 12).

Table 1: Number of Children in the Sample, by Age, Cycle 4

Age	No. of children	Age	No. of children
0[3]	2 358	9	1 329
1[4]	2 673	10	1 285
2	3 154	11	1 183
3	4 963	12	1 090
4	1 627	13	1 091
5[5]	6 255	14	1 151
6[3]	1 979	15	1 101
7	1 928	16	1 173
8	1 368	17	1 081
		Total	36 789

Table 2: Number of Children in the Sample, by Province, Cycle 4

Province	No. of children
Newfoundland and Labrador	2 168
Prince Edward Island	1 246
Nova Scotia	2 718
New Brunswick	2 517
Quebec	6 312
Ontario	9 046
Manitoba	2 928
Saskatchewan	2 954
Alberta	3 661
British Columbia	3 218
Outside the 10 provinces	21
Total	36 789

3 For children aged 0 and 1, the sample is a household sample. More than one child of the same age (twins) may be selected.

4 The sample of fiveyear-olds consists of 1,856 children selected in Cycle 2 and 4,399 children selected from the Birth Register in Cycle 4.

5 The sample of six-year-olds consists of 1,852 children selected in Cycle 1 and 127 children selected in Cycle 2.

6.0 Data Collection

Data for Cycle 4 of the NLSCY were collected between the fall of 2000 and the spring of 2001. They were collected in two main settings: households and schools.

6.1 Household collection

The survey combines computer-assisted interviewing methods and the use of paper questionnaires. Depending on the composition of the household and the nature of the required components, the interview was conducted partly or completely by telephone and/or field visit. This section provides a brief description of the "Collection tools" or the "Survey instruments", in other words the computer-assisted and paper questionnaire components used in NLSCY collection.

6.1.1 Entry/exit component

The first part of the interview was used to prepare a list of all household members, determine the relationships between them, gather tracing information and record basic demographic characteristics such as sex, date of birth, marital status, relationships between household members and dwelling information.

The Person Most Knowledgeable about the child was also identified in this component. This was done once the information about the relationships between household members had been collected.

6.1.2 Child component

A child component was created for each selected child between 0 and 17 years of age. The Person Most Knowledgeable (PMK) about the children and youth answered the child component questions. The PMK was usually the child's mother, but it could also be the father, a step-parent or an adoptive parent who lived in the same dwelling. Only the PMK or his/her spouse was permitted to answer the questions in this component.

At the end of this component, the respondent was asked to provide the name, address and telephone number of two people (friends, relatives) who would be able to help us trace the family in two years, when the survey will be repeated.
For the children in primary school (aged up to 11 years old, excluding those in kindergarten), in preparation for data collection in the schools, the PMK was asked to consent to the collection of information from the child's teacher and school principal.

List of subjects covered

- Education
- Health
- Medical and biological information
- Mother's work after the child's birth
- Child's development
- Temperament
- Literacy
- Communication
- Activities
- Behaviour

- Positive behaviour
- Sleep habits
- Motor and social development
- Relationships
- Parenting
- Custody
- Expectations (Aspirations)
- Socio-demographic characteristics

Note:
For households in which the only child selected was in the 16-17 subgroup and was living with his/her parents, only two subjects were covered: Expectations (Aspirations) and Socio-demographic characteristics. If the child was no longer living with his/her parents, the component was not created.

6.1.3 Adult component

An adult component was created for the PMK and his/her spouse or partner. Only the PMK or his/her spouse was permitted to answer the questions in this component. There was never more than one adult component per household, even if more than one child was selected in the household.

List of subjects covered

- Education
- Labour force
- Income
- Health
- Family functioning
- Neighbourhood safety
- Social support
- Socio-demographic characteristics

Note:
For households in which the only child selected was in the 16-17 subgroup and was living with his/her parents, only four subjects were covered: Education, Labour force, Income and Socio-demographic characteristics. If the child was no longer living with his/her parents, the component was not created.

6.1.4 Adolescent component

This component was added to the interview in Cycle 4. It was used only for 16- and 17-year-olds. The adolescent was the only person permitted to answer the questions in this component, whether he/she was living in the family home or not. This component was followed by a self-administered questionnaire and a pencil-and-paper cognitive test.

List of subjects covered

- Education
- Labour force
- Income
- Health
- Activities

6.1.5 Peabody component (revised Peabody Picture Vocabulary Test – PPVT-R)

The revised Peabody Picture Vocabulary Test (PPVT-R) was used to assess the child's level of receptive vocabulary. It was administered to each selected child between 4 and 6 years old and to children aged 7 or over who were not yet in Grade 2. After obtaining the parent's oral consent, the interviewer asked the child the Peabody component questions directly.

6.1.6 Number Knowledge Component

This component was added to the interview in Cycle 4. It is a direct measure which assesses the child's understanding of the concept of quantities and of the system of whole numbers. This component was administered to children aged 4 and 5 years old. After obtaining the parent's oral consent, the interviewer asked the child the Number Knowledge component questions directly.

6.1.7 Control screen component

A control screen was created for each questionnaire or paper form required, to remind the interviewer to administer the appropriate child or adolescent questionnaire. Information (identification code, given name, etc.) was also transcribed from the screen to the paper questionnaire, and the questionnaire's sequence number was entered in the computer.

6.1.8 Ages and Stages questionnaires

Parents of children aged 3 to 71 months were asked to complete the section of the booklet that corresponded to the selected child's age in months. Five development measures were assessed in the questionnaire: communication skills, gross motor skills, fine motor skills, problem-solving skills, and personal and social development. Each booklet contained specific sections for each age group.

6.1.9 Direct assessments questionnaire for four- and five-year-olds

Direct assessments were added to NLSCY to facilitate the collection of data on child development and learning: the "Who am I?" questionnaire. This booklet was administered to four- and five-year-olds.

6.1.10 Mathematics test (computation exercise)

The mathematics test is an objective indicator of the child's academic performance in mathematics. It was administered to children in Grade 2 or above, ranging in age from 7 to 15.
It consisted of a set of nine booklets of varying levels of complexity. The level was determined by the child's grade.

LEVEL	Usually used for Grade
2	2
3	3
4	4
5	5
6	6
7	7 (Secondary 1 in Quebec)
8	8 (Secondary 2 in Quebec)
9	9 (Secondary 3 in Quebec) Secondary 1 in Manitoba)
10	10 (Secondary 4 in Quebec) Secondary 2 in Manitoba) (Level I in Newfoundland)

In previous cycles, the mathematics tests were administered as part of the school collection. To reduce the amount of work that teachers were required to do for this survey, and to avoid disrupting classroom activities at the end of the school year, it was decided to administer the test at home rather than at school.

To minimize the impact of this decision on the length of the household interview, it was decided to administer only the mathematics test and drop the math and reading skills indicator questions. The indicator was a placement test to determine what level of math test should be administered. Instead, the level was determined on the basis of the child's grade.

6.1.11 Self-complete Questionnaires – ages 10-11, 12-13, 14-15 and 16-17

Respondents between 10 and 17 years of age completed a paper questionnaire on various aspects of their lives. For the 10-15 subgroup, the child was given the questionnaire during the interview and asked to complete it himself/herself. To ensure confidentiality, the child placed the completed questionnaire in an envelope, sealed the envelope and gave it to the interviewer. For the 16-17 subgroup, the questionnaire was mailed out in advance and was to be completed before the interview. If the questionnaire had not been completed before the interview, the respondent had to complete it during the interview, seal it in an envelope and give it to the interviewer.

The Self-complete Questionnaires consisted of a set of four booklets, one for each of the four age groups. The table below shows the subjects covered by each age-group section in the booklet. The questions for each subject were different for each age group. The booklets are reproduced in Book 2 of "National Longitudinal Survey of Children and Youth, Cycle 4 Survey Instruments 2000-2001".

Subject	Section in the booklet			
	10-11 Booklet 20	12-13 Booklet 21	14-15 Booklet 22	16-17 Booklet 23
Friends and family	A	A	A	A
School	B	B	B	-
About me	C	C	C	B
Feelings and behaviour	D	D	D	C
My parents	E	G	G	E
Puberty	F	H	H	F
Smoking, drinking and drugs	G	F	F	D
Activities	H	E	E	-
Dating	-	H	H	G
Health	-	H	H	F
Work and money	-	I	I	-
Decision-making	-	-	-	H

6.1.12 Cognitive test for 16- and 17-year-olds

The test was administered on a paper questionnaire to be completed by the respondent. It covered reading and mathematics.

Two versions of the test were developed based on aptitude. Each booklet contained 18 questions designed to measure mathematical aptitude.

The mathematics questions dealt with the use of mathematics in everyday activities, such as interpreting graphs and spatial diagrams and solving equations in order to make decisions.

6.1.13 List of components and questionnaires for each age group and interview type

Age 0-3

Interview type	Components	Approximate length of interview
Telephone	Entry/exit	75 minutes
	Adult	
	Child	
	Ages and stages questionnaire and its Control Screen component	

Age 4-6

Interview type	Components		Approximate length of interview
Telephone and face-to-face interview	Entry/exit		140 minutes
	Adult		
	Child		
	Peabody component	Children aged 4-6 years	
	Number Knowledge Component	Children aged 4-5 years	
	Development questionnaire and its Control Screen component	Children aged 3-71 months	
	"Who am I?" booklet and its Control Screen component	Children aged 4-5 years	

Age 7-9

Interview type	Components		Approximate length of interview
Telephone and face-to-face interview	Entry/exit Adult		83 minutes
	Child		
	Peabody component	if in Grade 1 or below	
	Mathematics test and its Control Screen component	if in Grade 2 or above	

Age 10-15

Interview type	Components		Approximate length of interview
Telephone and face-to-face interview	Entry/exit		90 minutes
	Adult		
	Child		
	Mathematics test and its Control Screen component	if in Grade 2 or above	
	Self-complete Questionnaire and its Control Screen component		

Age 16-17

Interview type	Components		Approximate length of interview
Telephone and face-to-face interview	Entry/exit		105 minutes
	Adult	Only if child lives with parents	
	Child		
	Adolescent		
	Self-complete Questionnaire and its Control Screen component		
	Cognitive test for 16-17-year-olds and its Control Screen component		

6.1.14 Households in which all the selected children were aged 3 or under

The computer-assisted interview and the paper questionnaire on Ages and Stages were completed by telephone since neither the child nor the parent's consent and signature were required for questionnaire administration.
The interview was conducted in two stages. During the initial call, the interviewer completed the computer-assisted interview and determined which version of the Ages and Stages questionnaire should be used. The interviewer told the respondent that a questionnaire would be mailed to him/her, and made an appointment to call one or two weeks later to collect the responses.

6.1.15 Households in which the selected children were aged 4 or over

The first few components of the computer-assisted interview were completed by telephone; the rest of the interview, which had both computer-assisted and paper components, had to be completed during a field visit.
Between the initial call and the field visit, the parents of the 4-5 subgroup also received the appropriate version of the Ages and Stages questionnaire by mail so that they could complete it before the interviewer's visit.

6.1.16 Information kit

Before the collection period, Statistics Canada's regional offices mailed an information kit on the survey to all parents and 16- and 17-year-olds.
Along with their information kit, the 16- and 17-year-olds received a Self-complete Questionnaire, which the interviewer was to collect at the time of the field visit.

6.1.17 Collection period

The collection period was divided into five phases between September 2000 and June 2001:

Phase 1 Sep.-Oct. 2000 Age 0-3

Phase 2	Nov.-Dec. 2000	Age 6-17
Phase 3	Jan.-Feb. 2001	Age 5
Phase 4	Mar.-May 2001	Age 4-17 (continued)
Phase 5	June - July 2001	Age 0-3 (continued)

6.1.18 Non-response follow-up

All cases not processed in Phase 1 for reasons such as no contact, hard refusal or language barriers were returned to the interviewers for inclusion in a new Phase 5 sample. Similarly, all cases not processed in Phases 2 and 3 were returned in Phase 4.

6.1.19 Collection personnel (training, supervision and control)

The NLSCY was conducted by Labour Force Survey (LFS) interviewers. A number of them had worked on one or more previous cycles of the NLSCY. All LFS interviewers report to a staff of senior interviewers who are responsible for ensuring that interviewers are familiar with the survey's concepts and procedures. The senior interviewers ensure that prompt follow-up action is taken for refusal and other non-response cases. If necessary, non-response cases are transferred to a senior interviewer and reassigned. The senior interviewers in turn report to the LFS program managers, located at Statistics Canada's regional offices.

For the NLSCY, a combination of classroom training and self-study materials was used to ensure that interviewers and supervisors had a proper understanding of the survey concepts. In the self-study portion, which preceded the classroom training, the program managers, senior interviewers and interviewers read the Interviewer's Manual prepared for the survey and completed a case study exercise.
The classroom training was two-day course given by a program manager or senior interviewer. In all, each interviewer received nearly 20 hours of training.

DAY 1	Overview of the survey and its components "Contact/Demo" component Child component * - various ages Ages and Stages questionnaire (age 0-5) Set of Peabody Test pictures (six-year-olds and older children not yet in Grade 2) Math and reading skills indicator (ages 6-15) Self-complete questionnaires (ages 10-17)
DAY 2	Adolescent component (ages 16 and 17) Cognitive test (ages 16 and 17) Adult component Assessment of interview conditions Tracing Non-response Incentives

6.2 School collection

This phase of data collection took place in the schools between April and June 2001.

For children in the sample who were attending school at a level higher than kindergarten and aged 11 or less, the PMK was asked to consent to the collection of information from the child's teacher and principal. School collection involved three questionnaires, which were mailed out to teachers and principals. They were asked to complete the questionnaires and mail them back to Statistics Canada in the envelopes provided. If a child had more than one teacher, the package was to be sent to the current teacher who knew the child best, preferably a language or math teacher.

6.2.1 Teacher's questionnaire

Two questionnaires were developed, a first one for cases where the child had one teacher for the basic academic subjects and a second one for cases where the child had different teachers for the basic academic subjects. These questionnaires dealt with the child's academic performance and behaviour at school, the teacher's methods of instruction and the atmosphere in the classroom.

6.2.2 Principal's questionnaire

This questionnaire collected information about the teaching methods used in the school, the availability of educational resources, and the social atmosphere in the school. Hence, the Principal's Questionnaire was about school policies and the educational environment and not about a specific child.

7.0 Data Processing

7.1 Editing

The main output of the NLSCY is a "clean" master data file. This section presents a brief summary of some of the processing steps involved in producing this file.

Computer Generated Edits

As discussed earlier, all of the information for the household collection (except for the 10-11 year old and 12-13 and 14 -15 and 16 -17 year old self-completed questionnaires) was collected in a face-to-face or telephone interview using computer-assisted interviewing (CAI). As such, it was possible to build various edits and checks into the questionnaire for the various household CAI components, in order to ensure high quality of the information collected.

Types of Computer Edits
Various types of computer generated edits were used to check data while the interviewer was completing the interview.

The NLSCY computer generated survey used the following:

- Review Screens,
- Range Edits,
- Flow Patterns Edits,
- Consistency Edits.

Review Screens
Review screens were created for important and complex information.

Example:
The selection procedures for the PMK, a critical element of the survey, were based on the household roster. The household roster screen showed the demographic information for each household member and his/her relationship to every other household member. The collected information was displayed on the screen for the interviewer to confirm with the respondent before continuing the interview.

Range Edits
Range edits were used for continuous variables, to confirm or correct unusual answers during collection.

Example:
For the question regarding the weight of a child at birth, if a weight entered into the computer was either significantly high or low, a pop-up message would appear asking the interviewer to confirm the answer with the respondent.

Flow Pattern Edits
All flow patterns were automatically built into the CAI system.

Example:
In the Child Care Section, the PMK is asked if he/she used daycare or babysitting in order that he/she (or a partner/spouse) could work or study. Based on the response given the flow of the questions could be different. If Child Care was used, the CAI system continued with a series of questions about the specific care method(s) used for the child. If not, the CAI system automatically skipped this series of questions.

General Consistency Edits
Some consistency edits were included as part of the CAI system, and interviewers were able to "slide back" to previous questions to correct for inconsistencies. Instructions were displayed to interviewers for handling or correcting problems such as incomplete or incorrect data.

Example:
In the collection of the Labour Force Section, the number of weeks working, not working, and looking for work should not total more than 52 weeks. If this was the case, the system generated a pop-up window which stated the error and instructed the interviewer to slide back to the appropriate question to confirm the data and make corrections as required.

Consistency Edits Between Cycles
Edits were also performed to ensure consistency between cycles for data that was not expected to change. Data from the previous cycle (feedback variables) were included in the CAI system for the current cycle. When inconsistencies were identified, the interviewer was asked by the system to confirm the Cycle 4 data with the respondent through a series of questions.

Example:
For the Chronic Conditions questions, if a chronic condition such as asthma was reported in the previous cycle but not indicated as being present in the current cycle, the system prompted the interviewer to ask questions to determine if the current data was in fact correct, or if the condition had changed since the previous cycle.

7.2 Data Capture

Paper and Pencil Questionnaires
Some questionnaires for the NLSCY were completed on paper and pencil questionnaires (PAPI). The 10-11, 12-13 and 14-15 year old and 16 -17 Self-Completed Questionnaires, the Teachers' Questionnaires and the Principals' Questionnaire were all completed by PAPI. All of these documents were completed directly by a survey respondent.

Data Capture for PAPI Questionnaires
Data capture for these questionnaires were accomplished at a centralized area at Statistics Canada's Head Office.

Any document containing at least one respondent-completed item was captured and a file containing each record was provided to Head Office processing staff for further processing. As part of the capture system, some quality checks were built in to flag unusual entries to warn the operators of potentially incorrect entries.

In cases where more than one response was checked off by the respondent, the operators were instructed to accept the first response. Errors remaining within the questionnaires were then edited at a later stage.

7.3 Minimum Completion Requirements

Defining Requirements
One of the first steps in the NLSCY processing was to define the requirements for a responding household.

No Information Collected
In some cases, no NLSCY information was collected for a sampled household. This happened, for example, when an interviewer was unable to make contact with a selected household for the entire collection period, in other cases the household refused to participate in the survey, special circumstances such as an illness or death in a family or extreme weather conditions sometimes prevented an interview from taking place.

Partial Information
In other cases, it was possible to carry out some of the interview, but a complete interview was not obtained for a variety of reasons. Some respondents were willing to give only a certain amount of time to the completion of the survey. In some cases an interviewer completed a portion of the survey with the respondent and made an appointment to continue at another time but was unable to re-contact the respondent.

Criteria for Partial Response
It was necessary to come up with criteria for deciding what to do with these "partial" interviews. If the majority of the survey had been completed, obviously the preference was to keep this case and label it as a responding household. However, if only very minimal information was collected the decision was made to drop the household and treat it as a non-responding household. An assessment was made as to whether or not there was an adequate amount of information collected for at least one child, adult or youth component in each household. If there was, the household was maintained in the responding sample.

Missing Variables
All missing variables for households were set to not-stated or were imputed. If there was not adequate information then the household was dropped from the responding sample and treated as a non-response.

The longitudinal file also contains 195 records that were created for some longitudinal children for whom no data was collected in this cycle. These are children who are now deceased or who have moved out of the country, but who will be kept on the longitudinal file for weighting purposes. For these records, all variables except for the longitudinal weight (DWTCW01L) have been set to ‘not stated’.

7.4 Head Office Editing

Stages of Editing
For CAI questionnaires for the NLSCY, two stages of editing were conducted.

- Pre-edit
- Consistency Editing

The purpose of the Pre-edit was to carry out some basic formatting and preliminary editing.

Step	Action	Done to the:
1	✍ Non-response values from the CAI system were recoded to standard non-response codes for refusals, don't know and not-stated. ✍ Mark All That Apply' questions were destrung and values converted to Yes (1) or No (2) responses. ✍ Databases files were created for each section of the Adult and Child questionnaires and Youth questionnaires.	complete Adult and Child file and Youth file
2	✍ Small data base files were created for each section of each questionnaire ✍ Within several sections, different wording was used for different age groups. For example, in the Activities section, Question 3 asks "In the past 12 months, outside of school hours, how often has (the child) taken part in any clubs, groups or community programs with leadership....". The wording for 4 to 5 year-olds (ACTQ3D1) was "such as Beavers, Sparks or church groups ?". The wording for 6 to 9 year olds (ACTQ3D2) was "such as Brownies, Cubs or church groups?" Initially these questions were stored as separate variables. As part of the pre-edit the two variables were collapsed into one output variable DAACQ3D. ✍ The flow patterns for each section were processed and valid skips were assigned 'not applicable' codes (6, 96, 996..).	Separate DBF files from Step 1

7.5 Consistency Editing

After the pre-edit, consistency editing was carried out to verify the relationship between two or more variables.

Example:
In the Socio-Demographic Section, for children who were not born in Canada, Question DSDCQ2B asks on what year they first immigrated to Canada. There was a consistency edit which compared this question to the year of birth of the child. If the year of immigration was before year of birth then year of immigration was set to not-stated in the edit.

Consistency Between Cycles
Editing was also performed to ensure consistency between cycles.

Example: The responding child's school grade in Cycle 4 should not be less than the grade reported in Cycle 3.

Flags were set for inconsistencies between cycles. These variables contain 'Z' in the variable name.

Data File for 10 to 17 year olds
One data file was produced for the 10-11,12-13 ,14-15 and 16 -17 year old questionnaires. For questions that did not apply to an age group, the variables were set to 'not applicable' codes (6,96,996..).

Data File for Teacher's File
In this cycle there were 2 Teachers' questionnaires and a Principal's questionnaire. These are to be released in the Fall of 2003.

7.6 Naming Convention and Coding Structure for NLSCY Variables

The NLSCY microdata file documentation system has employed certain standards to label variable names and values. The intent is to make data interpretation more straight-forward for the user.

Naming Convention for Variables

A naming convention has been used for each variable on the NLSCY data file in order to give users specific information about the variable. All variable names are at most eight characters long so that these names can easily be used with analytical software packages such as SAS or SPSS.

Format for Variable Names

D SE C Q nnx or D SE C d Q nnx

"D" refers to the NLSCY Cycle
"A" indicates the first cycle,
"B" the second cycle,
"C" the third etc...

SE - refers to the section of the questionnaire where the question was asked or the section from which the variable was derived.

C - refers to the collection unit or the unit to which the variable refers.

There are five possibilities[6] :
"C" is the variable refers to the child,
"P" the PMK.
"S" the spouse/partner
"H" the household
"Y" is the variable refers to the youth,

d - the lower case letter refers to the NLSCY Cycle in which the variable first appeared on the file or cycle in which changes to a previously asked question were made.

Example: "b" indicates the variable was new in Cycle 2. In subsequent cycles, new variables will also be identified using the lowercase letter representing the cycle. New variables in Cycle 3 will contain a "c", in Cycle 4 a "d", etc. Some revisions were made to

6 It should be noted that while variables do exist for various units of analyses (i.e., the PMK, the spouse/partner and the household), it will only be possible to produce "child estimates" from the NLSCY file. The characteristics of the PMK, spouse/partner and household can be used to describe attributes of the child. For example it will be possible to estimate the number of children living in a household with low income, or the number of children for whom the PMK has scored high on the depression scale etc. However it will **not** be possible to produce estimates of the number of low income households or depressed PMKs.

the content of the questionnaire between cycles. If the revision resulted in a change to the meaning or the values of a question, the variable was treated as new and contains a "d".

Q refers to the variable for a question that was asked directly on one of the NLSCY questionnaires

“S” refers to a score calculated for one of the scales used on the questionnaire

“D” means the variable was derived from other questions that were asked on the questionnaire

“I” means the variable is a flag created to indicate that an item has been imputed

“Z” means the variable is a flag created to indicate an inconsistency in reported data between the current and previous cycles

“nnx” refers to the question or variable identification. Generally nn is a sequential number assigned to the variable; and x is a sequential alphabetic indicator for a series of variables of a similar type

7.7 Acronym Names for Questionnaire Sections

The following table gives the acronyms that were used for each section of the various NLSCY questionnaires. This acronym is embedded in the variable name for all variables on the NLSCY data file. The acronym is the second and third characters of the variable name.

	Variable	**Collected or Derived from the:**
GE	Geographic	sample information
HH	Household	dwelling characteristics
MM	Variables collected as part of the household roster. .	Basic demographic variables for each household member. These variables are included on the NLSCY data file for the child, the PMK and the spouse/partner
DM	Demographic- derived to explain the living arrangements of the child:	information of the household roster and relationship grid
SD	Socio-demographic	child on the Child's Questionnaire and for the PMK and spouse/partner on the Adult Questionnaire.
HL	Health	PMK and Spouse on the Adult questionnaire, and for the Child on the Child questionnaire, and for the Youth on the Youth questionnaire
CH	Adult Chronic Conditions	PMK and Spouse in the Health section of the Adult questionnaire
RS	Restriction of Activities :	PMK and Spouse in the Health section of the Adult questionnaire
DP	Depression scale	Parent Questionnaire (this scale was administered to the PMK)
ED	Education	children 4 to 15 years old on the Child's Questionnaire and about the PMK and spouse/partner on the Adult Questionnaire and for the Youth on the Youth questionnaire
LF	Labour force	PMK and spouse/partner on the Adult Questionnaire and for the Youth on the Youth questionnaire
IN	Income	household income and personal income of the PMK, collected on the Adult Questionnaire and for the Youth on the Youth questionnaire
FN	Family functioning	Adult Questionnaire (section asked to the PMK or spouse
MD	Medical/biological .	Child's Questionnaire (0 to 3 years of age)
TM	Temperament	Child's Questionnaire (3 months to 35months)
LT	Literacy	Child's Questionnaire (0 to 9 years)
AA	Activities	Child's Questionnaire (0 to 15 years) and for the Youth on the Youth questionnaire
BE	Behaviour	Child's Questionnaire (0 to 11 years)
MS	Motor and social development	Child's Questionnaire (0 to 47 months)
RL	Social relationship	Child's Questionnaire (4 to 9 years)

PR	Parenting style	Child's Questionnaire (0 to 15 years)
CR	Child care	Child's Questionnaire (0 to 13 years)
PP	PPVT test:	4 to 6 years old (if child in grade 1 or less included those over 6 years of age)
FF	Friends and Family	10 to 17 Self-complete Questionnaires
SC	School	10 to 17 Self-complete Questionnaires
AM	About Me	10 to 17 Self-complete Questionnaires
FB	Feelings and Behaviour	10 to 17 Self-complete Questionnaires
PM	My Parents and Me	10 to 17 Self-complete Questionnaires
PU	Puberty	10 to 17 Self-complete Questionnaires
DR	Smoking, drinking and drugs	10 to 17 Self-complete Questionnaires
AT	Activities	10 to 17 Self-complete Questionnaires
HT	Health	10 to 17 Self-complete Questionnaires
WK	Work and Sources of Money	10 to 17 Self-complete Questionnaires
DA	Dating	10 to 17 Self-complete Questionnaires
DC	Decision Making	10 to 17 Self-complete Questionnaires
MA	Math computation test	children in grade 2 and over and Cognitive Math test for 16 -17 year olds.
SF	Neigbourhood Safety	Adult Questionnaire (section asked to the PMK or spouse
SP	Social Support	Adult Questionnaire (section asked to the PMK or spouse
SL	Sleep	Child's Questionnaire (0 to 13 years)
PB	Positive Behaviour	Child's Questionnaire (3 to 5years)
AS	Aspirations	Child's Questionnaire (16 to 17 years)
AG	Ages & Stages	Direct Measure (3 to 71 months)
WM	Who am I	Direct Measure (4 to 5 years)
KN	Number Knowledge	Direct Measure (4 to 5 years)
OB	Observation Assessment	All children who receive Direct Measure

7.8 Examples of Variables Names

In order to illustrate the naming convention used for variables included on the NLSCY data file the following examples are given.

Variable Name Refers to:

DLFSQ2	Q2 in the Labour Force Section for the spouse/partner
D	a Cycle 4 variable
LF	the Labour Force Section
S	the spouse/partner
Q	an item asked directly on the questionnaire
2	the ID of the item.

DPRCS03	a positive interaction score on the parenting scale for a 2 to 15 year-old child
D	a Cycle 4 variable
PR	the Parenting Section
C	the child.
S	a score
3	ID of the variable

7.9 Coding Structure for NLSCY Variables

Some standards have been developed for the coding structure of NLSCY variables in order to explain certain situations in a consistent fashion across all variables. The following describes these various situations and the code used to describe the situation.

Refusal
During a CAI interview, the respondent may choose to refuse to provide an answer for a particular item. The CAI system has a specific function key that the interviewer presses to indicate a refusal. This information is recorded for the specific item refused and transmitted back to Head Office.

On the NLSCY data file an item which was refused is indicated by a code "8".

For a variable that is one digit long the code will be "8", for a 2 digit variable "98" for a three digit variable "998" etc.

Don't Know
The respondent may not know the answer to a particular item. Again the CAI system has a specific function key to describe this situation.

On the NLSCY data file, the code used to indicate that the respondent did not know the answer to an item is "7". For a variable that is one digit long the code will be "7", for a two-digit variable "97" for a three-digit variable "997" etc.

Not Applicable
In some cases a question was not applicable to the survey respondent. A code "6", "96" "996" ... has been used on the data file to indicate that a question or derived variable is not applicable.

In some cases a single question or series of questions was not applicable. For example, the question on number of hours per week the child is cared for in a daycare centre (DCRCQ1G1) is only applicable for children for whom this type of care is used (DCRCQ1G=1). Otherwise there will be a code 996 for this question.

In other cases an entire section of the questionnaire was not applicable or even an entire questionnaire. For example, the Motor and Social Development Section was applicable only to children 0 to 3 years old. For all children outside of this age group (i.e., 4 years and

older) the motor and social development variables have been set to not-applicable ("6", "96", "996" etc.).

For cases where the PMK did not have a spouse or common-law partner residing in the household, all "spouse" variables (e.g., the Labour Force Section and the Education Section for the spouse) have been set to not applicable.

Not-Stated
In some cases, as part of Head Office processing the answer to an item has been set to not-stated. The not-stated code indicates that the answer to the question is unknown. Not-stated codes were assigned for three main reasons.

- As part of the CAI interview, the interviewer was permitted to enter a refusal or don't know code, as described above. When this happened the CAI system was often programmed to skip out of this particular section of the questionnaire. In the case of refusal, it was assumed that the line of questioning was sensitive and it was likely that the respondent would not answer any more questions on this particular topic area. In the case of a "don't know", it was assumed that the respondent was not well enough informed to answer further questions. As part of the NLSCY processing system, it was decided that all of these subsequent questions should be assigned a not-stated code. A not-stated code means that the question was not asked to the respondent. In some cases it is not even known if the question was applicable to the respondent.

- In some cases a questionnaire was not started or it was started but ended prematurely. For example, there may have been some kind of an interruption, or the respondent decided that she/he wished to terminate the interview. If there was enough information collected to establish this household as a responding household, then all remaining items on the questionnaire (and on questionnaires that had not yet been started) were set to not-stated. The one exception was that if it was known that a certain section or a certain questionnaire was not applicable, then these questions were set to not applicable.

- The third situation in which not-stated codes were used was as a result of consistency edits. When the relationship between groups of variables was checked for consistency, if there was an error, often one or more of the variables was set to not-stated.

For derived variables if one or more of the input variables to the derived variable had a refusal, don't know or not-stated code, then the derived variable was set to not-stated.

7.10 Coding of Open-ended Questions

Open-ended Format
A few data items on the NLSCY questionnaire were recorded by interviewers in an open-ended format. For example, in the Labour Force Section, a PMK who had worked in the previous 12 months was asked a series of open-ended questions about the current or most recent job:

- What kind of business, service or industry is/was this?
- What kind of work are/were you doing?
- At this work, what are/were your most important duties or activities?

How they are recorded
The interviewer recorded in words the answer provided by the PMK. At Head Office, these written descriptions were coded into industry and occupation codes to describe the nature of the work of the PMK. Similar information was collected for the spouse/partner and codes assigned to describe the nature of the work.

How they are coded
The coding systems used were the 1991 Standard Occupational Classification codes (SOC) and the North American Industrial Classification System (NAICS). Grouped versions of these codes are available on the data file (DLFPcD7A and DLFPcD8A for the PMK, and DLFScD7A and DLFScD8A for the spouse/partner).

7.11 Imputation flags

Missing Variables

For various reasons there are certain variables that may be missing for responding households on the NLSCY file. This is usually referred to as item non-response.

Imputation
For some variables on the NLSCY file, however, rather than using a special non-response code, imputation has been carried out. Imputation is the process whereby missing or inconsistent items are "filled in" with plausible values. For the NLSCY, imputation was carried out for household income, PMK income and youth income. See chapter 11 for more details on imputation.

Imputation flags have been included on the NLSCY file so that users will have information on the extent of imputation and what specific items have been imputed on what records.

All imputation flags on the NLSCY data file have an "I" as the sixth character of the variable name. For example, the name of the imputation flag for the income of the PMK is DINPcI1A.

7.12 Derived Variables

Combining Items
A number of data items on the data file have been derived by combining items on the questionnaire in order to facilitate data analysis. For example, in the Labour Force section, one of the questions is on the Number of Weeks Worked but in the Adult Education section, the question is Whether They Are Presently Going To School. The combination of these two questions forms a variable that is based on the Actual Situation Of Work And Study.

Longitudinal derived variables
Longitudinal derived variables were created to indicate changes between data reported in the current and previous cycles for family structure and PMK and Spouse changes.

Derived Variable Name
All derived variables on the NLSCY data file have a "D" as the fifth character of the variable name. The name of the variable for the primary care arrangement is DLFPD51.

8.0 Content of the Survey

The NLSCY was designed to follow an ecological or holistic approach to measuring child development. The survey captures the diversity and dynamics of the factors affecting children. To ensure that all relevant topic areas affecting child development were adequately addressed by the survey, a multidisciplinary consultation was carried out at the inception of the survey. The selection of specific subject areas, priorities and survey questions was very much a group effort with input and advice from:

- the NLSCY expert advisory group that consists of researchers in the area of child development and the social sciences;
- federal departments;
- representatives from the provinces and territories responsible for child development programs.

It was recommended that the NLSCY cover a broad range of characteristics and factors affecting child growth and development. Extensive information was gathered about the child, as well as the child's parent(s), characteristics of the family and the neighbourhood as well as the child's school and school experiences. This section provides an outline of the content for each section of the questionnaire included in the NLSCY data.

8.1 Processing System

As part of the NLSCY processing system, there are some basic quality checks performed for each section of the questionnaire. Any items for which there was a high level of non-response or that were frequently involved in edit failures were looked at in detail. Where appropriate, comparisons were made to external data sources and analyses were carried out to investigate possible reasons for differences from these other sources. Any concerns about potential data quality problems for any items in a particular section of the questionnaire are discussed in this section of the documentation.

For a discussion on the validation of the scale scores, please see Chapter 9.

8.2 Survey components

The NLSCY is divided into several components; these are described in Chapter 6, Data Collection. Below is a summary of each component.

Household This is the first part of the interview. The household roster asks for basic demographic information for each household member and their relationship to everyone else in the household.

Adult Questions asked about the PMK and spouse. For children aged 16 and 17, not all the sections in the adult component are asked. The adult component is completed once even if there are two children in the household.

Child Questions about the selected child asked to the PMK. A child component is completed for each selected child. The only sections of the Child Questionnaire asked about youth aged 16 and 17 are the Aspirations and Expectations section and the Socio-Demographics section.

Ages and Stages	Questions about the selected child aged zero to five years. This questionnaire was completed on paper by the PMK before the interview.
Youth	Questions asked about the selected child, if he/she is 16 or 17 years old. In this section respondents answer questions about themselves in a computer assisted interview (CAI)
Self-completes	Respondents aged 10-17 answer questions about themselves in a paper questionnaire.
Direct assessments	Several direct assessments are done with the children and youth; these are described in Chapter 16.
Teacher and Principal Questionnaires	For children who are in school (for children aged up to 15 years old, except those in kindergarten), a questionnaire is sent to their teacher and the principal of the school. This information will be included in a second release.

8.3 Demographic Variables

The demographic variables are collected in the household questionnaire. As part of the household questionnaire some basic demographic information (e.g., age, gender, and marital status) is collected for all members of the child's household. The relationship grid is also completed as part of this questionnaire i.e., the relationship of everyone in the household to the PMK. Using this information it is possible to create an extensive set of variables to describe the child's family situation. Most of these derived variables are critical to the analyses of NLSCY data and are described in Chapter 4.

It is necessary to perform an extensive series of edits on the data that were collected.

The following are some examples of the types of editing that are carried out.

? a birth parent should be at least 12 years older (and not more than 55 years older) than a birth child

? the difference in age between a husband and wife should be less than 29 years.

8.4 Adult Questionnaire

Education (Parent)

The Education Section is completed for both the PMK and spouse/partner. The objective is to gather information on the years of school completed, educational attainment, and current attendance at an educational institution.

Research has indicated a link between maternal educational attainment, the home environment and child development. The questions on full-time and part-time school attendance provide an indicator of the main activities of the PMK and the spouse/partner.

Labour Force

Employment stability impacts the home environment, both in terms of income and stress levels. Research indicates that parental unemployment can adversely impact child outcomes.

The Labour Force Section is completed for both the PMK and spouse/partner. The main objective of the section was to determine employment stability as an indicator of the continuity of employment income. Questions include periods of absence from work, reason for the most recent absence, hours worked, and work arrangements (e.g. shifts) during the previous year. A series of questions were asked about the PMK and spouse/partner's current or most recent job held.

A complete description is recorded for the current or most recent job. Industry and occupation coding was carried out using North American Industry Classification System (NAICS) 1997 and 1991 Standard Occupational Classification codes.

Labour Force Derived Variables

Several labour force derived variables have been created for the PMK and spouse/partner of the PMK. They include:

DLFPcD5A / DLFScD5A:	NAICS code for PMK's /Spouse's current job
DLFPcD6A/DLFScD6A:	SOC91 for PMK's /Spouse's main job
DLFPcD7A/DLFScD7A:	Standard industry code for current job – grouped
DLFPcD8A/DLFScD8A:	Standard occupation code for current job – grouped

Income

In the Income section of the survey, the sources of income and the income are collected for each household. Income range is also collected for the PMK and for the PMK's spouse. This information provides an indicator of the family's economic situation, an essential component of the child's environment.

As family income is an important part of many studies on child development, we impute a value for household income if the respondent did not answer these questions. See Chapter 11 for a detailed explanation of how income is imputed.

Two derived variables (DINHD04A and DINHD05A) have been created to compare the household income to the low income cut-offs (LICOs). LICOs are used to distinguish "low income" family units from "other" family units. A family unit is considered "low income" when its income is below the cutoff for its family size and its community. A family at or above the cutoff falls into the "other" category[7]. The variable DINHD03A gives the value of the LICO by geographic area.

Also included in the income variables are two new questions (DINHDdQ06 and DINHDdQ07) that ask the respondent about how they feel about their family's income security. These questions were suggested by Dr. Harvey Krahn of the University of Alberta. They are similar to questions used by the Population Research Laboratory, Department of Sociology, University of Alberta to measure aspects of quality of life.

7 For more information about Statistics Canada's low-income measures, please see Low Income Cutoffs from 1992 to 2001 and Low Income Measures from 1991 to 2000. Catalogue no. 75F0002MIE no. 005

Adult Health

This section asks the PMKs and their spouses about general health, chronic conditions, restriction of activities as well as questions on smoking and drinking. The smoking questions have been included because research has indicated that parental smoking behaviours may be predictive of the use of cigarettes by children. Alcohol consumption is covered because of potential impacts on the adult's physical or mental health, the family's economic situation, and family relationships.

Chronic Conditions

PMKs and their spouses are asked whether or not they have any long-term conditions (e.g. allergies, asthma, and high blood pressure). A derived variable (DCHPdD01 or DCHSdD01**)** indicates that the respondent answered "yes", they have at least one of the long-term conditions.

Restriction of Activities

PMKs and their spouses are asked a series of questions about whether or not their activities are restricted at home, work, school etc.. A derived variable (DRSPdD01 or DRSSdD01) is also created stating whether or not the PMK or spouse reported an activity restriction.

In Cycle 4, an answer category was added: "Yes, sometimes". This change matches the activity restriction questions asked on the 2001 Census.

Maternal History

This section is asked to determine the pregnancy history of mothers of children less than two years of age. These questions on pregnancy and birth were provided by Dr. J.-F. Saucier, Ste. Justine Hospital, Montreal, and later modified by the Project Team.

Depression Scale

A Depression scale (DDPPS01) was administered to the PMK as part of the Adult Questionnaire; see Chapter 9 for information about this scale.

Neighbourhood Safety

This section gathers information about the respondent's satisfaction with his/her neighbourhood as a place to raise children, including perception of the extent of danger and problems, and of social cohesion or "neighbourliness". Two scales are created in this section: Neighbourhood Safety Score (DSFHdS5), indicating the degree of perceived neighbourhood safety and the Neighbours Score (DSFHS6), indicating the degree of neighbour cohesiveness. See Chapter 9, for more information about these scales.

Socio-demographic Characteristics

The objective of the Socio-demographic Section is to gather information on immigration, ethnic background and the language profile of household members. This will allow for analysis of various components of the Canadian population and will permit identification of visible minorities. As well, there is a question on religious affiliation. These questions are asked of PMK, spouse and the child.

Due to a technical problem, no socio-demographic questions were asked to returning respondents in Cycles 3 or 4. The data on the file reflects the respondent's answers in Cycle 2. Users should be aware that information about language(s) spoken, in particular,

may have changed since we last asked these questions and should use these variables with caution.

8.5 Child Questionnaire

Education (Child)

The objective of this section is to get some basic information about the child's educational experiences. The amount and type of information collected varied depending upon the age of the child, with more information being collected for the older children who have had greater school experience.

Basic information is collected for all age groups, such as: the child's grade level, type of school and language of instruction, whether the child looks forward to school, absenteeism, number of school changes and residential moves.

For children in grade 1 or higher, additional questions are asked concerning other aspects such as skipping and repeating grades, achievement and special education.

Health (Child)

The objective of this section is to provide information on the child's physical health – general health, injuries, limitations and chronic conditions – and use of health services and medications.

For children four and five years old who report having a chronic condition, health status information on topics such as hearing, sight, speech and overall mental well-being is also collected. From this information a Health Status Index (HUI3) is calculated (DHLCcD2A). The HUI3 is a generic health status index that is able to synthesize both quantitative and qualitative aspects of health. The index, developed at McMaster University's Centre for Health Economics and Policy Analysis, is based on the Comprehensive Health Status Measurement System (CHSMS). It provides a description of an individual's overall functional health, based on eight attributes: vision, hearing, speech, mobility (ability to get around), dexterity (use of hands and fingers), cognition (memory and thinking), emotion (feelings), and pain and discomfort.

The scores of the HUI3 embody the views of society concerning health status. Each person's preferences are represented as a numerical value (typically between 0 and 1) for a given health state (Some of the worse states of health are often given values less than 0, indicating that the individual considers them to be worse than death.) This index is also used by the National Population Health Survey.

In Cycles 1 and 2 of the NLSCY, the HUI2 was used to calculate the Health Utility Index. In Cycle 3, the HUI was released using both HUI2 and HUI3. The HUI2 was a provisional measure and has been replaced by the HUI3. For Cycle 4, only the index calculated with the HUI3 has been released.

Medical/Biological

The Medical /Biological Section was completed for children in the zero to three age group. The major objective is to collect information on factors such as gestational age and birth weight. These factors have been shown to have a direct impact on a child's growth and development. For example, in the long term, underweight babies face higher risks of poor health as well as longer-lasting developmental difficulties.

For each child under two, the nature of the delivery, general health of the child at birth and the use of specialized services following the birth were collected in this section. The NLSCY also investigates the biological mother's pregnancy and delivery history, including policy-relevant topics such as the mother's breast-feeding experiences and prenatal lifestyle.

There were derived variables created for this section that bear note. Two variables were derived to indicate the gestational age of the child. DMDCD06 gives the gestational age in days and DMDCD07 indicates if the child was born prematurely (gestational age 258 days or less), in the normal range (gestational age 259 to 293 days) or late (gestational age 294 days or later).

A variable was derived (DMDCD08) to indicate if the child was of normal birth weight (2500 grams), moderately low birth weight (1500 to 2499 grams) or very low birth weight (< 1500 grams).

Work After Birth

These questions are asked to determine the time interval after which mothers returned to work following the birth of a child and the extent to which these mothers participated in the labour force upon their return.

Milestones

These questions were added in Cycle 4 to provide a better measure of early child development. Taken as a package, developmental milestones, such as when the child first said words or took first steps, provide a general sense of a child's development. Experts with the Dunedin study in New Zealand recommended to the Project Team that developmental milestones be used as a measure of development. The items are from the draft questionnaires for the Early Childhood Longitudinal Study Program (Birth Cohort) of the National Center for Education Statistics in the United States.

Temperament

This section measures the temperament of young children by asking the parent about the degree of difficulty their child presents them. This measure is based on the assumption that a child's temperament is influenced by the parent's perception of the difficulty of the child, and that temperament is not solely based on biological origins. More information about this section can be found in Chapter 9.

Literacy

This section measures children's exposure to books and their interest in reading and learning-related activities that parents do with their children. The focus of this section is the stimulation young children receive at home.

For children aged zero to two, several questions were added in Cycle 4 to measure how often the children do certain activities with their parents, such as tell stories, sing songs and teach new words. These questions are adapted from the Early Childhood Longitudinal Study in the United States.

Similar questions are asked about children aged three to five, with changes to reflect age appropriate activities. A question about number activities has been added on the suggestion of colleagues of Dr. Robbie Case, the developer of the Number Knowledge Test, who state that numeracy is a crucial factor when it comes to learning and literacy. This question was added to provide information on children's numeracy activities that may correlate with results on the Number Knowledge Test.

Communications

The items have been modified from the New Zealand Competent Children Study. They cover a child's ability to understand oral messages and to pass a message on to someone else, as well as to communicate verbally. The final question, about speech being easily understood, is only asked of three year olds. Four and five year olds are asked a similar question as part of the Health Status Index in the Health section.

Activities

This section measures the child's participation in various non-school activities and the amount of household responsibility taken on by 10 and 11 year olds at home. The latter questions are used to create the Home Responsibilities Score (DACCS6), indicating the degree of home responsibilities. The section will give some sense of how the child spends his/her time, of personal interests, as well as the degree of interaction with peers.

In Cycle 4, several questions were added for children six to nine years old to determine how often parents get to do certain activities with their children, such as eating a meal, playing a game, doing chores together. When there is a spouse/partner in the household, these questions are asked about both the PMK and his/her spouse/partner. These questions were added in Cycle 4 to help measure the type and frequency of activities the child does with both his/ her mother and father.

Behaviour

The objective of this section is to assess aspects of the behaviour of children two years of age and older and of feeding patterns for one to three year olds.

The questions in this section are used to measure the prevalence of behaviours such as hyperactivity and physical aggression. The scales derived from these questions are described in detail in Chapter 9.

Positive Behaviour

The objective of this new section is to assess positive behaviour of children aged three to five, including perseverance and independence. The New Zealand's Competent Children's Study has found that perseverance and independence were among a cluster of competencies that are good indicators of a child's overall performance.

Questions have been adapted from the New Zealand study and the behaviour questions used for other ages in the NLSCY.

Sleep

Research suggests that sleeping difficulties are predictive of a child's potential difficulties. Conversely, absence of such difficulties has been correlated with easy temperament and positive outcomes.

Several new questions, adapted from the Étude longitudinale du développement des enfants du Québec (ELDEQ) – Santé Québec, were added in Cycle 4. These questions asked about hours of sleep, hours of uninterrupted sleep at night, how often the parents sleep was disturbed by the child and so on.

Motor Social Development

The Motor and Social Development (MSD) Scale measures dimensions of the motor, social and cognitive development of children from birth to three years; the questions vary by the age of the child. Three scores (DMSCS01, DMSCS02 and DMSCdS03) are derived from these questions; the scores are explained in Chapter 9.

Relationships

The objective of this section is to provide information about the child's relationships with others. Positive relationships with other children and adults may help to counteract other factors that place a child at risk.

Questions about doing things with friends, and getting along with parents, teachers and friends are based on those in Ontario Child Health Survey.

Parenting

Parenting style is considered to have an important influence on child behaviour and development. The objective of this section is to measure certain parenting behaviours. Scales are created from the questions in this section. For more information about these scales, see Chapter 9.

In Cycle 4, a question was added asking PMKs who have a spouse/partner in the house how often the PMK and spouse/partner agree with each other about parenting decisions. This question was developed by the Project Team and is similar to questions in the Strayhorn and Weidham scale, from which the other parenting questions have been adapted.

Custody

This section was designed to provide information on the child's family arrangements; whether or not his or her parents are married, whether the child's parents were separated or divorced, the age of the child when parents divorced/separated and so on. This data will not be released in Cycle 4 as technical problems prevented it from being collected properly. This section has been redesigned for the next cycle of collection and data on children's family arrangements should be available in Cycle 5.

Child care

This section provides basic information about the methods of care currently provided for the child while the parents are working or studying, plus some information on previous care. Concepts measured include both the amount of time spent by the child in child care and the methods of care used for each child. In addition, information is obtained on the number of changes in child care arrangements that the child has experienced and the reason(s) for changes in the past 12 months. The section also identifies whether or not a child care centre is profit or non-profit and whether home care is licensed or unlicensed. In Cycle 4, questions were added to assess parental satisfaction with child care arrangements, the options parents considered by parents for child care, and a question on the ratio of care givers to children.

Ages and Stages Questionnaires

The *Ages and Stages Questionnaires (ASQ)* are parent-report instruments, developed by Jane Squires, LaWanda Potter, and Diane Bricker, at the University of Oregon, designed to identify infants and young children who show potential developmental problems. There are

19 questionnaires that cover the age range from 4 to 60 months. Each questionnaire includes about 30 items covering five domains of development:

- Communication: e.g., babbling, vocalizing, listening, and understanding
- Gross Motor: arm, body, and leg coordination
- Fine Motor: hand and finger co-ordination
- Problem Solving: doing different activities with objects, drawing
- Personal-Social: solitary and social play, dressing and feeding self

The questionnaires also include an overall section that asks about general parental concerns but is not used in the NLSCY, as these questions are similar to those already included in the survey.

The NLSCY is using the *ASQ's* for children aged four months to five years, inclusive. In consultation with the publisher, Statistics Canada has grouped the questionnaires into booklets to facilitate administration in the NLSCY. In Cycle 4, respondents received the appropriate booklet with their introductory letter. Respondents were asked to complete the correct questionnaire for their child before the interview. Responses from parents of four month to three year olds were collected over the telephone at the same time as the parent interview; questionnaires for four and five year olds were collected during the home interview.

8.6 Youth Questionnaire (ages 16 and 17)

Parent Report

Aspirations and Expectations

These questions were added in Cycle 4 to assess parental aspirations and expectations for their youth, and parental views on their youth's school experiences. Providing help with school work, discussing school experiences and future educational plans has been linked to school success.

These questions were developed by the Centre for Education Statistics at Statistics Canada, using NLSCY questions and questions from other education surveys, such as, the Youth in Transition Survey and the School Leavers Survey.

Youth Report (CAI)

Youth Education

This new section looks at the youth's education experience. This section is divided into four parts: school leavers (those who are not in school and have not graduated from high school), school finishers (those who are not in school and have graduated), currently in school (for youth still in high school), and post-secondary (for youth who are attending post-secondary education). The questions were developed by the Centre for Education Statistics at Statistics Canada using NSLCY questions and questions from other education surveys, such as, the Youth in Transition Survey and the School Leavers Survey.

One of the objectives of this section is to help determine the factors involved in youth choosing to continue their schooling or to leave school.

Youth Labour Force

The youth Labour Force section is intended to measure youth experience in the labour market. Some youth may be working part-time while attending school, while others may

have made the transition to the workforce. These questions are a mix of NLSCY questions from the self-completes for 14-15 year olds and of the adult labour force questions.

Youth are asked to report about current work, work during the current school year and work last summer.

Youth Income

The youth Income section asks the youth about their income from various sources in the last 12 months. These questions are similar to those asked of the parents. Income information can measure how much spending money youth have to make autonomous decisions.

Youth Health

This section asks about the youth's general health, injuries, chronic conditions and restriction of activities. These questions are similar to the child and adult health questions. New questions have been added about exercise and sleep.

Exercise and sleep are important indicators of the youth's attitude toward their bodies and how they take care of themselves. The amount of sleep reported can be used to help understand if youth are successfully balancing the demands of work, school, volunteering, sports, etc.

Youth Activities

Adolescence can be a time of high involvement in a variety of activities that are not school related. It is important to measure these activities to understand how this involvement can contribute to good outcomes. This section includes questions about physical activities, literacy activities, T.V. watching, computer use and community involvement. These questions have been adapted from the questions asked of younger adolescents.

Questions were also added about the youth's access to a vehicle and whether or not they have a driver's license. Driving is an important 'coming of age' activity for this age group.

Neighbourhood

Neighbourhood factors have been shown to influence child and adolescent outcomes in a variety of domains (school achievement, behaviour, emotional and social functioning, motor and social development). These effects increase as children move through the life course, increasing their interactions and exposure to extra-familial environments. This has been evidenced in the academic literature, as well as by research conducted using NLSCY data (Boyle and Lipman, Kohen et. al, Offord and Lipman). In Cycle 4, youth are asked about their perceptions of their neighbourhood. Some of the questions are similar to those asked in the Adult Questionnaire about neighbourhood safety. Others questions, about people in the neighbourhood, come from the UCLA Survey of Adolescent Experience.

The Neighbourhood Structure Score (DACYDS01) is calculated using these questions. A high score indicates a high degree of neighbourhood structure and a low score indicates a low degree of neighbourhood structure.

8.7 Self-completes (ages 10-17)

The objective of these questionnaires is to collect information directly from the youth on a variety of aspects of his/her life to supplement information obtained from the parent. The

questionnaire also collects information from the youth on subjects about which only the youth could reliably report. For 16-17 year olds, some information is still collected on the self-completes even though these youth are reporting their own information in the CAI portion of the interview. We felt that youth may be more comfortable answering sensitive questions on a paper questionnaire rather than face to face with an interviewer.

Friends and Family (ages 10 to 17)

The objective of this section is to determine how well the youth feels he/she gets along with others.

The section collects information on the extent and quality of the youth's social support network, such as numbers of close friends, time spent with friends, presence of someone the youth can confide in. The questions vary depending on the age of the youth. The questions were adapted from the Ontario Child Health Study and the NLSCY Child Questionnaire.

The Friends Scale (DFFCS01) is constructed from these questions; see Chapter 9 for more information.

This section also contains a measure of intimacy for the 14-15 year olds. This question, about how often the youth shared secrets and private feelings with close friends, was adapted from Furnman and Buhmester's Network of Relationships Inventory.

School (ages 10 to 15)

This section asks about the youth's attitude towards school, how well he/she is doing at school, importance of good grades and feelings of safety and acceptance at school, perception of the teacher with respect to fairness and providing extra help. For 14 and 15 years olds, there is a series of questions about school based extra-curricular activities, such as sports or drama. These questions have been modified by the Project Team from the following sources: Western Australia Child Health Survey, Northwest Territories' Health Attitudes, Knowledge and Behaviours Study, Marsh Self-Description Questionnaire, World Health Organization (WHO) Survey on Health Behaviours in School Children.

Attitudes about school may be an important influence on a youth's educational accomplishments. Research shows that a negative attitude towards school may be associated with poor school performance.

About Me (ages 10 to 17)

These questions are used to determine the youth's overall self-esteem. A score is calculated (DC1CS02) based on the answers to these questions. Please see Chapter 9 for more information on this scale.

For youth aged 12 to 17, additional questions are asked about youths' feelings about life now and in the future. These questions are from the Western Australia Child Health Survey.

Youth aged 14-17 are also asked about painful events, such as a break-up with boyfriend/girlfriend or death of someone close to them.

Feelings and Behaviours (ages 10 to 17)

Behaviour Checklist (ages 10 to 15)

This section replicates the behaviour checklist used in the parent-report CAI Child Questionnaire. It provides indicators of the following behaviours: conduct disorder, hyperactivity, inattention, physical aggression, indirect aggression, emotional disorder, anxiety, and prosocial behaviours. Scores for these behaviours are also created; see Chapter 9 for more information about the scales.

Risky Behaviours (ages 10 to 17)

These questions about risky behaviours, such as staying out all night without permission, are also replicated from the Child Questionnaire. The questions are expanded for the older age groups to capture behaviours that may become more common as the youth get older. These questions were adapted by the Project Team from the following sources: National Longitudinal Survey of Youth at Ohio State University, Western Australia Child Health Survey and Dr. Richard Tremblay of the University of Montreal.

Suicide (ages 12 to 17)

This section includes questions about suicide, including whether the youth knows anyone who has committed suicide, whether they have seriously considered or attempted suicide. These questions were adapted from the 1992 British Columbia Adolescent Health Survey.

Depression (ages 16 and 17)

These youth are asked about feelings of depression, using the same questions asked of the PMKs. These questions and the scale derived from them are described in Chapter 9.

My Parent(s) (ages 10 to 17)

The objective of this section is to gather information from the youth about his/her perception of the relationship with his/her parents.

My Parents and Me (ages 10 to 15)

Three scales are created using these questions: Parental Nurturance (DPMCcS1), Parental Rejection (DPMCbS2b) and Parental Monitoring (DPMCcS3). These scales are described in Chapter 9.

Conflict Resolution Scale (ages 16 to 17)

These questions replicate those asked of parents of 12 to 15 year olds. For 16 and 17 year olds, the questions are asked separately about the youth's mother and father. Two scores are derived from these questions – Conflict Resolution Scale – Mother (DPMCdS4) and Conflict Resolution Scale – Father (DPMCdS5). These scales are described in the next chapter.

Puberty (ages 10 to 17)

Puberty is an important marker of physical development. This section asks the youth about key physiological indicators and their perceptions of their own puberty. These questions were provided by Dr. Richard Tremblay at the University of Montreal.

Note: For youth aged 12-17, these questions are included in the Health section

Smoking, Drinking and Drugs (ages 10 to17)

This section asks questions to determine if the youth has used cigarettes, alcohol or drugs and the extent of usage. The behaviours have been correlated with negative behaviours and outcomes, such as delinquent behaviours and poor school performance. The questions vary by age.

The smoking questions are adapted from the Youth Smoking Survey, the WHO Survey on Health Behaviours in School Children and the Western Australia Child Health Survey.

The questions on alcohol were adapted from the Western Australia Child Health Survey and from questions provided by Dr. Richard Tremblay of the University of Montreal.

The questions on the use of drugs and addictive substances were adapted from the N.W.T. Health Attitudes, Knowledge and Behaviours Study.

Questions on driving under the influence of drugs and alcohol and being a passenger in a car with a drunk driver have been added for 16-17 year olds, as it is the first time they have the ability to drive. These are important risk-taking activities in this age group. The questions have been adapted from the North Carolina Evaluation of School-Based Health Centers.

Activities (ages 10 to 15)

The objective is to determine the youth's extent of participation in activities outside of school hours and use of free time. Activities include sports, arts, dance or music, Guides or Scouts, jobs and volunteering. Reading for pleasure, using a computer and watching T.V. are also covered. Generally, the activities are also covered on the CAI parent-report Child Questionnaire for children under 10.

Literacy activities (ages 14 and 15)

These questions ask about how often youth engage in literacy activities outside of school, such as using a library or reading. These questions are similar to those asked of the PMK for younger children and of the 16-17 year olds in the CAI questionnaire.

Health (ages 12 to 17)

Youth are asked to report on their height and weight, symptoms of stress, use of seatbelts and helmets, healthy eating and dating. The questions vary with age.

The questions on physical indicators of stress were adapted from the WHO Health Behaviours in School Children Survey.

Use of seatbelts and helmets questions were modified from the U.S. Youth Risk Behaviour Survey, which were used in the 1992 British Columbia Adolescent Health Survey.

Questions on dating and sexual activity were adapted by the Project Team from various adolescent questionnaires such as the Minnesota Adolescent Health Survey and the British Columbia Adolescent Health Survey.

Work and Money (ages 12 to 15)

Youth are asked about their work during the school year and those aged 14-15 are asked about work last summer. The 14 and 15 year olds are asked more detailed questions about their job(s), such as hours worked and pay. They are also asked about whether work reduces the amount of time they spend studying.

Youth are also asked about how much money they received from various sources, such as parents and work.

These questions were developed by the Project Team after reviewing several other surveys.

Dating (ages 12 to 17)

This section asks youth about their experiences with a boyfriend / girlfriend and their sexual activity. The question about sexual behaviour on the 12-13 year old questionnaire was modified from the Youth and Aids Survey. Questions are also asked about contraceptive use and, for the 16-17 year olds, reasons for abstaining from sex or reasons for not using birth control. These questions were designed by the content team in consultation with experts from youth surveys such as BC Adolescent Health Survey and Minnesota Adolescent Health Survey.

Decision Making (ages 16 to 17)

These questions were added to measure the youth's decision making style. The items in these scales come from the 30-item Identity Style Inventory (ISI3) developed by Michael D. Berzonsky in the Department of Psychology at the State University of New York at Cortland. See Chapter 9, for information about these scales.

9.0 Validation of the survey scales

9.1 Validation of scale data

9.1.1 Scale Definition

For some of the concepts deemed important to measure in the NLSCY it was decided that the concept would most appropriately be measured through the use of a scale. A scale is simply a group of questions or items that measure a certain concept when the answers to the items are put together.

For example, on the child's questionnaire, it was determined that it was important to have an assessment of certain parenting behaviours. The scale is intended to measure three different constructs or factors related to parenting; positive interaction, ineffective parenting and consistent parenting.

9.1.2 Scales and Calculations

For each factor measured by a scale, a score is calculated. The score for a particular factor can be used to give an ordering of individuals. For example, for the Parenting Scales, for children with higher scores for the "positive interaction" factor, the PMK reported having more positive encounters with the child (e.g., laughed with them more, praised them more etc.). The score for a particular factor is usually based on a series of items, since one single item usually cannot measure the factor or construct with adequate precision.

During the development of the NLSCY, when consideration was being given to what scales should be used to measure a particular concept, an attempt was made to select scales that had been used in other studies. In this way, the psychometric properties of the measures produced by each scale were available with complete references.

9.1.3 Evaluation of Scale Data

In many instances, the wording of certain questions in the original scale was modified and in some cases new questions were added. Sometimes the scale that was used had not previously been used for children in Canada, or had only been used for very small samples. Given these concerns and further concerns regarding interviewing conditions, it was felt that the factor structures of the scales used in the NLSCY could be different from the ones given in the literature. Therefore the project team felt the need to carry out an extensive evaluation of the scale data to ensure that the psychometric properties found in other studies also held true for the NLSCY experience.

There were three major steps in the analyses of the scale data. First a new factor analysis was performed on all scales to determine the constructs or factors inherent in each scale. Then scale scores were calculated based on this factor structure. Finally reliability measures were produced. The general procedures followed for each of these steps are described in detail in the following pages.

NOTE: Many of the scales were developed and validated in Cycle 1. In subsequent cycles, the same factor structure which emerged from the Cycle 1 analysis was

imposed. Imposing the same factor structure ensures that the scales are consistent across time to allow for longitudinal analysis and cross-sectional comparisons. Each scale has a note indicating in which cycle the validation was performed.

9.2 Factor analyses

9.2.1 Factor Analysis for Scales

The factor structure of each scale was determined based on data from the first cycle. The factor structure imposed on the scales already used in the first cycle and repeatedly utilized in the second, third and fourth cycles of the survey was the result of analyses of data from the first cycle. This analysis was redone after cycle 4 and it is these results that are summarized below. For the results from the cycle 1 factor analysis please refer to the cycle 1 User Guide.

1. The sample of respondents for each scale (and age group, if the scale used different questions for different groups), was randomly divided into two half-samples. This was done to find out whether different samples would yield the same results.

2. Principal component analysis was carried out separately on each half-sample to find out how many factors should be extracted in the subsequent factor analysis. In principle, the same number of factors as was found in the literature was expected. In practice, however, some scales showed a different number of factors because in some cases factors combined while in others new factors emerged.

3. Factor analysis was done on each half-sample and the factor structure and loading of each factor were compared across the half-samples.

4. In the factor analysis, the items for each child in the appropriate age group were used, multiplied by the child's normalized weight. An individual's statistical weight is normalized by dividing his/her weight (DWTCW01C) by the average weight for all individuals. Thus, the sum of the normalized weights is equal to the sample size.

5. Once the factor structures were analysed and the items included in each factor were determined, scores were calculated. To produce the scores, 1 was subtracted from each item so that the lowest possible score would be 0. A score of 0 indicates that the child has no problems for all factors in the behaviour scale except for the Prosocial factor, where a score of 0 indicates the absence of prosocial behaviour. Some items were imputed. The imputed values were computed by a procedure (the SAS PRINQUAL procedure) that determines which of the possible values for an item is the most plausible for an individual in view of his/her response profile, the response profiles of others in the sample, and the number of factors included in the analysis.

6. The score for each factor on the scale was derived by totaling the values of the items that made up that factor (including imputed values). The score was set to "missing" if too many of the values of any items included in the factor were unreported. A value may be missing if the parent refused to answer or did not know the answer to the item.

9.2.2 Data Transformation Using Optimal Scaling

Before performing the factor analysis for each of the NLSCY scales, the data were transformed using optimal scaling. The method used was one proposed by Young and several associates (Young, 1981) which is a variant of Fisher's optimal scaling technique. The method is presented as a means of transforming data that are fundamentally nominal or ordinal in nature to interval or ratio level data so that statistical techniques which are appropriately applied only to interval and ratio data may be utilized.

9.2.3 Factor Analysis Using Weighted Data

Factor analysis requires that the data have the property of interval or ratio data, meaning that the distance between each answer category of the question should be the same. For example, in scales where the answer choices are: Never, Sometimes, Often, and Always, one must assume that the distance between Never and Sometimes is the same as that between Sometimes and Often in the respondent's perception. It was felt that this was not necessarily true in the case for the scales used in the NLSCY.

9.3 Calculation of scores and item imputation

9.3.1 Calculation of Scores for Each Factor

The results of the factor analyses were used to determine what items "loaded" into each factor (i.e. were a part of each factor). The next step was to calculate a score for each factor. This was done by summing the values for each individual item that made up the factor. In some cases some rescaling of values was done before the final score was calculated. The following example illustrates how factor scores were computed.

9.3.2 Example of Factor Score Computation

One of the constructs that emerged in the factor analysis for the Parenting Scale on the Child's Questionnaire was the ineffective parenting factor (Age 2-11 Years). In the factor analysis on cycle 1 data seven items were found to load into this factor.

DPRCQ04	How often do you get annoyed with your child for saying or doing something he/she is not supposed to?
DPRCQ08	Of all the times you talk to your child about his/her behaviour, what proportion is praise?
DPRCQ09	Of all the times you talk to your child about his/her behaviour, what proportion is disapproval?
DPRCQ13	How often do you get angry when you punish your child?
DPRCQ14	How often do you think the kind of punishment you give your child depends on your mood?
DPRCQ15	How often do you feel you have problems managing your child in general?
DPRCQ18	How often do you have to discipline your child repeatedly for the same thing?

The answer categories for these items were of two types:

1 - never
2 - about once a week or less
3 - a few times a week
4 - one or two times a day
5 - many times each day

1 – never
2 - less than half the time
3 - about half the time
4 - more than half the time
5 - all the time

In the calculation of the score for this ineffective parenting factor, the categories were rescaled to 0 to 4 (i.e., the category "never" was scored as 0, the category "about once a week or less/less than half the time" was scored as 1, ... and the category "many times each day/all the time" was scored as 4). In order to compute the score, these values were summed across the seven items involved in the factor resulting in a ineffective parenting score in the range 0 to 28. A low score of 0 represents the absence of a problem and a high score of 28 indicates a high degree of problems. For most of the scores calculated for the NLSCY, a score of 0 represents the absence of a problem. However there are exceptions to this which are noted in the documentation for each particular scale.

9.3.3 Negative Loading

Note that the second item that loaded into the ineffective parenting factor, DPRCQ08 (Of all the times you talk to your child about his/her behaviour, what proportion is praise?) is in the opposite direction compared to the other items. In fact the item loaded "negatively" into the factor. Therefore when computing the score the values for this item were reversed - all the time was scored as 0, more than half the time as 1, ... and never as 4. In the documentation for each scale any item that was reversed for the scoring algorithm due to a negative loading is indicated.

9.3.4 Non-Response Code

The score for the ineffective parenting factor is labelled as DPRCS04 on the record layout for the micro data file. An "S" in the 5th position of the variable name indicates a score.

When the score was being calculated for each factor there was a possibility that one or more of the items making up the score had a non-response code (don't know, refusal or not-stated). If the number of items with a non-response code was above a certain threshold, the factor score was set to not-stated. Generally this threshold value was set at 10% of the items.

9.3.5 Raw Items

It should be noted that in addition to the scores, the raw items for each scale are included on the micro data file. This will allow researchers to consider alternate factor structures if desired. For the raw items the original values (in the 1 to 5 range for the parenting scale) have been retained before any rescaling or reversal of values took place.

9.4 Reliability measures for scales

Reliability refers to the accuracy, dependability, consistency or ability to replicate a particular scale. In more technical terms, reliability refers to the degree to which the scale scores are free of measurement error. There are many ways to measure reliability.

9.4.1 Cronbach's Alpha

One of the most commonly used reliability coefficients is Cronbach's alpha (Cronbach, 1951). Alpha is a measure of the internal consistency of the items within the factor. It is based on the average covariance of items within the factor. It is assumed that items within a factor are positively correlated with each other because they are attempting to measure, to a certain extent, a common entity or construct.

9.4.2 Interpretations of Cronbach's Alpha

Cronbach's alpha has several interpretations. It can be viewed as the correlation between the scale or factor and all other possible scales containing the same number of items, which could be constructed from a hypothetical universe of items that measure the characteristic of interest. For example, in the ineffective parenting factor, the seven questions included in the scale can be viewed as a sample from the universe of all possible items. Parents could also have been asked: "How often do you raise your voice when you discipline your child?" or "How often do you threaten punishment more often than you use it?" Cronbach's alpha indicates how much correlation can be expected between the scale which was used and all other possible seven-item scales measuring the same thing.

Another interpretation of Cronbach's alpha is the squared correlation between the score an individual obtains on a particular factor (the observed score) and the score he/she would have obtained if questioned on all possible items in the universe (the true score). Since alpha is interpreted as a correlation coefficient, it ranges from 0 to 1.

Generally, it has been shown that alpha is a lower bound to the reliability of a scale of n items (Novick and Lewis, 1967). In other words, in most situations alpha provides a conservative estimate of a score's reliability.

9.4.3 What is a satisfactory level of reliability?

It is difficult to specify a single level that should apply in all situations. Some researchers believe that reliabilities should not be below 0.8 for widely used scales. At that level, correlations are affected very little by random measurement error. At the same time, it is often very costly in terms of time and money to obtain a higher reliability coefficient. It should be noted that for some of the factors for which scores were computed for the NLSCY, the reliabilities are below this level. The Cronbach's alpha is given in the documentation for each score that has been calculated. Researchers can determine for themselves whether or not the score has adequate reliability for their specific purposes.

Finally, it should be mentioned that for the NLSCY the Cronbach's alpha for each factor score was computed using SAS. Typically, the alpha coefficients calculated using SAS are lower than those calculated using SPSS.

9.5 Parent-reported and child-reported scales

The remainder of this chapter provides and in-depth description of the sources of the NLSCY scales and all analytical results of factor and reliability analysis. Changes made to the scales across cycles are also described. For convenience, the scales are listed in alphabetical order. The table below provides a brief summary of the NLSCY scales followed by individual descriptions of each scale.

9.5.1 Parent-reported scales

Behaviour Scale
The objective of the behaviour scale is to assess aspects of the behaviour of children two years of age and older.

Initially, an attempt was made to measure the following behaviours for children aged 2 and 3:

- hyperactivity,
- emotional disorder,
- anxiety,
- physical aggression,
- inattention,
- prosocial behaviour,
- separation anxiety and
- opposition

For children between 4 and 11 years of age, an attempt was made to measure similar behaviours to the 2 to 3 year olds; separation anxiety and opposition behaviours were omitted, while indirect aggression, conduct disorder and property offences were added.

Theoretical Constructs
Below are the theoretical constructs that were used for the factor analysis. The actual scales that emerged from the analysis vary from these constructs.

Separation anxiety (2 and 3 year olds)
Includes items DBEC6CC1, 6DD1, 6LL1, 6PP1 and Q6TT1 from Achenbach's Child Behavior Checklist (CBCL).

Opposition (2 and 3 year olds)
Includes items DBECQ8E1, Q6G1, Q6R1 and Q8T1 drawn from Achenbach's CBCL.

Conduct disorder (2 - 11 year olds)
Includes items DBECQ6G, Q6X, Q6AA, Q6FF, Q6JJ and Q6PP from the Ontario Child Health Study (OCHS).

Hyperactivity (2 - 11 year olds)
Includes items DBECQ6B, Q6I, Q6P, Q6S and Q6W from the OCHS and Q8HH and DBEC6QQ from the Montreal Longitudinal Survey. In previous cycles, item CBECQ6N was included in this construct. A decision was made to drop this item from cycle 4 and all future cycles as respondents found it to be too repetitive.

Emotional disorder (2 - 11 year olds)
Includes items DBECQ6F, Q6K, Q6Q, Q6V, Q6CC, Q6MM and Q6RR from the OCHS. Anxiety includes NLSCY items taken from OCHS emotional disorder items (DBECQ6F, Q6Q, Q6V and Q6CC). In previous cycles the items CBECQ6II and CBECQ6Y were included. A decision was made to remove both items from cycle 4 and all future cycles.

Indirect aggression (2 - 11 year olds)
Includes items DBECQ6J, Q6R, Q8Z1, Q6LL and Q6TT from Lagerspetz, Bjorngvist and Peltonen of Finland.

Physical aggression (2 - 3 year olds and 8 - 11 year olds)
Includes items DBECQ6X from the Montreal Longitudinal Survey and DBECQ6G, Q6AA and Q6NN from the OCHS.

Inattention (2 - 11 year olds)
Includes items DBECQ6P from the OCHS and DBECQ6QQ from the Montreal Longitudinal Survey.

Prosocial behaviour (6 - 11 year olds)
Includes items DBECQ6A, Q6H, Q6M, Q6GG and Q6OO from the OCHS and DBECQ6D, Q6U, Q6BB, Q6SS and Q6UU from the Montreal Longitudinal Survey; the last four items are from a scale devised by K. Weir and G. Duveen. In cycles 1 through 3, these items were asked of all children aged 4 to 11. In Cycle 4, all 4 and 5 year olds were excluded from this scale and were asked the questions in the positive behaviour section instead.

Factor Analysis for the Behaviour Scale
The following indicates the items that were included on the questionnaire to measure these various constructs of behaviour. A complete factor analysis was carried out for the behaviour scale to assess the psychometric properties of this scale for the NLSCY population. As part of this analysis, the items that loaded into each construct or factor were compared to the expected result described below. The results of this analysis are presented later on in this section.

Analysis of children aged 2 and 3 years
There were 7,122 two- and three-year-olds in the sample. The group was split into two sub-samples of 3,477 and 3,645 individuals, and the analysis for this age group was performed separately for each sub-sample. The non-response rate for most items was about 2.0%. Some individuals were excluded from the analysis that produced the factors. The exclusion criteria were as follows: individuals with 10% or more items coded "missing" were not included in the analysis. After the criteria were applied, there were 3,413 and 3,565 individuals left in the sub-samples to be analysed.

FACTOR	SCORE	ITEMS
Hyperactivity – inattention	DBECdS01	DBECQ6B, DBECQ6I, DBECQ6P, DBECQ6S, DBECQ6QQ, DBECQ8HH
Emotional disorder – anxiety	DBECdS03	DBECdQ6F, DBECQ6K, DBECQ6Q, DBECQ6V, DBECQ6CC, DBECQ6MM, DBECQ6RR
Physical aggression – opposition	DBECS04	DBECQ6G, DBECQ6W, DBECQ6X, DBECQ6NN, DBECQ6R1, DBECQ8T1, DBECQ8Z1, DBECQ8E1
Separation anxiety	DBECS05	DBECQ6CC, DBECQ6DD1, DBECQ8PP1, DBECQ8LL1, DBECQ8TT1

Cronbach's alpha for children 2 and 3 years
Cronbach's alpha (raw value) was computed with SAS using normalized weighted data.

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Hyperactivity-inattention (DBECdS01)	0.739	DBECQ6P	0.684
Emotional disorder-anxiety (DBECdS03)	0.600	DBECQ6MM	0.518
Physical aggression-opposition (DBECS04)	0.716	DBECQ8Z1	0.677
Separation anxiety (DBECS05)	0.584	DBECQ6DD1	0.460

Analysis of children aged 4 to 11 years
There were 13,765 children in the 4 to 11 age group. Two sub-samples of 6,830 and 6,935 were created for analysis. The item non-response rate was approximately 3.5% for most of the 47 items involved in the analysis. Individuals were excluded from the analysis if there were 10% or more items coded "missing," or refused. After this criteria was applied 6,574 and 6,681 individuals remained in the sub-samples to be analysed

FACTOR	SCORE	ITEMS
Hyperactivity – inattention	DBECdS06	DBECQ6B, DBECQ6I, DBECQ6P, DBECQ6S, DBECQ6W, DBECQ6QQ and DBECQ8HH
Emotional disorder – anxiety	DBECdS08	DBECQ6F, DBECQ6K, DBECQ6Q, DBECQ6V, DBECQ6CC, DBECQ6MM and DBECQ6RR
Physical aggression – conduct disorder	DBECdS09	DBECQ6G, DBECQ6X, DBECQ6AA, DBECQ6FF, DBECQ6JJ and DBECQ6NN
Indirect aggression	DBECS10	DBECQ6J, DBECQ6R, DBECQ6Z, DBECQ6LL and DBECQ6TT

Cronbach's alpha for children 4 to 11 years
Cronbach's alpha (raw value) was computed with SAS using normalized weighted data.

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Hyperactivity-inattention (DBECS06)	0.815	DBECQ6P	0.774
Emotional disorder – anxiety (DBECdS08)	0.736	DBECQ6V	0.684
Physical aggression – conduct disorder (DBECdS09)	0.772	DBECQ6AA	0.714
Indirect aggression (DBECS10)	0.766	DBECQ6LL	0.711

NOTE: The scores for these factors could not be computed in, 510, 501, 501, and 1,031 cases respectively because of unreported values.

Analysis of children aged 6 to 11 years
There were 7,687 children in the 6 to 11 age group. Two sub-samples of 3,751 and 3,936 were created for analysis. The item non-response rate was approximately 3.8% for the items involved in the analysis. Individuals were excluded from the analysis if there were 10% or more items coded "missing," or refused. After this criteria was applied 3,407 and 3,578 individuals remained in the sub-samples to be analysed

FACTOR	SCORE	ITEMS
Prosocial Behaviour	DBECdS07	DBECQ6A, DBECQ6D, DBECQ6H, DBECQ6M, DBECQ6U, DBECQ6BB, DBECQ6GG, DBECQ6OO, DBECQ6SS and DBECc6UU

Cronbach's alpha for children aged 6 to 11 years

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Prosocial behaviour (DBECdS07)	0.831	DBECQ6SS	0.804

Analysis of children aged 8 to 11 years
There were 4,399 children in the 8 to 11 age group. Two sub-samples of 2,186 and 2,213 were created for analysis. The item non-response rate was approximately 2.8% for the 6 items involved in the analysis. Individuals were excluded from the analysis if there were 10% or more items coded "missing," or refused. After this criteria was applied 2,081 and 2,093 individuals remained in the sub-samples to be analysed

FACTOR	SCORE	ITEMS
Property offences (8-11years)	DBECdS11	DBECQ6C, DBECQ6E, DBECQ6L, DBECQ6T, DBECQ6DD and DBECQ6PP

Cronbach's alpha for 8 to 11 Year Olds
Cronbach's alpha (raw value) was computed with SAS using normalized weighted data.

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Property offences (DBECdS11)	0.623	DBECQ6C	0.544

NOTE: The scores for these factors could not be computed in 225 cases because of unreported values.

9.5.2 Depression scale (PMK)

Objectives & Overview
The depression scale was administered to the PMK as part of the Parent Questionnaire. Questions for this scale (DDPPQ12A to DDPPQ12L) are a shorter version of the depression rating scale (CES-D), comprising 20 questions, developed by L. S. Radloff of the Epidemiology Study Center of the National Institute of Mental Health in the United States. This rating scale is used to measure the frequency of symptoms in the public at large. The occurrence and severity of symptoms associated with depression during the previous week are measured. The rating scale was reduced to 12 questions by Dr. M. Boyle of the Chedoke-McMaster Hospital of McMaster University.

This rating scale is aimed at gathering information about the mental health of respondents, with particular emphasis on symptoms of depression. Several members of the NLSCY advisory group of experts pointed out that the best way of proceeding was to measure one particular aspect of the PMK's mental health instead of trying to measure overall mental health. It was proposed that this section focus on depression for the following reasons: depression is a prevalent condition; it has been demonstrated that depression in a parent affects the children; present research on this subject is generally based on demonstration groups and not on population samples; and it is felt that introducing policies in this area could make a difference.

Items Included in the Depression Rating Scale
The depression rating scale includes twelve questions, each of which contains four response categories. In order for the lowest score value to be 0, the value for each question was reduced by 1 in calculating the score. As well, the answer categories were reversed for questions having a negative loading (DDPPQ12F, DDPPQ12H, and DDPPQ12J). The total score (DDPPS01) may therefore vary between 0 and 36, a high score indicating the presence of depression symptoms.

The factor structure of each scale was determined based on data from the first cycle. The factor structure imposed on the scales already used in the first cycle and repeatedly used in subsequent cycles of the survey was the result of analyses done based on data from the first cycle.

Analytical Results (Based on Cycle 1 data)
In analysing this scale, unweighted data were used. The sample size was 13,439 PMK's. However, once the observations containing mostly missing values were eliminated, the analysis dealt with only 13,140 PMK's. The non-response rate for the various questions in the rating scale was roughly 2.0%, whereas for the total score, a non-response rate of 2.2% was obtained. There was no imputation for the variables in this rating scale.

Cronbach's alpha results (Based on Cycle 1 data)
In spite of the possibility of extracting more than one factor from the depression rating scale, a single-factor analysis was used since the interest was in developing a global depression index. Following the analysis, the 12 variables of the scale were all kept as components of this factor since all 12 loading values met the established threshold.

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Depression (DDPPS01)	0.820	DDPPQ12D	0.790

9.5.3 Family functioning scale

Objectives & Overview
Questions related to family functioning, i.e., DFNHQ01A to DFNHQ01L, were developed by researchers at the Chedoke-McMaster Hospital of McMaster University and have been used widely both in Canada and abroad. This scale is used to measure various aspects of family functioning, (e.g. problem solving, communications, roles, affective involvement, affective responsiveness and behaviour control).

Question DFNHQ01M, drawn from the Follow-up to the Ontario Child Health Study, was added to the original scale to determine whether alcohol consumption had an effect on global family dynamics. However, it was not used in the analysis of the scale.

This scale is aimed at providing a global assessment of family functioning and an indication of the quality of the relationships between parents or partners. For this reason and because of the small number of questions, no attempt was made to measure the various aspects of family functioning.

Other surveys have shown that the relationship between family members has a considerable effect on children. The results of the Ontario Child Health Study have shown, for example, that there is an important link between family dysfunction and certain mental conditions in children.

Administering the Family Functioning Scale
The family functioning scale was administered to either the PMK or the spouse/partner as part of the Parent Questionnaire. The scale includes twelve questions, each of which contains four response categories. In order for the lowest score value to be 0, the value of the categories was reduced by 1 in calculating the score. The order of the categories was reversed for questions having a negative loading (DFNHQ01A, Q01C, Q01E, Q01G, Q01I, and Q01K). The total score (DFNHS01) may therefore vary between 0 and 36, a high score indicating family dysfunction.

Analytical Results (based on Cycle 1 data)
In analysing this scale, unweighted data were used. The non-response rate for the different variables was between 1.3 and 1.4%, whereas for the total score, a non-response rate of 1.9% was obtained. There was no imputation for the variables in this scale.

Cronbach's Alpha for Family Functioning Scale (based on Cycle 1 data)

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Family Functioning (DFNHS01)	0.880	DFNHQ01L	0.870

Distribution of Values for the Family Functioning Scale
When the values for the factor score for the family functioning scale are examined for the NLSCY children, the distribution that is observed is not a continuous one. In fact, the most common score is 12. This is a result of the fact that there are 12 items in the scale and four possible rescaled values (0 to 3). Many respondents had a rescaled score of 1 for every item in the scale and thus an overall score of 12. This means that the respondent answered "agree" to all of the items in the scale which were positive and "disagree" to all of the negative items, as opposed to the more extreme answers of "strongly agree" or "strongly disagree."

9.5.4 Home responsibilities scale

Objectives & Overview
The object of the activities scale is to measure the child's participation in home responsibilities.

This set of questions about responsibilities is from the Home Observation for Measurement of the Environment-Short Form questionnaire in the National Longitudinal Survey of Youth, Ohio State University.

NOTE: In Cycle 1, these questions were only asked of 10 and 11 year olds as they were the eldest age group. In subsequent cycles these questions were asked of all children 10 to 13 years of age.

Analytical Results (Based on Cycle 1 data)
In the cycle 1 sample there were 3,434 children aged 10 or 11 years. They were divided into two sub samples of size 1,705 and 1,729 and an analysis was done on each sample. The non-response rate for the 5 items was 1.3%. Individuals with missing values were excluded from the analysis that was conducted for the purpose of constructing the factor. After these exclusions, the sub-samples contained 1,680 and 1,709 individuals respectively. No imputation took place. As a result of factor analysis, one factor was identified: the activities factor (DACCS6). Items DACCQ6A to DACCQ6E loaded into the factor.

Scale Score
To produce the score, 1 was subtracted from each item so that the lowest score would be 0. The values for each item were reversed so that a high score would indicate a high degree of home responsibilities. The final score was derived by totalling the values of all items with non-missing values. The score ranges from 0 to

15. A score of 0 indicates the respondent does not participate in home responsibilities.

Once the factor structures were analysed and the items included in the factor was determined, the score was calculated. No imputation was done on the values. If any values were missing the final score was set to missing. A value may be missing if the child refused to answer or did not know the answer to the question.

Cronbach's Alpha for Home Responsibilities Scale*

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Home responsibilities (DACCS6)	0.778	DACCQ6B	0.705

NOTE: The final activities score could not be calculated for 45 (1.3%) individuals, due to missing values for the items comprising this factor.
**The variable names have been changed to indicate that they are in Cycle 4 but the analysis presented above was done in Cycle 1.*

9.5.5 Cycle 4 Motor and social development score

Motor and Social Development Section

The Motor and Social Development Section of the Child's Questionnaire was completed for children in the 0 to 3 age group. The objective was to measure motor, social and cognitive development of young children. A scale, made up of 48 questions (DMSCQ01 to DMSCQ48), was used to assess these concepts. According to the age in months, 15 questions were asked of each child.

The Motor and Social Development (MSD) Scale

The Motor and Social Development (MSD) Scale was developed by Dr. Gail Poe of the U.S. National Center for Health Statistics. The MSD scale consists of a set of 15 questions that vary by the age of the child, which measure dimensions of the motor, social and cognitive development of young children from birth through 3 years. Each item asks whether or not a child is able to perform a specific task. The scale has been used in collections of the National Longitudinal Survey of Youth in the United States and in recent versions of the National Child Development Survey in England.

The following table shows which questions were asked to each age group.

Age in Months	MDS Questions
0 to 3	DMSCQ01 to DMSCQ15
4 to 6	DMSCQ08 to DMSCQ22
7 to 9	DMSCQ12 to DMSCQ26
10 to 12	DMSCQ18 to DMSCQ32
13 to 15	DMSCQ22 to DMSCQ36
16 to 18	DMSCQ26 to DMSCQ40
19 to 21	DMSCQ29 to DMSCQ43
22 to 47	DMSCQ34 to DMSCQ48

Standardized Scores

A raw score was calculated for each child by summing the number of "yes" answers to each item in the scale (DMSCS01). Due to a problem with the application question 26 was not asked of the 7 and 8 month olds. As a result these children have a raw score that has a maximum of 14. Using data from previous cycles it was noted that in at least 93% of cases children of these ages responded “no” to this question. As well, the children who would have responded “yes” would still have the highest scores for this age group even without taking that question into consideration. Therefore no adjustment was done to compensate for this error.

Although there were different sets of questions depending on the age in months of the child, differences were observed when comparing score within these age bands. For example, there was a specific set of questions for children 4 to 6 months old. It was found that children who were 6 months old had scores that were on average higher than those 4 months olds. Therefore a decision was made to produce standardized scores. These scores, calculated for each age in month, would make it possible to compare scores across ages. All children, aged 3 years or less, received a standardized score based on Cycle 1 data and the children aged 3 – 47 months also received a standardized score based on the Cycle 4 data.

Standardized Scores Using Cycle 4 Norms

Each child 3 to 47 months old was assigned a standard score. This standardization was done by 1 month age groups. For each month age group the mean and standard deviation of the raw score was found and were used to produce a normalized score with a mean of zero and a standard deviation of one. This score was adjusted such that the mean MSD score was 100 and the standard deviation was 15. Therefore children who are 3 months old have an average MSD score of 100, children who are 4 months old have an average MSD score of 100, ..., and children 47 months old have an average MSD score of 100.

Once these scores were calculated children who were more than 3 standard deviations away from the mean (scores smaller than 55 or greater than 145) were identified, and the norms were recalculated not including these children. These children were considered outliers and are not representative of other children their age. Therefore the average of MSD scores on the data file by age in months may not be exactly 100. Using this standardized score (DMSCdS03) makes it possible to compare scores of children across the 0 to 3 age group, not controlling for age.

This score was not calculated for children aged 0-2 months as there were not enough respondent children by age in months to establish a norm.

Standardized Scores Using Cycle 1 Norms

A second standardized score (DMSCS02) was calculated for all children 0-47 months. This score was calculated in the same way as mentioned above; only the norms were derived using the data from Cycle 1, and then applied to the Cycle 4 data.

Overall there are no major differences found when comparing the scores found using Cycle 4 norms and the scores found using Cycle 4 norms. The score calculated using Cycle 1 norms should be used in order to compare scores over cycles. This score is available for all cycles of data.

The Motor and Social Development Scale question have remained unchanged throughout the four cycles of the National Longitudinal Survey of Children and

Youth, but there have been changes to the calculation of the final scores. For more information on these changes please refer to the *Appendix on Revisions to Previous Cycles.*

9.5.6 Neighbourhood safety scale

Objectives & Overview
To gather information on the respondent's satisfaction with his/her neighbourhood as a place to raise children, including perception of the extent of danger and problems, and of social cohesion or "neighbourliness". Recent research by Dr. Jacqueline Barnes at the Judge Baker Children's Centre, Harvard University in Boston has found that parents' fear of danger and perception of social disorder in the neighbourhood affected their sense of attachment to the neighbourhood and their disciplinary strategies. The information on the parent's perception of the neighbourhood is supplemented by the interviewer's observation of several aspects of the block where the respondent lives.

DSFHQ01, DSFHQ02, DSFHQ05A to DSFHQ06E: These questions cover length of residency in the neighbourhood, satisfaction with the neighbourhood as a place to bring up children, safety, social cohesion and neighbourhood problems. They represent a revised version of specific sections of the Simcha-Fagan Neighbourhood Questionnaire used by Dr. Jacqueline Barnes in her studies of neighbourhoods in Boston and Chicago. Revisions were made based on the factor analysis of the sections, in consultation with Dr. Barnes. DSFHQ03: This question on volunteer involvement is based on a question in the National Population Health Survey.

Changes to Neighbourhood Section across cycles
These scales have been used intermittently over the four cycles of the NLSCY. In Cycle 1, three scales were created: neighbourhood safety (ASFHQ05A to ASFHQ05C), neighbours (ASFHQ06A to ASFHQ06E) and neighbourhood problems. The entire Neighbourhood section was not asked of survey participants in Cycle 2. In Cycle 3, the Neighbourhood section was reintroduced without questions ASFHQ05A to ASFHQ05C and without questions ASFHQ07A to ASFHQ07F.

The Cycle 4 scale questions are the same as the Cycle 1 questions with the exception of DSFHQ05C where there has been a small wording change. Also, the questions that made up the neighbourhood problems scale in Cycle 1 (ASFHQ07A to ASFHQ07F) have not been included in Cycle 4.

Analytical Results

In the sample there were 31,744 children. They were divided into two sub samples and analysis was done on each sub-sample. Individuals with missing values were excluded from the analysis conducted for the purpose of constructing the factor. After these exclusions the sub-samples contained 15,720 and 16,024 individuals respectively. As a result of the factor analyses, two factors were identified: the neighbourhood safety factor and the neighbours factor. The items that comprised each factor are described in the following table:

FACTOR	SCORE	ITEMS
Neighbourhood Safety Score	DSFHS5	DSFHQ05A, DSFHQ05B, DSFHQ05C
Neighbours Score	DSFHS6	DSFHQ06A, DSFHQ06B, DSFHQ06C, DSFHQ06D, DSFHQ06E

Scale Score
If too many values were missing the final score was set to missing. To produce the final scores, 1 was subtracted from each item so that the lowest score would be 0. All the score values were reversed. The final score was derived by totalling the values of all items with non-missing values. A score of 0 indicates the following for the two neighbourhood scales:

- a low degree of neighbourhood safety
- a low degree of neighbourhood cohesiveness

Cronbach's Alpha results
Cronbach's alphas for these factors are given in the table below (calculated using SAS):

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Neighbourhood Safety (DSFHS5)	0.701	DSFHQ05B	0.544
Neighbours (DSFHS6)	0.883	DSFHQ06C	0.848

NOTE: Scores could not be calculated for 3,571 (11.2%) and 6,534 (20.6%) individuals due to missing values.

9.5.7 Parenting scales

Objectives & Overview
The objective of this scale is to measure certain parenting practices. Specifically, two scales were used. The first was designed to measure the positive interaction, hostility/ ineffectiveness and consistency of the parenting of the child. The second scale was designed to measure parental practices that may or may not provoke aversion.

The questions from the Child's Questionnaire used to measure these aspects of parenting are identified in the following paragraphs. A complete factor analyses was done on the parenting scales to evaluate the psychometric properties of these scales for the NLSCY population. The make-up of each factor obtained during these analyses was compared to that which had been indicated in the literature. The results of these analyses are presented later in this section.

Questions DPRCQ01 to DPRCQ18 on positive interaction, hostility or ineffectiveness and on coherence were provided by Dr. M. Boyle of the Chedoke-McMaster Hospital, based on the work of Dr. Ken Dodge (Vanderbilt University) and an adaptation of the Parent Practices Scale of Strayhorn and Weidman. (For children ages 0 to 23 months, only questions DPRCQ01 to DPRCQ07 were asked.)

Questions DPRCQ19 to DPRCQ25 which measure parental practices which may or may not cause aversion, these were provided by Dr. M. Boyle.

Calculation of Parenting Scores
Once the factor structures were analysed and the items included in each factor were determined, scores were calculated. To produce the scores, 1 was subtracted

from each item so that the lowest possible score value would be 0. For each of the four factors, a score of 0 indicates:

- the absence of positive interaction for the positive interaction factor;
- the absence of ineffective interaction for the ineffective factor;
- the absence of consistent parenting for the consistency factor;
- the existence of punitive interaction or aversion producing practices for the hostility/ineffective parenting factor.
- a low degree of parent-child conflict (12-15 years only)

Analytical Results for Children aged 0 to 23 months

There were 4,008 children in the sample for the age group 0 to 23 months. The group was split into two sub-samples of 1,987 and 2,021 individuals, and the analysis for this age group was performed separately for each sub-sample. The non-response rate for the seven items ranged from 1.82% to 2.07%. Individuals were excluded from the analysis that produced the factors when 10% of the data for that factor was missing. After the criterion was applied, there were 1,922 and 1,943 individuals left in the sub-samples to be analysed. No imputation was done. The factor analysis derived two factors for this age group: positive interaction (DPRCS01), and ineffective (DPRCS02). The items making up each factor are listed below.

FACTOR	SCORE	ITEMS
Positive interaction	DPRCS01	DPRCQ1, DPRCQ02, DPRCQ03, DPRCQ06 and DPRCQ07
Ineffective	DPRCS02	DPRCQ 04 and DPRCQ05

Cronbach's Alpha for Children Aged 0 to 23 Month

Cronbach's alpha (raw value) was computed with SAS using normalized weighted data.

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Positive interaction (DPRCS01)	0.661	DPRCQ07	0.587
Ineffective (DPRCS02)	0.339	N/A – only 2 items incl.	N/A – only 2 items incl.

NOTE: The scores for these factors could not be computed in 163 and 145 cases respectively because of unreported values.

Analytical Results for Children Aged 2 to 11

There were 21,777 children in the sample for the age group 2 to 11. The group was split into two sub-samples of 10,784and 10,993 individuals, and the analysis for this age group was performed separately for each sub-sample. The non-response rate for each of the twenty-one items ranged from 3.07%to 3.69%. Individuals missing 10% or more of the items were excluded from the analysis. After the criteria were applied, there were 10,321 and 10,554 individuals left in the sub-samples to be analysed. The factor analysis derived four factors for this age group: positive interaction (DPRCS03), and hostility (DPRCS04), consistency (DPRCS05) and punitive/aversive (rational) parenting (DPRCS06). The items making up each factor are listed below.

FACTOR	SCORE	ITEMS
Positive interaction	DPRCS03	DPRCQ01, DPRCQ02, DPRCQ03, DPRCQ06 and DPRCQ07
Ineffective	DPRCS04	DPRCQ04, DPRCQ08*, DPRCQ09, DPRCQ013, DPRCQ014, DPRCQ0 15and DPRCQ018
Consistency	DPRCS05	DPRCQ10, DPRCQ11, DPRCQ12*, DPRCQ16* and DPRCQ17*
Rational	DPRCS06	DPRCQ21*, DPRCQ22, DPRCQ23* and DPRCQ24

** Indicates that the item value was reversed when computing the score.*

Cronbach's Alpha for Children Aged 2 to 11 Years

Cronbach's alpha (raw value) was computed with SAS using normalized weighted data (in general, Cronbach's alphas computed by SAS are lower than those produced by SPSS).

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Positive interaction (DPRCS03)	0.778	DPRCQ02	0.715
Ineffective (DPRCS04)	0.701	DPRCQ13	0.654
Consistency (DPRCS05)	0.664	DPRCQ12	0.576
Rational (DPRCS06)	0.551	DPRCQ22	0.359

NOTE: The scores for these factors could not be computed in 902, 1,074, 1,488 and 1,016 cases respectively because of unreported values.

Conflict Resolution Scale for Children 12 to 15 Years

The conflict resolution score was created for children aged 12-15. The following items were used in the factor analysis: DPRCBb30A, DPRCBb30B, DPRCBb30C, DPRCBb30D, DPRCBb30E, DPRCBb30F, DPRCBb30G, and DPRCBb30H.

The factor structure of this scale was determined based on data from cycle 3. To produce the scores, 1 was subtracted from each item so that the lowest possible score value would be 0. For this factor, a high score indicates a high level of conflict between parent and child.

Analytical Results for Children Aged 12 to 15 Years

There were 4,155 children in the sample for the age group 12 to 15 years. The group was split into two sub-samples of 2,090 and 2,065 individuals, and the analysis for this scale was performed separately for each sub-sample. The non-response rate for the eight items ranged from 4.52% to 5.12%%. In total 262 cases that had one or more missing value and were excluded from the analysis. These cases were given a missing value for the overall score since no imputation was completed. Items DPRCBb30A and DPRCBb30H were reversed in the calculation of the score. All values were recoded from 1-5 to 0-4. The final score ranges from

0-30 with a high score indicating a higher degree of parent-child disagreements. The Cronbach's Alpha value for the score is 0.745.

9.5.8 Social support

Objectives & Overview
The original scale contains 24 items from Robert Weiss's Social Provisions Model that describes six different social functions or 'provisions' that may be acquired from relationships with others. Due to the length of the scale, and on the advice of Dr. M. Boyle at Chedoke-McMaster Hospital, the survey uses the shortened version (containing 6 items) that was derived for the Government of Ontario's Better Beginnings, Better Futures Project. This measures guidance (2 questions), reliable alliance (2 questions), and attachment (2 questions). Furthermore, in Cycle One, 4 additional questions on different types of social support (i.e. religious, community services) were added as suggested by Dr. Tom Hay. These questions were not included for Cycle Three, however, due to a lack of variability in response. Questions similar to those suggested by Dr. Hay were taken from F-COPES (Family Crisis Oriented Personal Evaluation Scales) and included in Cycle 4. F-COPES draws upon the coping dimensions of the Resiliency Model of Family Adjustment and Adaptation (McCubbin, Olson & Larsen: 1981). The total social support measurement includes 6 questions and not only focuses on the quantity of social support but on the quality of social supports as well.

In Cycle 2 the entire social support section was dropped due to a belief that there would be little temporal variation in the amount individuals received and concerns regarding response burden.

In Cycle 4, this section is asked of all PMK's with children/youth less than 16 years of age and includes the following items: DSPHQ01A, DSPHQ01B, DSPHQ01C, DSPHQ01D, DSPHQ01E, DSPHQ01F, DSPHQ01H, and DSPHQ01I.

Changes to Social Support Section
In cycle 4 the following changes were made to the social support section:

- Kept the 6 items, from the original 24, that we have used for Cycle 1 and Cycle 3; however, replaced items ASPHQ02A-D used in Cycle 1 by the F-COPES items

- Attached two additional questions from the above measurement from the social integration sub-scale (questions items DSPHQ01H & DSPHQ01I). The questions on social integration are significant because they assess one's feeling of belonging to a group that shares similar interests, concerns, and activities thus measuring another factor of social support.

- Added four supplementary questions from the F-COPES, and one question based on the F-COPES framework, that center on the same reasoning as those questions used in Cycle 1. However, the suggested questions steer away from the simple 'yes' and 'no' responses that fail to indicate variability and instead use the response categories of 'Strongly disagree', 'Disagree', 'Agree', 'Strongly Agree'.

Analytical Results
There were 30,325 children in the sample for the age group 0 to 16 years. The group was split into two sub-samples of 15,001 and 15,324 individuals, and the analysis for this scale was performed separately for each sub-sample. The non-response rate for the eight items averaged about 2.6%. In total 2,033 cases had more than one missing value and were excluded from the analysis. These cases

were given a missing value for the overall score since no imputation was done. Items DSPHQ01A, DSPHQ01E, DSPHQ01F and DSPHQ01I were reversed in the calculation of the score. All values were recoded from 1-4 to 0-3. The final score ranges from 0-24 with a high score indicating a higher degree of social support. The factor structure imposed was the same as that used in Cycle 1.

This analysis was done using normalized weighted data.

Cronbach's alpha for Social Support Scale

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Social Support (DSPHS01)	0.877	DSPHQ01G	0.853

9.5.9 Temperament Scale

This scale was used in the NLSCY to measure the various aspects of the temperament of young children from ages 0-3 (fussy/difficult, unpredictable, dull, inadaptable, persistent/unstoppable, irregular, negative adaptation and affect). This section was completed by the PMK. The scale was developed by Dr. John Bates of the University of Indiana and was originally known as the Infant Characteristics Questionnaire (ICQ). Dr. Jo-Anne Finegan of Toronto's Hospital for Sick Children created a revised version of the ICQ to be used for three-year-olds. The Temperament Scale has been used in large-scale studies and is considered by specialists to be the best available measure for use in household surveys.

This particular scale went through many alterations from cycles 1 through 3. After the validation of cycle 1 results, only the fussy/difficult construct was used in cycle two. In cycle three, the unadaptable construct was again included in the survey. The following section will show specifically what variables were kept, discarded and reintroduced and in what cycles these changes were made. This following section will outline which questions have been removed and added in each cycle, according to the age group.

In Cycle 4, it was decided to ask the temperament questions of children aged 0 to 2 years rather than 0 to 3 years as had been done in all previous cycles. In addition three news items were added to the temperament scale (see description by age breakdown listed below). These changes were made based on factor breakdown and reliability analyses undertaken by J. Douglas Willms[8].

Changes to 3-11 Month Old Questions:

Action Taken	Variable Name
Questions removed from cycle 2 and 3	ATMCQ02, ATMCQ03, ATMCQ04, ATMCQ10, ATMCQ15, ATMCQ16, ATMCQ18, ATMQC23
Questions added in cycle 4	DTMCQ15, DTMCQ23

Note: Variable names reflect the last cycle in which the variable was included

[8] J. Douglas Willms is a Professor at the University of New Brunswick and the Director of the Canadian Research Institute for Social Policy.

Changes to 6-11 Month Old Questions:

Action Taken	Variable Name
Questions removed from cycle 2 and 3	ATMCQ13, ATMCQ24
Questions removed from cycle 3 only	DTMCQ11, DTMCQ12, DTMCQ14
Questions removed for cycle 2 but reintroduced in cycle 3	CTMCQ27
Questions added to cycle 3 for the first time	CTMCQ25, CTMCQ26
Questions reintroduced in cycle 4	DTMCQ15, DTMCQ23

Note: Variable names reflect the last cycle in which the variable was included

Changes to 1-3 Year Old Questions:

Action Taken	Variable Name
Questions removed from cycle 2 and 3	ATMCQ2, ATMCQ3, ATMCQ4, ATMCQ10, ATMCQ13, ATMCQ15, ATMCQ18, ATMCQ21, ATMCQ22, ATMCQ23, ATMCQ24, ATMCQ28, ATMCQ32
Questions removed from cycle 2 but reintroduced in cycle 3	CTMCQ25, CTMCQ26, CTMCQ27, CTMCQ29, CTMCQ30, CTMCQ31
Question reintroduced in cycle 4	DTMCQ15, DTMCQ23a

Note: Variable names reflect the last cycle in which the variable was included

As previously mentioned, the Temperament Scale has been used in large-scale studies and is considered by specialists to be the best available measure for use in household surveys. Unfortunately, however, when used in the context of the NLSCY data, the scale proved to be problematic. Subsequently, there is no Cronbach's Alpha value available.

A more detailed description of the Temperament scale can be found in an article entitled, "Re-Visiting the Bates Temperament Scales: Is there a need to refine the measures for use in future cycles of the NLSCY" by Mr. Tony Haddad. For the purposes of this discussion, however, there are three main explanations as to why the Temperament scale did not fare so well. Firstly, data analysis shows that the distributions of these items are highly skewed and multi-modal. Secondly, the focus group study of the temperament questions found that some people find it difficult to understand the concept of the "average" child, which is included in the anchor and in the introduction to the temperament scale in the survey. Finally, exploratory factor analysis undertaken by Statistics Canada found that the items in the scale load strongly onto difficult/fussy, but that the remaining theoretical factors are unstable with regard to their loading (remaining items do not load on expected theoretical concepts) (Haddad 1999:4). For these reasons this scale was not created in either Cycle 3 or Cycle 4 although the questions making up the scale are included on the release file.

Children 3 to 11 Months
For children aged 3 to 5 months, the scale is made up of questions DTMCQ01, DTMCQ05 to DTMCQ08, DTMCQ14, DTMCQ15, DTMCQ17, DTMCQ19,

DTMCQ20, DTMCQ23 and DTMCQ33 is intended to measure the extent to which the child is fussy, unadaptable, unpredictable and dull.

For children 6 to 11 months old, the scale is made up of questions DTMCQ01, DTMCQ05 to DTMCQ09, DTMCQ15, DTMCQ17, DTMCQ19, DTMCQ20, DTMCQ23, DTMCQ25, DTMCQ26, DTMCQ27 and DTMCQ33. This expanded list of questions measures the same four aspects of temperament as for children 3 to 5 months old.

Children 1 to 2 Years
For children 1 and 2 years-old, questions DTMCQ1, DTMCQ01 to DTMCQ08, DTMCQ11, DTMCQ12, DTMCQ15, DTMCQ17, DTMCQ19, DTMCQ23a, DTMCQ25, DTMCQ26, DTMCQ27, DTMCQ29, DTMCQ30, DTMCQ31 and DTMCQ33 should theoretically measure the degree to which the child is difficult, irregular, unadaptable, affectively negative and persistent/unstoppable.

Meanings of Ratings for Specified Behaviours
The respondent, in most cases a parent, is required to answer each question in the scale by assigning a rating between 1 and 7. For all questions except DTMCQ14, a 1 means that the child has a favourable response or usually exhibits the specified behaviour, while a 7 indicates that the child reacts negatively or seldom displays the behaviour in question. If the child is in the middle, a 4 is assigned. In question DTMCQ14, the meanings of the ratings are reversed.

9.6 Child-reported scales

9.6.1 General Self Score

Objectives & Overview
The objective of the General Self Score is to measure the child's overall self-esteem. The self-esteem scale was expanded each year to include the oldest cohort. This means that by Cycle 4 the items making up this scale are asked of all youths between 10 and 17 years of age.

In Cycle 2 and subsequent cycles, the factor scores were derived based on the factorial structure identified in Cycle 1. Below is a description of the items that were included on the questionnaire to measure these scales. The analysis used to construct the scale and the results of these analyses, based on Cycle 1 data.

Questions DAMCQ01A to DAMCQ01D on overall self esteem were taken from the General-Self Scale of the Marsh Self Description Questionnaire developed by H.W Marsh.

Once the factor structures were analysed and the items included in each the factor were determined, the scores were calculated. No imputation was done for missing values. If any values were missing, the final score was set to missing. To produce the final scores, 1 was subtracted from each item so that the lowest score would be 0. The final score was derived by totalling the values of all items with non-missing values. A score of 0 indicates the following:

- a lack of general self esteem for the general self scale

Analytical Results (Based on Cycle 1 data)
In the sample there were 3,434 children aged 10 or 11 years. They were divided into two sub-samples of sizes 1,705 and 1,729 and analysis was done on each sample. The non-response rates for the 8 items ranged from 14% to 15.8%. Individuals with missing values were excluded from the analysis conducted for the purpose of constructing the factor. After these exclusions, the sub-samples contained 1,371 and 1,413 individuals respectively, for analysis purposes. As a result of factor analysis, the general self factor was identified. The items that make up this factor are described below.

FACTOR	SCORE	ITEMS
General Self	DAMCS02	DAMCQ01A, DAMCQ01B, DAMCQ01C, and DAMCQ01D

Cronbach's Alpha
Cronbach's alpha coefficients (raw values) were calculated with SAS, using the normalized weighted data.

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
General Self (DAMCS02)	0.728	DAMCQ01C	0.629

NOTE: Scores could not be calculated for 555 individuals (16.2%), due to missing values.

9.6.2 Behaviour scale

This section replicates the behaviour checklist included in the Child's Questionnaire completed by the PMK. All children aged 10-15 answers these questions in the self-complete portion of the survey. It is intended to provide indicators of the following behaviours: conduct disorder, hyperactivity, inattention, physical aggression, indirect aggression, emotional disorder, anxiety and prosocial behaviours. For a more detailed explanation refer to Section 7.6.2 of this chapter. All analysis presented below was done in Cycle 1.

Analytical Results (based on Cycle 1 data)
In the sample there were 3,434 children aged 10 or 11 years. They were divided into two sub samples of size 1,705 and 1,729 and analysis was done on each sample. The non-response rates for the 8 items ranged from 13.6% to 16.7%. Individuals with missing values were excluded from the analysis conducted for the purpose of constructing the factor. After these exclusions, the sub-samples contained 1,352 and 1,398 individuals respectively, for analysis purposes. As a result of imposed factor analysis, five factors were identified: hyperactivity-inattention, prosocial behaviour, emotional disorder-anxiety, physical aggression-conduct disorder, and indirect aggression. The items that comprised each factor are described in the following table.

FACTOR	SCORE	ITEMS
Indirect aggression	DFBCS01	DFBCQ01J, DFBCQ01R, DFBCQ01Z, DD1CQ1LL and DD1CQ01TT
Anxiety and emotional disorder	DFBCdS02	DFBCQ01F, DFBCQ01K, DFBCQ01Q, DFBCcQ1V, DFBCQ1CC, DFBCQ1MM and DFBCQ1RR
Conduct disorder and physical aggression	DFBCS03	DFBCQ01G, DFBCQ01X, DFBCQ1AA, DFBCQ1FF, DFBCQ1JJ** and DFBCQ1NN**
Hyperactivity/inattention	DFBCdS04	DFBCc01B, DFBCQ01I, DFBCQ01P, DFBCQ01S, DFBCQ01W, DFBCQ1HH and DFBCQ1QQ
Prosocial behaviour	DFBCS05	DFBCQ1A, DFBCQ1D, DFBCQ1H, DFBCQ1M, DFBCQ1U, DFBCQ1BB, DFBCQ1GG, DFBCQ1OO, DFBCQ1SS, and DFBCc1UU
Property Offences	DFBCS07	DFBCQ01C, DFBCQ01E, DFBCQ01L, DFBCQ01T, DFBCQ1DD and DFBCQ1PP

** The analysis of these scales was done in Cycle 1, but the variable names have been changed to reflect the current cycle of data.*
*** Wording changed, see description provided earlier in this section.*

Cronbach's Alpha for Behaviour Scale
Cronbach's alphas for these factors are given below (calculated using SAS):

FACTOR	ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Indirect aggression (DFBCS01)	0.728	DFBCQ1LL	0.657
Anxiety and emotional disorder (DFBCdS02)	0.760	DFBCQ1II	0.717
Conduct disorder and physical aggression (DFBCS03)	0.738	DFBCQ1AA	0.678
Hyperactivity / inattention (DFBCdS04)	0.751	DFBCQ1QQ	0.717
Prosocial behaviour (DFBCS05)	0.766	DFBCQ1SS	0.741
Property Offences (DFBCS07)			

NOTE: The scores for these factors could not be computed in, 566 (16.5%), 597 (17.4%), 585 (17%), 621 (18.1%) and 587 (17.1%) cases respectively because of unreported values.
The variable names have been changed to indicate they are in Cycle 4, but this analysis presented above is based on Cycle 1 data.

9.6.3 Depression scale

General Information about the Depression Scale
The depression scale used to measure PMK depression was also used for the 12/13 year olds in Cycle 2 and again for the 16/17 year olds in Cycle 4. The factor structure that was used for the PMK scale was also imposed on the youth scale.

In order to produce the score, 1 was subtracted from each item so that the lowest score would be 0. The final score was derived by totalling the values of all items with non-missing values. As well, the answer categories were reversed for questions having a negative loading (DFBCd10F, DFBCd10H, and DFBCd10J). The total score (DHTCbS1B) may therefore vary between 0 and 36, a high score indicating the presence of depression symptoms.

FACTOR	SCORE	ITEMS
Depression	DHTCbS1B	DFBCd10A, DFBCd10B, DFBCd10C, DFBCd10D, DFBCd10E, DFBCd10F, DHTCb10G, DFBCd10H, DFBCd10I, DFBCd10J, DFBCd10K, DFBCd10L

For more information on the creation of this scale and related analysis, refer to the parent-reported depression scale discussed earlier in this chapter.

9.6.4 Friends Scale (DFFCS01)

Objectives & Overview
The object of the friends scale is to measure how well the child feels he/she gets along with his/her peers. In order to understand how the factorial structure was determined in Cycle 1, below is a description of the items that were included on the questionnaire in Cycle 1 to measure peer relations, the analysis used to construct the scale and the results of these analyses.

Score Calculation
Once the factor structures were analysed and the items included in the factor was determined, the score was calculated. No imputation was done on the values. If any values were missing the final score was set to missing. A value may be missing if the child refused to answer or did not know the answer to the question.

To produce the score, 1 was subtracted from each item so that the lowest score would be 0. The final score was derived by totalling the values of all items with non-missing values. The score ranges from 0 to 16. A score of 0 indicates the respondent does not have a lot of friends and does not have positive relations with other children.

Analytical Results for Friends Scale (based on Cycle 1 data)
In the sample in Cycle 1, there were 3,434 children aged 10 or 11 years. They were divided into two sub samples of size 1,705 and 1,729 and analysis was done on each sample. The non-response rates for the 4 items ranged from 10.9% to 11.5%. Individuals with missing values were excluded from the analysis conducted for the purpose of constructing the factor. After these exclusions, the sub-samples contained 1,508 and 1,529 individuals respectively, for analysis purposes. No imputation took place. As a result of factor analysis, one factor was identified: the friends factor (DFFCS01). All items, DFFCQ01 to DFFCQ04, loaded into the factor.

Cronbach's Alpha results (based on Cycle 1 data)
Cronbach's alpha coefficients (raw values) were calculated with SAS:

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Friends (DFFCS01)	0.779	DFFCQ04	0.779

NOTE: Scores could not be calculated for 397 (11.6%) individuals due to missing values. The analysis presented here was done in Cycle 1, but the variable names have been changed to indicate Cycle 4.

9.6.5 Identity style inventory

Objectives & Overview
This measure is new to the NLSCY and was added for 16 and 17 year olds only. The items in these scales come from the 30-item Identity Style Inventory (ISI3) developed by Michael D. Berzonsky in the Department of Psychology at the State University of New York at Cortland. Based on the premise that the style of decision-making an adolescent adopts will greatly influence their life choices, it was felt that it was important to include a measure to capture this information. Using the 30-item ISI3, three different sub-scales were created to describe the adolescent's identity style. In order to provide a title that would be clearly understood by the youths, this section is referred to as decision making in the questionnaires. This section provides a summary of the ISI3, for a more detailed description of the Identity Style Inventory see the article entitled *"Identity Status, Identity Processing Style, and the Transition to University"* by Berzonsky, Michael D. and Kuk, Linda S. in the Journal of Adolescent Research (2000).

Description of Identity Styles
***Informative-Orientation*:** These individuals actively seek out, evaluate, and utilize self-relevant information. They are sceptical about their self-constructions and willing to test and revise aspects of their self-identity when confronted with discrepant feedback. Research indicates that utilization of an informational identity orientation is positively associated with self-reflection, problem-focused coping efforts, high need for cognition, cognitive complexity and vigilant decision making.

Normative-Orientation: These individuals deal with identity questions and decisional situations by conforming to the prescriptions and expectations of significant others. Research indicates that they are also conscientious and agreeable, but they have a low tolerance for ambiguity and a strong need for structure and cognitive closure.

***Diffuse-Orientation*:** These individuals are reluctant to face up to and confront personal problems and decisions. If one procrastinates and delays long enough, behavioural reactions will be dictated and controlled by situational demands and incentives. This orientation has been found to be positively associated with avoidant coping, self-handicapping, other-directedness, and maladaptive decisional strategies and negatively correlated with self-reflection, conscientiousness, and cognitive persistence.

FACTOR	SCORE	ITEMS
Diffuse-Orientation	DDECdS01	DDECdQ02, DDECdQ06, DDECdQ08, DDECdQ10,DDECdQ15,DDECdQ18, DDECdQ20, DDECdQ22, DECdQ27,DDECdQ29
Information-Orientation	DDECdS02	DDECdQ01, DDECdQ04, DDECdQ05, DDECdQ09, DDECdQ11, DDECdQ16, DDECdQ17, DDECdQ21, DDECdQ24, DDECdQ26, DDECdQ28
Normative-Orientation	DDECdS03	DDECdQ03, DDECdQ07, DDECdQ12, DDECdQ13, DDECdQ14, DDECdQ19, DDECdQ23, DDECdQ25, DDECdQ30

Creation of Scale Score*

As with the other scales used in the NLSCY, a score was created for each of the three identity style orientations. In order to produce the scores, 1 was subtracted from each item so that the lowest score would be 0. The original item values ranged from 1 to 5. After the re-ordering the item values ranged from 0 to 4. The final score was derived by totalling the values of all items with non-missing values.

Diffuse-Orientation (DDECdS01): The lowest value for this score is zero and the highest value is 40. A high score indicates the presence of a diffuse-orientation identity.

Information-Orientation (DDECdS02): The lowest value for this score is zero and the highest value is 44. A high score indicates the presence of an information-orientation identity.

Normative-Orientation (DDECdS03): The lowest value for this score is zero and the highest value is 36. A high score indicates the presence of a normative-orientation identity.

**NOTE: The NLSCY calculation of scores differs from that of the original author.*

Analytical Results

In the sample there were 1,856 children aged 16 and 17 years. For the factor analysis the sample was divided into two sub samples of sizes 907 and 949 and analysis was done on each sample. The non-response rate for the 30 items was approximately 27%. Individuals with missing values (in excess of 10%) were excluded from the analysis conducted for the purpose of constructing the factors. After these exclusions, the sub-samples contained 633 and 676 individuals respectively, for analysis purposes. Factor structures were imposed based on the groupings of the Identity Style Inventory. The analysis was done using normalized weighted data. The table below presents the Cronbach's alpha coefficients for each factor.

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Diffuse-Orientation (DDECdS01)	0.737	DDECdQ20	0.701
Information-Orientation (DDECdS02)	0.767	DDECdQ05	0.743
Normative-Orientation (DDECdS03)	0.640	DDECdQ14	0.591

9.6.6 My Parents and Me

Objectives & Overview
This section was part of the self-complete questionnaire given to children in the 10 to 15 age group. The objective was to complement the Parenting Section on the Child's Questionnaire completed by the PMK by gathering information directly from the child regarding his/her perception of his/her relationship with parents. For the self-completed questionnaire, it was also considered important to obtain a measure of parental supervision (i.e., monitoring), as this has been shown to be linked to child outcomes - there is a correlation between a lack of supervision and negative outcomes, such as juvenile delinquency and other risk-taking behaviours.

The scale that was used was also used in the Western Australia Child Health Survey. It was developed by Lempers et al (1989) based on work of Schaefer (1965) and Roberts et al (1984) and measures parental nurturance, rejection and monitoring. This information will complement the constructs measured in the parent-completed Child's Questionnaire (positive child-parent interaction, ineffective child-parent interaction, and consistent child-parent interaction, aversive and non-aversive parent management techniques.)

The objective of the My Parents and Me scale is to measure the child's perception of his/her relationship with his/her parents and parental supervision. Below is a description of the items that were included on the 10-15 year old questionnaires to measure family relations, the analysis used to construct the scale and the results of these analyses.

Questions DPMCcQ1A to DPMCcQ1Q were taken from the Western Australia Child Health Survey. In addition to these questions, questions DPMCcQ1R to DPMCcQ1T were also used. The scale was developed by Lempers et al. (1989) based on work of Schaefer (1965) and Roberts et al. (1984) and measures parental nurturance, rejection and monitoring.

To construct the My Parents and Me Scale for the NLSCY, a factor analysis was conducted to test the theoretical construct. In the factor analysis the items were multiplied by the child's normalized weight. An individual's statistical weight is normalized by dividing his/her weight (DWTCW01C) by the average weight of all individuals. Consequently, the sum of the normalized weights is equal to the sample size.

Analytical Results (Based on Cycle 3 data)
In the sample of 10-15 year olds there were 5,539 children. The sample was divided into two sub-samples and an analysis was done on each sample. Individuals with missing values were excluded from the analysis conducted for the purpose of constructing the factor. After these exclusions, the two sub-samples contained 2,509 and 2,584 individuals respectively.

As a result of the factor analyses, three factors were identified for the 10-15 year olds: the parental nurturance factor, the parental rejection factor and the parental monitoring factor. The items that comprised each factor are described in the following table.

FACTOR	SCORE	ITEMS*
Parental Nurturance	DPMCcS1	DPMCcQ1A, DPMCcQ1D, DPMCcQ1K, DPMCcQ1H, DPMCcQ1I, DPMCcQ1M, DPMCcQ1Q
Parental Rejection	DPMCbS2b	DPMCcQ1C, DPMCcQ1G, DPMCcQ1J, DPMCcQ1L, DPMCcQ1O, DPMCcQ1P, DPMCcQ1R
Parental Monitoring	DPMCcS3	DPMCcQ1B, DPMCcQ1F, DPMCcQ1N, DPMCcQ1E, DPMCcQ1T

** Variable names have been changed to reflect the current cycle although the analysis described was done using Cycle 3 data.*

Cronbach's Alpha for My Parents and Me Scale
Cronbach's alpha coefficients (raw values) were calculated with SAS, using the normalized weighted data.

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Parental Nurturance (DPMCcS1)	0.88	DPMCQ1M	0.855
Parental Rejection (DPMCbS2B)	0.73	DPMCcQ1R	0.680
Parental Monitoring (DPMCcS3)	0.57	DPMCcQ1T	0.459

** Variable names have been changed to reflect the current cycle although the analysis described was done using Cycle 3 data.*

9.6.7 Neighbourhood structure score

Objectives & Overview
To gather information on the respondent's satisfaction with his/her neighbourhood, including perception of the extent of danger and problems, and of social cohesion or "neighbourliness".

The items included in the score represent a revised version of specific sections of the Simcha-Fagan Neighbourhood Questionnaire used by Dr. Jacqueline Barnes in her studies of neighbourhoods in Boston and Chicago.

Analytical Results
In the sample there were 2,057 children. They were divided into two sub samples and analysis was done on each sub-sample. Individuals with missing values were excluded from the analysis conducted for the purpose of constructing the factor. After these exclusions the sub-samples contained 1,008and 1,049 individuals respectively. As a result of the factor analyses, two factors were identified: the neighbourhood safety factor and the neighbours factor. The items that comprised each factor are described in the following table:

FACTOR	SCORE	ITEMS
Neighbourhood Structure Score	DACYdS01	DACCYd31A to DACCYd31D, DACCYd31F, DACCYd31G

Scale Score
If too many values were missing the final score was set to missing. To produce the final scores, 1 was subtracted from each item so that the lowest score would be 0.

All the score values were reversed. The final score was derived by totaling the values of all items with non-missing values. A score of 0 indicates the following:

- a low perception of neighbourhood structure

Cronbach's Alpha results
Cronbach's alphas for these factors are given in the table below (calculated using SAS):

FACTOR	CRONBACH'S ALPHA (RAW)	ITEM THAT LOWERS CRONBACH'S ALPHA THE MOST IF IT IS EXCLUDED	CRONBACH'S ALPHA IF THE ITEM IS EXCLUDED
Neighbourhood Structure Score (DACYdS01)	0.729	DACCYd31G	0.665

NOTE: Scores could not be calculated for 495 (24.1%) individuals due to missing values.

9.6.8 Conflict Resolution Scales for Youth

Two conflict resolution scores were created for youth aged 16 and 17 years. One score relates to the relationship between the youth and their mother (DPMCdS4) and the other score refers to the relationship between the youth and their father (DPMCdS5). A high score indicates an elevated number of disagreements between youth and parent. The following items were used in the factor analysis:

Factor	Score	Items
Conflict resolution - Mother	DPMCdS4	DPMCdQ06C, DPMCdQ06D, DPMCdQ6E, DPMCdQ6F, DPMCdQ6G, DPMCdQ6F, DPMCdQ6G, DPMCdQ6H, DPMCdQ6I, DPMCdQ6J, DPMCdQ6K et DPMCdQ6L
Conflict resolution - Father	DPMCdS5	DPMCdQ9C, DPMCdQ9D, DPMCdQ9E, DPMCdQ9F, DPMCdQ9G, DPMCdQ9H, DPMCdQ9I, DPMCdQ9J, DPMCdQ9K et DPMCdQ09L

Analytical Results - Mother
There were 1,856 youth in the sample for the age group 16 to 17 years. The group was split into two sub-samples of 907 and 949 individuals, and the analysis for this scale was performed separately for each sub-sample. The non-response rate for the included items was approximately 26%. In total 530 cases that had one or more missing value and were excluded from the analysis. These cases were given a missing value for the overall score since no imputation was completed. All values were recoded from 1-5 to 0-4. The final score ranges from 0-32 with a high score indicating a higher degree of parent-child disagreements. The Cronbach's Alpha value for the score is 0.840.

Analytical Results - Father
There were 1,856 youth in the sample for the age group 16 to 17 years. The group was split into two sub-samples of 907 and 949 individuals, and the analysis for this scale was performed separately for each sub-sample. The non-response rate for

the included items was approximately 32%. In total 593 cases that had one or more missing value and were excluded from the analysis. These cases were given a missing value for the overall score since no imputation was completed. All values were recoded from 1-5 to 0-4. The final score ranges from 0-32 with a high score indicating a higher degree of parent-child disagreements. The Cronbach's Alpha value for the score is 0.885.

9.6.9 Ages and Stages

The Ages & Stages questionnaires were provided to PMK's of all children aged 3 months to 5 years. One or two paper questionnaires (depending on the child's age) containing ten sections labelled A to K were provided to the PMK. The cover page of the paper questionnaire indicated to the respondent which section to complete. The table below indicates which age group was assigned to each section.

Booklet 1	Section
3 – 4 months	A
5 – 7 months	B
8 – 11 months	C
12 – 17 months	D
18 -23 months	E
24 – 29 months	F
30 – 35 months	G
36 – 47 months	H
Booklet 2	**Section**
48 – 59 months	J
60 – 71 months	K

The questions included in the Ages and Stages questionnaires are grouped into the five categories listed below with each respondent receiving a score in the range of 0 to 60. For this measure, a high score indicates that the child is at or above the normal range for their age group. For more information on this measure please refer to Chapter 8.

Factor	Score	Range of scores
Problem solve score	DAGCdS01	0 to 60
Personal score	DAGCdS02	0 to 60
Communication score	DAGCdS03	0 to 60
Fine motor score	DAGCdS04	0 to 60
Gross motor score	DAGCdS05	0 to 60

9.7 Cycle 4 Scales

9.7.1 Parent-Reported Scales

Variable	Scale Name	Universe
DDPPS01	Depression Score (refers to PMK)	PMK of children 0 to 15 years
DFNHS01	Family Functioning Score	PMK **or** Spouse of children 0 to 15 years
DSFHS5	Neighbourhood Safety Score	PMK **or** Spouse of children 0 to 15 years
DSFHS6	Neighbours Score	PMK **or** Spouse of children 0 to 15 years
DSPHS01	Social Support Score	PMK **or** Spouse of children 0 to 15 years
DACCS6	Home Responsibilities Score	PMK of children 10 to 13 years

Variable	Scale Name	Universe
Behaviour		
DBECDS01	Hyperactivity – Inattention Score	PMK of children 2 to 3 years
DBECDS03	Emotional Disorder – Anxiety Score	PMK of children 2 to 3 years
DBECS04	Physical Aggression and Opposition Score	PMK of children 2 to 3 years
DBECS05	Separation Anxiety Score	PMK of children 2 to 3 years
DBECDS06	Hyperactivity – Inattention Score	PMK of children 4 to 11 years
DBECDS07	Prosocial Behaviour Score	PMK of children 6 to 11 years
DBECDS08	Emotional Disorder – Anxiety Score	PMK of children 4 to 11 years
DBECDS09	Conduct Disorder – Physical Aggression Score	PMK of children 4 to 11 years
DBECS10	Indirect Aggression Score	PMK of children 4 to 11 years
DBECDS11	Property Offences Score	PMK of children 8 to 11 years

Variable	Scale Name	Universe
MSD		
DMSCS01	Raw Score for Motor and Social Development	PMK of children 0 to 47 months
DMSCS02	Standardized Score for Motor and Social Development – based on cycle 1 norms	PMK of children 0 to 47 months
DMSCSD03	Standardized Score for Motor and Social Development – based on Cycle 4 norms	PMK of children 0 to 47 months

Variable	Scale Name	Universe
Parenting		
DPRCS01	Positive Interaction	PMK of children 0 to 23 months
DPRCS02	Ineffective Parenting Style	PMK of children 0 to 23 months
DPRCS03	Positive Interaction	PMK of children 2 to 11 years
DPRCS04	Ineffective Parenting Style	PMK of children 2 to 11 years
DPRCS05	Consistency	PMK of children 2 to 11 years
DPRCS06	Rational Parenting Style	PMK of children 2 to 11 years
DPRCBS09	Conflict Resolution Scale	PMK of children 12 to 15 years

9.7.2 Child-Reported Scales (Self-Complete)

Variable	Scale Name	Universe
DFFCS01	Friends Score	Children 10 to 17 years
DAMCS02	General Self Score	Children 10 to 15 years
Behaviour		
DFBCS01	Indirect Aggression Score	Children 10 to 15 years
DFBCdS02	Anxiety and Emotional Disorder Score	Children 10 to 15 years
DFBCS03	Conduct Disorder – Physical Aggression Score	Children 10 to 15 years
DFBCdS04	Hyperactivity – Inattention Score	Children 10 to 15 years
DFBCS05	Prosocial Score	Children 10 to 15 years
DFBCS07	Property Offences Score	Children 10 to 15 years
Parenting		
DPMCcS1	Parental Nurturance Score	Children 10 to 15 years
DPMCbS2B	Parental Rejection Score	Children 10 to 15 years
DPMCcS3	Parental Monitoring Score	Children 10 to 15 years

9.7.3 Youth-Reported Scales (Self-Complete)

Variable	Scale Name	Universe
DACYDS01	Neighbourhood Structure Score	Youth 16 to 17 years
DEPS01	Depression Score	Youth 16 to 17 years
DPMCdS4	Conflict Resolution Scale - Mother	Youth 16 to 17 years
DPMCdS5	Conflict Resolution Scale - Father	Youth 16 to 17 years
DECS01	Decision Making - Diffusion	Youth 16 to 17 years
DECS02	Decision Making - Informative	Youth 16 to 17 years
DECS03	Decision Making - Normative	Youth 16 to 17 years

10.0 Survey methodology - Response rates

In a survey, there are 2 types on non-response: total non-response where the selected unit do not answer any question and partial non-response where the selected unit answers enough questions to be a respondent, but without answering every question. The first part of this chapter describes the response rates for NLSCY Cycle 4. A survey's response rates are a measure of the effectiveness of the collection process and a good indicator of the quality of the estimates produced. For a longitudinal survey, the cross-sectional response rates are useful for measuring the effectiveness of collection. The longitudinal response rates tend to be used to measure the quality of the estimates produced. In Chapter 10, only the cross-sectional response rates will be examined. Special attention will be paid to households requiring more work, such as refusals and traces. Chapter 13 will take a closer look at longitudinal response rates in the context of quality analysis.

The second part of the chapter describes partial non-response. Response rates for different portions of the questionnaire and for the direct measures component will be exposed.

10.1 Total Non-response

10.1.1 Definitions

Some definitions are needed in order to understand the tables in this chapter.

A **respondent household** is a household for which an adult component or **a** child or youth component was completed.

A **respondent child** is a child for whom an adult component or **his** child or youth component was completed. A respondent household that does not have a complete adult component may have one respondent child and one non-respondent child. There were 19 children in that situation.

An **out-of-scope household** is a household whose children are all deceased or living outside Canada's 10 provinces.

The **response rate** is defined as the children response rate, the number of respondents over the number of in-scope children.

Refusals are identified during the processing of response files. The interviewers' notes are used, and a keyword search is performed to complete the coding.

The **refusal rate** is defined as the children refusal rate, the number of refusals over the number of in-scope children.

A household is **to be traced** when a call made in connection with the case results in a trace response code.

The **to-be-traced rate** is defined as the number of cases to be traced over the total sample (in scope and out of scope).

A household is considered **not traced** when it is non-respondent and the final response code is a trace response code.

The **successful trace rate** is defined as the successful children trace rate, the number of successfully traced children over the number of children to be traced.

10.1.2 Response rates

Children that are part of the cycle 4 sample are from many sources. Children were added at every cycle. Table 1 shows the response rate by sample origin for Canada.

Table 1: Unweighted Cross-Sectional Response Rate for Children, NLSCY Cycle 4, Canada

Source	Sample Total	Out-of-scope	In-scope total	Response rate	Refusal rate	To-be-traced rate	Successful trace rate
Longitudinal children selected in cycle 1	15,632	44	15,588	84.53%	9.90%	15.55%	91.98%
Longitudinal children selected in cycle 2	3,610	25	3,585	85.13%	6.22%	21.75%	84.33%
Longitudinal children selected in cycle 3	8,117	47	8,070	88.12%	5.37%	21.97%	85.70%
0-1 year old children selected from LFS	5,075	119	4,956	77.54%	10.96%	7.57%	91.41%
5 years old children selected from The Birth Register	4,390	177	4,213	74.20%	7.60%	35.19%	73.79%
Total longitudinal children	27,359	116	27,243	85.67%	8.06%	18.27%	88.54%
Total[9]	36,824	412	36,412	83.24%	8.07%	18.81%	85.40%

Longitudinal households are more likely to respond than households just introduced into the sample. The reason for non-response also varies by sample source. Longitudinal households are easier to trace than households selected from the Birth Register. The source's age and the available information are the main reasons. For longitudinal households, the most recent contact was two years before; the households have responded and provided additional contacts. The most recent contact for households selected from the LFS dates back at least one year. Since the Birth Register was also used to select five-year-olds, the information too must be five years old.

The tables below present the provincial response rates for children. As in the case of Table 2, the total includes 20 children now living outside Canada's 10 provinces.

9 Twenty children included in the canadian total are outside one of the 10 provinces.

Table 2: Unweighted Cross-Sectional Response Rate for Children, by Province, NLSCY Cycle 4

Province	Sample Total	Out-of-scope	In-scope total	Response rate	Refusal rate	To-be-traced rate	Successful trace rate
Newfoundland and Labrador	2,117	11	2,106	86.7%	6.7%	19.3%	87.3%
Prince Edward Island	1,227	7	1,220	84.0%	8.3%	15.1%	77.3%
Nova Scotia	2,707	30	2,677	84.4%	7.3%	20.7%	81.4%
New Brunswick	2,497	18	2,479	82.2%	8.8%	22.2%	84.5%
Quebec	6,310	40	6,270	85.2%	9.0%	18.2%	92.6%
Ontario	9,143	57	9,086	82.2%	8.9%	18.2%	84.8%
Manitoba	2,920	78	2,842	82.9%	7.8%	17.2%	75.9%
Saskatchewan	2,921	59	2,862	82.2%	7.7%	16.7%	87.1%
Alberta	3,743	48	3,695	80.8%	8.7%	19.7%	86.1%
British Columbia	3,219	47	3,172	83.8%	8.5%	20.8%	84.6%
TOTAL[1]	36,824	412	36,412	83.2%	8.4%	18.8%	85.4%

The table below shows the response rate by age group for children introduced in Cycle 1. For the other components of the sample, Table 2 also provides the age-group results since children introduced in Cycle 2 are all four or five years old and children introduced in Cycle 3 are all two or three years old. Response rates tend to decline as children get older.

Table 3: Unweighted Cross-Sectional Response Rate for Children Introduced in Cycle 1, by Age Group, NLSCY Cycle 4, Canada

Age group	Sample Total	Out-of-scope	In-scope total	Response rate	Refusal rate	To-be-traced rate	Successful trace rate
06-07	3,780	18	3,762	84.9%	8.6%	19.3%	90.2%
08-09	2,697	10	2,687	85.8%	8.5%	17.1%	93.1%
10-11	2,468	6	2,462	84.5%	9.8%	14.8%	92.1%
12-13	2,181	1	2,180	85.1%	9.7%	14.5%	90.8%
14-15	2,252	4	2,248	83.2%	11.6%	12.4%	93.2%
16-17	2,254	5	2,249	82.5%	12.3%	12.2%	94.9%
TOTAL	15,632	44	15,588	84.5%	9.9%	15.6%	92.0%

The table below presents the response rates for children introduced in Cycles 1 and 2 based on whether they responded in every cycle or failed to respond in at least one cycle. Children introduced in Cycle 3 and contacted in Cycle 4 were all respondents in Cycle 3. The response rates for the two categories are very different.

Table 4: Unweighted Cross-Sectional Response Rate for Children Introduced in Cycles 1 and 2, by Age Group And Response History, NLSCY Cycle 4, Canada

Age Group	In-scope total		Response rate		Refusal rate	
	Always a respondent	At least 1 non-response	Always a respondent	At least 1 non-response	Always a respondent	At least 1 non-response
04-05	3,108	395	88.7%	55.7%	4.5%	20.8%
06-07	3,472	372	88.8%	54.8%	6.1%	29.3%
08-09	2,430	257	90.0%	51.4%	5.8%	33.5%
10-11	2,222	240	87.8%	53.8%	7.2%	33.8%
12-13	1,991	189	88.5%	49.2%	7.3%	35.4%
14-15	2,036	212	87.2%	45.3%	8.7%	39.6%
16-17	2,012	237	86.3%	50.6%	9.1%	38.8%
TOTAL	17,271	1,902	88.3%	52.3%	6.7%	31.6%

Age Group	In-scope total		To-be-traced rate		Successful trace rate	
	Always a respondent	At least 1 non-response	Always a respondent	At least 1 non-response	Always a respondent	At least 1 non-response
04-05	3,108	395	19.1%	42.6%	89.3%	65.7%
06-07	3,472	372	17.5%	35.8%	92.8%	76.9%
08-09	2,430	257	15.5%	32.9%	95.0%	84.7%
10-11	2,222	240	13.4%	28.2%	94.0%	83.8%
12-13	1,991	189	13.6%	24.3%	93.7%	73.9%
14-15	2,036	212	11.3%	22.6%	96.1%	79.2%
16-17	2,012	237	11.2%	21.5%	96.9%	86.3%
TOTAL	17,271	1,902	15.1%	31.5%	93.3%	76.3%

The table below provides the response rates for children for whom tracing was done. The table shows the impact that tracing has on response rates. As a result of tracing, 3,550 children from longitudinal households and 834 cross-sectional five-year-olds responded to the survey, which boosted the response rates by 17.0% and 19.8% respectively. It should be noted that interviewers were instructed not to conduct interviews for traced children aged 0 and 1 who were selected from the LFS. This instruction was not followed in all cases, and some households responded; this accounts for the low response rate for children aged 0 and 1.

Table 5: Unweighted Cross-Sectional Response Rate for Traced Children, NLSCY Cycle 4, Canada

Source	Sample Total	Out-of-scope	In-scope total	Response rate	Refusal rate	To-be-traced rate
Longitudinal children selected in cycle 1	2,430	35	2,395	73.0%	8.3%	92.0%
Longitudinal children selected in cycle 2	785	23	762	69.7%	5.8%	84.6%
Longitudinal children selected in cycle 3	1,784	38	1,746	72.9%	5.3%	85.8%
0-1 year old children selected from LFS	395	25	370	31.4%	4.1%	91.6%
5 years old children selected from The Birth Register	1,545	61	1,484	56.2%	4.5%	73.9%
Total longitudinal children	4,999	96	4,903	72.4%	6.9%	88.6%
Total	6,939	182	6,757	66.6%	6.2%	85.5%

10.2 Partial Non-response

We saw in the sections on complete non-response that even when a person gave enough information in order to be considered a respondent, some of the variables may still be not answered. There are various reasons why this might happen. Examples of such reasons - in no particular order – are: unwillingness to answer sensitive questions, respondent fatigue, accidentally missing parts of the questionnaire, or operational difficulties.

It is reasonable to assess partial non-response by components. This is because the variables within a component share a common subject matter and/or are being used together for deriving variables about the same subject matter. One usually finds that the partial non-response has different characteristics – depending on the subject matter being treated.

For instance, Motor-Social Development - for children of ages 0 to 3 - is quite thoroughly answered, since parents have more interest for this topic. The questions on income, on the other hand, may be considered too personal by some respondents, and some partial non-response will arise. Note that income - in particular - is imputed – see Chapter 11 – so even though we show in the table the partial non-response to the survey questions, the derived variable has a value for any respondent.

By the nature of the survey – dealing with children of different ages – some of the components are only applicable to respondents of certain ages. Also, within components, the same topic may be addressed with different questions – again, due to age differences. Hence in some cases even within a component there are subgroups of variables that can be treated separately.

We describe the extent of partial non-response taking the primary release file as example. It is intended to warn the user on what can be expected in the way of partial non-response for different components. If such information is needed for only one variable at a time, the code book should be consulted.

The description is of general nature, and it may be necessary to gather more information for every situation a user needs to deal with. Also, different users may prefer different sub-groupings of questions, and we only present one set of choices. Only selected subgroups of variables are shown – even though very few are left out.

10.2.1 Partial non-response for questionnaire component

In Table 6 we present some of the components, and where applicable, subgroups of variables within components. For each we present the ages for which the variables are applicable, and the number of variables involved. Even though the age is a major factor that separates variables within components and their domain of applicability, occasionally other criteria are important – for instance in the section on the chronic condition of the spouse.

In the fourth column the count of records for which at least some of the variables apply is presented. Then we present the count of the records where all of the variables apply. You will note that there are situations where the second count is much smaller than the first, indicating that the variables considered are part of a flow of questions with a lot of valid skips. However, we can still use the subgroups of variables to assess the extent of non-response.

For each subgroup of variables, within its domain there are some records where no applicable question has been answered. We call these 'Complete Subgroup Non-response', with the understanding that the non-response is complete only for the subgroup of variables in question. The ratio of the complete subgroup non-response to the count of applicable records is the 'Fraction of Total Subgroup Non-response' column.

Other than complete subgroup non-response there is always non-response to only some of the questions within the subgroup. When reporting average non-response, for each survey respondent we count the number of questions that are being answered versus the number of questions that are applicable. Due to the flow of the questionnaire, some of the respondents have more questions that are applicable to them than others. Still, the ratio answered/applicable is a reasonable indicator. What we report in the table is the average of this ratio among all records – under the label 'Average Fraction of Non-response'.

It should always be the case that the average fraction of non-response is larger than the fraction of total subgroup non-response. This is because the first also measures the complete subgroup non-response. Furthermore, the difference between the two is an indication of how much partial non-response there is within that subgroup and for that subset of records.

For instance, in the first line of table 6 we can see that Child Activities for 10 – 13 year olds has 7 questions, and all are applicable to every one of the 3941 records. Furthermore, 180 of those records have a non-response code for all variables. The average fraction of non-response is very close to the fraction of complete non-response. This indicates that for the records other than the 180 with total subgroup non-response there is almost no partial non-response.

Table 6: Partial Non-Response – NLSCY Cycle 4

Component	Age of Child	Number of Variables	At least one variable is Applicable	All variables Applicable	Complete Subgroup Non-response	Average Fraction of Non-response	Fraction of Total Subgroup Non-response
Child Activities	10-13	7	3941	3941	180	4.59%	4.57%
	7-9	2	3957	3957	140	3.61%	3.54%
	6-15	4	11403	1186	440	3.89%	3.86%
	4-9	13	11675	9772	292	3.46%	2.50%
	3-9	5	16089	16089	362	2.32%	2.25%
	0-5	7	17065	6166	273	1.64%	1.60%
	16-17	37	1858	505	313	17.34%	16.85%
Child Behaviour	10-11	6	2086	2086	103	4.95%	4.94%
	8-11	7	4393	4393	194	4.61%	4.42%
	6-11	11	7678	7678	295	5.03%	3.84%
	4-11	12	13758	13758	412	4.24%	2.99%
	2-3	13	6977	6977	120	2.24%	1.72%
	2-11	17	20732	20732	520	2.74%	2.51%
Chronic Condition - PMK and Spouse	All Ages	18	29310	29310	704	2.40%	2.40%
	All Ages	18	24804	24804	1410	5.68%	5.68%
Child Care Arrangement	6-11	2	3757	650	311	8.30%	8.28%
	6-13	12	7184	1958	423	5.89%	5.89%
	4-11 1-11	2 9	6732 15710	1865 1769	968 797	14.39% 5.54%	14.38% 5.07%
	0-5	21	9178	5902	291	8.53%	3.17%
	0-11	29	24740	3	741	3.98%	3.00%
Demographic Derived Variables – Child, Household and PMK	All Ages	22	30320	30309	3	2.14%	0.01%
	All Ages	3	30320	30144	0	0.00%	0.00%
	All Ages	2	30309	30309	0	0.00%	0.00%
Depression Scale - PMK	All Ages	13	30320	29310	2572	8.59%	8.48%
Child Education	12-15 6-15	2 3	3674 11189	3549 10631	159 465	4.49% 4.24%	4.33% 4.16%
	4-11	2	7616	7350	388	5.15%	5.09%
	4-15	46	17483	492	481	3.27%	2.75%
Education - - Household	All Ages	5	30320	30320	568	1.87%	1.87%
Education - PMK	All Ages	13	30320	1833	248	3.78%	0.82%
Education - Spouse	All Ages	13	25560	1732	547	6.28%	2.14%
Education - Youth	16-17	128	1858	291	275	51.92%	14.80%
Family Functioning	All Ages	15	29314	24808	800	3.22%	2.73%
Geographical	All Ages	4	30320	30320	0	0.43%	0.00%
Household	All	5	30320	23253	15	0.24%	0.05%

	Ages						
Health - Child	4-5	32	6083	94	96	16.94%	1.58%
	4-15	15	17483	5606	543	3.13%	3.11%
	3-15	3	21897	21897	602	2.75%	2.75%
	0-3	3	10985	6471	136	1.32%	1.24%
	0-11	2	24740	24738	1416	15.29%	5.72%
	0-15	40	28465	58	618	2.37%	2.17%
	All Ages	2	30192	30191	4659	15.92%	15.43%
Maternal History	All Ages	3	25954	25877	2541	9.91%	9.79%
Health - PMK	All Ages	6	30320	6206	677	3.35%	2.23%
Health - Spouse	All Ages	6	25560	6524	1363	6.87%	5.33%
Health - Youth	16-17	39	1858	373	306	16.62%	16.47%
Income - Household	All Ages	11	30320	25571	0	1.52%	0.00%
Income - PMK	All Ages	11	30320	30320	0	9.12%	0.00%
Income - Spouse	All Ages	11	30320	25560	4760	23.36%	15.70%
Income - Youth	16-17	7	1858	1858	290	19.54%	15.61%
Labour Force - Household	All Ages	2	30320	30320	1478	7.51%	4.87%
Labour Force - PMK	All Ages	34	30320	0	191	3.30%	0.63%
Labour Force - Spouse	All Ages	36	30320	0	11	5.77%	0.04%
Labour Force - Youth	16-17	31	1858	30	306	20.82%	16.47%
Literacy - Child	5-9	2	5595	5595	188	3.37%	3.36%
	3-5	12	10495	10402	166	1.85%	1.58%
	2-5	2	13055	13055	211	1.63%	1.62%
	0-3	9	6990	6776	115	1.83%	1.65%
Math Assessment	7-15	3	9083	9083	1776	19.55%	19.55%
Medical - Child	0	9	1742	771	752	43.47%	43.17%
	0-1	53	3883	699	824	22.00%	21.22%
	0-3	16	4799	3413	267	5.78%	5.56%
	0-5	2	14631	9173	448	3.24%	3.06%
Age, Sex, DoB for Child, PMK or Spouse	All Ages	6	30320	30320	0	0.00%	0.00%
	All Ages	5	30320	28382	0	0.00%	0.00%
	All Ages	5	25560	25553	0	0.00%	0.00%
Motor - Social Development	1-3	8	7726	7250	389	5.55%	5.03%
	0	11	310	72	9	3.18%	2.90%
	0-1	22	2859	3	53	3.41%	1.85%
	0-3	10	10106	8242	520	6.66%	5.15%
PPVT-R	4-6 +	4	8007	8007	3	10.53%	0.04%

Parenting	12-15	10	3704	3704	157	4.29%	4.24%
	2-11	22	20732	20210	711	4.05%	3.43%
	2-15	13	24305	1081	845	3.50%	3.48%
	0-1	2	4011	4011	144	3.96%	3.59%
	0-11	7	24551	24551	631	2.69%	2.57%
Relationships	8-15	2	8054	8054	364	4.62%	4.52%
	4-9	5	11675	8810	323	2.83%	2.77%
Activity Restrictions - PMK and Spouse	All Ages	6	11805	9834	683	7.91%	5.79%
	All Ages	6	8793	8461	1378	16.04%	15.67%
Socio - Demographics for PMK, Spouse, and Child	All Ages	81	30320	30320	182	2.96%	0.60%
	All Ages	89	30320	3576	312	4.31%	1.03%
	All Ages	89	25560	3635	0	6.18%	0.00%
Safety	All Ages	14	29314	29314	730	5.24%	2.49%
	0-2	6	6571	6056	133	2.93%	2.02%
Social Support	All Ages	14	30320	29314	1765	6.44%	5.82%
Temperament - Child	0-1	3	1122	986	26	2.73%	2.32%
	0-3	18	6960	5841	145	2.77%	2.08%
Work After Birth	4-5	2	3319	2427	232	7.20%	6.99%
	0-5	2	11378	9173	282	4.83%	2.48%

10.2.2 Direct measures components response rates

As in the previous cycles, questions asked in cycle 4 can be grouped in different components. These components are group of questions generated by the application according to the child's age. The principal components are: the adult component, child, youth, self-complete, knowledge number, PPVT-R, Who am I, and maths. The tables below show the response rates for those components.

Table 7: Number of Respondent Children with Complete Adult, Child or Youth Components, NLSCY Cycle 4, Canada

Cycle	Age Group	Adult and child or youth		Adult only		Child or youth only		Total
		Count	%	Count	%	Count	%	
1	06-07	3,115	96.9%	60	1.9%	38	1.2%	3,213
1	08-09	2,251	97.1%	43	1.9%	24	1.0%	2,318
1	10-11	2,008	96.5%	50	2.4%	23	1.1%	2,081
1	12-13	1,799	97.0%	40	2.2%	16	0.9%	1,855
1	14-15	1,808	96.6%	46	2.5%	18	1.0%	1,872
1	16-17	1,779	95.9%	29	1.6%	48	2.6%	1,856
2	04-05	2,996	98.2%	23	0.8%	33	1.1%	3,052
3	02-03	6,906	97.1%	112	1.6%	94	1.3%	7,112
4	00-01	3,690	96.0%	39	1.0%	114	3.0%	3,843
4	05-05	3,013	96.4%	58	1.9%	55	1.8%	3,126

Total	29,365	96.8%	500	1.6%	463	1.5%	30,328

The self-complete component is aimed at children 10 years old or older. Response rates tend to diminish with age.

Table 8: Number of Respondent Children with Complete Self-Complete Component, NLSCY Cycle 4, Canada

Age Group	Self-complete completed		Self-complete not completed	
	Count	%	Count	%
10-11	1,660	79.7%	423	20.3%
12-13	1,470	79.3%	385	20.7%
14-15	1,442	77.0%	430	23.0%
16-17	1,384	74.6%	472	25.4%

The Knowledge Number component is aimed only at 4 and 5 years old children. Among the respondents, the response rate is slightly higher for previously surveyed children.

Table 9: Number of Respondent Children with Complete Knowledge Number Component, NLSCY Cycle 4, Canada

GrouAge Group	Knowledge Number completed		Knowledge Number not completed	
	Count	%	Count	%
Longitudinal 4 and 5 years old	2,623	88.9%	329	11.1%
Cross-sectional 5 years old	2,579	82.5%	547	17.5%

The PPVT-R component is aimed only at 4 to 6 year-olds and at some exceptions if the child does not attend school. The counts are practically the same as in the previous component.

Table 10 : Number of Respondent Children with Complete PPVT-R Component, NLSCY Cycle 4, Canada

Age Group	PPVT-R completed		PPVT-R not completed	
	Count	%	Count	%
Longitudinal 4 and 5 years old	2,627	89.0%	325	11.0%
Longitudinal 6 and 7 years old	1,615	86.2%	258	13.8%
Other Longitudinal children	46	71.9%	18	28.2%
Cross-sectional 5 years old	2,579	82.5%	547	17.5%

The Who Am I component is aimed only to 4 and 5 year-olds. Once again, the longitudinal component has a response rate slightly higher. This component is the least answered of the 3 components aimed at 4 and 5 years old children, but the difference is negligible, probably because this assessment is, in order, the third one given to the child. In almost every case, if an assessment is answered, the 3 of them are answered.

Table 11: Number of Respondent Children with Complete Who Am I Component, NLSCY Cycle 4, Canada

Age Group	Who Am I completed		Who Am I not completed	
	Count	%	Count	%
Longitudinal 4 and 5 years old	2,337	79.2%	615	20.8%
Cross-sectional 5 years old	2,390	76.5%	736	23.5%

The Mathematics component is aimed at children from 2nd to 9th grade. Six and 7 years old children are either in 1st or 2nd grade. Hence, they are not all in-scope. The same thing goes for other age groups. Children of a specific age should be in a specific grade to be in-scope. More details about the Mathematics Assessment available in Chapter 13.

Table 12: Number of Respondent Children with Complete Mathematics Component, NLSCY Cycle 4, Canada

Age Group	Mathematics completed		Mathematics not completed	
	Count	%	Count	%
06-07	1,093	78.0%	308	22.0%
08-09	1,806	81.9%	400	18.1%
10-11	1,607	81.5%	364	18.5%
12-13	1,400	79.2%	367	20.8%
14-15	1,378	77.8%	393	22.2%

Conclusion

The cross-sectional response rates are generally lower than expected for Cycle 4. By definition, non-response is total non-response. During weighting, we adjusted for total non-response in order to avoid creating a bias in the estimates. That adjustment always depends on the assumption that non-respondents have the same characteristics as respondents. The higher the non-response is, the greater the chance that that assumption is false, which would reduce the representativeness of the sample.

Partial non-response varies by component, and in some cases even within a component. Within each component there may be complete subgroup non-response or just non-response to some questions. The first situation is easier to address since corrections may be possible for all variables simultaneously, while the second situation may need a more detailed approach. Even though in most situations the partial non-response is a few percents, one needs to assess the extent on a case by case basis and where necessary handle it properly in the data analysis.

11.0 Imputation

The definition of a respondent, as given in the previous chapter, is a child or youth, who has at least one of the child/youth component or the adult component completed. For the respondents, there exist many cases of partial non-response. In some cases this may be for an entire component or only for certain questions. Imputation is the process whereby missing or inconsistent items are "filled in" with plausible values. In the NLSCY, imputation is only carried out for certain variables in the adult and youth income section and the adult labour force section. For all the other variables non-response codes are used.

Imputation flags have been included on the NLSCY file so that users will have information on the extent of imputation and what specific items have been imputed on what records. All imputation flags on the NLSCY data file have an "I" as the fifth character of the variable name. For example, the name of the imputation flag for household income (DINHQ03) is DINHI03.

11.1 Adult Income Imputation

Several income questions were asked during the NLSCY household interview. Information on income, broken down in seven sources, was asked for the PMK as well as his/her spouse. Those seven income sources are: wages and salary, self-employment net income, Employment insurance benefits, Child Tax Benefit/National Child Benefit, social assistance, child spousal support and other sources. The respondents were also asked to provide their best estimate for the household income.

Income is a sensitive topic. As a result, some respondents refused to provide answers to the detailed income questions. Among those, some provided an estimate of their income using ranges. Finally, for those who responded, amounts declared in the income sections were sometimes incoherent with answers provided in the labour force section (for example, an individual might have worked for an employer according to answers provided in the labour force section but no wages and salaries were reported in the income section). Income imputation was carried out to fill out the holes resulting from partial non response as well as to rectify, when possible, these incoherencies.

We had to impute income for 7928 individuals among the 45 821 PMK/spouse of children included in the data file. Since the records on the NLCSY data file represent children, not adults or households, this represents 7291 children. These children had at least one parent whose income was imputed.

Imputation for six out of the seven income sources was done using a nearest neighbour approach. This method first identifies a respondent to the income section who has the same characteristics as the individual with incomplete income data. Once the nearest neighbour has been identified, all income amounts reported by the respondent are used to impute the non-respondent record. It should be noted that all amounts are imputed even if only one source of income was missing on the receiver record. This approach was implemented to ensure the coherence between the income sources. The remaining income source, child tax benefit, was then imputed deterministically with a simplified version of the formulae used by Revenue Canada to compute the benefit.

11.2 Youth Income Imputation

Information on income, broken down in four sources, was asked as part of the youth component. The youth were asked their income from odd jobs, employers, parents and other sources. Cycle 4 was the first cycle in which income questions were asked of the youth.

Income is a sensitive topic. As a result, some respondents refused to provide answers to the detailed income questions. Among those, some provided an estimate of their income using ranges. Finally, for those who responded, amounts declared in the income sections were sometimes incoherent with answers provided in the labour force section (for example, a youth might have worked for odd jobs for pay, either in the school year or in the summer, according to answers provided in the labour force section but no income from odd jobs was reported in the income section). Income imputation was carried out to fill out the holes resulting from partial non response as well as to rectify, when possible, these incoherencies.

Of the 1855 youth who were respondents to the survey, only 1568 had information from the youth component. Only these youth were considered for imputation. This is possible as it only requires an adult component for the youth to be a respondent. There was not enough information about the youth who did not complete the youth component to be able to impute them. Also, if these youth were imputed, the only information on the data file about them would be income, which would not make sense. From those who completed the youth component 396 had their income imputed.

Imputation was carried out for each of the 4 sources of income. Imputation for most cases was done using a nearest neighbour approach. This method first identifies a respondent to the income section who has the same characteristics as the individual with incomplete income data. Once the nearest neighbour has been identified, the missing sources of income are copied on the non-respondent record. The imputation assigned incomes to cases where sources of income were not reported by the respondent, or where the amounts reported in the income section did not agree with what was reported in the labour force section. In the remaining cases, where only one of the four income sources was missing, and there was a total income provided in ranges, a plausible value was imputed.

11.3 Household Income Imputation

Finally, the household income is computed using the PMK, spouse and youth income, whether they are reported or imputed, and the reported household income. A study comparing the household income reported by the respondent to the sum of the PMK, spouse and youth income revealed that declared household income was of poor quality. This is why the household income was calculated using the sum of the personal incomes.

The household income needed to be at least $10,000 or else the adults in the household were imputed. As well, if the household income was greater than $150,000, then the sum of the income of the adults in the household needed to be at least $100,000 or else these cases were also imputed.

11.4 Adult Labour Force Imputation

In Cycle 4 a respondent is defined as a child or youth who has at least one of the following components completed for their household: adult, child or youth. This means that there are cases where we only have a child or youth component and no adult component.

There was no adult component for 463 children or youth. In these cases the following information was imputed: number of weeks worked in the past 12 months, number of hours usually worked per week, and number of weeks in the past 12 months, without work and looking for work. This information was imputed for the PMK and the spouse where applicable. The information was imputed to the non-respondent record from the nearest neighbour found during income imputation.

12.0 Weighting and treatment of non-response

The National Longitudinal survey of Children and Youth (NLSCY) is a probability survey. As is the case with any probability survey, the sample is selected so as to represent a reference population – the Canadian population at a specific date within the context of the survey – as accurately as possible. Each unit in the sample must therefore represent a number of units in the population. In the NLSCY, several populations are represented. The sample is a combination of samples selected in the survey's first three cycles (1994, 1996 and 1998) and a new sample selected in Cycle 4 (2000).

Because the NLSCY is a longitudinal survey – that is, units are selected at a specific point in time and followed over time, it has more than one set of weights. The **longitudinal weighting of Cycle 4 respondents** represents the original population at the time of selection. The weights are computed using all Cycle 4 respondents to represent the population at the time of the original selection. This weighting was also calculated in the first three cycles. Therefore, there is a longitudinal weight for children selected in Cycle 1, a longitudinal weight for children selected in Cycle 2 and a longitudinal weight for children selected in Cycle 3.

In Cycle 4, a second longitudinal weight, a "funnel" weight, which uses only Cycle 4 longitudinal respondents who were also respondents in Cycles 1, 2 and 3, was added. The **longitudinal weighting of all-cycle respondents** affects only children selected in Cycle 1. It represents the same population as the weighting of Cycle 4 respondents for the sample selected in Cycle 1.

These three samples combined with the new sample selected in Cycle 4 also represent the population at the time of collection. **Cross-sectional weighting** is the third set of weights produced for the NLSCY. They represent the Canadian population on January 1, 2001.

Thus, there are three weight variables for the NLSCY:

- DWTCW01C contains the cross-sectional weight
- DWTCW01C contains the ordinary longitudinal weight
- DWTCW01C contains the funnel weight

12.1 Longitudinal weighting of Cycle 4 respondents

The NLSCY weighting strategy is based on a series of cascaded adjustments applied to a basic (or initial) weight. Conceptually, each child's basic weight is approximately equal to the inverse of his/her probability of selection. For example, for households selected from the LFS in 1996, the basic weight is the LFS subweight. For longitudinal children, such as children selected in 1994, the basic weight is determined using the weight computed for the original cycle. The final weight, either cross-sectional or longitudinal, is obtained by applying various adjustments to the basic weight.

This section explains the corrections made to the basic weights and the procedures for weighting the longitudinal samples.

12.1.1 Definition of a longitudinal respondent

As defined in Chapter 4, a longitudinal respondent is a child who was introduced in a previous cycle and whose adult component or child or youth component is complete. Children who were introduced in a previous cycle and died or moved outside Canada's 10 provinces are also longitudinal respondents. They represent children in the reference population who have similar life courses.

12.1.2 Weighting method for longitudinal samples

The first step in computing the longitudinal weight of Cycle 4 respondents is to determine the initial weight to use. An important adjustment in the weighting process is the non-response adjustment. The weight of respondents is adjusted for non-respondents using the characteristics of all responding children. In Cycle 3, we use the characteristics of Cycle 1 to model non-response. In Cycle 4, however, that information is six years old. As a result, the Cycle 1 information may no longer be representative of the situation in Cycle 4. The initial weight to be used depends on the source of the characteristics of respondents and non-respondents used to model non-response.

To represent the characteristics of non-respondents more accurately, the most recent information, which is from Cycle 3, is given preference. However, some Cycle 4 respondents did not respond in Cycle 3. Consequently, they have no initial weight for that cycle. The first step, then, is to determine that weight. It will have to be based on the child's weight in an earlier cycle.

Using the initial weight, two steps are required to obtain the longitudinal weight for the children selected in Cycles 1, 2 and 3. These adjustment factors are applied to the basic weight to produce the final longitudinal weight.

12.1.3 Determining the initial weight

Cycle 4 respondents may or may not have responded in Cycle 3. There were 309 Cycle 4 respondents who did not respond in Cycle 3. These respondents will be referred to as **converts**. To use the Cycle 3 information to model non-response, each Cycle 4 respondent must have a Cycle 3 weight. Converts do not have such a weight.

These 309 Cycle 4 respondents have a Cycle 3 weight of 0, while other respondents have a non-zero weight. The sum of the weights represents and must always represent the Canadian population. If a weight is assigned to each of these respondents, the weight of the units who responded in both Cycles 3 and 4 will have to be reduced. Some of the weight of the Cycle 3 and 4 respondents is transferred to the units converted in Cycle 4. The sum of the weights of all units in Cycle 3 does not change. The adjustment is computed for each age-province combination.

The initial weight of the **children introduced in Cycle 1** is defined as:

for Cycle 4 respondents:

$$\text{Initial weight} = \frac{\text{Cycle 3 weight before post-stratification} \times \sum_{\text{Respondents}} \text{Cycle 1 initial weight}}{\sum_{\text{Respondents}} \text{Cycle 1 initial weight} + \sum_{\text{Converts}} \text{Cycle 1 initial weight}}$$

for Cycle 4 converts:

$$\text{Initial weight} = \frac{\text{Cycle 1 weight} \times \sum_{\text{Respondents}} \text{Cycle 3 weight before post-stratification}}{\sum_{\text{Répondants}} \text{Cycle 1 initial weight} + \sum_{\text{Converts}} \text{Cycle 1 initial weight}}$$

For **children introduced in Cycle 2**, there are 226 converts. Therefore, the Cycle 3 initial weight is computed in the same way, except that the Cycle 2 initial weight is used instead of the Cycle 1 initial weight.

For **children introduced in Cycle 3**, there are no converts. All children sampled in Cycle 4 were respondents in Cycle 3. Hence the initial weight is the Cycle 3 weight before post-stratification.

12.1.4 Special situation of children introduced in Cycle 2

A preliminary adjustment is required for children introduced in Cycle 2. In Cycle 4, some siblings of children selected in Cycle 1 who joined the survey in Cycle 2 were excluded from the sample to reduce the response burden. Longitudinally, those children are still part of the target population. An adjustment is needed to take them into account. Since they were not part of the sample, they had no opportunity to respond, and therefore the non-response adjustment cannot be applied to them.

The adjustment factor is defined as equal to 1 if the children were selected from the LFS sample. If the children are siblings of children introduced in Cycle 1, the adjustment is as follows:

$$\text{Adjustment factor} = \frac{\sum_{Excluded? not\_excluded} \text{initial weights}}{\sum_{Not\_excluded} \text{initial weights}}$$

This adjustment factor is computed for each province. The adjusted weight is the initial weight used for children introduced in Cycle 2.

12.1.5 First adjustment: non-response adjustment

The initial weights reflect the attrition (non-response) observed in Cycles 1, 2 and 3. Now we have to take Cycle 4 attrition into account. We therefore compute an adjustment factor, which reflects the characteristics of Cycle 3 respondents and non-respondents.

Using Cycle 3 variables, response homogeneity groups (RHGs) are created. The RHG method involves grouping individuals with the same propensity to respond. Then an adjustment factor is computed for each RHG. That factor is defined as follows:

$$\text{Non-response adjustment} = \frac{\sum_{\text{Respondents? non? respondents}} \text{adjusted weights in the RHG}}{\sum_{\text{Respondents}} \text{adjusted weights in the RHG}}$$

Certain constraints (range of the adjustment factor and minimum size of each RHG) are imposed when the RHGs are formed so that reasonable, reliable adjustment factors can be obtained.

There are separate sets of RHGs for the children introduced in Cycles 1, 2 and 3. The three sets are required because the samples do not necessarily have the same non-response behaviour. There is every reason to believe that that behaviour varies with the

number of times the individual has been interviewed. The non-response adjustment model needs to take that into account.

12.1.6 Second adjustment: post-stratification adjustment

The second adjustment factor ensures consistency between the estimates produced by the survey and Statistics Canada's population estimates. This method is called post-stratification. For the sample of children selected in Cycle 1, the target population is the set of all children between the ages of 0 and 11 at the beginning of 1995. The post-stratification adjustment of that sample ensures consistency between the sum of the weights and the January 1995 population estimate for each province-age-sex combination. For the sample of children selected in Cycle 2, the population estimates for January 1997 are used, and for children introduced in Cycle 3, the estimates for January 1999 are used.

12.2 Longitudinal Weighting of All-Cycle Respondents – Children Introduced In Cycle 1

The longitudinal weighting of all-cycle respondents (funnel weighting) is very similar to the longitudinal weighting of Cycle 4 respondents. In fact, the all-cycle respondents are a subset of the Cycle 4 respondents. The general approach is the same: an initial weight is multiplied by a non-response adjustment factor and a post-stratification adjustment factor. However, the initial weight and the formation of the response homogeneity groups are different.

12.2.1 Weighting method for the sample of all-cycle respondents

The first step again is to determine the initial weight. Since all Cycle 4 respondents also responded in Cycles 1, 2 and 3, there are no converts. In the weighting for Cycle 3, the initial weight for Cycle 1 was used. RHGs were formed using Cycle 1 variables. The adjustment factor corrected for attrition in Cycles 2 and 3. However, that adjustment factor was computed using Cycle 3 respondents who may or may not have responded in Cycle 2. Consequently, a Cycle 3 non-response adjustment that relates solely to children who responded in Cycles 1 and 2 must be computed.

Once the weight has been adjusted for attrition in Cycles 2 and 3, it can be adjusted for the attrition observed in Cycle 4. The post-stratification adjustment will round out the longitudinal weighting for all-cycle respondents.

12.2.2 Determining the initial weight

The initial weight is the Cycle 1 weight before post-stratification. That weight has been corrected for Cycle 1 attrition.

To take into account the attrition observed in Cycle 2, information from Cycle 1 is used to form RHGs. The adjustment factor is computed for each RHG.

The initial weight from the preceding cycle is used. It is adjusted to account for the attrition observed in Cycle 3. All the Cycle 3 respondents of interest to us have an adjusted weight in Cycle 2. Information from Cycle 2 is used to form RHGs. The

adjustment factor is computed for each RHG. The weight adjusted for attrition in Cycles 1, 2 and 3 is the initial weight for Cycle 4 weighting.

12.2.3 First adjustment: non-response adjustment

Once again, non-response adjustment is based on the creation of RHGs. The amount of information is even more important than in the longitudinal weighting of all the Cycle 4 respondents, since information from the first two cycles is available, in addition to information from Cycle 3. The RHGs are created using Cycle 3 variables and Cycle 1-to-2 and Cycle 2-to-3 change variables. The adjustment factor is computed for each RHG.

12.2.4 Second adjustment: post-stratification adjustment

As in the case of the weighting of Cycle 4 responding children, an adjustment factor is computed to ensure consistency between Statistics Canada's population estimates and the NLSCY's estimates. The target population is still children aged 0 to 11 in January 1995. The adjustment is computed for each age-sex-province combination.

12.2.5 Comparison of the longitudinal weights of Cycle 4 respondents and all-cycle respondents

Some 20 variables were used to compare the estimates made using each set of weights. The proportions for each combination were compared. No significant difference was observed for the variables considered. No matter which set of weights was used, the conclusions were the same. For more details on choosing the right set of weights for the type of analysis being performed, see Chapter 17 – Analytic Issues.

12.3 Cross-sectional weighting

Cross-sectional weighting involves representing the population at the time of collection, i.e., in January 2001. The cross-sectional sample comprises children introduced in Cycles 1, 2 and 3 and children aged 0, 1 and 5 who were first surveyed in Cycle 4.

In the paragraphs that follow, we describe the adjustment factors that, when multiplied by the basic weights, produce the cross-sectional sample weights. Those adjustment factors vary depending on whether the child was first interviewed in Cycle 4 or not.

First, cross-sectional weights were calculated independently for the children selected in 1994, 1996, 1998 and 2000. After that, each of those components represented its own target population. However, those target populations were not entirely separate. It was therefore necessary to use other adjustment factors to correct for that overlap. The last step, post-stratification, ensures consistency between the survey estimates and Statistics Canada's population estimates.

12.3.1 Definition of a cross-sectional respondent

As defined in Chapter 4, a cross-sectional respondent is a child whose adult component or child or youth component is complete. In contrast to longitudinal

respondents, children who were introduced in a previous cycle and died or moved outside Canada's 10 provinces are out of scope. They are not in the January 2001 target population.

12.3.2 Cross-sectional weight of children first surveyed in Cycle 3

1) Children selected from the LFS sample

For children selected from the LFS sample, the weighting strategy is similar to the approach taken in Cycles 1, 2 and 3.

Adjustment 1: Adjustment for the number of rotation groups

The LFS sample is composed of six rotation groups. Each group is a representative subsample of the LFS target population. In the NLSCY, we used 12 rotation groups. Hence the first adjustment is 6/12. The adjusted weight is obtained by multiplying the LFS weight by 6/12.

Adjustment 2: Non-response adjustment

In surveys such as the NLSCY, some households fail to provide responses for a variety of reasons (refusal, special circumstances, language problems, temporary absence). To compensate for the reduction in sample size due to non-response, the weights of respondents are increased. This adjustment is made by multiplying the subweights of respondent households by the following factor:

$$\text{Adjustment factor} = \frac{\sum_{Sample} \text{adjusted weights of households in the NLSCY stratum}}{\sum_{Respondent} \text{adjusted weights of households in the NLSCY stratum}}$$

In this equation, the adjusted weight is the weight obtained after adjustment 1. A different factor is computed for each of the strata defined by the LFS specifically for non-response. Those strata are defined using the following information: province, economic region, census metropolitan area, urban or rural area, apartment frame, special region or not. Each stratum has a response rate of at least 60% and a minimum of 30 children. Strata that are too small or have response rate under 40% were combined until the above requirements were met.

Adjustment 3: Adjustment for households with more than one economic family

Some households contain more than one economic family. In such cases, one economic family must be chosen at random before a child can be selected. This adjustment is the inverse of the family's selection probability. Only one household was affected by this factor.

Adjustment 4: Adjustment for households with more than two in-scope children

In Cycle 4, only one child was to be surveyed in new households, with the exception of twins, who were both surveyed. Consequently, when there was more than one in-scope child in a family, one of them was selected at random. This adjustment compensated for the selection process; 95 households were affected.

This was the last adjustment made for these children before the weights were integrated.

2) Weighting of children selected from the Birth Register in 2000

The initial weight of children selected from the Birth Register was the inverse selection probability. The required adjustment factor were similar to the ones computed for children selected from the LFS sample.

Adjustment 1: Non-response adjustment

This factor boosted the sampling weight to compensate for non-response encountered in Cycle 4 collection. The RHG method was used for this adjustment. However, little information about non-responding children in this sample is available. Consequently, in this particular case, the RHGs correspond to the strata used to select the sample from the Birth Register.

Adjustment 2: Adjustment for households with more than one economic family and more than two in-scope children

This second adjustment took into account the fact that twins were surveyed. The basic weight of each twin was altered because households containing twins had a higher selection probability than households with only one in-scope child. The adjustment affected 76 households.

3) Weighting of children selected in Cycles 1, 2 and 3

It was not necessary to apply all the adjustments described in the preceding section, since that had already been done in Cycles 1, 2 and 3. The basic weight was the same as the one used for longitudinal weighting of Cycle 4 respondents. Hence it reflected the converts among the children introduced in Cycles 1 and 2. Two adjustments were required.

Special situation of children introduced in Cycle 2

Though they are included in the longitudinal population, the children who were introduced in Cycle 2 and dropped from the Cycle 4 sample to reduce the response burden are not included in the cross-sectional target population.

When those children were introduced in Cycle 2, an adjustment was made to offset the fact that children with older siblings were overrepresented in the sample. That adjustment was affected by the exclusion of some children from Cycle 4. As a result, the Cycle 3 basic weight was divided by the 1997 adjustment, and a new adjustment was made to correct for the number of households with older siblings. This adjustment was computed

for each province. The adjusted weight was the initial weight used for children introduced in Cycle 2.

Adjustment 1: Non-response adjustment

This factor boosted the basic weight to compensate for non-response. The adjustment used at this stage was computed with cross-sectional respondents. Out-of-scope children were not included in the numerator or the denominator.

Adjustment 2: Adjustment for interprovincial migration

The purpose of the second adjustment was to minimize the impact of rare types of interprovincial migration. Some children selected in 1994, 1996 and 1998 moved to another province after the first interview. In some cases, that created excessive weights for the new province of residence. For example, the weight of a child selected in Ontario was much larger than the weight of a child selected in Prince Edward Island. If a child selected in Ontario moved to Prince Edward Island and kept his/her initial weight, it would have a huge impact on the estimates for Prince Edward Island. This type of migration is very rare in the target population. Accordingly, it is unreasonable to assume that the child who moved from Ontario to Prince Edward Island represented a large number of children who did likewise in the target population. This would be a very unusual event. As a result, the weight of such children was adjusted downward.

12.4 Integration of weights

The three weight computation methods described above were used to produce estimates for their respective target populations. In some cases, however, those target populations were not separate. Consequently, an adjustment factor had to be computed to compensate for the overlap. A final adjustment was also needed to ensure that the weights produced estimates consistent with population estimates derived from other data.

12.4.1 Adjustment for target population overlap

There were four types of households in the NLSCY sample: those selected in Cycle 1, those selected in Cycle 2, those selected in Cycle 3, and those selected in Cycle 4. However, the target populations of those four samples overlapped in the case of children selected in 1996 who were five years old in 2000 and five-year-olds selected from the Birth Register in 2000. It was necessary to correct for the overlaps to ensure that the sample did not systematically overestimate the characteristics of the population of five-year-olds selected from the LFS sample.

To take into account the relative contribution of each sample, a multiplier was computed for each province. To obtain the optimum combination of the two samples, the multiplier would depend on the precision of the estimates produced from each sample. For example, an estimate from a high-precision sample is given more weight than an estimate from a low-precision sample. In this example, the former sample would have a large adjustment factor, and the latter a small factor. This approach is illustrated below.

Suppose that 30 longitudinal five-year-olds were sampled in New Brunswick in 1996 and 10 five-year-olds were selected from the Birth Register in 2000. Suppose also that the design of the Birth Register sample was twice as effective as the

design of the 1996 sample. In that case, the adjustment factor for the longitudinal children would be

$\frac{30/2}{30/2 ? 10} ? 0.6$ and the factor for the Birth Register sample would be 0.4.

Note that the sum of the two adjustment factors is 1.

12.4.2 Post-stratification adjustment

The weights computed to that point were post-stratified to ensure that the national and provincial estimates were consistent with the population estimates for aged 0-17 in January 2001. For Cycle 4, the post-strata were defined by province, age and sex. This adjustment factor was computed for each post-stratum. It was defined as the ratio of the population estimates to the sum of the post-stratum weights.

This adjustment was the final step in weighting the cross-sectional sample for Cycle 4 of the NLSCY.

13.0 Data quality and coverage

This chapter provides the user with information about the various factors affecting the quality of the survey data. There are two main types of error: sampling error and non-sampling errors. We will pay special attention to non-sampling errors in this chapter. We will conclude with an overview of quality indexes and an assessment of the quality of the direct measures components.

13.1 Sampling error

The estimates derived from this survey are based on a sample of children. If we had done a census of the target population with the same questionnaires, interviewers, supervisors, processing methods and so on, we might have obtained slightly different values. The difference between the estimates produced by a sample and the estimates obtained through complete enumeration under similar conditions is known as the sampling error of the estimates.

Sampling error can be estimated using the variance. For more details on calculating the estimated sampling error, see Chapters 14 and 15.

13.2 Non-sampling errors

There are many sources of non-sampling errors in any survey. Interviewers may misunderstand survey instructions; respondents may make mistakes in answering the questions; responses may be recorded in the questionnaire incorrectly; and errors may be made in processing and tabulating the data. These examples of non-sampling errors are difficult to quantify. Other kinds of error (or bias), especially non-response and the coverage of the intended population, are more easily quantifiable.

13.3 Non-response

Non-response is a type of error that can lead to bias in the survey estimates. Biased estimates can result if non-respondents have significantly different characteristics from respondents. In Chapter 10, cross-sectional response rates were computed for various domains: provinces, sample frame, age groups, response history and to-be-traced households.

In this section, we will discuss correction of this potential bias and longitudinal response rates for children introduced in Cycle 1.

Characteristics of non-response models

To minimize the risk associated with potential non-response bias, non-response models were used during weighting (see Chapter 12). This technique adjusts the sampling weights to correct for the potential bias due to non-response. However, it does not guarantee that there is no bias due to non-response. The danger stays hidden, and we have to remain vigilant. For that reason, considerable effort has been devoted to minimizing and studying non-response during both collection and processing.

For the households introduced in Cycle 4 of the survey, these models are based on the households' demographic characteristics, i.e., strata, for households selected from the Birth

Register, and a geographic area whose size depends on the number of children selected, for households selected from the LFS.

For households introduced in earlier cycles, a substantial amount of information is available to assess non-response. Consequently, we tried to model the event of non-response in Cycle 4 with variables obtained during Cycle 3 collection. Separate models were developed for each region (the Atlantic Provinces, Quebec, Ontario, the Prairies and British Columbia). We will not go into detail about each regional model, but here are a few of the conclusions we reached.

- Low-income people have a lower response rate than higher-income people.
- Poorly educated people have a lower response rate than better-educated people.
- People who work more have a higher response rate than people who work less.
- People who own their homes or live in single-family homes have a higher response rate than renters and people living in other types of accommodation.
- Parents who have a better perception of their child's academic performance have a higher response rate than parents who do not have as good a perception of their child's academic performance.

13.4 Longitudinal response rates

In a longitudinal survey, the cross-sectional response rate reflects the effectiveness of collection in the current cycle. One indicator of quality is the longitudinal response rate, as it shows how much attrition has taken place since the survey started.

Unfortunately, because of the survey method used in the first two collection cycles, it is impossible to obtain an accurate longitudinal response rate that takes all the components of attrition into account. Ideally, this rate would be the simple ratio of the number of longitudinal children who responded in Cycle 4 to the number of children contacted in Cycle 1. However, the actual number of children in non-responding households is unknown for Cycle 1. Similarly, the number of children from non-responding households in the LFS was also unknown. Since the exact denominator needed to compute the rate is unknown, an exact rate cannot be calculated.

As is common practice for longitudinal surveys, the response rate is based on Cycle 1 respondents only. The following tables show the Cycle 1 response rates by province and age group. The percentages shown are the ratio of the number of respondents in the cycle concerned to the number of respondents in Cycle 1.

13.4.1 Children selected in Cycle 1

Table 1: Unweighted Longitudinal Response Rate of Children Selected in NLSCY Cycle 1, by Province

Province	Number of respondents in cycle 1	Respondents in cycle 2		Respondents in cycle 3		Respondents in Cycle 4	
		Number	% of cycle 1	Number	% of cycle 1	Number	% of cycle 1
Newfoundland & Labrador	950	892	93.9%	844	88.8%	777	81.8%
Prince Edward Island	467	443	94.9%	434	92.9%	392	83.9%
Nova Scotia	1 191	1 068	89.7%	1 081	90.8%	989	83.0%
New Brunswick	1 070	958	89.5%	953	89.1%	839	78.4%
Quebec	3 182	2 944	92.5%	2 836	89.1%	2 529	79.5%
Ontario	4 342	3 899	89.8%	3 736	86.0%	3 326	76.6%
Manitoba	1 232	1 161	94.2%	1 107	89.9%	1 020	82.8%
Saskatchewan	1 413	1 305	92.4%	1 255	88.8%	1 077	76.2%
Alberta	1 599	1 465	91.6%	1 413	88.4%	1 245	77.9%
British Columbia	1 457	1 333	91.5%	1 273	87.4%	1 146	78.7%
Total	16 903	15 468	91.5%	14 932	88.3%	13 340	78.9%

Table 2: Unweighted Longitudinal Response Rate of Children Selected in NLSCY Cycle 1, by Age Group

Age Group[10]	Number of respondents in cycle 1	Respondents in cycle 2		Respondents in cycle 3		Respondents in Cycle 4	
		Number	% of cycle 1	Number	% of cycle 1	Number	% of cycle 1
00-01	4042	3726	92.2%	3603	89.2%	3249	80.5%
02-03	2930	2676	91.3%	2590	88.4%	2349	80.2%
04-05	2656	2425	91.3%	2353	88.6%	2095	78.9%
06-07	2401	2179	90.8%	2104	87.6%	1879	78.3%
08-09	2455	2243	91.4%	2155	87.8%	1893	77.1%
10-11	2419	2219	91.7%	2127	87.9%	1875	77.5%
Total	16903	15468	91.5%	14932	88.3%	13340	78.9%

Attrition accelerated in Cycle 4. Depending on the province, we lost 7% to 12% of respondents between Cycles 3 and 4. The loss between Cycles 2 and 3 was no more than 4%. The situation was similar for age groups; the losses between Cycles 3 and 4 were about 10%, compared with only 4% between Cycles 2 and 3.

13.4.2 Children selected in Cycle 2

The children selected in Cycle 2 were all aged 0 or 1 at the time of their selection. The table below provides the longitudinal response rates for those children by province.

10 Age is the age calculated in cycle 1 according to birth year.

Table 3: Unweighted Longitudinal Response Rate of Children Selected in NLSCY Cycle 2, by Province

Province	Number of respondents in cycle 2	Respondents in cycle 3		Respondents in Cycle 4	
		Number	% of cycle 2	Number	% of cycle 2
Newfoundland & Labrador	146	92	63.0%	115	78.8%
Prince Edward Island	110	70	63.6%	82	74.5%
Nova Scotia	252	165	65.5%	194	77.0%
New Brunswick	241	225	93.4%	185	76.8%
Quebec	819	778	95.0%	639	78.0%
Ontario	1 281	1 147	89.5%	968	75.6%
Manitoba	327	311	95.1%	228	69.7%
Saskatchewan	295	273	92.5%	184	62.4%
Alberta	353	317	89.8%	234	66.3%
British Columbia	328	302	92.1%	249	75.9%
Total	4152	3680	88.6%	3078	74.1%

There are a number of reasons for the observed attrition. Between Cycles 2 and 3, 164 children in the Atlantic were omitted from the sample as a result of a transmission problem. This accounts for the very low response rates for Newfoundland and Labrador, Prince Edward Island and Nova Scotia compared with other provinces. Those children were reintroduced to the sample in Cycle 4, which explains the improvement in response rates between Cycles 3 and 4. For Cycle 4, 439 children were dropped from the sample to reduce the response burden; that alone accounts for 10% of the observed attrition. Nevertheless, the attrition rate was very high between Cycles 3 and 4. In any case, those children have now completed their stay in the survey.

13.4.3 Children selected in Cycle 3

The children selected in Cycle 3 were all aged 0 or 1 at the time of their selection. Only those who responded in Cycle 3 were followed in Cycle 4, and only those who responded in Cycles 3 and 4 will be followed in Cycle 5. This is different from the approach we took in Cycles 2 and 3, which was to try to convert those who did not respond in previous cycles.

Table 4: Unweighted Longitudinal Response Rate of Children Selected in NLSCY Cycle 3, by Province

Province	Number of respondents in cycle 3	Respondents in Cycle 4	
		Number	% of cycle 3
Newfoundland & Labrador	568	527	92.8%
Prince Edward Island	273	242	88.6%
Nova Scotia	602	535	88.9%
New Brunswick	601	528	87.9%
Quebec	1360	1218	89.6%
Ontario	1985	1689	85.1%
Manitoba	656	566	86.3%
Saskatchewan	627	546	87.1%
Alberta	771	665	86.3%
British Columbia	682	600	88.0%
Total	8125	7116	87.6%

Because of the lower response rates in Cycle 4, the rate of retention of children after two collection cycles is much lower for children introduced in Cycle 3 than it was for children introduced in Cycles 1 and 2. Since there will be no conversion effort in Cycle 5, attrition can only get worse. Owing to the large sample size, however, the precision of the estimates will not be significantly affected. Although we adjusted the weights to compensate for non-response, potential bias may still be a problem.

13.5 Coverage

A sample is selected from a frame. That frame is intended to represent the target population. For the NLSCY, the sample is intended to represent the population for both longitudinal and for cross-sectional purposes, at the time of collection. However, when the frame does not represent the population accurately, there may be coverage errors. If the frame does not contain all the units in the target population, there is under-coverage. If the frame contains units that are not in the target population, there is over-coverage.

13.5.1 Coverage of the longitudinal sample

Coverage problems for the longitudinal sample may arise at the time of sample selection or during the second, third or fourth contact.

13.5.2 Coverage at the time of selection

1) Respondents from the LFS

Most of the children surveyed in the NLSCY were selected from households that had taken part in the LFS. This selection method leads to three problems that could produce a problem with coverage.

The first problem stems from the fact that only LFS respondents were considered in forming the NLSCY sample. Some households that were LFS non-respondents may

have had eligible children. These households were excluded from the NLSCY sample, which could lead to coverage errors.

The second problem relates to the fact that only households that had children at the time of the LFS were included in the NLSCY sample. Some households may not have been considered because the dwelling was vacant at the time of the LFS interview. Similarly, if the household membership did include eligible children at the time of the LFS, they were out of scope for the NLSCY. These households may change between the times of the LFS interview and when the NLSCY started interviewing. The exclusion of households that may have become out-of-scope from the NLSCY sample may be another source of coverage error.

The third problem is an extension of the later case. For certain selected households, members were no longer residing at the address of the initial LFS interview. Since the area frame of the LFS is based on the selection of a prescribed physical address, the cases were considered invalid and excluded from the interview. The new residents, if any, of that initial household were never contacted to determine if they would be in scope. This under-coverage of the sample frame is due to the length of time between the LFS interview and the NLSCY interview. This situation occurred when the selected occupants moved before collection for the NLSCY started. As a result, the NLSCY sample may under-represent the population of children who lived in more mobile families.

2) Respondents from the Birth Register

In Cycle 3, the Birth Register was used to select a large sample of one-year-olds. The selection method used could lead to two problems that could lead to coverage errors

One problem stems from the fact that the Birth Register does not have full coverage of every child born in 1998 who reside in Canada at sample selection time. Some births may not have been reported or were reported after the sample was selected and one-year-olds from immigrant families would not be on the Birth Register and therefore had no chance of being selected.

Another problem is due to the length of time between birth and sample selection. Some children may no longer be in the target population because they have died or have moved out of the country.

Nevertheless, these potential errors are generally minor, and their impact is negligible.

13.5.3 Coverage issues in subsequent contacts

Even if we could exclude any potential error at the time of collection, another type of error can surface when households are contacted for a second, third or more times.

This error is due to non-response. As noted in the previous section, some household with specific characteristics are less likely to respond in each succeeding cycle. A good example would be Households with low-income. Although non-response adjustment can compensate for the loss of these respondents in the short term, the accrued loss of low-income households may lead to biased estimates when the non-respondents can no longer be compensated by similarly profiled respondents.

13.5.4 Coverage of the cross-sectional sample

For the cross-sectional sample, more than one source was used: LFS respondents for children aged 0 and 1, and the Birth Register for five-year-olds. Then added to the

sample were three returning longitudinal samples: responding children from each of the first three cycles.

13.5.5 Coverage of cross-sectional sources

The sources for the cross-sectional sample are the same as for previous cycles. The LFS respondents have the same sources of potential error as discussed previously. The five-year-olds from the Birth Register also have the potential errors as discussed earlier. However, the five-year span between registration of the births and selection of the sample increase the possibility of errors. Children born in 1995 outside the country, no matter when they arrived in Canada, will not be on the Birth Register. In addition, children born in Canada have had five years to become out of the scope of this survey, for example if they were deceased or had moved out of the country.

13.5.6 Coverage of longitudinal sources

Children in previous cycles were selected to represent their respective 1995, 1997 and 1999 target populations. By using the same children to represent the 2001 population, we introduce two potential coverage and estimation problems: international immigration and inter-provincial migration.

13.6 International immigration

Between the time when the sample was selected and when the Cycle 4 collection started, a period of six, four and two years had elapsed, depending on which cycle the sample was selected. Immigrant children who arrived in Canada during that period were not eligible for selection. The following tables show the differences between the target populations with and without immigrants. Note that these estimates do not account for immigrants who entered Canada after the sample was selected and who emigrated before the Cycle 4 collection period.

Table 5: Difference Between Estimates of Total Population For The 2-17 Age Group Using Population Counts that Ignore Post-Selection Immigration, by Province

Province	Total number of people	Total number of people without immigrants	Difference	Relative difference after 4 cycles	Relative difference after 3 cycles
British Columbia	810,007	749,932	60,075	7.42%	5.59%
Ontario	2,483,067	2,302,712	180,355	7.26%	5.04%
Quebec	1,442,609	1,395,615	46,994	3.26%	2.35%
Alberta	687,666	667,360	20,306	2.95%	2.13%
Nova Scotia	189,493	184,439	5,054	2.67%	2.68%
Manitoba	260,212	253,592	6,620	2.54%	1.87%
Saskatchewan	240,499	237,519	2,980	1.24%	1.00%
Prince Edward Island	30,283	29,975	308	1.02%	0.67%
New Brunswick	150,152	148,953	1,199	0.80%	0.91%
Newfoundland and Labrador	106,103	105,305	798	0.75%	0.67%

The differences are fairly large for British Columbia and Ontario (over 7%). For some age-sex combinations in those provinces (we used age-sex-province combinations as our post-strata), the differences are more than 8%. Since recent immigrants and people who have been Canadians for over six years have different income and labour force characteristics, we run the risk of bias. The chances for bias increase over time. It is very difficult to adjust the weights to offset this bias.

13.7 Inter-provincial migration

The second source of error is inter-provincial migration. The Cross-sectional sample is intended to represent the population of children for each province at the time of collection. The province of residence may be different from the province at the time of selection if the child has moved. The weight represents the province at time of selection and has to be modified to reflect the province of residence at the time of collection. This deviation from the original sample design impacts the known probability of selection associated with the individual that has jumped stratum and can cause serious coverage and estimation issues. Over time, there will be more and more cases where there is a difference between the province at time of selection and province of residence at time of collection.

The problem can be described using this scenario... Children, who were selected in provinces with low probability of selection, such as in Ontario and Quebec, moved to a small province with a much higher probability of selection. Such children might dominate the estimation because of their excessively large sampling weight. To reduce their impact on estimation, their sampling weight was reduced in line with the other sampled units in the same area where they now reside.

Alternately, children who migrated from small provinces to large provinces had much smaller sampling weights than children originally selected in the large provinces. Their impact on the provincial estimates is therefore reduced (perhaps becoming insignificant)

compared to what it would have been if they had stayed in their original province of selection.

Coverage error due to inter-provincial migration is controlled. The technique of limiting the weight of children who migrate from a large province to a small province appears to be effective.

13.7.1 Quality indexes

Sampling and non-response errors were taken into account in planning the survey, and sample sizes were set high enough so that reliable estimates could be obtained after several cycles. The coefficients of variation will reflect those two types of errors.

The coefficients of variation can be used in combination with the sample size to produce quality indexes by province and two-year age cohort. For more details on the development of quality indexes, see section 13.8 of the present chapter. The indexes help determine whether the survey's initial objectives have been achieved. Some extrapolations can be performed to forecast the quality indexes of future cycles.

13.7.2 Quality indexes for the cross-sectional sample

For the purpose of analyzing the overall quality of estimation, the coefficients of variation of about 100 variables were computed for each province, each region and for Canada as a whole.

The quality indexes are based on CVs (with limits of 16.5% and 33.3%) and sample sizes (minimum of 30 respondents). The latter criterion is much more stringent.

13.7.3 Coefficients of variation

For the Cycle 4 cross-sectional sample, the CVs are less than 33.3% for all age cohorts for all provinces for proportions of 15% or less, except for Prince Edward Island, for which the minimum proportion has to be 20%. The CVs are between 16.5% and 33.3% for proportions under 5% in Canada, the Atlantic Provinces, the Prairies, Quebec and Ontario. In summary, the current sample produces estimates of good or acceptable quality.

The following tables show the proportions for which the quality is marginal (yellow - M) or unacceptable (red - X) for each province, the Prairies, the Atlantic and Canada. For example, for Canada (Table 6), the quality of the estimates of a 1% proportion for 12- and 13-year-olds is marginal, and the quality of the estimates of a 1% proportion for eight- and nine-year-olds is acceptable. Where a proportion for an age group is left blank, the quality is good.

Table 6: Quality Indexes, Canada, by Estimated Proportion and Age Group

Age	1%	5%	10%	15%	20%	30%	40%
0-1	M						
2-3							
4-5	M						
6-7	M						
8-9	M						
10-11	M						
12-13	X						
14-15	X						
16-17	X						
5 y.o.	M						

Table 7: Quality Indexes, Atlantic Provinces and Prairies, by Estimated Proportion and Age Group

Age	Atlantic							Prairies						
	1%	5%	10%	15%	20%	30%	40%	1%	5%	10%	15%	20%	30%	40%
0-1	X	M						X	M					
2-3	M							M						
4-5	X	M						X	M					
6-7	X							X	M					
8-9	X	M	M					X	M	M				
10-11	X	M	M					X	M	M				
12-13	X	M	M					X	M	M				
14-15	X	M	M					X	M	M				
16-17	X	M	M					X	M	M				
5 y.o.	M							X						

Table 8: Quality Indexes, Newfoundland and Labrador and Prince Edward Island, by Estimated Proportion and Age Group

Age	Newfoundland and Labrador							Prince Edward Island						
	1%	5%	10%	15%	20%	30%	40%	1%	5%	10%	15%	20%	30%	40%
0-1	X	X	M	M	M			X	X	M	M	M		
2-3	X	X	M	M	M			X	X	M	M			
4-5	X	M	M	M	M			X	X	M	M	M		
6-7	X	X	M	M	M			X	X	X	M	M	M	
8-9	X	X	X	M	M	M		X	X	X	X	M	M	M
10-11	X	X	X	M	M	M		X	X	X	X	M	M	M
12-13	X	X	M	M	M	M		X	X	X	X	M	M	M
14-15	X	X	M	M	M			X	X	X	X	M	M	M
16-17	X	X	M	M	M			X	X	X	X	M	M	M
5 y.o.	X	M						X	M	M				

Table 9: Quality Indexes, Nova Scotia and New Brunswick, by Estimated Proportion and Age Group

Age	Nova Scotia							New Brunswick						
	1%	5%	10%	15%	20%	30%	40%	1%	5%	10%	15%	20%	30%	40%
0-1	X	M	M	M				X	M	M	M			
2-3	X	M						X	M					
4-5	X	M	M					X	M	M				
6-7	X	X	M	M				X	X	M	M	M		
8-9	X	X	M	M	M			X	X	M	M	M		
10-11	X	X	M	M	M			X	X	M	M	M		
12-13	X	X	X	M	M	M		X	X	M	M	M		
14-15	X	X	X	M	M	M		X	X	X	M	M	M	
16-17	X	X	X	M	M			X	X	X	M	M	M	
5 y.o.	X	M						X	M					

Table 10: Quality Indexes, Quebec and Ontario, by Estimated Proportion and Age Group

Age	Quebec							Ontario						
	1%	5%	10%	15%	20%	30%	40%	1%	5%	10%	15%	20%	30%	40%
0-1	X	M						X						
2-3	M							M						
4-5	X	M						X	M					
6-7	X	M						X	M					
8-9	X	M	M	M				X	M	M				
10-11	X	M	M	M				X	M	M				
12-13	X	M	M	M				X	M	M				
14-15	X	M	M	M				X	M	M				
16-17	X	M	M	M				X	M	M				
5 y.o.	X	M						X	M					

Table 11: Quality Indexes, Manitoba and Saskatchewan, by Estimated Proportion and Age Group

Age	Manitoba							Saskatchewan						
	1%	5%	10%	15%	20%	30%	40%	1%	5%	10%	15%	20%	30%	40%
0-1	X	M	M					X	M	M				
2-3	X	M						X	M					
4-5	X	M	M					X	M	M				
6-7	X	X	M	M				X	X	M	M			
8-9	X	X	M	M	M			X	X	M	M	M		
10-11	X	X	X	M	M	M		X	X	M	M	M		
12-13	X	X	X	M	M	M		X	X	M	M	M		
14-15	X	X	M	M	M	M		X	X	M	M	M		
16-17	X	X	X	M	M	M		X	X	M	M	M		
5 y.o.	X	M						X	M	M				

Table 12: Quality Indexes, Alberta and British Columbia, by Estimated Proportion and Age Group

Age	Alberta							British Columbia						
	1%	5%	10%	15%	20%	30%	40%	1%	5%	10%	15%	20%	30%	40%
0-1	X	M	M					X	M	M				
2-3	X	M						X	M					
4-5	X	M	M					X	M	M				
6-7	X	M	M					X	X	M	M			
8-9	X	X	M	M	M			X	X	M	M	M		
10-11	X	X	M	M	M			X	X	M	M	M		
12-13	X	X	M	M	M			X	X	M	M	M		
14-15	X	X	M	M	M			X	X	M	M	M	M	
16-17	X	X	M	M	M			X	X	M	M	M	M	
5 y.o.	X	M						X	M					

13.7.4 Minimum size

There must be at least 30 respondents in any given category to ensure that the estimates meet a minimum reliability criterion. Table 13 provides the lower limits on estimates for proportions that can be released. For example, for Nova Scotia, a 10% estimate for six- and seven-year-olds would not be reliable and would not be released because it would generally be based on fewer than 30 respondents (10% is below the 13% limit). A 15% estimate subject to constraints on the CV could be reliable (15% is above the 13% limit).

Table 13: Lower Limits on the Estimated Proportion to Obtain Releasable Results (Minimum Of 30 Respondents) for Cycle 4, by Province and Cohort

Province	Age Group					
	6-7	8-9	10-11	12-13	14-15	16-17
Newfoundland and Labrador	20%	30%	29%	29%	24%	24%
Prince Edward Island	33%	51%	47%	51%	53%	51%
Nova Scotia	13%	20%	18%	23%	23%	21%
New Brunswick	16%	19%	21%	27%	31%	29%
Quebec	4%	7%	8%	9%	9%	8%
Ontario	3%	5%	6%	6%	6%	7%
Manitoba	12%	16%	20%	24%	20%	24%
Saskatchewan	12%	16%	16%	23%	19%	21%
Alberta	9%	13%	15%	15%	16%	14%
British Columbia	12%	15%	17%	16%	19%	21%

13.8 Quality indexes for future longitudinal samples

It is important to keep in mind that the NLSCY is a longitudinal survey. The initial sample was selected to be large enough to produce reliable estimates for each cohort up to the age of 25. Assuming a 90% response rate in every cycle and that households that did not respond in every cycle have a 50% response rate, we will obtain CVs and

sample sizes in the final collection cycle (from Cycle 8 for those who were 16 and 17 years old in Cycle 4, to Cycle 13 for those who were six and seven years old in Cycle 4) that will enable us to measure the quality indexes and the survey's success.

13.8.1 Coefficients of variation

The minimum proportions required to obtain indexes of good or acceptable quality will not change much from what we observed in Cycle 4. For some provinces, the minimum proportion may rise as much as 5%. In general, the constraint on the CV does not alter the quality index between Cycle 4 and the final collection cycle.

The following table provides quality indexes for each age group of the children introduced in Cycle 1, by province. The letter X denotes an estimate of unacceptable quality, with a CV of more than 33.3%; the letter M denotes an estimate of marginal quality, with a CV of between 16.5% and 33.3%. The Cycle 4 quality indexes are in red for estimates of unacceptable quality and in yellow for estimates of marginal quality. Cells with a slash show the change between the quality index for Cycle 4 and the quality index for the final collection cycle.

Table 14: Quality Indexes, Provinces of Canada, by Estimated Proportion and Age Group for the Final Collection Cycle (Cycle 8 for Ages 16 And 17, Cycle 9 for Ages 14 and 15, …, Cycle 13 for Ages 6 and 7)

Age	Newfoundland and Labrador							Prince Edward Island						
	1%	5%	10%	15%	20%	30%	40%	1%	5%	10%	15%	20%	30%	40%
6-7	X	X	X	M	M	M		X	X	X	X	M	M	M
8-9	X	X	X	M	M	M		X	X	X	X	M	M	M
10-11	X	X	X	M	M	M		X	X	X	X	M	M	M
12-13	X	X	X	M	M	M		X	X	X	X	M	M	M
14-15	X	X	M	M	M	M		X	X	X	X	X	M	M
16-17	X	X	M	M	M			X	X	X	X	M	M	M
	Nova Scotia							New Brunswick						
	1%	5%	10%	15%	20%	30%	40%	1%	5%	10%	15%	20%	30%	40%
6-7	X	X	M	M	M			X	X	M	M	M		
8-9	X	X	X	M	M	M	M	X	X	M	M	M		
10-11	X	X	M	M	M	M		X	X	M	M	M		
12-13	X	X	X	M	M	M		X	X	X	M	M	M	
14-15	X	X	X	M	M	M		X	X	X	M	M	M	
16-17	X	X	M	M	M			X	X	X	M	M	M	
	Quebec							Ontario						
	1%	5%	10%	15%	20%	30%	40%	1%	5%	10%	15%	20%	30%	40%
6-7	X	M	M					X	M	M				
8-9	X	M	M	M				X	M	M	M			
10-11	X	X	M	M	M			X	M	M	M			
12-13	X	X	M	M	M			X	M	M	M			
14-15	X	X	M	M				X	M	M	M			
16-17	X	X	M	M				X	M	M	M			
	Manitoba							Saskatchewan						
	1%	5%	10%	15%	20%	30%	40%	1%	5%	10%	15%	20%	30%	40%
6-7	X	X	M	M	M			X	X	M	M	M		
8-9	X	X	X	M	M	M		X	X	M	M	M		
10-11	X	X	X	M	M	M		X	X	M	M	M		
12-13	X	X	X	M	M	M	M	X	X	X	M	M		
14-15	X	X	X	M	M	M		X	X	M	M	M		
16-17	X	X	X	M	M	M		X	X	M	M	M		
	Alberta							British Columbia						
	1%	5%	10%	15%	20%	30%	40%	1%	5%	10%	15%	20%	30%	40%
6-7	X	X	M	M				X	X	M	M	M		
8-9	X	X	M	M	M			X	X	X	M	M	M	
10-11	X	X	M	M	M			X	X	X	M	M	M	
12-13	X	X	M	M	M			X	X	X	M	M	M	
14-15	X	X	M	M	M			X	X	X	M	M	M	
16-17	X	X	M	M	M			X	X	X	M	M	M	

We also see little change at the national and provincial levels, as indicated in the tables below.

Table 15: Quality Indexes, Atlantic and Prairies, by Estimated Proportion and Age Group for the Final Collection Cycle (Cycle 8 for Ages 16 And 17, Cycle 9 for Ages 14 And 15, …, Cycle 13 for Ages 6 and 7)

Age	Atlantic							Prairies						
	1%	5%	10%	15%	20%	30%	40%	1%	5%	10%	15%	20%	30%	40%
6-7	X	M	M					X	M					
8-9	X	M	M					X	M	M				
10-11	X	M	M					X	M	M				
12-13	X	M	M					X	M	M				
14-15	X	M	M					X	M	M				
16-17	X	M	M					X	M	M				

For the Prairies, the only difference occurs in the 6-7 age group. The estimates for 5% and 10% proportions drop from good to marginal quality by Cycle 13. For the 10% proportion, this change does not occur until Cycle 13.

Table 16: Quality Indexes, Canada, by Estimated Proportion and Age Group for the Final Collection Cycle (Cycle 8 for Ages 16 and 17, Cycle 9 for Ages 14 and 15, …, Cycle 13 for Ages 6 and 7)

Age	1%	5%	10%	15%	20%	30%	40%
6-7	M						
8-9	X						
10-11	X	M					
12-13	X	M					
14-15	X						
16-17	X						

At the national level, the estimates for 1% proportions for the 8-9 and 10-11 age groups fall from marginal to unacceptable quality. For 10- and 11-year-olds, the quality of 5% proportions becomes marginal in Cycles 10 and 11. For 12- and 13-year-olds, the quality of 5% proportions drops to marginal in Cycles 8, 9 and 10.

Consequently, if the assumptions we have made concerning response rates in future cycles hold – i.e., a 90% response rate for all-cycle respondents and a 50% rate for others – the quality indexes will remain within the limits set by the survey's provincial objectives, except in the case of Prince Edward Island, which has already failed to meet those quality objectives in Cycle 4. At the national level, only the 6-7 age group maintains good-quality estimates for 1% proportions. For the 10-11 and 12-13 age groups, the quality of 5% estimates falls from good to marginal two or three cycles before the end of their stint in the sample.

13.8.2 Minimum size

The attrition of the sample combined with the small samples in the Atlantic Provinces places increasing limits on the release of good- or acceptable-quality results for those provinces. The lower limits for each province are presented in Table 17 below.

Table 17: Lower Limits on the Estimated Proportion to Obtain Releasable Results (Minimum Of 30 Respondents) for the Final Collection Cycle, by Province and Cohort

Province	Age Group					
	6-7	8-9	10-11	12-13	14-15	16-17
Newfoundland and Labrador	26%	35%	34%	33%	28%	24%
Prince Edward Island	50%	63%	60%	62%	68%	58%
Nova Scotia	18%	27%	24%	29%	28%	24%
New Brunswick	23%	25%	26%	30%	33%	31%
Quebec	6%	9%	10%	11%	11%	10%
Ontario	5%	7%	8%	8%	8%	8%
Manitoba	16%	21%	26%	29%	24%	27%
Saskatchewan	17%	21%	21%	27%	21%	24%
Alberta	14%	19%	20%	20%	20%	18%
British Columbia	16%	20%	22%	21%	23%	24%

13.9 Data Quality for direct measures components

In this section, we will assess the quality of the direct measures components, by essentially aiming on characteristics affecting non-response.

13.9.1 Self Complete component

The self complete component is aimed at children 10 years old or older. It's made up of different sections. For most of the shared sections, response rate tends to diminish with the child's age. Response rates are similar from one section to another. Hence, response rates are around 78% for the younger children (10-11 years old) and go down to around 73% for the older ones (16-17 years old). Most of the children who did not answer one section do not answer the other ones (95% to 98% of children who did not answer to one section do not answer another one).

Table 18: Approximate Response Rate by Section (one question is assess by section) of the Self Complete Component, NLSCY Cycle 4, Canada

Section	Response rate			
	10-11 y.o.	12-13 y.o.	14-15 y.o.	16-17 y.o.
Friends and family	78.4%	78.6%	76.3%	73.4%
About Myself	78.6%	77.9%	72.8%	74.1%
Feelings and behaviour	77.0%	78.3%	76.8%	73.3%
My parents	76.8%	76.9%	76.5%	73.9%
Tobacco, alcohol and drugs	76.0%	76.1%	75.7%	73.7%
School	78.9%	78.9%	78.3%	X*
Puberty	72.3%	X	X	X
Health	X	72.3%	73.5%	72.4%
Activities	76.1%	77.2%	76.5%	X
Work and money	X	74.0%	63.0%	X
Relationships	X	X	X	72.0%
Decision making	X	X	X	71.6%
Number of surveyed children	2083	1855	1872	1856

* X : section not asked to the respondent

Response behaviour may vary according to some children's characteristics. When these data are used for analysis purposes, one should take into account these response behaviours in the comments, especially when analyzing small domains. Since weighting was used with the final data the analysis can be very volatile.

The province is a variable with highly related to non-response. As with all direct measures, response rates in Alberta are especially low at 61%.

Table 19: Self Complete Component Weighted Response Rate by Province, NLSCY Cycle 4

Province	Response rate
Newfoundland and Labrador	88%
Prince Edward Island	89%
Nova Scotia	75%
New Brunswick	78%
Quebec	86%
Ontario	79%
Manitoba	75%
Saskatchewan	72%
Alberta	61%
British Columbia	74%

Moreover, in Cycle 4, girls responded slightly more than boys. Differences are quite small between the response rates (1% to 3%) for any age or section. However, one could note that for the "My parents" section, the response rate for girls in the sample is around 78% and 73% for boys. For this section, if we look at the 10-11 year-olds, respondent girls make up 80% of the 10-11 years old girls and the boys answer in a 70% proportion (weighted counts).

13.9.2 Number Knowledge component

This assessment is aimed at 4 and 5 years old children. For the data quality assessment, we investigated some variables that are related with non-response.

Response rates differ according to the province. Using weighted data, one obtains rates ranging from 76% in Alberta to 96% in Prince Edward Island. The small response rate in Alberta is consistent with the other direct measures components, establishing a more general problem with this province.

Table 20: Number Knowledge Component Weighted Response Rate by Province, NLSCY Cycle 4

Province	Response rate
Newfoundland and Labrador	85%
Prince Edward Island	96%
Nova Scotia	90%
New Brunswick	90%
Quebec	92%
Ontario	82%
Manitoba	88%
Saskatchewan	79%
Alberta	76%
British Columbia	86%

Another variable related with non-response was if children were not yet in school. They responded more (92%) than children who were in school (83%).

Non-response adjustments for total non-response contribute to lowering the potential biases due to non-response. Using survey weights is therefore quite significant. The user should remain cautious and assess, on a case by case basis, if it is necessary to evaluate the impact of non-response of this component.

13.9.3 Peabody Picture and Vocabulary Test component (PPVT)

This assessment is aimed at children 4 to 6 years of age. However, some older children not in school are also eligible for this assessment. The response rate for the test is around 86%. This response rate was broken down by other variables and some highlights are presented here.

The table shows the break-down of the non-response by province. There is a notably large level of non-response for Alberta. We suspect that the non-response in Alberta may be due to operational or is endemic to the population as similar response rates were measured for other components of the survey.

Table 21: Knowledge Number Component Weighted Response Rate by Province, NLSCY Cycle 4

Province	Response rate
Newfoundland and Labrador	88%
Prince Edward Island	97%
Nova Scotia	89%
New Brunswick	89%
Quebec	92%
Ontario	82%
Manitoba	87%
Saskatchewan	81%
Alberta	78%
British Columbia	86%

Among the francophone populations in Canada we observed a higher response rate. We also found that if the dwelling of the respondent was owned by a member of the household, the non-response rate was about 14%, while the non-response rate rose to 20% otherwise. The number of rooms in the dwelling was also seen to be related to the response rate, respondents living in dwellings with fewer rooms tended to have lower response.

Of note, we also found that where the PMK reported that they talked about things together with the child every day, the non-response was 13.3%. In contrast, where the frequency was less than every day, the non-response was as high as 15.8% to 18.0%, depending on the category reported.

Some other factors that we found to be affected by the non-response are: the marital status of the PMK, income, aspects of the family functioning and PMK depression scales. The list is not exhaustive, and we only give these examples to illustrate that there is potential for measurement error, depending on what variables are being used and how non-response is dealt with in the analysis.

13.9.4 Who Am I component

The "Who Am I" is an assessment directly administered to children of ages 4 and 5 in home – at the time of the interview. It consists of ten tasks that the subject has to complete with pen and paper, under the directions of the interviewer. All children of the appropriate age for whom an interview has taken place should respond to this assessment. For various reasons in some cases the assessment was not completed. In what follows we give some details on the extent of this non-response and possible relationships with other factors.

Of the 6,078 eligible children, around 27% did not return a test, or returned a blank test. This non-response is not uniform. Using data mining software and contextual knowledge we determined some characteristics that seem to be related to the non-response. One factor that stands in a category of its own is the province where the child lives. On one hand, there are inherent differences between provinces, which might explain the low non-response in P.E.I.. In our case it is at 9% and previous cycles had similar non-response rates. On the other hand operational difficulties which were non-uniform across regional offices created a non-response component that may only appear in this cycle – such as the 31% rate in Saskatchewan and the 28% rate in Alberta. For Quebec the rate was 18%, while for Ontario it was 35%.

Table 22: Who Am I Component Weighted Response Rate by Province, NLSCY Cycle 4

Province	Response rate
Newfoundland and Labrador	78%
Prince Edward Island	91%
Nova Scotia	81%
New Brunswick	79%
Quebec	82%
Ontario	65%
Manitoba	81%
Saskatchewan	69%
Alberta	72%
British Columbia	78%

Other factors were found by inspecting variables collected at the time of the interview, but for which the non-response was much lower. They are either basic questions, which place little burden on the respondent, or were asked at the early stages of the interview. For instance, the non-response appeared lower for those children not yet in school.

Some factors were related to the home environment. The children that play sports together with their parents only once or twice a month have a non-response rate of about 22% , while those that play sports together everyday have a non-response rate of 29%. Also, children that never play sports with their parents have a non-response rate of 29%, so even if the interaction is there, the possible explanation is not always clear. When categorizing by the proportion of praise that the parent uses when they talk to the child, the non-response rate is 34% where praise is used about half the time, and 23% where praise is used more than half the time.

The communication ability of the child could be expected to have an impact on the non-response, as the assessment might be less of a burden where the child communicates easily. Indeed, where the parent reports that the child can explain things often, the non-response is 24%, compared to 33% where the child can only rarely explain things. Similar rates of 24% and 31% are obtained for the classification by whether the child can follow a conversation often or rarely. One should keep in mind though that the communication ability might be related to the assessment outcome as well, hence there is the possibility that the non-response is causing bias.

The non-response to the “Who Am I” assessment should be taken into account when analyzing the data, since it appears to be related to several factors, some of which may be included in the particular analysis. We only gave here a few examples of factors from different areas – the user needs to make a detailed assessment for the particular covariates he/she intends to use.

13.9.5 Mathematics component

13.9.5.1 Collection

The mathematical ability of survey respondents in school grades 2 through 10 - or national equivalents – is measured by using booklets of multiple choice questions. Each grade level receives a separate booklet, except for grades 9 and 10, which both have the same booklet. The number of

questions is 20 for all booklets, except again for the grade 9-10 booklet, which only has 15 questions.

In previous cycles the booklets were mailed to the school of the child, were administered by the school staff to the respondent, and then mailed back. A decision was made during the collection of Cycle 4 to administer the booklets at the personal interview. Because of the timing of the decision, operational difficulties have arisen, and we present here the potential impact on data quality.

One issue is that at the time of the decision some personal interviews were already completed, and the interviewer needed to go back to the household in order to administer the booklet. In such a situation the appropriate entry/exit procedures could not be used, and we lack auxiliary information, such as the exact date when the assessment was completed. There was also an increase in the number of invalid identifiers – each test carries an interviewer recorded identifier – which is consistent with non-standard operating conditions. It is likely that a mistake in recording the identifier would be easier to catch if the interviewer had the laptop with the case in progress, and the identifier already displayed on the screen. The majority of the invalid identifiers were recovered by using auxiliary information. Presented here are the counts related to the booklets:

There were 8071 booklets captured. Of the 8071 records, 962 had an invalid identifier. In most cases at least part of the identifier could be used together with the auxiliary information, in the recovery process. All but 66 identifiers have been recovered, so we have 8005 captured booklets. We tried to see what circumstances could explain the broken identifiers. The only clear difference was made by the break-down by province. In New-Brunswick, 77% of the booklets had a bad identifier, while the corresponding percentage in other provinces ranged from 3.6% to 11%. Since historically the regional office responsible for New Brunswick had finished collection earlier than other offices, it is possible that they were the most affected by the timing of the decision to change the collection procedures.

13.9.5.2 Non-response

Of the 9077 children that were respondents and were eligible for the math assessment, 88% had a captured booklet. Some of the booklets were returned blank, while others did not have all questions answered. There were 627 blank booklets, while booklets with at least one but less than five questions answered were 106 in number.

The method used to assign scores can work around missing questions, and we derived a score for all the non-blank booklets. It is true that a score based on only some questions is less reliable than a score based on all the questions. However, since statistical analysis always uses subgroups of records and not individual records, even the less reliable scores carry useful information – the perturbations tend to cancel each other as group size increases. Also note that the records that have between 1 and 4 questions answered account for less than 1.5% of the number of tests. The underlying assumption for the tests, were a prescribed time limit is imposed, is that the respondents inability to complete the test is part of the ability assessment.

The overall response rate is 81% and we could therefore assign a score for these eligible children. The response rate by province is presented in Table 23.

Table 23: Mathematics Component Weighted Response Rate by Province, NLSCY Cycle 4

Province	Response rate
Newfoundland and Labrador	87%
Prince Edward Island	89%
Nova Scotia	81%
New Brunswick	69%
Quebec	89%
Ontario	77%
Manitoba	84%
Saskatchewan	83%
Alberta	72%
British Columbia	79%

It can be seen that in Alberta and New Brunswick the response rates were the lowest. Given the collection issues mentioned above, the high non-response rate in New Brunswick is consistent with the fact that the bad identifiers were predominant – it is possible that the 66 unsolved identifiers are from the same province. Also the high non-response in Alberta is consistent with similar non-response for other components in Alberta. In this cycle, for this province there seems to be a more general problem – perhaps operational as well or endemic to that population.

Apart from operational problems, population characteristics are often related with the non-response. We found this to be the case for the math assessment. Here are a few such characteristics: marital status of the PMK, highest schooling level of the PMK, child being in good health, income adequacy. It may be that income adequacy alone is the major influence among these four. There is also association of the non-response with items from the adult depression scale, the language of the test – French speaking children have a lower non-response – and the school grade the child is in was also related to non-response.

Users should be aware that the non-response to the math tests can create a bias and should take into account its impact when analyzing the data.

13.9.6 Response rate to the cognitive instrument

This component is aimed at youth aged 16 and 17 regardless of their school level. The response rate for the Cognitive Measure test was 67% overall. Youth reaching their late teens experience increased independence, mobility and time demands (such as exams, employment, and homework), all of which have a negative effect on response. Some subgroups with notably lower response rates were males and Anglophones, each at 64%. Only 4% of non-respondents and 46% of partial respondents to the Cycle 4 self-complete questionnaire responded to the cognitive test. Youths in our longitudinal sample born in 1982 (of whom less than 30 are still responding) were inadvertently excluded from testing even though their effective age is 17.

Those using the Cognitive Measure in analysis should be aware of the geographic disparity in response. In particular, the province of Alberta (43%), especially Edmonton (30%), had a low response rate to the cognitive test. When doing regional comparisons, analysts should be wary of statements that focus on regional differences as non-response error may in part contribute to theses differences.

Conclusion

Data quality depends on various sources of error. Some types, such as sampling error, are planned for from the beginning, and the required adjustments are made. Others are more difficult to correct for, and we have identified the different types of potential biases. It is important to note that those biases are generally small and never compromise the quality of the estimates. Sufficient information is known about the children from information of other components, that corrective procedures can be applied to component non-response to lessen the impact of measured biases.

Non-response should be assessed when analyzing direct measures components. Non-response and some child's characteristics can be correlated. Survey weights are not adjusted for the specific non-response to a direct measure component. They only take into account total non-response.

14.0 Guidelines for Tabulation, Analysis and Release

This section of the documentation outlines the guidelines to be adhered to by users tabulating, analyzing, publishing or otherwise releasing any data derived from the survey microdata file. With the aid of these guidelines, users of microdata should be able to produce the same figures as those produced by Statistics Canada and, at the same time, will be able to develop currently unpublished figures in a manner consistent with these established guidelines.

14.1 Rounding Guidelines

In order that estimates for publication or other release derived from the NLSCY microdata file correspond to those produced by Statistics Canada, users are urged to adhere to the following guidelines regarding the rounding of such estimates:

a) Estimates in the main body of a statistical table are to be rounded to the nearest hundred units using the normal rounding technique. In normal rounding, if the first or only digit to be dropped is 0 to 4, the last digit to be retained is not changed. If the first or only digit to be dropped is 5 to 9, the last digit to be retained is raised by one. For example, in normal rounding to the nearest 100, if the last two digits are between 00 and 49, they are changed to 00 and the preceding digit (the hundreds digit) is left unchanged. If the last digits are between 50 and 99 they are changed to 00 and the preceding digit is incremented by 1.

b) Marginal sub-totals and totals in statistical tables are to be derived from their nearest 100 units using normal rounding.

c) Averages, proportions, rates and percentages are to be computed from unrounded components (i.e., numerators and/or denominators) and then are to be rounded themselves to one decimal using normal rounding.

d) Sums and differences of aggregates (or ratios) are to be derived from their corresponding unrounded components and then are to be rounded themselves to the nearest 100 units (or the nearest one decimal) using normal rounding.

e) In instances where, due to technical or other limitations, a rounding technique other than normal rounding is used resulting in estimates to be published or otherwise released which differ from corresponding estimates published by Statistics Canada, users are urged to note the reason for such differences in the publication or release document(s).

f) Under no circumstances are unrounded estimates to be published or otherwise released by users. Unrounded estimates imply greater precision than actually exists.

14.2 Sample Weighting Guidelines for Tabulation

Sample Design
The sample design used for the NLSCY was not self-weighting. When producing simple estimates, including the production of ordinary statistical tables, users must apply the proper survey weight. If proper weights are not used, the estimates derived from the microdata file cannot be considered to be representative of the survey population, and will not correspond to those produced by Statistics Canada. The weight assigned to each child reflects the number of children represented by a particular respondent.

For any analysis dealing with correlation analysis or any other statistics where a significance measure is required, it is recommended that a "sample" weight be used. This weight is obtained by multiplying the survey weight by the sample size and dividing this total by the total estimated population. This produces a mean weight of 1 and a sum of weights equal to the sample size.

Benefit of Using an Adjusted Weight
The benefit of this adjusted weight is that an over estimation of the significance (which is very sensitive to sample size is avoided while maintaining the same distributions as those obtained when using the sampling weight. The disadvantage is that the numerator is not weighted up to the target population and the Coefficient of Variance Tables described in section 12 and presented in Appendix 3 are no longer useful as a measure of data quality.

Software Differences
Users should also note that some software packages may not allow the generation of estimates that exactly match those available from Statistics Canada, because of their treatment of the weight field.

14.3 Definitions of Types of Estimates: Categorical vs. Quantitative

Unit of Analysis
The NLSCY file has been set up so that the child is the unit of analysis. The weight that can be found on each record (DWTCW01C for the cross-sectional sample, DWTCW01L for the longitudinal sample of children introduced in one of the first three cycles and respondent in Cycle 4, and DWTCdW1L for the longitudinal sample introduced in cycle 1 and respondent to all four cycles) is a "child" weight. Estimates of parents or families cannot be made from the NLSCY microdata file.

14.3.1 Categorical Estimates

Categorical estimates are estimates of the number, or percentage of the surveyed population possessing certain characteristics or falling into some defined category. An estimate of the number of persons possessing a certain characteristic may also be referred to as an estimate of an aggregate.

Examples of Categorical Questions
Q: Was (the child) born before, after or on the due date?

R: Before
After
On due date

Q: Compared to other babies in general, would you say the (the child's) health at birth was:

R: Excellent
Very good
Good
Fair
Poor

14.3.2 Quantitative Estimates

Quantitative estimates are estimates of totals or of means, medians and other measures of central tendency of quantities based upon some or all of the members of the surveyed population. They also specifically involve estimates of the form $\frac{\hat{X}}{\hat{Y}}$

where $\hat{X}$ is an estimate of the surveyed population total quantity and $\hat{Y}$ is an estimate of the number of people in the surveyed population contributing to that total quantity.

An example of a quantitative estimate is the average number of days of care received by babies who required special medical care following birth. The numerator is an estimate of the total number of days for which babies required special care. The denominator is the number of babies who required special care at birth.

Example of a Quantitative Question
Q: For how many days, in total, was this care received?

R: Days

Q: What was the child's weight at birth in pounds and ounces?

R: Pounds Ounces

14.3.3 Tabulation of Categorical Estimates

Estimates of the number of children with a certain characteristic can be obtained from the microdata file by summing the final weights of all records possessing the characteristic(s) of interest. These estimates may be cross-sectional or longitudinal.

Proportions and ratios of the form $\frac{\hat{X}}{\hat{Y}}$ are obtained by:

(a) summing the final weights of records having the characteristic of interest for the numerator ($\hat{X}$);

(b) summing the final weights of records having the characteristic of interest for the denominator ($\hat{Y}$), then;

(c) dividing the numerator estimate by the denominator estimate.

14.3.4 Tabulation of Quantitative Estimates

Estimates of quantities can be obtained from the microdata file by multiplying the value of the variable of interest by the final weight for each record, then summing this quantity over all records of interest.

For example, to obtain an estimate of the total number of days of special care received by infants who were born prematurely

multiply the number of days for which special care was received by the final weight then sum this value over all records for which the child was born prematurely.

To obtain a weighted average of the form $\frac{\hat{X}}{\hat{Y}}$, the numerator ($\hat{X}$) is calculated as

for a quantitative estimate and the denominator (Y), is calculated as for a categorical estimate. For example, to estimate the average number of days spent in special care by premature babies:

(a) estimate the total number of days as described above;
(b) estimate the number of children in this category by summing the final weights of all records for babies which were premature; then
(c) divide estimate (a) by estimate (b).

14.4 Guidelines for Statistical Analysis

Sample Design
The NLSCY is based upon a complex sample design, with stratification, multiple stages of selection, and unequal probabilities of selection of respondents. Using data from such complex surveys presents problems to analysts because the survey design and the selection probabilities affect the estimation and variance calculation procedures that should be used. In order for survey estimates and analyses to be free from bias, the survey weights must be used.

Variance Estimates
While many analysis procedures found in statistical packages allow weights to be used, the meaning or definition of the weight in these procedures differ from that which is appropriate in a sample survey framework, with the result that while in many cases the estimates produced by the packages are correct, the variance estimates that are calculated are not adequate. Variances for simple estimates such as totals, proportions and ratios (for qualitative variables) are provided in the accompanying Sampling Variability Tables.

Rescaling the Weights
For other analysis techniques (for example linear regression, logistic regression and analysis of variance), a method exists which can make the variances calculated by the standard packages more meaningful, by incorporating the unequal probabilities of selection. The method rescales the weights so that there is an average weight of 1.

Example of Rescaling the Weights
For example, suppose that analysis of all male children is required. The steps to rescale the weights are as follows:

-Select all respondents from the file with SEX = male (variable DMMCQ02).
-Calculate the AVERAGE weight for these records by summing the original person weights (DWTCW01C) from the microdata file for these records and then dividing by the number of records with SEX = male.
-For each of these records, calculate a RESCALED weight equal to the original person weight divided by the AVERAGE weight.
-Perform the analysis for these respondents using the RESCALED weight.

However, because the stratification and clustering of the sample's design are still not taken into account, the variances calculated in this way are likely to be under-estimated.

Calculation of Variance Estimates
The calculation of truly meaningful variance estimates requires detailed knowledge of the design of the survey. Such detail cannot be given in this microdata file because of confidentiality. Variances that take the complete sample design into account can be calculated for many statistics by Statistics Canada on a cost-recovery basis.

14.5 Coefficient of Variation Release Guidelines

Release Guidelines

Before releasing and/or publishing any estimate from the NLSCY, users should first determine the quality level of the estimate. The quality levels are acceptable, marginal and unacceptable. As discussed in Chapter 13, sampling and non-sampling errors both influence data quality. For the purposes of this document, however, estimate quality is based solely on the sampling error illustrated by the coefficient of variation, as shown in the table below.

First, the number of children who contribute to the calculation of the estimate should be determined. If this number is less than 30, the weighted estimate should be considered to be of unacceptable quality.

For weighted estimates based on sample sizes of 30 or more, users should determine the coefficient of variation of the estimate and follow the guidelines below. These quality level guidelines should be applied to weighted rounded estimates.

All estimates can be considered releasable. However, those of marginal or unacceptable quality level must be accompanied by a warning to caution subsequent users.

14.6 Quality Level Guidelines

Quality Level of Estimate	Guidelines
1. Acceptable	Estimates have: a sample size of 30 or more, and low coefficients of variation in the range 0.0% to 16.5%. No warning is required.
2. Marginal	Estimates have: a sample size of 30 or more, and high coefficients of variation in the range 16.6% to 33.3%. Estimates should be flagged with the letter M (or some similar identifier). They should be accompanied by a warning to caution subsequent users about the high levels of error, associated with the estimates.
3. Unacceptable	Estimates have: a sample size of less than 30, or very high coefficients of variation in excess of 33.3%. Statistics Canada recommends not to release estimates of unacceptable quality. However, if the user chooses to do so then estimates should be flagged with the letter U (or some similar identifier) and the following warning should accompany the estimates: "The user is advised that…(specify the data)…do not meet Statistics Canada's quality standards for this statistical program. Conclusions based on these data will be unreliable, and most likely invalid. These data and any consequent findings should not be published. If the user chooses to publish these data or findings, then this disclaimer must be published with the data."

15.0 Variance calculation

NLSCY is a probabilistic survey. A sample has been selected to represent the target population. A given variability is inherent to the sample selection, which would be different if a different sample would have been selected. This variability is known as the sampling error, as described in Chapter 13. In addition, adjustments have been made to take into account the non-response. The measure of this variability is known as the variance. In this chapter, we will explain why it is important to calculate the variance and we will present different tools to do so.

15.1 Importance of the variance

The variance of an estimate is a good indicator of the quality of the indicator. A high variance estimate is considered non reliable. In order to quantify a large variance, a relative measure of the variability is used, namely the coefficient of variation (CV). The coefficient of variation is defined as the ratio of the square root of the variance over the estimate. The square root of the variance is also known as the standard deviation. The coefficient of variation, as opposed to the variance, allows the analyst to compare estimates of different magnitude on the same scale. As a result, it is possible to assess the quality of any estimates with the CV.

Also, the variance or the CV is required for statistical tests such as hypothesis tests, which determine if two estimates are statistically different. Consequently, variance or CV calculation is mandatory.

15.2 Variance and CV calculation

It is almost impossible to derive an exact formula to calculate the variance for the NLSCY, due to the complex sample design, non-response adjustments and the post-stratification. A very good way to approximate the true variance is to use the Bootstrap method. A set of 1000 Bootstrap weights is available. Variance calculation using these 1000 Bootstrap weights involves calculating the estimates with each of these 1000 weights and then, calculating the variance of these 1000 estimates.

Three tools, all making use of the bootstrap weights, have been developed to help the users to calculate the variance and the CV for their estimates. These tools are:

- CV look-up tables, using a representative design effect, to get approximate CVs for some domains, by age cohort or by province.
- Excel spreadsheet for proportions, giving approximate CVs for a large number of domains, by age and by province.
- Macros to calculate the variance, using the Bootstrap weights.

The use of one or more of these tools depends on the type of analysis and the level of precision required.

15.2.1 CV look-up tables

In order to supply coefficients of variation which would be applicable to a wide variety of categorical estimates produced from this microdata file and which could be readily accessed by the user, a set of Approximate Sampling Variability Tables has been produced. These "look-up" tables can be found in the Research Data Centres, in Word format, under the filename of "TabVar_C4_F.doc". They allow the user to obtain an approximate coefficient of variation based on the size of the estimate calculated from the survey data.

15.2.2 Computation of Coefficients of Variation

The coefficients of variation (c.v.) are derived using the variance formula for simple random sampling and incorporate a factor which reflects the multi-stage, clustered nature of the sample design. This factor, known as the design effect, was determined by first calculating design effects for a wide range of characteristics and then choosing from among these a conservative value to be used in the look-up tables which would then apply to the entire set of characteristics.

15.2.3 Sample Requirements

For the NLSCY, the sample was constructed taking account the following requirements.

- A sufficient sample was required in each of the 10 provinces to allow for the production of reliable estimates for all longitudinal children who were 0 to 11 years of age in Cycle 1.

- It was also necessary to have a large enough sample to produce estimates for Cycle 1 at the Canada level by seven key age groupings or cohorts: 0 to 11 months, 1 year, 2 to 3 years, 4 to 5 years, 6 to 7 years, 8 to 9 years, and 10 to 11 years.

- In each province, a sufficient sample size was required for Cycle 2 to produce reliable estimates for all children who were 0 to 11 years of age in Cycle 1.

15.2.4 Design Effect, Sample Size, Population

The tables that follow show the design effects, sample sizes and population counts by province and age groupings used to produce the Approximate Sampling Variability Tables.

First, the tables for the cross-sectional samples:

Cross-Sectional Sample			
Province	**Design effect**	**Sample size**	**Population**
Newfoundland and Labrador	3.4	1,826	116,080
Prince Edward Island	3.2	1,025	33,311
Nova Scotia	3.8	2,259	208,160
New Brunswick	3.4	2,037	165,078
Quebec	3.9	5,337	1,590,325
Ontario	4.1	7,468	2,747,236
Manitoba	4.0	2,356	289,266
Saskatchewan	3.4	2,353	265,221
Alberta	3.3	2,986	763,858
British Columbia	4.1	2,659	892,908
Atlantic Provinces	3.9	7,147	522,629
Prairies	3.8	7,695	1,318,345
Total	4.9	30,307	7,071,581

Cross-Sectional Sample			
Age group	**Design effect**	**Sample Size**	**Population**
0 to 23 months	2.0	4,008	698,830
2 to 3 years	2.1	6,970	672,184
4 to 5 years	3.4	6,078	781,425
6 to 7 years	2.4	3,284	792,014
8 to 9 years	2.8	2,306	833,959
10 to 11 years	2.8	2,081	841,556
12 to 13 years	2.7	1,855	806,656
14 to 15 years	2.7	1,870	819,163
16 to 17 years	2.6	1,855	825,794
Total (0 to 17 years)	4.9	30,307	7,071,581

Design effects for the longitudinal sample according to each cycle are as follows:

Longitudinal Sample for Children Introduced in Cycle 1			
Province	**Design effect**	**Sample Size**	**Population**
Newfoundland and Labrador	2.6	703	83,306
Prince Edward Island	2.4	383	22,859
Nova Scotia	3.6	947	141,133
New Brunswick	3.0	796	113,100
Quebec	3.3	2,515	1,096,131
Ontario	3.8	3,351	1,759,224
Manitoba	3.6	988	192,461
Saskatchewan	3.0	1,044	185,565
Alberta	3.1	1,346	521,231
British Columbia	3.4	1,116	569,468
Atlantic Provinces	3.1	2,829	360,398
Prairies	3.4	3,378	899,257
Total[11] Canada	4.2	13,340	4,751,437

Longitudinal Sample for Children Introduced in Cycle 1			
Age group	**Design effect**	**Sample size**	**Population**
6 to 7 years	2.6	3,184	694,038
8 to 9 years	2.8	2,307	806,780
10 to 11 years	2.8	2,084	813,570
12 to 13 years	2.7	1,855	777,888
14 to 15 years	2.7	1,870	790,231
16 to 17 years	2.5	2,012	859,828
Total[1] (6 to 17 years)	4.2	13,340	4,751,437

11 Some longitudinal respondents are not cross-sectional respondents (for example, deceased children). Neither province nor age are assigned to these cases. Even so, they are part of the Canadian total.

Longitudinal Sample for Children Introduced in Cycle 2			
Province	**Design effect**	**Sample size**	**Population**
Newfoundland and Labrador	1.9	116	11,897
Prince Edward Island	3.9	81	3,693
Nova Scotia	2.2	188	20,029
New Brunswick	2.2	178	16,815
Quebec	2.1	638	173,505
Ontario	2.1	964	288,921
Manitoba	1.8	223	30,718
Saskatchewan	6.3	181	29,741
Alberta	2.0	251	78,948
British Columbia	1.7	232	85,149
Atlantic Provinces	1.7	563	52,435
Prairies	2.5	655	139,406
Total[1] (4 to 5 years)	2.1	3,078	749,824

Longitudinal Sample for Children Introduced in Cycle 3			
Province	**Design effect**	**Sample size**	**Population**
Newfoundland and Labrador	8.8	516	10,717
Prince Edward Island	2.5	240	3,155
Nova Scotia	1.9	539	20,053
New Brunswick	2.4	525	15,888
Quebec	1.9	1,209	154,651
Ontario	1.4	1,711	267,171
Manitoba	3.0	556	29,608
Saskatchewan	1.8	536	25,682
Alberta	1.6	687	77,769
British Columbia	1.4	591	85,103
Atlantic Provinces	2.9	1,820	49,813
Prairies	2.1	1,779	133,059
Total[1] (2 to 3 years)	2.0	7,116	690,471

15.2.5 Approximate Sampling Variability Tables

All coefficients of variation in the Approximate Sampling Variability Tables are approximate and, therefore, unofficial. The use of actual variance estimates would likely result in estimates with lower variances; for example, estimates listed as "unacceptable" in the Approximate Sampling Variability Tables could move up to the "marginal" category.

Remember: If the number of observations on which an estimate is based is less than 30, the weighted estimate should be classified as "unacceptable" regardless of the value of the coefficient of variation for this estimate. This is because the formulas used for estimating the variance do not hold true for small sample sizes.

15.3 How to Use the C.V. Tables For Categorical Estimates

The following rules should enable the user to determine the approximate coefficients of variation from the Sampling Variability Tables for estimates of the number, proportion or percentage of the surveyed population possessing a certain characteristic and for ratios and differences between such estimates.

Rule 1: Estimates of Numbers Possessing a Characteristic (Aggregates)

The coefficient of variation depends only on the size of the estimate itself. On the Sampling Variability Table for the appropriate geographic area or age group, locate the estimated number in the left-most column of the table (headed "Numerator of Percentage") and follow the asterisks (if any) across to the first figure encountered. This figure is the approximate coefficient of variation.

Rule 2: Estimates of Proportions or Percentages Possessing a Characteristic

The coefficient of variation of an estimated proportion or percentage depends on both the size of the proportion or percentage and the size of the total upon which the proportion or percentage is based. Estimated proportions or percentages are relatively more reliable than the corresponding estimates of the numerator of the proportion or percentage, when the proportion or percentage is based upon a sub-group of the population. For example, the proportion of female babies who were of low birth weight is more reliable than the estimated number of "female babies who were of low birth weight". Note that in the tables the c.v.'s decline in value reading from left to right.

When the proportion or percentage is based upon the total population of the geographic area or age group covered by the table, the c.v. of the proportion or percentage is the same as the c.v. of the numerator of the proportion or percentage. In this case, Rule 1 can be used.

When the proportion or percentage is based upon a subset of the total population, reference should be made to the proportion or percentage (across the top of the table) and to the numerator of the proportion or percentage (down the left side of the table). The intersection of the appropriate row and column gives the coefficient of variation.

Rule 3: Estimates of Differences Between Aggregates or Percentages

The standard error of a difference between two estimates is approximately equal to the square root of the sum of squares of each standard error considered separately.

That is, the standard error of a difference $\hat{d} = (\hat{X}_1 - \hat{X}_2)$ is :

$$\sigma_{\hat{d}} = \sqrt{(\hat{x}_1 \sigma_1)^2 + (\hat{x}_2 \sigma_2)^2}$$

where $\hat{X}_1$ is estimate 1, $\hat{X}_2$ is estimate 2, σ_1 and σ_2 are the coefficients of variation of $\hat{X}_1$ et $\hat{X}_2$ respectively. The coefficient of variation of $\hat{d}$ is given by $\sigma_{\hat{d}} / \hat{d}$. This formula is accurate for the difference between separate and uncorrelated characteristics, but is only approximate otherwise.

Rule 4: Estimates of Ratios

Where the numerator is not a subset of the denominator (for example, the ratio of the number of low birth-weight female babies to that of low-birth weight male babies), the standard deviation of the ratio of the estimates is approximately equal to the square root of the sum of squares of each coefficient of variation considered separately multiplied by the ratio itself.

The standard error of ratio $\hat{R} = \hat{X}_1 / \hat{X}_2$ is therefore:

$$\sigma_{\hat{R}} = \hat{R} \sqrt{\sigma_1^2 + \sigma_2^2}$$

where σ_1 and σ_2 are coefficients of variation of $\hat{X}_1$ (the number of low-birth weight female babies) and $\hat{X}_2$ (the number of low birth-weight male babies) respectively. The coefficient of variation of $\hat{R}$ is given by $\sigma_{\hat{R}} / \hat{R}$. The formula will tend to overstate the error, if $\hat{X}_1$ and $\hat{X}_2$ are positively correlated and understate the error if $\hat{X}_1$ and $\hat{X}_2$ are negatively correlated.

Rule 5: Estimates of Differences of Ratios

In this case, Rules 3 and 4 are combined. The c.v.'s for the two ratios are first determined using Rule 4, and then the c.v. of their difference is found using Rule 3.

Warning Note on Confidence Intervals

Release guidelines applying to estimates also apply to confidence intervals. For example, if the estimate is "marginal", then the confidence interval is marginal and should be accompanied by a warning note to caution subsequent users about high levels of error.

15.4 Spreadsheet with approximate CVs

A second tool available to users to calculate the variance is the spreadsheet with approximate CVs. Available in Excel format, it has over 6,000 domains. These domains are cross-tabulations of age, age groups, provinces, or regions. The sample sizes for these domains are also available. Predicted sample sizes and CVs are also given for cycles 5 and 6.

15.5 CVs computation

15.5.1 In Cycle 4

Simulations have been run to calculate variances, coefficients of variation and also confidence intervals at the 95% level for different proportions ranging from 1 to 50%. These were based on the cross-sectional population for Cycle 4 using bootstrap weights. Note that the use of bootstrap weights from the longitudinal population can produce slightly different estimates for predicting cycle 5 and 6 sample sizes. In practical terms, we simulated a dichotomous variable in proportions of 1, 5, 10, 15, 20, 30, 40 and 50%. In doing so, we obtained a good approximation for the complete spectrum of proportions since having one proportion, we also know the corresponding 100% minus the calculated proportion. The variance and standard error will remain the same, but not the c.v. An approximate c.v. is obtained by dividing the standard error by the proportion. However the user should note that for disclosure issues, for a dichotomous variable, both variables should be publishable simultaneously. You should always ensure the quality of the smaller proportion. For a given repetition, the observed proportion in the random sample can be different from that of the targeted proportion. We therefore use the mean of 100 repetitions to be able to account for that variability.

15.5.2 In cycles 5 and 6

We projected the cycle 5 and 6 populations by assuming a uniform response rate of 90% for each year and repeated the exercise on those estimated populations. With the mean from a 100 repetitions we can adjust for the randomness in the selection of respondents and for non-response.

15.5.3 Available domains

We've studied numerous domains of estimation, in particular for various geographical levels. We calculated proportions for each province, for regions - Atlantic (Newfoundland and Labrador, Prince Edward Island, Nova Scotia, New Brunswick), Québec, Ontario, Prairies (Manitoba, Saskatchewan, Alberta) and British Columbia; and for Canada as a whole. For demographic characteristics, we used individual ages, and the following age groups at Cycle 4.

Age groups available

0-1	2-7	5-6	9-11	14-15
0-3	2-8	6-13	10-11	16-17
0-5	4-11	6-7	10-15	
0-6	4-5	6-8	10-17	
2-3	4-6	7-8	12-13	
2-5	4-7	7-9	12-17	

15.5.4 Fields in the spreadsheet

The fields used in the spreadsheet are:

Province: The province or ATLANTIC or PRAIRIES for these specific regions or CANADA for the country as a whole.

C4 Age: Age at Cycle 4. Can take values from 0 to 18 and different age groups.

C5 Age: Age at cycle 5. Can take values from 2 to 20 and different age groups.
C6 Age: Age at cycle 6. Can take values from 4 to 22 and different age groups.
Target prop. : The theoretical proportion used to simulate a variable. Can take the values 1%, 5%, 10%, 15%, 20%, 30%, 40% and 50%.
Cycle: C4 (observed), C5 (simulated response rate), C6 (simulated response rate) for every cycle.
Yhat: The mean from100 calculated proportions. Should be close to Target prop.
n: The average sample size of the specified domain from 100 repetitions.
Bs_var: The mean of 100 variances for the specified domain.
Bs_sd: The mean of 100 standard errors for the specified domain.
Bs_cv: The mean of 100 coefficients of variation for the specified domain.
Cil95: The mean of 100 95% confidence interval lower boundaries.
Ciu95: The mean of 100 95% confidence interval upper boundaries.

Note that according to the sampling design for cycle 5, there shouldn't be any 6-7 year-old kids selected (4-5 year-olds at Cycle 4). The results are only projections.

For example, to estimate the proportion of 3 year-old boys or girls in Newfoundland and Labrador, we select province « NEWFOUNDLAND & LABRADOR», C4 Age « 4 » and Target prop. « 50% » (since a reasonable estimate of that proportion should be close to 1 out of 2). For every cycle, the sample size drops from 472 to 425 and then to 382 kids. The coefficients of variation are stable ranging from 5,64% to 6,26% in cycles 4 through 6. The confidence intervals reiterate the stability of the variance estimate of that variable even with the smaller sample size.

Another way to use the table is to select only one province and one age group and to look at the variability of the coefficient of variation to determine the proportion and the sample size to get a reliable estimate. If we fix a coefficient of variation threshold of 16,5%, we can see, for example, in selecting province of Quebec for the age group 0-5 year-olds, we get estimates for proportions above 5% in every cycle.

As a reference, many surveys in Statistics Canada use the following quality standards:

1) An estimate is said **acceptable** if the sample size is at least 30 and the coefficient of variation is lower than 16,5%
2) An estimate is said **marginal** if the sample size is at least 30 and the coefficient of variation is between 16,5% and 33,33%. This estimate should be accompanied by a warning to emphasise the high level of error.
3) An estimate is said **unacceptable** if the sample size is lower than 30 or if the coefficient of variation is greater than 33,33%. This estimate should not be released. Reminder: Statistics Canada Quality Level Guidelines

15.6 Examples of using C.V. Tables for Categorical Estimates

The following are examples using actual NLSCY data to illustrate how to apply the foregoing rules.

Example 1: Estimates of Numbers Possessing a Characteristic (Aggregates)

Using NLSCY data, 31,038 babies were estimated to be of low birth weight (i.e., less than 2,500 grams). How does the user determine the coefficient of variation of this estimate?

1. Refer to the c.v. table for children in 0 to 23 months group. It should be noted that, because the question on birth weight applied only to children in this age group, this table should be used to determine the c.v. for this estimate.

2. The estimated aggregate (31,038) does not appear in the left-hand column (the "Numerator of Percentage" column), so it is necessary to use the figure closest to it, namely 30,000.

3. The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 10,5 %.

4. The approximate coefficient of variation of the number of low birth-weight babies is estimated to be 10,5 %. The finding that there were 31,038 babies that were of low birth weight is "acceptable" and no warning message is required to produce this estimate since the c.v. for the estimate is in the 0.0% to 16.5% range.

These 31,038 children represent around 5% of the 0 to 23 months population. With the spreadsheet, we select the province "CANADA", age in Cycle 4 of "0-1" and a target proportion of 5%. For Cycle 4, a c.v. of 9.09% is obtained. The spreadsheet is usually less conservative than c.v. look-up tables. The conclusions remain unchanged.

The following examples are solely about the c.v. look-up tables. However, it is quite easy to adapt the examples to the spreadsheet.

Example 2: Estimates of Proportions or Percentages Possessing a Characteristic

Using NLSCY data, it is estimated that 68.2% (21,155 / 31,038) of low birth-weight babies were born prematurely (gestational age 258 days or less). How does the user determine the coefficient of variation of this estimate?

1. Refer to the c.v. table for children in 0 to 23 months age group. It should be noted that, because the questions on birth weight and delivery time applied only to children in this age group, this table should be used to determine the c.v. for this estimate.

2. Because the estimate is a percentage which is based on a subset of the total population (i.e., low birth-weight babies who were born prematurely), it is necessary to use both the percentage (68.2%) and the numerator portion of the percentage (21,155) in determining the coefficient of variation.

3. The numerator, 21,155, does not appear in the left-hand column (the "Numerator of Percentage" column) so it is necessary to use the figure closest to it, namely 21,000. Similarly, the percentage estimate does not appear as any of the column headings, so it is necessary to use the figure closest to it, 70.0%.

4. The figure at the intersection of the row and column used, namely 7.0% is the coefficient of variation to be used.

5. The approximate coefficient of variation of the percentage of low birth-weight babies who were premature is estimated to be 7.0%. Since the c.v. for the estimate falls in the 0.0% to 16.5% range, this estimate is "acceptable", and the finding that 68.2% of low birth-weight babies were born prematurely requires no warning note.

Example 3: Estimates of Differences Between Aggregates or Percentages

Using NLSCY data, it is estimated that 7.0% (23,901 / 339,782) of female babies were born prematurely, while 9.2 % (33,210 / 359,048) of male babies were born prematurely. How does the user determine the coefficient of variation of the difference between these two estimates?

1. Using the c.v. table for the 0 to 23 months age group in the same manner as described in example 2 gives the c.v. of the estimate for female babies as 11.7%, and the c.v. of the estimate for male babies as 9.4%.

2. The standard deviation of the difference estimate is therefore given by Rule 3, which is $?_{\hat{d}} ? \sqrt{?\hat{x}_1?_1?^2 ? ?\hat{x}_2?_2?^2}$, where $?_1$ and $?_2$ are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively.

That is:

$?_{\hat{d}} ? \sqrt{?0.07*.117?^2 ? ?0.092*.094?^2}$

$?_{\hat{d}} ? 0.012$

Therefore, the approximate c.v. is 0.012 / 0.022 = 54.1%.

Example 4: Estimates of Ratios

Suppose now a user wants to compare the number of low birth-weight female babies to the number of low birth-weight male babies. The user is interested in comparing these estimates in the form of a ratio. How does the user determine the coefficient of variation of this estimate?

1. First of all, this estimate is a ratio estimate, where the numerator of the estimate = ($\hat{X}_1$) is the number of low birth-weight female babies and denominator = ($\hat{X}_2$) of the estimate is the number of low birth-weight male babies.

2. Refer to the table for the 0 to 23 months age group. The questions on birth weight were applicable only to children in the 0 to 23 months age group.

3. The numerator of this ratio estimate is 23,901. The figure closest to it is 24,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 11.7%.

4. The denominator of this ratio estimate is 33,210. The figure closest to it is 35,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 9.4%.

5. The standard deviation of the ratio estimate is therefore given by Rule 4, which is $?_{\hat{R}} ? \hat{R}\sqrt{?_1^2 ? ?_2^2}$, where $?_1$ and $?_2$ are the coefficients of variation of $\hat{X}_1$ and $\hat{X}_2$ respectively.

That is:

$?_{\hat{R}} ? 0.72*\sqrt{(0{,}117)^2 ? (0{,}094)^2}$

$?_{\hat{R}} ? 0.108$

The ratio of low birth-weight female babies versus low birth-weight male babies is 23,901 / 33,210, or 0.72 : 1. Since the c.v. for the estimate falls in the 0.0% to 16.5% range (0.108 / 0.72 = 15.0%), this estimate is "acceptable" and the requires no warning note.

15.7 How to Use the C.V. Tables to Obtain Confidence Limits

Although coefficients of variation are widely used, a more intuitively meaningful measure of sampling error is the confidence interval of an estimate. A confidence interval constitutes a statement on the level of confidence that the true value for the population lies within a specified range of values. For example a 95% confidence interval can be described as follows:

If sampling of the population is repeated indefinitely, each sample leading to a new confidence interval for an estimate, then in 95% of the samples the interval will cover the true population value.

Using the standard error of an estimate, confidence intervals for estimates may be obtained under the assumption that under repeated sampling of the population, the various estimates obtained for a population characteristic are normally distributed about the true population value. Under this assumption, the chances are about 68 out of 100 that the difference between a sample estimate and the true population value would be less than one standard error, about 95 out of 100 that the difference would be less than two standard errors, and about 99 out 100 that the differences would be less than three standard errors. These different degrees of confidence are referred to as the confidence levels.

Confidence intervals for an estimate are generally expressed as two numbers, one below the estimate and one above the estimate, as where k is determined depending upon the level of confidence desired and the sampling error of the estimate.

Confidence intervals for an estimate can be calculated directly from the Approximate Sampling Variability Tables by first determining from the appropriate table the coefficient of variation of the estimate and then using the following formula to convert to a confidence interval CI:

$$IC_X = [\hat{X} - t\hat{X}\sigma_{\hat{X}}, \hat{X} + t\hat{X}\sigma_{\hat{X}}]$$

where $\sigma_{\hat{X}}$ is the determined coefficient of variation for $\hat{X}$ and
t = 1 if a 68% confidence interval is desired
t = 1.6 if a 90% confidence interval is desired
t = 2 if a 95% confidence interval is desired
t = 3 if a 99% confidence interval is desired.

Note Regarding Release Guidelines

Release guidelines applying to estimates also apply to confidence intervals. For example, if the estimate is "marginal", then the confidence interval is marginal and should be accompanied by a warning note to caution subsequent users about high levels of error.

Example of Using the C.V. Tables to Obtain Confidence Limits

A 95% confidence interval for the estimated proportion of babies who were of low birth weight would be calculated as follows.

Estimate of X = 4.4%, t = 2

The alpha estimate of X = 10.5% (.105 expressed as a proportion)
is the coefficient of variation of this estimate as determined by the tables

CIx = {0,044 - (2)(0,044)(0,105), 0,044 + (2)(0,044)(0,105)}
CIx = {0,044 - 0,009, 0,044 + 0,009}
CIx = {0,035, 0,053}

With 95% confidence it can be said that between 3.5% and 5.3% of babies who were 0 to 23 months old at the time of the survey were of low birth weight.

15.8 How to Use the C.V. Tables to Do a T-test

Hypothesis Testing

Standard errors may also be used to perform hypothesis testing, a procedure for distinguishing between population parameters using sample estimates. The sample estimates can be numbers, averages, percentages, ratios, etc. Tests may be performed at various levels of significance, where a level of significance is the probability of concluding that the characteristics are different when, in fact, they are identical.

Let $\hat{X}_1$ and $\hat{X}_2$ be sample estimates for two characteristics of interest. Let the standard error on the difference $\hat{X}_1 ? \hat{X}_2$ be $? _{\hat{d}}$.

If $t ? \frac{\hat{X}_1 ? \hat{X}_2}{? _{\hat{d}}}$ is between -2 and 2, then no conclusion about the difference between the characteristics is justified at the 5% level of significance. If however, this ratio is smaller than -2 or larger than +2, the observed difference is significant at the 0.05 level. That is to say that the characteristics are significantly different.

Example of Using C.V. Tables to do a T-Test
Let us suppose we wish to test, at 5% level of significance, the hypothesis that there is no difference between the proportion of low birth-weight female babies and that of low birth-weight male babies. From example 3, the standard error of the difference between these two estimates was found to be = .012.

Hence,

$$t ? \frac{\hat{X}_1 ? \hat{X}_2}{? _{\hat{d}}} ? \frac{0{,}070 ? 0{,}092}{0{,}012} ? \frac{? 0{,}022}{0{,}012} ? ? 1.8$$

Since t = -1.8 is between -2 and 2, no conclusion at the 0.05 level of significance can be made regarding the difference in proportions of low birth-weight male or female babies.

15.9 Coefficients of Variation for Quantitative Estimates

Quantitative Estimates

For quantitative estimates, special tables would have to be produced to determine their sampling error. Since most of the variables for the NLSCY are categorical in nature, this has not been done.

As a general rule, however, the coefficient of variation of a quantitative total will be larger than the coefficient of variation of the corresponding category estimate. If the corresponding category estimate is not releasable, the quantitative estimate will not be either. For example, the coefficient of variation of the total number of days of special medical care received for low birth-weight babies would be greater than the coefficient of variation of the corresponding proportion of babies who were of low birth weight. Hence if the coefficient of variation of the proportion is not releasable, then the coefficient of variation of the corresponding quantitative estimate will also not be releasable.

Pseudo Replication

Coefficients of variation of such estimates can be derived as required for a specific estimate using a technique known as pseudo replication. This involves dividing the records on the microdata files into subgroups (or replicates) and determining the variation in the estimate from replicate to replicate. Users wishing to derive coefficients of variation for quantitative estimates may contact Statistics Canada for advice on the allocation of records to appropriate replicates and the formulae to be used in these calculations.

15.10Release Cut-offs for the NLSCY

In the tables that follow, cut-off numbers are given for NLSCY estimates in order for them to be of "acceptable", "marginal" or "unacceptable" quality. Users are encouraged to use these cut-offs when publishing data from the NLSCY. First a table is given to show the cut-offs at the provincial, regional and Canada level. Then a table is given to show the cut-offs for the various age cohorts. An interpretation of what is meant by the various cut-off levels can be found in Section 11.4.

For example, an estimate for Nova Scotia of 5,000 would fall into the "marginal" range. This would mean that the estimate should be flagged and a note of caution would be attached for subsequent users about the high level of error associated with the estimate.

Geographical Release Cut-Offs
Cross-Sectional Sample

Province	Acceptable – Estimates at or above:	Marginal - Estimates between:	Unacceptable – Estimates at or below :
Newfoundland and Labrador	7500	2000 - 7500	2000
Prince Edward Island	3500	1000 - 3500	1000
Nova Scotia	12500	3500 - 12500	3500
New Brunswick	9500	2500 - 9500	2500
Quebec	42000	10500 - 42000	10500
Ontario	54000	13500 - 54000	13500
Manitoba	17000	4500 - 17000	4500
Saskatchewan	13500	3500 - 13500	3500
Alberta	30000	7500 - 30000	7500
British Columbia	48000	12500 - 48000	12500
Atlantic Provinces	10500	2500 - 10500	2500
Prairies	23500	6000 - 23500	6000
Total Canada	42000	10500 - 42000	10500

Release Cut-Offs by Age Group
Cross-Sectional Sample

Age Group	Acceptable – Estimates at or above:	Marginal - Estimates between:	Unacceptable – Estimates at or below :
0 to 23 months	12500	3500 - 12500	3500
2 to 3 years	7500	2000 - 7500	2000
4 to 5 years	16000	4000 - 16000	4000
6 to 7 years	21000	5500 - 21000	5500
8 to 9 years	36500	9500 - 36500	9500
10 to 11 years	40000	10500 - 40000	10500
12 to 13 years	41500	11000 - 41500	11000
14 to 15 years	41000	10500 - 41000	10500
16 to 17 years	40500	10500 - 40500	10500
Total Canada	42000	10500 - 42000	10500

Geographical Release Cut-Offs
Longitudinal Sample Introduced in Cycle 1

Province	Acceptable – Estimates at or above:	Marginal - Estimates between:	Unacceptable – Estimates at or below :
Newfoundland and Labrador	10500	3000 - 10500	3000
Prince Edward Island	4500	1500 - 4500	1500
Nova Scotia	17500	5000 - 17500	5000
New Brunswick	14000	4000 - 14000	4000
Quebec	50000	13000 - 50000	13000
Ontario	71000	18000 - 71000	18000
Manitoba	23000	6500 - 23000	6500
Saskatchewan	18000	5000 - 18000	5000
Alberta	40500	10500 - 40500	10500
British Columbia	57500	15500 - 57500	15500
Atlantic Provinces	14000	4000 - 14000	4000
Prairies	32500	8500 - 32500	8500
Total Canada	54500	13500 - 54500	13500

Release Cut-Offs by Age Group
Longitudinal Sample Introduced in Cycle 1

Age Group	Acceptable – Estimates at or above:	Marginal - Estimates between:	Unacceptable – Estimates at or below :
6 to 7 years	20500	5500 - 20500	5500
8 to 9 years	35000	9000 - 35000	9000
10 to 11 years	38500	10000 - 38500	10000
12 to 13 years	40500	10500 - 40500	10500
14 to 15 years	40500	10500 - 40500	10500
16 to 17 years	38000	10000 - 38000	10000
TOTAL CANADA	54500	13500 - 54500	13500

Geographical Release Cut-Offs
Longitudinal Sample Introduced in Cycle 2

Province	Acceptable – Estimates at or above:	Marginal - Estimates between:	Unacceptable – Estimates at or below :
Newfoundland and Labrador	4500	2000 - 4500	2000
Prince Edward Island	2500	1500 - 2500	1500
Nova Scotia	6000	2000 - 6000	2000
New Brunswick	5500	2000 - 5500	2000
Quebec	19000	5500 - 19000	5500
Ontario	22000	6000 - 22000	6000
Manitoba	7500	2500 - 7500	2500
Saskatchewan	17000	7500 - 17000	7500
Alberta	18000	5500 - 18000	5500
British Columbia	18500	5500 - 18500	5500
Atlantic Provinces	5500	1500 - 5500	1500
Prairies	17000	5000 - 17000	5000
Total Canada	19000	5000 - 19000	5000

Geographical Release Cut-Offs
Longitudinal Sample Introduced in Cycle 3

Province	Acceptable – Estimates at or above:	Marginal - Estimates between:	Unacceptable – Estimates at or below :
Newfoundland and Labrador	4000	1500 - 4000	1500
Prince Edward Island	1000	500 - 1000	500
Nova Scotia	2500	1000 - 2500	1000
New Brunswick	2500	1000 - 2500	1000
Quebec	8500	2500 - 8500	2500
Ontario	8000	2000 - 8000	2000
Manitoba	5000	1500 - 5000	1500
Saskatchewan	3000	1000 - 3000	1000
Alberta	6500	2000 - 6500	2000
British Columbia	7000	2000 - 7000	2000
Atlantic Provinces	3000	1000 - 3000	1000
Prairies	5500	1500 - 5500	1500
Total Canada	7500	2000 - 7500	2000

15.11SAS and SPSS macros to calculate the variance, using the Bootstrap weights.

SAS and SPSS macros have been developed to calculate the variance using the Bootstrap weights. From the methods available, this method is the best approach to obtain a very good approximation of the true variance. With this method, it is possible to calculate the variance of any estimates, for any domains. The variance calculated using this method takes into account the sample design and also, the specificities of the variable of interest. Finally, as opposed to the other methods, the user is not restricted to pre-defined domains.

This method has many advantages but requires more work from the researcher. Variance calculation using these macros is more time consuming than the other two methods presented earlier. The user must first get familiar with the macros before using them. However, these macros have been developed in such a way that they are easy to use.

The researcher must have access to the macros, to the data files and to the Bootstrap weight files. Access to these tools is possible in the Statistics Canada Data Research Centers (RDC). Also, detailed documentation on how to use these SAS or SPSS macros is available in the RDCs.

Despite the time required to run these macros, it is strongly recommended to use this method to calculate the variance of any estimates which must be published. This method provides a more precise and accurate measure of the true variance.

Naming convention for the Bootstrap weight files :

For the cross-sectional Bootstrap weight files:

BVC_xT, where x is the cycle. For example, for cycle 4: BVC_4T

For the Bootstrap weight files, "regular" longitudinal weights:

BVCx_yLa, where x = the initial cycle and y = the current cycle. For example, for longitudinal children introduced in cycle 1, in cycle 4 the filename would be: BVC1_4La.

The "a" at the end of the file name represents the regular longitudinal weight.

For the Bootstrap weight file, funnel longitudinal weights (children responding to all cycles):

BVCx_yLb, where x = the initial cycle and y = the current cycle.
For cycle 4, the only file is: BVC1_4Lb

16.0 Direct Assessment

Background

Research on early childhood development plays a significant role in the formulation of policy for young children. Adding early developmental assessment tools to the NLSCY will help to enhance the knowledge about developmental processes in early childhood and provide relevant data on which to base policy directions for this stage.

Choices about the assessment tools to be included in the NLSCY for Cycle 4 (collection in 2000-2001) were made on the basis of an extended literature review, development of a research framework on child development and learning, consultations with many experts in Canada and internationally, review of material on many different possible instruments and field testing of the most likely possibilities. The instruments selected for consideration were also reviewed using a number of criteria. The criteria included reliability and validity of the instrument, coverage of domains in the research framework, ability of the instrument to indicate normal development and developmental delays, the ease of administration by lay interviewers and the availability of the instrument in English or French (or ease of translation to French or English). The final decision was strongly influenced by key experts who had had a history of providing advice to the NLSCY Team.

Two early childhood assessments were selected to be administered to children four and five years of age in addition to the PPVT-R; the *Who Am I?* and the Number Knowledge Test. The Ages and Stages Questionnaires were also added in Cycle 4 to help assess the development level of children from four to 71 months old. These questionnaires are described in Chapter 8, Content of the Survey.

The NLSCY conducts direct assessments of older children with math tests. These tests are also described in this chapter.

16.1 The PPVT-R

The PPVT-R was designed to measure receptive or hearing vocabulary and in fact can be used for any age group, up to adult. The test was developed by Lloyd and Leota Dunn, at the University of Hawaii, and has been widely used in large-scale data collections as well as assessments. A French adaptation of the PPVT-R was developed by the test's authors and Claudia M. Thériault at St. Thomas University in Fredericton, N. B. The French test is called the Échelle de vocabulaire en images Peabody (EVIP).

For the NLSCY, the PPVT-R was used to measure school readiness for children in the four to five age group. Verbal parental consent was required before the test was administered. If permission was granted, the interviewer then administered the test to the child in the home. The child looked at pictures on an easel and identified the picture that matched the word the interviewer read out.

A total raw score was calculated for each child who completed the PPVT-R by computing correct responses. A standardized score was also assigned to each child. Standard scores allow for comparisons of scores to be made across age groups. Obviously a five year-old would be expected to perform better on the PPVT-R than a four year-old and thus have a higher score. The standard score takes into account the child's age.

Standard scores for a test are usually developed based on the distribution of scores obtained by some defined sample of individuals. This is called the norm sample. For the PPVT-R individuals in the norm sample were assigned standard scores so the mean of the standard scores was 100 and the standard deviation was 15 for all age groupings. This standardization was done by 2 month age groups.

The PPVT-R norm sample was based on a sample selected in the United States. It was decided that it would be appropriate to develop standardized scores for the Canadian context. Therefore, in collaboration with the developers of the test, Canadian norms have been developed for children in the four to five age group. It should be noted that the standardization was done separately for the PPVT-R and EVIP. Therefore, when global comparisons are made between children who completed the test in English vs. French, by definition, performance should be equivalent.

Reliability measures for the PPVT have been calculated based on the American norm sample (Dunn and Dunn, 1981).

16.1.1 Psychometric Properties of Scores

This section addresses the quality of the test itself as it applies to the survey population, as compared to the original population for which the test was developed. We find that the test is by and large still provides a reasonable assessment of the child's ability, and we outline below the reasons.

a) The Raw Score

One of the main advantages of the test in a survey context is that the test is tailored to the child's age and performance so that not all of the questions need to be asked to determine the ability level. Based on the age of the respondent a starting question is selected, and the test proceeds with increasingly difficult questions. When the subject appears to be answering at random – at least six out of the last eight questions are missed - the test stops and a score is derived based on the rank of the last question and the number of incorrect answers.

Questions are ranked in increasing order of difficult and are designed to be equally spaced on the 'difficulty scale'. Originally the PPVT-R test was calibrated by using a representative sample of about 5,000 English speaking children, similar efforts were undertaken to calibrate the French version. The item difficulties were calculated using the Rasch model. In the language of Item Response Theory, this is known as the one parameter logistic model.

Since the test calibration was done some years ago, it is natural to expect some drift of the item difficulties, as the language itself evolves and some words become more or less common. To check whether this is the case an IRT analysis of the items was done, and derived scores based on the new difficulties of the items were created. For some of the items we did find some deviation from what was to be expected in the original test.

However, no systematic - consistently up or consistently down in any portion of the test – deviations were found. Furthermore, the scores derived by using IRT were consistent with the raw scores. The correlation coefficient between the tests was 94% for the English version, and 96% for the French version, which is high by any standards. Therefore we are confident that the raw scores can be used as they are.

For a number of children (64) the test was not finished in the field as per the rules of the instrument and no score could be assigned by the application. These cases resulted in a score of 0 despite a significant number of test questions being

answered. We were able to derive proxy raw scores for these children as a by-product of the IRT analysis. For the rest of the respondents – that is, 99% of them – we used the score that was produced by the application using the original rules of the PPVT test.

b) The standardized score.

As described earlier, the standardized score is determined by using population distributions for each age. Strictly speaking we can never know the population distribution, since applying the test to the whole population is not feasible. One way to deal with this is to use the sample that we have within an age group as representative of the population in that age group, and derive the necessary percentiles.

However, certain sample limitations exist that need to be addressed before the score can be standardized. By inspecting the percentiles for different ages, we would expect an increasing trend in the ability measurement with age. While there is an overall increasing trend, for a lot of ages the trend is reversed. This is due to the fact that within each age group the sample is not large enough, and a lot of noise is introduced due to poor representation.

The better approach is to use the percentiles from the sample as a starting point, and smooth the progression with age until we are satisfied that we have a 'natural' progression. We used the progression of the original percentiles – from the PPVT-R handbook - as an example of what degree of smoothing should be expected. Then we used the resulting points as the percentiles for standardization. We should note here that even though features of the norms were similar, the percentiles drifted upwards over the years, which, according to the experts, can be expected.

The test is applied to children of ages 4 to 6, but also to some older children[12] of higher age that are not in school. Beyond the age of 8½ years the sample count was so low that we could not derive norms directly – even if smoothing was used. We were able to extrapolate the norms up to the age of 9 years and 4 months. Beyond that, a comparison with the original norms showed that simple linear extrapolation is not a good way to derive norms. For children of ages higher than 9 years and 4 months that took the test – 39 of them – we did not assign a standardized score. They do have a raw score however, and if the user finds some other external benchmark, a standardized score could be derived.

c) Final Note

The PPVT scores used in NLSCY are a valid measurement of ability. There is however non-response, which should be handled on a case by case basis when doing analysis to minimize the potential bias. For more information about non-response, please see Chapters 10 and 13.

16.1.2 PPVT Assessment

Once the entire NLSCY interview had been completed and the Interviewer had left the household she completed an assessment questionnaire to assess the conditions under which the PPVT-R or EVIP was administered to indicate factors that may have influenced the child's responses and his/her overall reaction to the test.

12 The PPVT is generally administered to children aged 4 to 6 years old. Children in Grade 2 and above are given a math test. In cases where the child is older than 6 but has not yet entered Grade 2, the PPVT is administered.

Responses to these assessment questions are included on the micro data file and can be grouped into four factors. A score was calculated for each of these factors.

The child's attitude to the test

The score for this factor is labeled as DPACS01 on the micro data file. It was derived using the following items; DPACQ1 (attitude) DPACQ2 (rapport), DPACQ3 (perseverance/persistence), DPACQ4 (cooperation), DPACQ5 (motivation/interest), and DPACQ9 (shy/anxious). These items were all ranked by the interviewer on a scale of 1 to 5. Before calculating the score, the order of values was reversed for items DPACQ1, DPACQ2, DPACQ3, DPACQ4 and DPACQ5 (i.e., 5 was recoded to 1, 4 was recoded to 2, etc.) so that the higher the value the more severe the problem. Then all values were rescaled to 0 to 4 by subtracting one from the value for all six items. Finally the sum of these transformed values was taken across the six items. The final score ranges from 0 to 24, where 0 means the absence of a problem and 24 the highest possible score with respect to problems.

Physical and health problems

The score for this factor is labeled as DPACS02 on the micro data file. It was derived using the following items; DPACQ6 (problems with visual sharpness), DPACQ7 (problems with hearing) and DPACQ8 (health problems). The score ranges from 0 to 3, where 0 means the absence of a problem and 3 means the highest score with respect to problems. In order to calculate the score, the number of 'yes' answers was summed.

The room environment

The score for this factor is labelled as DPACS03 on the micro data file. It was derived from DPACQ13 (light problems) and DPACQ14 (temperature problems). Before calculating the score the order of values was reversed and then the values were rescaled to 0 to 4. Finally the sum of these transformed values was taken across the two items. The final score ranges from 0 to 8, where 0 means the absence of a problem and 8 the highest possible score with respect to problems.

The level of distraction during the test

The score for this factor is labelled as DPACS04 on the micro data file. It was derived from DPACQ10 (noise interference), DPACQ11 (interruptions), DPACQ12 (distractions) and DPACQ15 (presence of others). Before calculating the score the order of values was reversed and then the values were rescaled to 0 to 4. Finally the sum of these transformed values was taken across the four items. The final score ranges from 0 to 16 where 0 means the absence of a problem and 16 the highest possible score with respect to problems.

To assess whether or not each of the above factors had an impact on the test score a threshold value was established for each score. A child with a score above this value was said to have a problem.

For the child's attitude to the test the threshold value was set at 13. Any child with an attitude score greater than or equal to 13 was said to have a problem. This corresponds to a "below average" value to at least one of the items that makes up the score for the factor. For physical and health problems the threshold was set at 1. Any child with at least one physical or health problem was said to have a problem. For the room environment score the threshold value was set at 5. For the distraction score the threshold was set at 9.

The following table shows the percentage of children who took the test that had a problem for any of the assessment factors (i.e., a factor score at or above the threshold). As well the average standard score of children with the problem are compared to those without the problem.

PPVT Assessment (Based on Cycle 1 Data)

PPVT Assessment factor	% of children with problem	Average standard score for children with problem[13]	Average standard score for children without problem
Attitude	6.4%	93.4	100.5
Physical and health problems	3.7%	96.2	100.1
Room environment problems [14]	-		
Distraction problems	6.4%	94.3	100.4

As can be seen in the table, the PPVT Assessment revealed that significantly lower scores were obtained for children who had a less positive attitude to the test, who had physical or health problems, or who were distracted by their surroundings during the test. At the same time only small proportions of children were affected by these problems.

16.2 Number Knowledge Assessment

The purpose of the Number Knowledge Assessment is to assess the development of children's understanding of numbers by examining their comprehension of the system of whole numbers. For the NLSCY, the assessment is administered to four and five year old children.

The assessment was developed by Dr. Robbie Case from the University of Toronto, with colleagues, including Yukari Okamoto at the University of Southern California, Santa Barbara. The assessment is constructed based on Dr. Case's theory of central conceptual structures for explaining the development of children's thought. Before his death in May 2002, Dr. Case was adapting the test for the NLSCY. Following Dr. Case's death, Yukari Okamoto assisted the NLSCY team in completing the adaptations of the assessment for the survey.

Theoretical Background

According to Dr. Case's theory, four developmental levels can be distinguished in children's understanding of numbers: pre-dimensional, uni-dimensional, bi-dimensional, and integrated bi-dimensional. Some degree of mastery of each level is required prior to continuing on to the next. Typically the four levels are attained at the ages of 4, 6, 8, and 10. The pre-dimensional level assesses the ability to count by rote and to quantify small sets, using concrete objects. This knowledge is important for the next level where children

13*The differences in scores for children with attitude problems, physical problems and distraction problems are all significant at the 95% confidence level.*

14*There were only 12 children for which there was a room environment problem. Therefore the numbers were too small to draw any conclusions about this factor.*

deal with changes in quantity without objects than can be touched or seen. The uni-dimensional level assesses children's knowledge of the number sequence and ability to handle simple arithmetic problems. To solve the items, children must rely on a "mental counting line" in their heads. This "line" integrates their understanding of numbers and quantities. This assessment measures the essential prerequisites for successful school learning.

For the purpose of the NLSCY, we are aiming to assess children's understanding at the first three levels – also referred to as levels 0, 1, and 2. Dr. Case felt that a child between the ages of four and five was unlikely to complete questions higher then level 2 therefore the top level of difficulty was omitted from the NLSCY assessment.

Assessment Description

In consultation with Dr. Case and Dr. Okamoto the test was revised for the NLSCY. The assessment has been made continuous with three levels; some items were revised or dropped to make the test somewhat shorter. The original version of the test was discontinuous (i.e. the child had to pass sufficient items at any one level to go on the next). Since it is accepted that we cannot expect a child to do well at a level without also doing well at the preceding level, it is sensible to stop administering the assessment after a certain number of missed items. The test was also programmed into the computer application, so that the stopping rule was automatically applied. The interviewer asked the child the question and then entered in the application whether or not the child answered correctly.

The test is composed of 22 questions. Some of them have two parts - a) and b). Children must pass both part a) and b) to earn a pass for these items. This convention was adopted because each two-part item gives children a choice between two alternatives and a child has a 50% chance of getting the right answer by guessing alone[15]. Requiring children to pass two such items before they get a point increases confidence that children have the knowledge tapped by the item.

Children are not permitted to use a pencil and paper to answer the questions, which are given orally. Instead, the children must rely on a "mental counting line", which integrates the child's understanding of numbers and quantities. Children do have access to the various manipulative aids such as chips and a number card to help solve the problems.

The administration of the Number Knowledge assessment should take approximately 10-15 minutes.

Scoring

An "Age Equivalent Score" (DKNCdS01) is derived based on the child's responses. The "Age Equivalent Score" assigns a point for each of the three levels passed and then the points are totalled. Passing a level means passing a certain number of items from that level – for instance, for the pre-dimensional level, 3 out of 5 items must be correct. The minimum is zero, and the maximum is three.

Level 1 represents the proportion of correct responses for the pre-dimensional level. There are 5 items in this level. To reach the age equivalent of this level the child must achieve a proportion of at least 0.6 (i.e. get 3 out of 5 correct responses). Level 2 represents the proportion of correct responses for the uni-dimensional level. There are 7 items in this level. To reach the age equivalent of this level the child must achieve a proportion of at least 0.6 (i.e. get 4 out of 7 correct responses). Level 3 represents the proportion of correct responses for the bi-dimensional level. There are 8 items in this level. To reach the age

15 *For example, part a) may ask which of two piles of counting chips is bigger and part b) asks which pile is smaller.*

equivalent of this level the child must achieve a proportion of at least 0.6 (i.e. get 5 out of 8 correct responses).

Evaluation of the assessment

Analysis was conducted on the Number Knowledge data to determine whether this assessment has good validity. The analyses included: comparing age equivalent scores to the child's age, comparison with the *Who Am I?,* an analysis of the items and of non-response. As the scoring procedures were being developed the NLSCY team consulted with Dr. Okamoto to ensure that the procedures were consistent with Dr. Case's theories.

All the evidence indicates that the test has good validity and should provide data users with information about the child's acquisition of the necessary skills to succeed at math in school. This assessment is not free of non-response bias. Please see Chapters 10 and 13, for more information on non-response.

16.3 WHO AM I?[16]

The purpose of the *Who Am I?* assessment is to evaluate the developmental level of young children from 3 to 7 years of age. For the NLSCY, the assessment is administered to four and five year old children.

The assessment was developed by Dr. Molly de Lemos and her colleagues at the Australian Council for Educational Research (ACER). The NLSCY team worked closely with Dr. de Lemos to make some modifications to the assessment for the NLSCY (mainly dropping the drawing task) and to enhance the administration and scoring procedures for the NLSCY context.

Theoretical Background

The *Who Am I?* instrument assesses the developmental level of young children from the ages of 3 years to 7 years. The *Who Am I?* involves copying and writing tasks. The copying tasks in the assessment are designed to assess the child's ability to conceptualize and reconstruct a geometrical shape. The writing tasks assess the ability of the child to understand and use symbolic representations such as numbers, letters and words. The child's ability to complete the tasks depends on many factors including maturation, culture, experiences, and language skills.

The use of the ability to copy geometrical figures to assess level of development in children has been long established. This type of assessment is included in measures of intelligence and development over a long period time. Piaget's research on the development of spatial concepts in young children also provides evidence of the validity of copying tasks as a measure of developmental level.

Because the *Who Am I?* assesses nonverbal language, it can be used to assess children whose knowledge of English or French is limited. These children could be allowed to complete tasks in their mother tongue as well as English and French. Their scores in their mother tongue would provide information on their developmental stage; the score in English or French would give some idea of their development in that language. The NLSCY chose to only assess children in English or French for two reasons. First, it was felt that an assessment of the child's development in one of the official languages was an important indicator of the child's ability to function in the Canadian school system. Secondly, it would

16 *For more information about the Who Am I?, please see. "Patterns of Young Children's Development: An International Comparison of Development as Assessed by Who Am I? By Molly de Lemos. Research paper published by the Applied Research Branch, Human Resources Development Canada. R-02-5E*

be operationally difficult to score questionnaires in the variety of languages spoken in Canada.

The tasks were developed based on research that indicates that copying skills are strongly associated with subsequent school achievement, are valid across different cultural groups and provide a reliable measure of development at the time of assessment. Also, children's attempts at early writing are linked to their growing understanding of the way spoken sounds are represented by print.

Assessment Description

The *Who Am I?* assessment is composed of three scales: a copying scale, a symbols scale and a drawing scale. The copying scale (DWIcdS02) is composed of a circle, cross, square, triangle and diamond which the child attempts to copy. The symbols scale (DWIcdS03) is composed of printing their name, printing some letters, numbers, words and a sentence. Children are only required to complete as much as they feel they can, but they are encouraged to at least attempt each task. For the drawing task, the child is asked to draw a picture of themselves. The drawing scale is not used in the NLSCY due to time constraints. Dr. Molly de Lemos was consulted before the drawing scale was dropped for the NLSCY.

The assessment consists of an appealing booklet in which the child completes the tasks as the assessor turns the pages and gives instructions. The booklet takes about 10 minutes to complete and is scored in head office. The child completes as much as he/she is able but is encouraged to produce at least a scribble for each task.

Scoring

In addition to the three scales, there is a combined total score (DWIcdS01). As mentioned, in the NLSCY, the drawing scale is not included and will not be discussed here. Each sub-scale (copying and symbols) is composed of four levels. The scorer uses detailed scoring instructions to determine the child's level for each task. Finally, the total of the copying and symbols score gives a general overview of the child's developmental level.

For the NLSCY, the *Who Am I?* assessment is hand-scored by trained individuals at Statistics Canada. These individuals have been trained to recognize signs of each level in a child's responses. Scorers who cannot make a decision on a child's level because the work does not fit clearly into one level are asked to make an judgment about the child's level based on the score on other items. Scoring was done by a small number of people and was subject to quality control procedures. The head of the scoring team also met with a staff member from ACER to consult on scoring procedures.

Imputation

In summing scores on the *Who Am I?* tasks to obtain a total score for the Copying and Symbols scales, as well as a total score, it is necessary according Dr. de Lemos to allocate a score in cases in which responses have been recorded as 0 (no attempt).

In most cases, it is assumed that no attempt indicates that the child is unable to do the task. From a developmental point of view, this is equivalent to a scribble. For the construction of norms, no attempt responses were considered to be equivalent to a scribble, and were allocated a score of 1.

It was, however, noted that, in some cases, children who were capable of more advanced responses on previous items did not attempt some of the more difficult items, particularly the diamond and the sentence. In such cases, allocating a score of 1 would lead to an underestimate of the child's developmental level. For this reason a procedure recommended by the author was used for dealing with cases in which the child makes no attempt. This involved assigning a score based on the score to other items. For example, if

a child had a score of 4 on the square and did not attempt the diamond then a score of 3 would be applied to the diamond.

Dr. de Lemos felt that imputation was necessary to make the NLSCY data more consistent with data collected with the *Who Am I?* in other studies. In most cases, the *Who Am I?* is administered by the child's teacher or an ACER researcher trained in child development. The NLSCY uses lay interviewers who only have a short time, in the interview setting, to develop rapport with the child. This made it harder for the interviewers to convince the children to attempt the more difficult items. The imputation rules attempt to adjust the scores to better reflect the child's developmental level.

Evaluation of the assessment

Analysis was conducted on the *Who Am I?* data to determine whether this assessment has good validity. The analyses included: comparing age equivalent scores to the child's age, comparison with the Number Knowledge, comparison with *Who Am I?* in other studies, an analysis of the items and of non-response. As the scoring procedures were being developed the NLSCY team consulted with Dr. de Lemos.

All the evidence indicates that the test has good validity and should provide data users with information about the child's developmental level. This assessment is not free of non-response bias. Please see Chapters 10 and 13, for more information on non-response.

16.4 Mathematics Computation Exercise

In Cycle 3, there was both a Math and a Reading test. In Cycle 4, only the math test was administered. The reading test was removed because of time constraints. It was decided that only one test could be administered. The math test was chosen as it has been administered in all previous cycles.

The Mathematics Computation Exercise administered to the child is a shortened version of the Mathematics Computation Test of the standardized Canadian Achievement Tests, Second Edition (CAT/2). The CAT/2 is a series of tests designed to measure achievement in basic academic skills.

This test is a shortened version of the CAT/2 mathematical operations test. The CAT/2 mathematical operations test measures the student's ability to do addition, subtraction, multiplication and division operations on whole numbers, decimals, fractions, negatives and exponents. Problem solving involving percentages and the order of operations are also measured. The short version of the test developed for the purposes of the NLSCY consists of 20 questions at each level, except the 9-10 level, which has 15 questions.

In Cycle 3 of the NLSCY, every child surveyed, from grades two to ten, was given mathematics and reading tests. For a test to be administered, the consent from parents and the school board are required. In Cycle 4, it was decided to change the administration of the test to take place in the home during the interview. This was done to reduce the burden on schools and with the hope of improving the response rate to the tests. Unfortunately, this decision was taken too late in the collection cycle to administer the test to all households during the interview. For households where the questionnaire had already been completed, a second visit was made to complete the math test. The implications of the change in administration are two-fold. All children are no longer tested at the same time of the year and the testing conditions can vary widely from one household to another. For the Cycle 4 collection, the math tests were administered between January and June of 2001.

Scoring

Each child who took the mathematics test was given a raw (gross) score, a scaled score referred to as the classical scaled score and an IRT scaled score. The raw (gross) score is obtained simply by adding the number of correct answers. The Classically derived scale score and the IRT scaled score are described as follows.

The approach of the item response theory (IRT) was used successfully in Cycle 2 to derive scores for the reading comprehension tests. IRT is a measurement system commonly used in psychometric and educational testing. IRT test scoring involves calculating either the most likely or the expected value of the ability of the examinee. The probability of a correct response to a question is assumed to be a certain logistic function of the examinee's ability. This probability is an S-shaped curve over the range of abilities. Its shape depends on the difficulty of the question, and sometimes also the discriminating power of the question (in the two-parameter IRT model) and the chance of a hypothetical no-ability examinee guessing correctly (in the three-parameter model, for multiple-choice questions).

Unlike the approach of the classical theory, the IRT makes it possible to scale the scores without preset population standards. Using common test items linking grades, standards are estimated from the entire population of children taking the test for this cycle. Scores are derived ranking each child within a level then the scores are vertically scaled to reflect the progression of scores throughout all the levels. In order to ensure comparability from year to year, each sample from each cycle must represent equivalent populations.

Three-parameter logistical model was chosen for the math tests. The three-parameter model takes into consideration both the difficulty and the discrimination of the item and also considers the pseudo-guessing component. In this way, the IRT takes into consideration the pattern of responses. Two children with the same raw (gross) score will not have the same scaled score unless they answered exactly the same way. For example, a child who only answered the 5 easiest questions correctly would have a lower scaled score than the one who only answered the 5 hardest questions correctly.

This score differs from the other scaled score reported for the math test as it provides a greater precision in the estimates of test performance. Unlike the other reported scale score, this score is not referenced to an external population of expected performance but is instead measured against the expected performance of the current population as estimated by all test subjects. Rank test analysis performed using both methods of scoring showed no significant difference between the two measurements. It is highly recommended that either scale scores be used in all analyses involving growth over time.

The classical scaled score is derived from standards (norms) established by the Canadian Test Centre (CTC). The CTC developed these standards from a sample of Canadian children from all 10 provinces (however, the test has been developed in English only and so in Quebec, the sample represents only the English schools), which is referred to as the normative sample. The children from the normative sample received the complete test. The scaled scores are units of a single scale with equidistant intervals that covers all of the grade levels. The scale was developed using a Thurstone procedure derived from the classical testing theory.

The fact that a short test was used for children in the NLSCY sample meant that it was not possible to directly associate the CTC scaled scores with the gross scores obtained in the survey. For this reason, the CTC normative sample was used to calculate the percentile rank for each gross score on our shortened version of the test. For example, using level 6, we find in the short test a percentile rank of 0.94% for a raw (gross) score of 1. On the complete test, the percentile ranks of 0.55% and 0.99% correspond to raw (gross) scores of

3 and 4 and to scaled scores of 315 and 319 respectively. After linear interpolation, we obtain a scaled score of 318 for the gross score of 1 on the short version of the test.

The table below shows the relation between the raw (gross) scores and the scaled scores by grade for the NLSCY mathematics test.

Relation Between Raw Scores and Scaled Scores (Classical) by Grade for the Cycle 4 Mathematics Test

Raw Scores	Scaled Scores							
	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grades 9 - 10
0	205	231	267	294	311	330	361	406
1	216	244	285	306	318	338	376	427
2	225	261	301	324	332	359	401	464
3	235	273	314	339	347	381	425	504
4	246	287	327	355	365	405	443	533
5	253	299	339	370	383	426	464	558
6	259	310	350	382	397	444	480	582
7	264	317	361	392	409	461	494	603
8	269	324	371	403	421	477	506	627
9	274	333	380	414	433	492	517	652
10	280	340	388	425	445	506	529	677
11	285	348	396	434	456	518	540	701
12	292	355	405	443	468	529	557	727
13	300	363	416	453	480	541	570	754
14	306	372	425	464	495	550	583	789
15	314	381	434	478	510	559	597	871
16	324	390	445	489	527	574	614	N/A
17	334	401	458	503	544	594	637	N/A
18	345	416	475	522	564	611	664	N/A
19	361	434	497	540	584	636	684	N/A
20	392	470	524	568	622	674	729	N/A

Cycle 3 scores
The cycle 3 maths scores (CMACS01, S02, S03) have been corrected on the data file. The values in CMACS03 were incorrect on the previous files and should not be used.
The values in CMACS01 or 02 were recalculated, see the Appendix II on Changes to previous cycles for more detail.

16.5 Cognitive Measure (16-17 year olds)

Discussions during the May, 1999 Expert Advisory Group Meeting clearly outlined the need to move away from curriculum-based testing. The advisory group recommended moving away from school based testing for this age group as:

? the NLSCY is not consistently gaining permission from school boards to do testing in the classroom;
? there are fewer youth in school; and,
? we will have to do testing in the home for 18-19 year olds.

Sixteen- and seventeen-year-olds in the NLSCY were asked to take a Cognitive Measure test. This test comprises questions (items) from the pilot of the Programme for International Student Assessment (PISA 2000) mathematics test that were not in the final version of the PISA test. Since the PISA test was designed for 15-year-olds, the theoretical accuracy of the NLSCY Cognitive Measure test in estimating the ability of advanced respondents was not ideal. However, some NLSCY respondents noted that the test was quite difficult for them, and the test seems to be at least as effective as the NLSCY mathematics tests from previous cycles.

There were actually two tests, one for higher-ability youths and a slightly easier one for lower-ability youths. Data from previous cycles was used to pre-select the respondents into the high-ability group and the low-ability group. Each test contained 18 items. Ten items were shared between the two levels of tests. Eight items were unique to the test intended for the high-ability group and eight items were unique to the test designed for the low-ability group.

Methodology for scoring

To obtain the Cognitive Measure score, the three-parameter model from Item Response Theory (IRT) was used. For free-response items, the pseudo-chance parameter (to model guessing in multiple-choice items) was fixed at a value of zero. Maximum Likelihood Estimation (MLE) of the three item parameters (discrimination, difficulty and pseudo-chance) followed by *Expected A Priori* (EAP) estimation of the Cognitive Measure score was performed in an iterative process until the Cognitive Measure score converged sufficiently. To make the ability score consistent with other NLSCY measures, each ability estimate used in the parameter estimation was weighted by the cross-sectional survey weight adjusted for component non-response. The statistical software SAS was used to performed these computations.

As a final step, a lower bound was placed on the Cognitive Measure score, which raised the scores of the bottom 31 respondents. Removing or lowering this bound decreased the correlation of the Cognitive Measure score with all of the math test scores from the first three cycles of NLSCY. These 31 overly low scores appear to be an artifact of guessing or the low-stakes nature of the test.

Treatment of attempted items

Before the iterative estimation process began, raw scores for each item were calculated. Incorrect answers were scored as zero and correct answers were scored as 1, which is standard for IRT. Partially correct answers were scored as marks received divided by the maximum possible mark; e.g., 1/2 for half-marks.

Treatment of items with no response

Items without a response ("unanswered items") can be either "omitted" items or "not-reached" items. Omitted items are those that the examinee probably saw but did not answer. In the scoring of the cognitive test, an unanswered item earlier in the test than the last item attempted was considered to be an omitted item. The first item after the last item attempted was also considered to be an omitted item. In this case, the respondent probably saw the question, decided that it was too difficult, and stopped taking the test.

Usually, examinees omit items because they do not know the correct answer. Therefore, an omitted free-response item was given a zero mark. An omitted multiple-choice item was given a mark of one divided by the number of choices.

Not-reached items are those that the examinee probably did not see. These items do not provide any information about the ability of the respondent. In the cognitive test scoring, all

items up to and including the item after the last attempted item were considered to be reached items. Items not reached by a respondent were ignored in the estimation of that respondent's ability.

Note: The cognitive measure is not timed. The respondent can take as much time as needed to complete the test.

Scores (DMAydS01)
The scores presented on the data file have a mean of 0 and a standard deviation of 1. This scale is standard in IRT.

17.0 Analytic Issues

This chapter provides users with an overview of the various analytic issues one should consider when analyzing NLSCY data. Some of the points mentioned in this chapter have been explored in greater detail in other chapters. The idea is to assemble in a single chapter all the important issues that might affect an analysis of the data.

17.1 How a complex sample design affects analysis

As described in the chapter 5 on the sample and the chapter 12 on weighting, the children in the NLSCY sample were selected using a complex sample design in order to meet clients' needs and address certain operational constraints. To make effective use of the data and produce valid estimates, it is important to use sample weights, either longitudinal or cross-sectional depending on the type of analysis.

17.2 Unit of analysis

In the NLSCY, the unit of analysis is always the child. Some household data were collected, but no estimates can be produced for the household level. All estimates must be for the child level.

17.3 Type of analysis: cross-sectional, longitudinal or repeated

With the NLSCY, users have their choice of different types of analysis: longitudinal analysis, cross-sectional analysis or even treating the NLSCY data as if they came from a repeated survey. Each type involves different issues.

17.3.1 Longitudinal analysis

In any longitudinal analysis, the population represented when using the longitudinal weights is always the population at the time of the child's initial selection. With the survey now in its fourth Cycle, the analyst must deal with the problems of sample attrition and lack of coverage due to total non-response and the problem of partial non-response. In addition, since the children selected in Cycle 1 have now responded four times, a conditioning bias may have developed over time. All these problems are described later in this chapter.

In addition, in Cycle 4, for the first time, two sets of longitudinal weights are available. The first set comprises longitudinal children who responded in Cycle 4, whether or not they responded in all previous cycles. The second set of longitudinal weights, computed for children introduced in Cycle 1, applies to children who have responded in every cycle.

17.3.2 Cross-sectional analysis

In any cross-sectional analysis, the population represented when using the cross-sectional weights is always the population at the time of collection of the current cycle (Cycle 4). In addition to the above-mentioned problems affecting longitudinal analysis, a further problem must be taken into consideration in cross-sectional analysis: the coverage of the current population. As described in Chapter 5, the cross-sectional sample is composed of children selected at different points in time. For example, children aged 6 to 17 were selected in 1994. The sample has not been replenished since. Consequently, people who have immigrated to Canada since 1994 have no chance of being included in the sample. The impact of this situation is described in the chapter on data quality.

17.3.3 Repeated analysis

A repeated analysis is an analysis in which a single target population is studied at different points in time. For example, with the NLSCY, one-year-olds can be studied in 1994-95, 1996-97, 1998-99 and 2000-2001.

Depending on the domain being analyzed, certain issues can arise in either longitudinal or cross-sectional analysis. Other issues should also be considered. First, for a particular age group, the number of children in the sample can vary substantially from one cycle to another. This is particularly true for one-year-olds and five-year-olds.

In addition, when choosing a domain, an analyst must determine whether the samples are independent or dependent. Take for example the one-year-olds selected in Cycles 1 through 4. By the very nature of the sample selection process, all the samples of one-year-olds would be independent [17]. On the other hand, a study about 11-year-olds would not yield independent sampled units. The 11-year-olds in Cycle 1 obviously come from the original sample of children aged 0 to 11. In Cycle 2, the 11-year-olds are actually the nine-year-olds from Cycle 1. The 11-year-olds in Cycle 3 are the seven-year-olds selected in Cycle 1, and the ones in Cycle 4 are the five-year-olds selected in Cycle 1. Thus, all these “samples” of 11-year-olds come from the same initial ample. In fact many might be siblings measured at different points in time sharing similar if not identical household characteristics. Consequently, they are statistically dependent. In computing the variance, the analyst must ensure to measure the covariance between the four samples of 11-year-olds.

17.4 Partial and total non-response

Like any other survey, the NLSCY is subject to non-response. There are two main types of non-response: total and partial. By definition, a respondent child has an adult component and at least one child/youth component completed for that household (it may be their own or a sibling’s component).

17.4.1 Total non-response

Total non-response is the complete lack of data for a selected and eligible child due to factors such as refusal to take part in the survey or inability to trace the child. Total non-response is taken into account and corrected within the sampling weights. This process is described in detail in the chapter on weighting. The greater the amount of

17 Except for 1 year old siblings of previous longitudinal households.

total non-response, the more difficult it is to adjust the weight of responding units to compensate for the non-response. Moreover, from a longitudinal standpoint, with the first set of longitudinal weights described above, a child who responded in Cycle 4 may not have responded in a previous cycle. The analyst must account for this in their analysis.

17.4.2 Partial non-response

For a child to be considered a respondent in Cycle 4, the child/youth component or the Person Most Knowledgeable (PMK) component must be completed. This means that a child can be deemed a respondent even if some questions or even entire components are not answered. For a given child, there are many sources of information. Depending on the child's age, the information may come from one or more of the following sources: themselves, a parent, a teacher, the school principal. Hence, even if one of these sources failed to provide the desired information, the child may still be considered a respondent.

In particular, the income variables are imputed. Most other variables are set to "don't know" or "not stated". The analyst must therefore determine the extent of non-response in the variables of interest before proceeding with the analysis.

17.5 Data processing

The NLSCY contains a huge amount of information. Although every effort is made to ensure that the data is of high quality with online editing while the data is being collected, if each variable were verified once captured, it would be impossible to make the data available within a reasonable length of time. As a result, data processing focuses on the survey's key variables, and errors may be left in the data. However, efforts are made to keep errors to a minimum.

17.6 Coverage of the cross-sectional and longitudinal samples

The goal of any survey is to be able to produce reliable estimates that are representative of the target population identified in the survey's planning phase. After each cycle, a quality evaluation is conducted to determine whether our sample is still representative of the population that we wanted to represent. The findings of that study are presented in the chapter on data quality.

Briefly, two factors in the NLSCY reduce the coverage of the sample over time. First, from a longitudinal standpoint, cycle non-response decreases the size of the originally selected sample. Since it has been shown that the characteristics of non-respondent units are different from those of respondent units, a significant amount of non-response by units with specific characteristics will tend to reduce the sample's overall coverage.

For the cross-sectional sample coverage is reduced by an additional factor: changes in the population. As previously mentioned, since the sample of older children (ages 6 to 17) was selected in Cycle 1 and has not been replenished since, all the children between the ages of 6 and 17 who have immigrated to Canada since 1994 not eligible for selection even though they're are part of the targeted cross-sectional population. If the characteristics of those excluded are substantially different from the characteristics of other children, then the

cross-sectional sample will be less representative. For a more detailed look at this subject, see the chapter on data quality (Chapter 13).

17.7 Sample sizes vary greatly from cycle to cycle for certain ages

The survey goals and therefore the sample design have changed from one cycle to another. When analysing domains with specific age groups, one must keep in mind that the sample size may vary substantially from cycle to cycle. As a result, the precision of the estimates will also be different from one cycle to another. This is particularly true for one-year-olds and five-year-olds. For more details, see the chapter on the sample (Chapter 5)

17.8 Sample reduction between Cycles 1 and 2

Sample reductions were made between Cycles 1 and 2. First, the sample of children in common with the National Population Health Survey (NPHS) was dropped. Then, to reduce the response burden on large households, the number of children selected within a household was limited to two. This meant that some children were removed from the longitudinal sample. Specifically, only 16,903 of the 22,831 children who responded in Cycle 1 were selected for the Cycle 2 sample. For further details, see the chapter on sample selection (Chapter 5).

17.9 Conditioning bias

In a longitudinal survey, the same respondents are re-contacted at predetermined intervals. Respondents who have already taken part in the survey anticipate a revisit on this topic in subsequent cycles. As a result, their behaviour or responses may be different from those of a unit responding who would have been selected for the first time. The longitudinal respondent may become influenced by being selected for the survey. Their behaviour and responses may be affected over time as they may feel personally evaluated.

17.10 Combining data over time

When analysing small domains or portion of the target population, the sample size may be very small. To increase the incidence or rare sample units, one might be tempted to combine the data from a number of cycles. For example, in a study 10-year-olds with a specific medical condition in Prince Edward Island, it is quite conceivable that the sample will not be large enough to yield reliable estimates. One might therefore combine the data from Cycles 1 to 4 for 10-year-olds with the required medical condition in Prince Edward Island. In performing this kind of analysis, one must be aware that this new domain is not a known or intuitive population but with a set of children who had specific characteristics at different points in time. The time of collection variable should be included in the analysis so that the cycle effect can be identified. Other issues, such as which survey weight to use and the dependence of the sample units must be considered.

17.11 Computing the variance with certain software applications

SAS and SPSS, software applications commonly used by analysts, are able to compute point estimates correctly using sampling weights. With the exception of some SAS procedures, these applications are not designed to take into account the sample design when calculating the variance. These applications do not take into account the estimation complexities of the NLSCY (complex sample design, non-response adjustment and post-stratification) in calculating the variance. As a result, they tend to underestimate the variance, sometimes substantially. Consequently, the analyst is strongly encouraged to use the Bootstrap weights for the variance estimation.

Some applications can use the Bootstrap weights. SUDAAN and WesVar take into account the sample design in calculating the variance, using the Balanced Repeated Replication (BRR) method. The creation of BRR weights differ in theory from the creation of Bootstrap weights, but the variance estimator is the same. As a result, the NLSCY Bootstrap weights can be used with these applications.

There is a stand alone version of SUDAAN, as well as an integrated version with SAS. The latter gives the flexibility to use the SUDAAN procedures within SAS. With its 9 procedures, SUDAAN can produce estimates of means, ratios and totals, independence tests, linear, log-linear and logistic regressions and survival tests. SUDAAN can read SAS and SPSS files, as well as other common types of files.

Wesvar uses a “point and click” approach, which makes it easy to learn. The types of analysis are more limited than SUDAAN, but WesVar still allows the estimates of means, ratios and totals, independence tests, as well as linear and logistic regressions. WesVar can read SAS, SPSS and other common type of files, but the application requires an additional step to create a WesVar file before processing with the analysis.

Bootstrap weights can be used with other applications which offer the required programming environment and the desired analytical tools. SAS and SPSS macros have been developed by NLSCY to use the Bootstrap weights to produce variance estimates based on the sample design.

17.12 Updating the weights for previous cycles

The most recent population estimates are used to calculate the NLSCY's longitudinal and cross-sectional weights. Using population estimates ensures that the target population is represented as accurately as possible. When the longitudinal and cross-sectional weights for the first three NLSCY cycles were produced and published, only preliminary population estimates were available. The population estimates were revised, and the final counts are now available. As a result, the longitudinal and cross-sectional weights for the first three cycles were recalculated to reflect the revised population estimates. For more details, see the appendix.

Appendix I - Record Layouts

This appendix describes the layout of the files. It presents for each variable its position within the file, its name, its length and if it is an alpha or numeric value.

Example:

Variable START	Variable name	Variable length	Variable characteristics
@00756	DWTCWd1L	12.4	Numeric value with four decimals and of a maximum of twelve bites including the decimal symbol.
@00768	DMMCQ02	$1.	The symbol: $ indicates an alpha value of a maximum of 1 bite.

1. Primary Layout

```
@00001   DGEHD03     2.
@00003   DGEHbD04    2.
@00005   DGEHbD06    3.
@00008   DGEHbD07    8.
@00016   DDMHPC    $6.
@00022   DMMPQ01     3.
@00025   DMMPQ02    $1.
@00026   DMMPQ03A     4.
@00030   DMMPQ03B     2.
@00032   DMMPQ03C     2.
@00034   DMMPQ04     2.
@00036   DMMSQ01     3.
@00039   DMMSQ02    $1.
@00040   DMMSQ03A     4.
@00044   DMMSQ03B     2.
@00046   DMMSQ03C     2.
@00048   DMMSQ04     2.
@00050   DMMCQ01     3.
@00053   DMMCbQ1A     3.
@00056   DMMCdQ1B     4.
@00060   DMMCQ02    $1.
@00061   DMMCQ03A     4.
@00065   DMMCQ03B     2.
@00067   DMMCQ03C     2.
@00069   DDMCD01     2.
@00071   DDMHD02     2.
@00073   DDMCD03     2.
@00075   DDMCD04     1.
@00076   DDMCD05     1.
@00077   DDMCD06     2.
@00079   DDMPD06A     1.
@00080   DDMCD06B     2.
@00082   DDMCD06C     1.
@00083   DDMPD06D     2.
@00085   DDMSD06E     2.
@00087   DDMHD06F     2.
@00089   DDMHD07     2.
@00091   DDMCD08     2.
@00093   DDMCD09     2.
@00095   DDMCD10     2.
@00097   DDMCD11     2.
@00099   DDMCD12     2.
@00101   DDMCD13     2.
@00103   DDMCD14     1.
@00104   DDMCD15     1.
@00105   DDMCD16     1.
@00106   DDMCD17     2.
@00108   DDMCD18     2.
@00110   DDMCD18B     2.
@00112   DDMCD19     2.
@00114   DDMCD19B     2.
@00116   DDMCD20     2.
@00118   DHHHQ01     1.
@00119   DHHHQ03     2.
@00121   DHHHQ06     2.

@00123   DHHHdQ09     1.
@00124   DHHHD06B     1.
@00125   DEDPQ05     1.
@00126   DEDPQ06     1.
@00127   DEDPQ01     2.
@00129   DEDPQ02     1.
@00130   DEDPQ03     1.
@00131   DEDPdQ4b     2.
@00133   DEDPcQ7A     2.
@00135   DEDPcQ7B     2.
@00137   DEDPcQ7C     2.
@00139   DEDPcQ7D     2.
@00141   DEDSQ05     1.
@00142   DEDSQ06     1.
```

```
@00143   DEDSQ01    2.
@00145   DEDSQ02    1.
@00146   DEDSQ03    1.
@00147   DEDSdQ4b    2.
@00149   DEDScQ7A    2.
@00151   DEDScQ7B    2.
@00153   DEDScQ7C    2.
@00155   DEDScQ7D    2.
@00157   DEDPD01    2.
@00159   DEDPD02    1.
@00160   DEDPD04    2.
@00162   DEDSD01    2.
@00164   DEDSD02    1.
@00165   DEDSD04    2.
@00167   DEDHcQ8A    1.
@00168   DEDHcQ8B    1.
@00169   DEDHcQ8C    1.
@00170   DEDHcQ8D    1.
@00171   DEDHcQ8E    1.
@00172   DLFPQ01    2.
@00174   DLFPQ02    1.
@00175   DLFPcQ3A    2.
@00177   DLFPbQ03    2.
@00179   DLFPbQ04    2.
@00181   DLFPbQ5A    1.
@00182   DLFPbQ5B    1.
@00183   DLFPbQ5C    1.
@00184   DLFPbQ5D    1.
@00185   DLFPbQ5E    1.
@00186   DLFPbQ5F    1.
@00187   DLFPbQ5G    1.
@00188   DLFPbQ5H    1.
@00189   DLFPbQ06    1.
@00190   DLFPcQ6A    1.
@00191   DLFPcQ6B    2.
@00193   DLFPcQ6C    2.
@00195   DLFPbQ13    1.
@00196   DLFPb14A    2.
@00198   DLFPb14B    1.
@00199   DLFPb14C    9.2
@00208   DLFP14CC    9.2
@00217   DLFPb14D    2.
@00219   DLFPb15A    2.
@00221   DLFPbQ16    1.
@00222   DLFPb17A    2.
@00224   DLFPb17B    2.
@00226   DLFSQ01    2.
@00228   DLFSQ02    1.
@00229   DLFScQ3A    2.
@00231   DLFSbQ03    2.
@00233   DLFSbQ04    2.
@00235   DLFSbQ5A    1.
@00236   DLFSbQ5B    1.
@00237   DLFSbQ5C    1.
@00238   DLFSbQ5D    1.
@00239   DLFSbQ5E    1.
@00240   DLFSbQ5F    1.
@00241   DLFSbQ5G    1.
@00242   DLFSbQ5H    1.
@00243   DLFSbQ06    1.
@00244   DLFScQ6A    1.
@00245   DLFScQ6B    2.
@00247   DLFScQ6C    2.
@00249   DLFSbQ13    1.
@00250   DLFSb14A    2.
@00252   DLFSb14B    1.
@00253   DLFSb14C    9.2
@00262   DLFS14CC    9.2
@00271   DLFSb14D    2.
@00273   DLFSb15A    2.
@00275   DLFSbQ16    1.
@00276   DLFSb17A    2.
@00278   DLFSb17B    2.
@00280   DLFPcD5A    4.
@00284   DLFPcD6A   $4.
@00288   DLFPcD7A    2.
@00290   DLFPcD8A    2.
@00292   DLFPD25    1.
@00293   DLFPD34    2.
@00295   DLFPbD38    2.
@00297   DLFPD51    1.
@00298   DLFScD5A    4.
@00302   DLFScD6A   $4.
@00306   DLFScD7A    2.
@00308   DLFScD8A    2.
@00310   DLFSD25    1.
@00311   DLFSD34    2.
@00313   DLFSbD38    2.
@00315   DLFSD51    1.
@00316   DLFHD49B    2.
@00318   DLFHD50    2.
@00320   DINPc1AA    6.
@00326   DINPc1AB    6.
@00332   DINPc1AC    6.
@00338   DINPc1AD    6.
@00344   DINPc1AE    6.
@00350   DINPc1AF    6.
@00356   DINPc1AG    6.
@00362   DINSc1AA    6.
@00368   DINSc1AB    6.
@00374   DINSc1AC    6.
@00380   DINSc1AD    6.
@00386   DINSc1AE    6.
@00392   DINSc1AF    6.
@00398   DINSc1AG    6.
@00404   DINHQ03    7.
@00411   DINHdQ05    1.
@00412   DINHdQ06    1.
@00413   DINHdQ07    1.
@00414   DINPD02    2.
@00416   DINPCD04    7.
@00423   DINScD02    2.
@00425   DINScD04    7.
@00432   DINHD01A    2.
@00434   DINHD01B    2.
@00436   DINHD03A    5.
```

```
@00441   DINHD04A    5.
@00446   DINHD05A    2.
@00448   DINHD03P    7.
@00455   DINHD07    2.
@00457   DLFSPL    1.
@00458   DLFSSL    1.
@00459   DINPCI1A    1.
@00460   DINSCI1A    1.
@00461   DHLPQ01    2.
@00463   DCHPQ1A    1.
@00464   DCHPd1AB    1.
@00465   DCHPQ1B    1.
@00466   DCHPQ1C    1.
@00467   DCHPQ1D    1.
@00468   DCHPQ1E    1.
@00469   DCHPQ1F    1.
@00470   DCHPQ1G    1.
@00471   DCHPQ1H    1.
@00472   DCHPQ1I    1.
@00473   DCHPQ1J    1.
@00474   DCHPQ1K    1.
@00475   DCHPQ1L    1.
@00476   DCHPQ1M    1.
@00477   DCHPQ1N    1.
@00478   DCHPQ1O    1.
@00479   DCHPQ1U    1.
@00480   DRSPd1AA    1.
@00481   DRSPQ01A    1.
@00482   DRSPdQ1F    1.
@00483   DRSPQ01E    1.
@00484   DRSPQ01D    1.
@00485   DHLPQ02    1.
@00486   DHLPQ03    2.
@00488   DHLPcQ05    2.
@00490   DHLPcQ5A    2.
@00492   DHLPQ06    3.
@00495   DHLSQ01    2.
@00497   DCHSQ1A    1.
@00498   DCHSd1AB    1.
@00499   DCHSQ1B    1.
@00500   DCHSQ1C    1.
@00501   DCHSQ1D    1.
@00502   DCHSQ1E    1.
@00503   DCHSQ1F    1.
@00504   DCHSQ1G    1.
@00505   DCHSQ1H    1.
@00506   DCHSQ1I    1.
@00507   DCHSQ1J    1.
@00508   DCHSQ1K    1.
@00509   DCHSQ1L    1.
@00510   DCHSQ1M    1.
@00511   DCHSQ1N    1.
@00512   DCHSQ1O    1.
@00513   DCHSQ1U    1.
@00514   DRSSd1AA    1.
@00515   DRSSQ01A    1.
@00516   DRSSdQ1F    1.
@00517   DRSSQ01E    1.
@00518   DRSSQ01D    1.
@00519   DHLSQ02    1.
@00520   DHLSQ03    2.
@00522   DHLScQ05    2.
@00524   DHLScQ5A    2.
@00526   DHLSQ06    3.
@00529   DCHPD01    1.
@00530   DRSPdD01    1.
@00531   DCHSD01    1.
@00532   DRSSdD01    1.
@00533   DHLMQ8    2.
@00535   DHLMQ9    2.
@00537   DHLMQ11    2.
@00539   DDPPQ12A    1.
@00540   DDPPQ12B    1.
@00541   DDPPQ12C    1.
@00542   DDPPQ12D    1.
@00543   DDPPQ12E    1.
@00544   DDPPQ12F    1.
@00545   DDPPQ12G    1.
@00546   DDPPQ12H    1.
@00547   DDPPQ12I    1.
@00548   DDPPQ12J    1.
@00549   DDPPQ12K    1.
@00550   DDPPQ12L    1.
@00551   DDPPS01    2.
@00553   DFNHQ01A    1.
@00554   DFNHQ01B    1.
@00555   DFNHQ01C    1.
@00556   DFNHQ01D    1.
@00557   DFNHQ01E    1.
@00558   DFNHQ01F    1.
@00559   DFNHQ01G    1.
@00560   DFNHQ01H    1.
@00561   DFNHQ01I    1.
@00562   DFNHQ01J    1.
@00563   DFNHQ01K    1.
@00564   DFNHQ01L    1.
@00565   DFNHQ01M    1.
@00566   DFNHQ02    2.
@00568   DFNHS01    2.
@00570   DSFHQ01    2.
@00572   DSFHQ02    2.
@00574   DSFHQ03    1.
@00575   DSFHcQ04    2.
@00577   DSFHQ05A    1.
@00578   DSFHQ05B    1.
@00579   DSFHQ05C    1.
@00580   DSFHQ06A    1.
@00581   DSFHQ06B    1.
@00582   DSFHQ06C    1.
@00583   DSFHQ06D    1.
@00584   DSFHQ06E    1.
@00585   DSFHS5    2.
@00587   DSFHS6    2.
@00589   DSPHQ01A    1.
@00590   DSPHQ01B    1.
@00591   DSPHQ01C    1.
```

```
@00592   DSPHQ01D    1.
@00593   DSPHQ01E    1.
@00594   DSPHQ01F    1.
@00595   DSPHQ01H    1.
@00596   DSPHQ01I    1.
@00597   DSPHdQ2A    1.
@00598   DSPHdQ2B    1.
@00599   DSPHdQ2C    1.
@00600   DSPHdQ2D    1.
@00601   DSPHdQ2E    1.
@00602   DSPHS01    2.
@00604   DSDPQ1    2.
@00606   DSDPQ2AA    1.
@00607   DSDPQ2AB    1.
@00608   DSDPQ2AC    1.
@00609   DSDPQ2AD    1.
@00610   DSDPQ2B    1.
@00611   DSDPQ3    4.
@00615   DSDPQ4A    1.
@00616   DSDPQ4B    1.
@00617   DSDPQ4C    1.
@00618   DSDPQ4D    1.
@00619   DSDPQ4E    1.
@00620   DSDPQ4F    1.
@00621   DSDPQ4G    1.
@00622   DSDPQ4H    1.
@00623   DSDPQ4I    1.
@00624   DSDPQ4J    1.
@00625   DSDPQ4K    1.
@00626   DSDPQ4L    1.
@00627   DSDPQ4M    1.
@00628   DSDPQ4N    1.
@00629   DSDPQ4O    1.
@00630   DSDPQ4P    1.
@00631   DSDPQ4Q    1.
@00632   DSDPQ4R    1.
@00633   DSDPQ4S    1.
@00634   DSDPb4AA    1.
@00635   DSDPb4AB    1.
@00636   DSDPb4AC    1.
@00637   DSDPb4AD    1.
@00638   DSDPb4AE    1.
@00639   DSDPb4AF    1.
@00640   DSDPb4AG    1.
@00641   DSDPb4AH    1.
@00642   DSDPb4AI    1.
@00643   DSDPb4AJ    1.
@00644   DSDPb4AK    1.
@00645   DSDPb4AL    1.
@00646   DSDPQ5A    1.
@00647   DSDPQ5B    1.
@00648   DSDPQ5C    1.
@00649   DSDPQ5D    1.
@00650   DSDPQ5E    1.
@00651   DSDPQ5F    1.
@00652   DSDPQ5G    1.
@00653   DSDPQ5H    1.
@00654   DSDPQ5I    1.
@00655   DSDPQ5J    1.
@00656   DSDPQ5K    1.
@00657   DSDPQ5L    1.
@00658   DSDPQ5M    1.
@00659   DSDPQ5N    1.
@00660   DSDPQ5O    1.
@00661   DSDPQ5P    1.
@00662   DSDPQ5Q    1.
@00663   DSDPQ5R    1.
@00664   DSDPQ5S    1.
@00665   DSDPQ6A    1.
@00666   DSDPQ6B    1.
@00667   DSDPQ6C    1.
@00668   DSDPQ6D    1.
@00669   DSDPQ6E    1.
@00670   DSDPQ6F    1.
@00671   DSDPQ6G    1.
@00672   DSDPQ6H    1.
@00673   DSDPQ6I    1.
@00674   DSDPQ6J    1.
@00675   DSDPQ6K    1.
@00676   DSDPQ6L    1.
@00677   DSDPQ6M    1.
@00678   DSDPQ6N    1.
@00679   DSDPQ6O    1.
@00680   DSDPQ6P    1.
@00681   DSDPQ6Q    1.
@00682   DSDPQ6R    1.
@00683   DSDPQ6S    1.
@00684   DSDPC6AA    1.
@00685   DSDPC6AB    1.
@00686   DSDPC6AC    1.
@00687   DSDPQ8    2.
@00689   DSDSQ1    2.
@00691   DSDSQ2AA    1.
@00692   DSDSQ2AB    1.
@00693   DSDSQ2AC    1.
@00694   DSDSQ2AD    1.
@00695   DSDSQ2B    1.
@00696   DSDSQ3    4.
@00700   DSDSQ4A    1.
@00701   DSDSQ4B    1.
@00702   DSDSQ4C    1.
@00703   DSDSQ4D    1.
@00704   DSDSQ4E    1.
@00705   DSDSQ4F    1.
@00706   DSDSQ4G    1.
@00707   DSDSQ4H    1.
@00708   DSDSQ4I    1.
@00709   DSDSQ4J    1.
@00710   DSDSQ4K    1.
@00711   DSDSQ4L    1.
@00712   DSDSQ4M    1.
@00713   DSDSQ4N    1.
@00714   DSDSQ4O    1.
@00715   DSDSQ4P    1.
@00716   DSDSQ4Q    1.
@00717   DSDSQ4R    1.
```

```
@00718   DSDSQ4S    1.
@00719   DSDSb4AA    1.
@00720   DSDSb4AB    1.
@00721   DSDSb4AC    1.
@00722   DSDSb4AD    1.
@00723   DSDSb4AE    1.
@00724   DSDSb4AF    1.
@00725   DSDSb4AG    1.
@00726   DSDSb4AH    1.
@00727   DSDSb4AI    1.
@00728   DSDSb4AJ    1.
@00729   DSDSb4AK    1.
@00730   DSDSb4AL    1.
@00731   DSDSQ5A    1.
@00732   DSDSQ5B    1.
@00733   DSDSQ5C    1.
@00734   DSDSQ5D    1.
@00735   DSDSQ5E    1.
@00736   DSDSQ5F    1.
@00737   DSDSQ5G    1.
@00738   DSDSQ5H    1.
@00739   DSDSQ5I    1.
@00740   DSDSQ5J    1.
@00741   DSDSQ5K    1.
@00742   DSDSQ5L    1.
@00743   DSDSQ5M    1.
@00744   DSDSQ5N    1.
@00745   DSDSQ5O    1.
@00746   DSDSQ5P    1.
@00747   DSDSQ5Q    1.
@00748   DSDSQ5R    1.
@00749   DSDSQ5S    1.
@00750   DSDSQ6A    1.
@00751   DSDSQ6B    1.
@00752   DSDSQ6C    1.
@00753   DSDSQ6D    1.
@00754   DSDSQ6E    1.
@00755   DSDSQ6F    1.
@00756   DSDSQ6G    1.
@00757   DSDSQ6H    1.
@00758   DSDSQ6I    1.
@00759   DSDSQ6J    1.
@00760   DSDSQ6K    1.
@00761   DSDSQ6L    1.
@00762   DSDSQ6M    1.
@00763   DSDSQ6N    1.
@00764   DSDSQ6O    1.
@00765   DSDSQ6P    1.
@00766   DSDSQ6Q    1.
@00767   DSDSQ6R    1.
@00768   DSDSQ6S    1.
@00769   DSDSC6AA    1.
@00770   DSDSC6AB    1.
@00771   DSDSC6AC    1.
@00772   DSDSQ8    2.
@00774   DSDPD01    3.
@00777   DSDPD02    3.
@00780   DSDPD02B    1.
@00781   DSDPD03    2.
@00783   DSDPD04    2.
@00785   DSDPD05    2.
@00787   DSDPD05B    1.
@00788   DSDPD06    2.
@00790   DSDPD06B    1.
@00791   DSDSD01    3.
@00794   DSDSD02    3.
@00797   DSDSD02B    1.
@00798   DSDSD03    2.
@00800   DSDSD04    2.
@00802   DSDSD05    2.
@00804   DSDSD05B    1.
@00805   DSDSD06    2.
@00807   DSDSD06B    1.
@00808   DEDCdQ0    2.
@00810   DEDCcQ0A    2.
@00812   DEDCD01    2.
@00814   DEDCdQ4A    2.
@00816   DEDCQ02    1.
@00817   DEDCQ03    1.
@00818   DEDCQ06    1.
@00819   DEDCD03    2.
@00821   DEDCdQ7F    2.
@00823   DEDCdQ9A    1.
@00824   DEDCdQ9B    2.
@00826   DEDCQ10    2.
@00828   DEDCd11A    2.
@00830   DEDCQ12A    1.
@00831   DEDCc12B    1.
@00832   DEDCc12C    1.
@00833   DEDCdQ13    2.
@00835   DEDCd13A    2.
@00837   DEDCQ14A    2.
@00839   DEDCb14A    2.
@00841   DEDCQ14B    2.
@00843   DEDCQ14C    2.
@00845   DEDCb14C    2.
@00847   DEDCQ14D    2.
@00849   DEDCb14E    2.
@00851   DEDCQ14F    2.
@00853   DEDCb14H    2.
@00855   DEDCd15A    1.
@00856   DEDCd15B    1.
@00857   DEDCd15C    1.
@00858   DEDCd15D    2.
@00860   DEDCc16A    1.
@00861   DEDCc16B    1.
@00862   DEDCQ17    2.
@00864   DEDCQ18A    1.
@00865   DEDCQ18B    2.
@00867   DEDCc18C    2.
@00869   DEDCc18D    2.
@00871   DEDCQ19A    1.
@00872   DEDCQ19B    1.
@00873   DEDCQ19C    1.
@00874   DEDCQ19D    1.
@00875   DEDCc19E    1.
```

```
@00876   DEDCb21A    1.
@00877   DEDCb21B    1.
@00878   DEDCc21C    1.
@00879   DEDCb21E    1.
@00880   DEDCb21F    1.
@00881   DEDCc21G    1.
@00882   DEDCb21I    1.
@00883   DEDCb21J    1.
@00884   DEDCb21K    1.
@00885   DEDCbQ22    2.
@00887   DEDCbQ23    2.
@00889   DHLCQ01    2.
@00891   DHLCQ02    2.
@00893   DHLCQ03B    4.2
@00897   DHLCQ04A    7.3
@00904   DHLCb4C1    5.3
@00909   DHLCQ05    2.
@00911   DHLCdQ5A    1.
@00912   DHLCdQ5B    1.
@00913   DHLCdQ5C    1.
@00914   DHLCdQ5D    1.
@00915   DHLCQ06    1.
@00916   DHLCQ07    1.
@00917   DHLCQ08    1.
@00918   DHLCQ09    1.
@00919   DHLCQ10    1.
@00920   DHLCQ11    1.
@00921   DHLCQ12    1.
@00922   DHLCQ13    1.
@00923   DHLCQ14    1.
@00924   DHLCQ15    1.
@00925   DHLCQ16    1.
@00926   DHLCQ17    1.
@00927   DHLCQ18    1.
@00928   DHLCQ19    1.
@00929   DHLCQ20    1.
@00930   DHLCQ21    1.
@00931   DHLCQ22    1.
@00932   DHLCQ23    1.
@00933   DHLCQ24    1.
@00934   DHLCQ25    1.
@00935   DHLCQ26    1.
@00936   DHLCQ27    1.
@00937   DHLCQ28    1.
@00938   DHLCQ29    1.
@00939   DHLCQ30    1.
@00940   DHLCQ31    2.
@00942   DHLCQ32    1.
@00943   DHLCQ33    2.
@00945   DHLCQ34    1.
@00946   DHLCQ35    1.
@00947   DHLCQ36    1.
@00948   DHLCQ37    1.
@00949   DHLCQ38    2.
@00951   DHLCQ39    2.
@00953   DHLCQ40    2.
@00955   DHLCQ41    2.
@00957   DHLCQ42    2.
@00959   DHLCQ43A    1.
@00960   DHLCQ43B    1.
@00961   DHLCQ43C    1.
@00962   DHLCQ44    1.
@00963   DHLCd45N    1.
@00964   DHLCd45O    1.
@00965   DHLCd45A    1.
@00966   DHLCQ45B    1.
@00967   DHLCQ45C    1.
@00968   DHLCQ45D    1.
@00969   DHLCQ45E    1.
@00970   DHLCQ45F    1.
@00971   DHLCQ45G    1.
@00972   DHLCQ45H    1.
@00973   DHLCd45M    1.
@00974   DHLCQ45I    1.
@00975   DHLCQ45J    1.
@00976   DHLCQ45K    1.
@00977   DHLCQ46    2.
@00979   DHLCQ47A    1.
@00980   DHLCQ47B    1.
@00981   DHLCQ48A    2.
@00983   DHLCQ48B    2.
@00985   DHLCQ48C    2.
@00987   DHLCQ48D    2.
@00989   DHLCQ48E    2.
@00991   DHLCQ48G    2.
@00993   DHLCQ48H    2.
@00995   DHLCQ48I    2.
@00997   DHLCQ49    1.
@00998   DHLCQ50    1.
@00999   DHLCQ51A    1.
@01000   DHLCQ51B    1.
@01001   DHLCQ51C    1.
@01002   DHLCQ51D    1.
@01003   DHLCQ51E    1.
@01004   DHLCQ52A    1.
@01005   DHLCQ52B    1.
@01006   DHLCQ52C    1.
@01007   DHLCQ52D    1.
@01008   DHLCQ52E    1.
@01009   DHLCQ52F    1.
@01010   DHLCQ52G    1.
@01011   DHLCQ52H    1.
@01012   DHLCQ52I    1.
@01013   DHLCQ52J    1.
@01014   DHLCQ52K    1.
@01015   DHLCQ52L    1.
@01016   DHLCQ52M    1.
@01017   DHLCQ52N    1.
@01018   DHLCQ52O    1.
@01019   DHLCcD2A    4.2
@01023   DHLCbD4C    5.3
@01028   DHLCbD4D    1.
@01029   DHLCbD45    1.
@01030   DHLCD51    1.
@01031   DMDCQ01A    1.
@01032   DMDCQ01B    1.
```

```
@01033   DMDCQ01C    1.
@01034   DMDCbQ2A    1.
@01035   DMDCbQ2B    1.
@01036   DMDCbQ2C    1.
@01037   DMDCbQ2D    1.
@01038   DMDCbQ2E    1.
@01039   DMDCQ03    1.
@01040   DMDCQ04    2.
@01042   DMDCQ05A    1.
@01043   DMDCQ05B    1.
@01044   DMDCQ05C    1.
@01045   DMDCQ05D    1.
@01046   DMDCQ06    2.
@01048   DMDCQ07    1.
@01049   DMDCQ08A    1.
@01050   DMDCQ08B    1.
@01051   DMDCQ08C    1.
@01052   DMDCQ08D    1.
@01053   DMDCQ09A    1.
@01054   DMDCQ09B    1.
@01055   DMDCQ09C    1.
@01056   DMDCQ09D    1.
@01057   DMDCQ09E    1.
@01058   DMDCQ10A    1.
@01059   DMDCQ10B    1.
@01060   DMDCQ10C    1.
@01061   DMDCQ10D    1.
@01062   DMDCQ10E    1.
@01063   DMDCQ12A    1.
@01064   DMDCQ12C    3.
@01067   DMDCQ13B    5.3
@01072   DMDCQ14B    2.
@01074   DMDCQ15    1.
@01075   DMDCQ16    1.
@01076   DMDCQ17    1.
@01077   DMDCQ18    1.
@01078   DMDCQ21A    1.
@01079   DMDCQ21B    1.
@01080   DMDCQ21C    1.
@01081   DMDCQ21D    1.
@01082   DMDCQ21E    1.
@01083   DMDCQ21F    2.
@01085   DMDCQ22    2.
@01087   DMDCQ23A    1.
@01088   DMDCQ23B    1.
@01089   DMDCQ23C    1.
@01090   DMDCQ23D    3.
@01093   DMDCQ23F    1.
@01094   DMDCQ24A    1.
@01095   DMDCQ24B    2.
@01097   DMDCQ25    1.
@01098   DMDCQ26    1.
@01099   DMDCQ27    2.
@01101   DMDCQ28A    1.
@01102   DMDCQ28B    1.
@01103   DMDCQ28C    1.
@01104   DMDCQ28D    1.
@01105   DMDCQ28E    1.
@01106   DMDCQ28F    1.
@01107   DMDCQ28G    1.
@01108   DMDCQ28H    1.
@01109   DMDCQ28I    1.
@01110   DMDCQ28J    1.
@01111   DMDCQ28K    1.
@01112   DMDCQ28L    1.
@01113   DMDCQ28M    1.
@01114   DMDCD01    1.
@01115   DMDCD02    2.
@01117   DMDCD03    2.
@01119   DMDCD04    2.
@01121   DMDCD05    2.
@01123   DMDCD06    3.
@01126   DMDCD07    1.
@01127   DMDCD08    1.
@01128   DMDCD09    1.
@01129   DMDCD10    1.
@01130   DMDCbQ29    1.
@01131   DWBCcQ1B    1.
@01132   DWBCcD2B    3.
@01135   DMDCbQ31    2.
@01137   DWBCcQ4A    1.
@01138   DWBCcQ4B    1.
@01139   DMLCdQ1    2.
@01141   DMLCdQ2    2.
@01143   DMLCdQ3    2.
@01145   DMLCdQ4    2.
@01147   DMLCdQ5    2.
@01149   DTMCQ01    2.
@01151   DTMCQ05    2.
@01153   DTMCQ06    2.
@01155   DTMCQ07    2.
@01157   DTMCQ08    2.
@01159   DTMCQ09    2.
@01161   DTMCQ11    2.
@01163   DTMCQ12    2.
@01165   DTMCQ14    2.
@01167   DTMCQ15    2.
@01169   DTMCQ17    2.
@01171   DTMCQ19    2.
@01173   DTMCQ20    2.
@01175   DTMCQ23    2.
@01177   DTMCQ23A    2.
@01179   DTMCcQ25    2.
@01181   DTMCcQ26    2.
@01183   DTMCcQ27    2.
@01185   DTMCcQ29    2.
@01187   DTMCcQ30    2.
@01189   DTMCcQ31    2.
@01191   DTMCQ33    2.
@01193   DLTCdQ1A    2.
@01195   DLTCdQ1B    2.
@01197   DLTCdQ1C    2.
@01199   DLTCdQ1D    2.
@01201   DLTCdQ1E    2.
@01203   DLTCdQ1G    2.
@01205   DLTCdQ1H    2.
```

```
@01207   DLTCdQ2    1.
@01208   DLTCQ03    2.
@01210   DLTCdQ4A    2.
@01212   DLTCdQ4B    2.
@01214   DLTCdQ4C    2.
@01216   DLTCdQ4D    2.
@01218   DLTCdQ4F    2.
@01220   DLTCdQ4H    2.
@01222   DLTCdQ4I    2.
@01224   DLTCdQ4P    2.
@01226   DLTCdQ05    1.
@01227   DLTCQ06B    2.
@01229   DLTCdQ7A    2.
@01231   DLTCdQ7C    2.
@01233   DLTCdQ7D    2.
@01235   DLTCdQ7E    2.
@01237   DLTCdQ11    2.
@01239   DLTCdQ12    2.
@01241   DLTCdQ13    2.
@01243   DLTCdQ14    2.
@01245   DCMCdQ1A    1.
@01246   DCMCdQ1B    1.
@01247   DCMCdQ1C    1.
@01248   DCMCdQ1E    1.
@01249   DCMCdQ1H    1.
@01250   DCMCdQ1I    1.
@01251   DACCd2AA    1.
@01252   DACCQ2AB    1.
@01253   DACCQ2AC    1.
@01254   DACCQ2AE    1.
@01255   DACCd2AH    1.
@01256   DACCd2AI    1.
@01257   DACCd2AJ    1.
@01258   DACCQ2B    2.
@01260   DACCQ3A    2.
@01262   DACCb3AA    2.
@01264   DACCQ3B    2.
@01266   DACCQ3C    2.
@01268   DACCQ3D    2.
@01270   DACCcQ4A    2.
@01272   DACCcQ4B    4.1
@01276   DACCbQ4C    2.
@01278   DACCQ5    1.
@01279   DACCQ6A    1.
@01280   DACCQ6B    1.
@01281   DACCQ6C    1.
@01282   DACCQ6D    1.
@01283   DACCQ6E    1.
@01284   DACCQ6F    1.
@01285   DACCQ7A    1.
@01286   DACCQ7B    2.
@01288   DACCQ8A    1.
@01289   DACCQ8B    2.
@01291   DACCdQ9A    2.
@01293   DACCdQ9C    2.
@01295   DACCdQ9D    2.
@01297   DACCdQ9E    2.
@01299   DACCdQ9F    2.
@01301   DACCdQ9G    2.
@01303   DACCd10A    2.
@01305   DACCd10C    2.
@01307   DACCd10D    2.
@01309   DACCd10E    2.
@01311   DACCd10F    2.
@01313   DACCd10G    2.
@01315   DACCS6    2.
@01317   DBECQ5    1.
@01318   DBECQ5A    2.
@01320   DBECQ6A    1.
@01321   DBECdQ6B    1.
@01322   DBECQ6C    1.
@01323   DBECQ6D    1.
@01324   DBECQ6E    1.
@01325   DBECQ8E1    1.
@01326   DBECdQ6F    1.
@01327   DBECQ6G    1.
@01328   DBECQ6H    1.
@01329   DBECQ6I    1.
@01330   DBECQ6J    1.
@01331   DBECQ8J1    1.
@01332   DBECQ6K    1.
@01333   DBECQ6L    1.
@01334   DBECQ6M    1.
@01335   DBECQ6P    1.
@01336   DBECQ6Q    1.
@01337   DBECQ6R    1.
@01338   DBECQ6R1    1.
@01339   DBECQ6S    1.
@01340   DBECQ6T    1.
@01341   DBECQ8T1    1.
@01342   DBECQ6U    1.
@01343   DBECQ6V    1.
@01344   DBECQ6W    1.
@01345   DBECQ6X    1.
@01346   DBECQ6Z    1.
@01347   DBECQ8Z1    1.
@01348   DBECQ6AA    1.
@01349   DBECQ6BB    1.
@01350   DBECQ6CC    1.
@01351   DBECQ6DD    1.
@01352   DBEC6DD1    1.
@01353   DBECQ6FF    1.
@01354   DBECQ6GG    1.
@01355   DBECd6JJ    1.
@01356   DBECQ6LL    1.
@01357   DBEC8LL1    1.
@01358   DBECQ6MM    1.
@01359   DBECQ6NN    1.
@01360   DBECQ6OO    1.
@01361   DBECQ6PP    1.
@01362   DBEC8PP1    1.
@01363   DBECQ6QQ    1.
@01364   DBECQ8HH    1.
@01365   DBECQ6RR    1.
@01366   DBECQ6SS    1.
@01367   DBECQ6TT    1.
```

```
@01368   DBEC8TT1    1.
@01369   DBECc6UU    1.
@01370   DBECQ7A    1.
@01371   DBECQ7B    1.
@01372   DBECcQ7C    1.
@01373   DBECcQ7D    1.
@01374   DBECQ7E    1.
@01375   DBECQ7F    1.
@01376   DBECdS01    2.
@01378   DBECdS03    2.
@01380   DBECS04    2.
@01382   DBECS05    2.
@01384   DBECdS06    2.
@01386   DBECdS07    2.
@01388   DBECdS08    2.
@01390   DBECdS09    2.
@01392   DBECS10    2.
@01394   DBECdS11    2.
@01396   DPBCdQ1A    1.
@01397   DPBCdQ1B    1.
@01398   DPBCdQ1C    1.
@01399   DPBCdQ1D    1.
@01400   DPBCdQ1E    1.
@01401   DPBCdQ1F    1.
@01402   DPBCdQ1G    1.
@01403   DPBCdQ1H    1.
@01404   DPBCdQ1I    1.
@01405   DPBCdQ1N    1.
@01406   DPBCdQ1O    1.
@01407   DPBCdQ1P    1.
@01408   DPBCdQ1Q    1.
@01409   DPBCdQ1R    1.
@01410   DSLCdQ1    2.
@01412   DSLCdQ2    2.
@01414   DSLCdQ3B    2.
@01416   DSLCdQ4    2.
@01418   DSLCDQ5    1.
@01419   DSLCdQ6    1.
@01420   DSLCdQ7    2.
@01422   DMSCQ01    1.
@01423   DMSCQ02    1.
@01424   DMSCQ03    1.
@01425   DMSCQ04    1.
@01426   DMSCQ05    1.
@01427   DMSCQ06    1.
@01428   DMSCQ07    1.
@01429   DMSCQ08    1.
@01430   DMSCQ09    1.
@01431   DMSCQ10    1.
@01432   DMSCQ11    1.
@01433   DMSCQ12    1.
@01434   DMSCQ13    1.
@01435   DMSCQ14    1.
@01436   DMSCQ15    1.
@01437   DMSCQ16    1.
@01438   DMSCQ17    1.
@01439   DMSCQ18    1.
@01440   DMSCQ19    1.
@01441   DMSCQ20    1.
@01442   DMSCQ21    1.
@01443   DMSCQ22    1.
@01444   DMSCQ23    1.
@01445   DMSCQ24    1.
@01446   DMSCQ25    1.
@01447   DMSCQ26    1.
@01448   DMSCQ27    1.
@01449   DMSCQ28    1.
@01450   DMSCQ29    1.
@01451   DMSCQ30    1.
@01452   DMSCQ31    1.
@01453   DMSCQ32    1.
@01454   DMSCQ33    1.
@01455   DMSCQ34    1.
@01456   DMSCQ35    1.
@01457   DMSCQ36    1.
@01458   DMSCQ37    1.
@01459   DMSCQ38    1.
@01460   DMSCQ39    1.
@01461   DMSCQ40    1.
@01462   DMSCQ41    1.
@01463   DMSCQ42    1.
@01464   DMSCQ43    1.
@01465   DMSCQ44    1.
@01466   DMSCQ45    1.
@01467   DMSCQ46    1.
@01468   DMSCQ47    1.
@01469   DMSCQ48    1.
@01470   DMSCS01    2.
@01472   DMSCS02    3.
@01475   DMSCdS03    3.
@01478   DRLCQ01    2.
@01480   DRLCQ02    2.
@01482   DRLCQ04    1.
@01483   DRLCQ06    2.
@01485   DRLCdQ07    2.
@01487   DRLCQ08    2.
@01489   DRLCQ09    2.
@01491   DPRCQ01    2.
@01493   DPRCQ02    2.
@01495   DPRCQ03    2.
@01497   DPRCQ04    2.
@01499   DPRCQ05    2.
@01501   DPRCQ06    2.
@01503   DPRCQ07    2.
@01505   DPRCQ08    2.
@01507   DPRCQ09    2.
@01509   DPRCQ10    2.
@01511   DPRCQ11    2.
@01513   DPRCQ12    2.
@01515   DPRCQ13    2.
@01517   DPRCQ14    2.
@01519   DPRCQ15    2.
@01521   DPRCQ16    2.
@01523   DPRCQ17    2.
@01525   DPRCQ18    2.
@01527   DPRCQ19    2.
```

```
@01529   DPRCQ20    2.
@01531   DPRCQ21    2.
@01533   DPRCQ22    2.
@01535   DPRCQ23    2.
@01537   DPRCQ24    2.
@01539   DPRCQ25    2.
@01541   DPRCQ25A    2.
@01543   DPRCb30A    2.
@01545   DPRCb30B    2.
@01547   DPRCb30C    2.
@01549   DPRCb30D    2.
@01551   DPRCb30E    2.
@01553   DPRCb30F    2.
@01555   DPRCb30G    2.
@01557   DPRCb30H    2.
@01559   DPRCb30I    2.
@01561   DPRCb30J    2.
@01563   DPRCQ26A    1.
@01564   DPRCQ26B    1.
@01565   DPRCQ26C    1.
@01566   DPRCQ26D    1.
@01567   DPRCQ26E    1.
@01568   DPRCQ26F    1.
@01569   DPRCQ26G    1.
@01570   DPRCQ26H    1.
@01571   DPRCQ26I    1.
@01572   DPRCQ26J    1.
@01573   DPRCQ26K    1.
@01574   DPRCQ27    1.
@01575   DPRCQ28    1.
@01576   DRLCQ05    1.
@01577   DRLCQ03    2.
@01579   DPRCS01    2.
@01581   DPRCS02    2.
@01583   DPRCS03    2.
@01585   DPRCS04    2.
@01587   DPRCS05    2.
@01589   DPRCS06    2.
@01591   DPRCbS09    2.
@01593   DCRCcQ1A    1.
@01594   DCRCQ1B    1.
@01595   DCRCQ1BA    3.
@01598   DCRCQ1BB    1.
@01599   DCRCQ1C    1.
@01600   DCRCQ1CA    3.
@01603   DCRCQ1CB    1.
@01604   DCRCQ1E    1.
@01605   DCRCQ1EA    3.
@01608   DCRCQ1EB    1.
@01609   DCRCQ1F    1.
@01610   DCRCQ1FA    3.
@01613   DCRCQ1FB    1.
@01614   DCRCQ1G    1.
@01615   DCRCQ1GA    3.
@01618   DCRCQ1GB    1.
@01619   DCRCQ1H    1.
@01620   DCRCQ1HA    3.
@01623   DCRCcQ1K    1.
@01624   DCRCc1KA    3.
@01627   DCRCcQ1J    1.
@01628   DCRCc1JA    3.
@01631   DCRCQ1D    1.
@01632   DCRCQ1DA    3.
@01635   DCRCcQ1I    1.
@01636   DCRCQ1IA    3.
@01639   DCRCcQ2A    2.
@01641   DCRCd2BB    1.
@01642   DCRCd10A    1.
@01643   DCRCd10B    1.
@01644   DCRCd10C    1.
@01645   DCRCd10D    1.
@01646   DCRCd10E    1.
@01647   DCRCd10F    1.
@01648   DCRCd10G    1.
@01649   DCRCd10H    1.
@01650   DCRCd10I    1.
@01651   DCRCc2AA    2.
@01653   DCRCc2AB    2.
@01655   DCRCc2AC    2.
@01657   DCRCc2AD    1.
@01658   DCRCc2AE    1.
@01659   DCRCc2AF    1.
@01660   DCRCc2AG    1.
@01661   DCRCc2AH    1.
@01662   DCRCc2AI    1.
@01663   DCRCc2CA    3.
@01666   DCRCc2CB    3.
@01669   DCRCQ03    2.
@01671   DCRCQ04    2.
@01673   DCRCc5AA    2.
@01675   DCRCQ05A    1.
@01676   DCRCQ05B    1.
@01677   DCRCQ05C    1.
@01678   DCRCQ05D    1.
@01679   DCRCQ05E    1.
@01680   DCRCQ05F    1.
@01681   DCRCQ05G    1.
@01682   DCRCQ06    1.
@01683   DCRCdQ6B    2.
@01685   DCRCQ07    2.
@01687   DCRCQ08A    1.
@01688   DCRCQ08B    1.
@01689   DCRCQ08C    1.
@01690   DCRCQ08D    1.
@01691   DCRCQ08E    1.
@01692   DCRCQ08F    1.
@01693   DCRCQ08G    1.
@01694   DCRCQ08H    1.
@01695   DCRCbQ8I    1.
@01696   DCRCbQ8J    1.
@01697   DCRCbQ8K    1.
@01698   DCRCcQ09    1.
@01699   DCRCD02    3.
@01702   DCRCD03    2.
@01704   DCRCD04    3.
@01707   DCRCD05    1.
```

```
@01708   DCRCD06    3.
@01711   DASCdQ01    1.
@01712   DASCdQ02    1.
@01713   DASCdQ03    1.
@01714   DASCdQ04    2.
@01716   DASCdQ05    2.
@01718   DASCdQ6A    1.
@01719   DASCdQ6B    1.
@01720   DASCdQ6C    1.
@01721   DASCdQ6D    1.
@01722   DASCdQ6E    1.
@01723   DASCdQ6F    1.
@01724   DASCdQ6G    1.
@01725   DASCdQ6H    1.
@01726   DASCdQ6I    1.
@01727   DASCdq7    1.
@01728   DASCdq9    2.
@01730   DASCd10A    1.
@01731   DASCd10B    1.
@01732   DASCd10C    1.
@01733   DASCd10D    1.
@01734   DASCd10E    1.
@01735   DASCd10F    1.
@01736   DASCd10G    1.
@01737   DASCd11A    1.
@01738   DASCd11B    1.
@01739   DASCd11C    1.
@01740   DASCd11D    1.
@01741   DASCd11E    1.
@01742   DASCd11F    1.
@01743   DASCdQ12    1.
@01744   DASCdQ13    2.
@01746   DASCdQ14    2.
@01748   DASCdQ15    2.
@01750   DASCdQ16    1.
@01751   DSDCQ1    2.
@01753   DSDCQ2AA    1.
@01754   DSDCQ2AB    1.
@01755   DSDCQ2AC    1.
@01756   DSDCQ2AD    1.
@01757   DSDCQ2B    1.
@01758   DSDCQ3    4.
@01762   DSDCQ4A    1.
@01763   DSDCQ4B    1.
@01764   DSDCQ4C    1.
@01765   DSDCQ4D    1.
@01766   DSDCQ4E    1.
@01767   DSDCQ4F    1.
@01768   DSDCQ4G    1.
@01769   DSDCQ4H    1.
@01770   DSDCQ4I    1.
@01771   DSDCQ4J    1.
@01772   DSDCQ4K    1.
@01773   DSDCQ4L    1.
@01774   DSDCQ4M    1.
@01775   DSDCQ4N    1.
@01776   DSDCQ4O    1.
@01777   DSDCQ4P    1.
@01778   DSDCQ4Q    1.
@01779   DSDCQ4R    1.
@01780   DSDCQ4S    1.
@01781   DSDCb4AA    1.
@01782   DSDCb4AB    1.
@01783   DSDCb4AC    1.
@01784   DSDCb4AD    1.
@01785   DSDCb4AE    1.
@01786   DSDCb4AF    1.
@01787   DSDCb4AG    1.
@01788   DSDCb4AH    1.
@01789   DSDCb4AI    1.
@01790   DSDCb4AJ    1.
@01791   DSDCb4AK    1.
@01792   DSDCb4AL    1.
@01793   DSDCQ5A    1.
@01794   DSDCQ5B    1.
@01795   DSDCQ5C    1.
@01796   DSDCQ5D    1.
@01797   DSDCQ5E    1.
@01798   DSDCQ5F    1.
@01799   DSDCQ5G    1.
@01800   DSDCQ5H    1.
@01801   DSDCQ5I    1.
@01802   DSDCQ5J    1.
@01803   DSDCQ5K    1.
@01804   DSDCQ5L    1.
@01805   DSDCQ5M    1.
@01806   DSDCQ5N    1.
@01807   DSDCQ5O    1.
@01808   DSDCQ5P    1.
@01809   DSDCQ5Q    1.
@01810   DSDCQ5R    1.
@01811   DSDCQ5S    1.
@01812   DSDCQ6A    1.
@01813   DSDCQ6B    1.
@01814   DSDCQ6C    1.
@01815   DSDCQ6D    1.
@01816   DSDCQ6E    1.
@01817   DSDCQ6F    1.
@01818   DSDCQ6G    1.
@01819   DSDCQ6H    1.
@01820   DSDCQ6I    1.
@01821   DSDCQ6J    1.
@01822   DSDCQ6K    1.
@01823   DSDCQ6L    1.
@01824   DSDCQ6M    1.
@01825   DSDCQ6N    1.
@01826   DSDCQ6O    1.
@01827   DSDCQ6P    1.
@01828   DSDCQ6Q    1.
@01829   DSDCQ6R    1.
@01830   DSDCQ6S    1.
@01831   DSDCC6AA    1.
@01832   DSDCC6AB    1.
@01833   DSDCC6AC    1.
@01834   DSDCQ8    2.
@01836   DSDCD01    3.
```

```
@01839   DSDCD02    3.
@01842   DSDCD02B    1.
@01843   DSDCD03    2.
@01845   DSDCD04    2.
@01847   DSDCD05    2.
@01849   DSDCD05B    1.
@01850   DSDCD06    2.
@01852   DSDCD06B    1.
@01853   DAGCdS01    2.
@01855   DAGCdS02    2.
@01857   DAGCdS03    2.
@01859   DAGCdS04    2.
@01861   DAGCdS05    2.
@01863   DEDYdQ01    1.
@01864   DEDYdQ02    1.
@01865   DEDYdQ04    1.
@01866   DEDYdQ06    2.
@01868   DEDYdQ07    2.
@01870   DEDYdQ08    1.
@01871   DEDYdQ09    1.
@01872   DEDYd10A    1.
@01873   DEDYd10B    1.
@01874   DEDYd10C    1.
@01875   DEDYd10D    1.
@01876   DEDYd10E    1.
@01877   DEDYdQ11    1.
@01878   DEDYdQ12    2.
@01880   DEDYdQ13    1.
@01881   DEDYdQ14    2.
@01883   DEDYdQ15    2.
@01885   DEDYdQ16    2.
@01887   DEDYdQ17    2.
@01889   DEDYdQ18    2.
@01891   DEDYdQ19    2.
@01893   DEDYdQ20    1.
@01894   DEDYdQ21    2.
@01896   DEDYdQ24    2.
@01898   DEDYdQ25    2.
@01900   DEDYdQ26    2.
@01902   DEDYdQ27    2.
@01904   DEDYd27A    4.
@01908   DEDYdQ28    2.
@01910   DEDYdQ29    1.
@01911   DEDYdQ30    2.
@01913   DEDYdQ31    1.
@01914   DEDYdQ32    2.
@01916   DEDYdQ33    2.
@01918   DEDYdQ34    2.
@01920   DEDYdQ35    1.
@01921   DEDYd36A    1.
@01922   DEDYd36B    1.
@01923   DEDYd36C    1.
@01924   DEDYd36D    1.
@01925   DEDYd36E    1.
@01926   DEDYd36F    1.
@01927   DEDYd36G    1.
@01928   DEDYd36H    1.
@01929   DEDYdQ37    1.
@01930   DEDYdQ38    1.
@01931   DEDYd39A    1.
@01932   DEDYd39B    1.
@01933   DEDYd39C    1.
@01934   DEDYd39D    1.
@01935   DEDYd39E    1.
@01936   DEDYdQ40    2.
@01938   DEDYdQ41    2.
@01940   DEDYdQ42    2.
@01942   DEDYdQ43    2.
@01944   DEDYdQ44    2.
@01946   DEDYdQ47    1.
@01947   DEDYdQ48    1.
@01948   DEDYdQ49    1.
@01949   DEDYdQ50    2.
@01951   DEDYdQ51    2.
@01953   DEDYdQ52    2.
@01955   DEDYd52A    2.
@01957   DEDYd52B    2.
@01959   DEDYd52C    2.
@01961   DEDYd52D    2.
@01963   DEDYd53A    1.
@01964   DEDYd53B    1.
@01965   DEDYd53C    1.
@01966   DEDYd53D    1.
@01967   DEDYd53E    1.
@01968   DEDYd53F    1.
@01969   DEDYd53G    1.
@01970   DEDYd53H    1.
@01971   DEDYd54A    1.
@01972   DEDYd54B    1.
@01973   DEDYd54C    1.
@01974   DEDYd54D    1.
@01975   DEDYd54E    1.
@01976   DEDYdQ55    2.
@01978   DEDYdQ56    1.
@01979   DEDYdQ57    1.
@01980   DEDYdQ58    2.
@01982   DEDYdQ59    2.
@01984   DEDYdQ60    2.
@01986   DEDYdQ61    2.
@01988   DEDYdQ62    1.
@01989   DEDYdQ63    1.
@01990   DEDYd64A    1.
@01991   DEDYd64B    1.
@01992   DEDYd64C    1.
@01993   DEDYd64D    1.
@01994   DEDYd64E    1.
@01995   DEDYdQ65    1.
@01996   DEDYdQ66    2.
@01998   DEDYdQ67    1.
@01999   DEDYdQ68    2.
@02001   DEDYdQ69    2.
@02003   DEDYdQ70    2.
@02005   DEDYdQ71    1.
@02006   DEDYdQ72    2.
@02008   DEDYdQ73    1.
@02009   DEDYdQ74    2.
```

```
@02011   DEDYdQ75    2.
@02013   DEDYdQ76    2.
@02015   DEDYdQ77    2.
@02017   DEDYdQ78    2.
@02019   DEDYdQ81    2.
@02021   DEDYdQ82    2.
@02023   DEDYdQ83    2.
@02025   DEDYdQ84    2.
@02027   DEDYdQ85    2.
@02029   DEDYdQ86    1.
@02030   DEDYdQ87    2.
@02032   DEDYdQ88    1.
@02033   DEDYdQ89    2.
@02035   DEDYdQ90    1.
@02036   DEDYdQ91    1.
@02037   DEDYdQ92    2.
@02039   DEDYdQ93    2.
@02041   DEDYdQ94    2.
@02043   DEDYdQ95    2.
@02045   DEDYdQ96    2.
@02047   DEDYdQ97    2.
@02049   DEDYdQ98    2.
@02051   DEDYdQ99    2.
@02053   DEDYd102    2.
@02055   DEDYd104    2.
@02057   DEDYd105    2.
@02059   DEDYd106    1.
@02060   DEDYd107    1.
@02061   DEDY108A    1.
@02062   DEDY108B    1.
@02063   DEDY108C    1.
@02064   DEDY108D    1.
@02065   DEDY108E    1.
@02066   DEDYd109    1.
@02067   DEDYd110    2.
@02069   DEDYd111    1.
@02070   DEDYd112    2.
@02072   DEDYd113    2.
@02074   DLFYdQ1A    1.
@02075   DLFYdQ1B    1.
@02076   DLFYdQ1C    1.
@02077   DLFYdQ1D    1.
@02078   DLFYdQ2A    2.
@02080   DLFYdQ3A    2.
@02082   DLFYdQ3B    2.
@02084   DLFYdQ4A    1.
@02085   DLFYdQ4B    1.
@02086   DLFYdQ4C    1.
@02087   DLFYdQ5    1.
@02088   DLFYdQ5A    7.2
@02095   DLFYdQ6    1.
@02096   DLFYdQ7A    1.
@02097   DLFYdQ7B    1.
@02098   DLFYdQ7C    1.
@02099   DLFYdQ7D    1.
@02100   DLFYdQ8A    2.
@02102   DLFYdQ8B    2.
@02104   DLFYdQ9a    1.
@02105   DLFYdQ9b    1.
@02106   DLFYdQ9c    1.
@02107   DLFYdQ10    1.
@02108   DLFYdQ11    2.
@02110   DLFYdQ12    2.
@02112   DLFYd13A    1.
@02113   DLFYd13B    1.
@02114   DLFYd13C    1.
@02115   DLFYd13D    1.
@02116   DLFYdQ14    2.
@02118   DLFYdQ15    2.
@02120   DINYdQ1A    6.
@02126   DINYdQ1B    6.
@02132   DINYdQ1C    6.
@02138   DINYdQ1D    6.
@02144   DINYdD01    6.
@02150   DINYdD1a    2.
@02152   DINYdi1a    1.
@02153   DHLYdQ01    2.
@02155   DHLYdQ2    2.
@02157   DHLYdQ3A    2.
@02159   DHLYdQ3B    2.
@02161   DHLYdQ3C    2.
@02163   DHLYdQ3D    2.
@02165   DHLYdQ4    1.
@02166   DHLYdQ4B    2.
@02168   DHLYdQ5    2.
@02170   DHLYdQ6    1.
@02171   DHLYdQ7    1.
@02172   DHLYdQ8    2.
@02174   DHLYdQ9    2.
@02176   DHLYdQ10    2.
@02178   DHLYdQ11    2.
@02180   DHLYdQ12    2.
@02182   DHLYd12A    1.
@02183   DHLYd12B    1.
@02184   DHLYd12C    1.
@02185   DHLYd12D    1.
@02186   DHLYd12E    1.
@02187   DHLYd13A    1.
@02188   DHLYd13B    1.
@02189   DHLYd13C    1.
@02190   DHLYdQ14    1.
@02191   DHLYd15A    1.
@02192   DHLYd15B    1.
@02193   DHLYd15C    1.
@02194   DHLYd15D    1.
@02195   DHLYd15E    1.
@02196   DHLYd15F    1.
@02197   DHLYd15G    1.
@02198   DHLYd15H    1.
@02199   DHLYd15I    1.
@02200   DHLYd15J    1.
@02201   DHLYd15K    1.
@02202   DHLYd15L    1.
@02203   DHLYd15M    1.
@02204   DHLYd15N    1.
@02205   DACYdQ1A    1.
```

```
@02206   DACYdQ1B    1.
@02207   DACYdQ1C    1.
@02208   DACYdQ1D    1.
@02209   DACYdQ1E    1.
@02210   DACYdQ1F    1.
@02211   DACYdQ02    1.
@02212   DACYdQ3A    1.
@02213   DACYdQ3B    1.
@02214   DACYdQ3C    1.
@02215   DACYdQ3D    1.
@02216   DACYdQ3E    1.
@02217   DACYdQ4A    2.
@02219   DACYdQ4B    2.
@02221   DACYdQ4C    2.
@02223   DACYdQ4D    2.
@02225   DACYdQ5    2.
@02227   DACYdQ6    1.
@02228   DACYdQ7    1.
@02229   DACYdQ8    2.
@02231   DACYdQ9    1.
@02232   DACYdQ10    1.
@02233   DACYd12A    1.
@02234   DACYd12B    1.
@02235   DACYd12C    1.
@02236   DACYd12D    1.
@02237   DACYd12E    1.
@02238   DACYd12F    2.
@02240   DACYd13A    1.
@02241   DACYd13B    1.
@02242   DACYd13C    1.
@02243   DACYd13D    1.
@02244   DACYd13E    1.
@02245   DACYd13F    1.
@02246   DACYd13G    1.
@02247   DACYdQ14    2.
@02249   DACYdS01    2.
@02251   DMACS01    2.
@02253   DMACS02    3.
@02256   DMACS03    3.
@02259   DMAYdS01    6.3
@02265   DPPCS01    3.
@02268   DPPCS02    3.
@02271   DPPCD01    3.
@02274   DPPCD02    1.
@02275   DOBCdQ1A    1.
@02276   DOBCdQ1B    1.
@02277   DOBCdQ1C    1.
@02278   DOBCdQ1D    1.
@02279   DOBCdQ1E    1.
@02280   DOBCdQ1F    1.
@02281   DOBCdQ1G    1.
@02282   DOBCdQ1H    2.
@02284   DOBCdQ2A    1.
@02285   DOBCdQ2B    1.
@02286   DOBCdQ2C    1.
@02287   DOBCdQ2D    1.
@02288   DOBCdQ2E    1.
@02289   DOBCdQ3A    1.
@02290   DOBCdQ3B    1.
@02291   DWICdS01    2.
@02293   DWICdS02    2.
@02295   DWICdS03    2.
@02297   DKNCdS01    1.
@02298   DEDCbZGD    1.
@02299   DCRCbZQ6    1.
@02300   DCRCbZQ7    1.
@02301   MEMCYCLE    1.
@02302   OUTFLAG    1.
@02303   XSECFLG    1.
@02304   LONGFLG    1.
@02305   FIELDRUK   $12.
@02317   PERSRUK   $14.
@02331   CHILDID    6.
@02337   DWTCW01C    12.4
@02349   DWTCW01L    12.4
@02361   DWTCWd1L    12.4;
```

2. Self Completes Record Layout

```
@00001   DFFCQ01    2.
@00003   DFFCQ02    2.
@00005   DFFCQ03    2.
@00007   DFFCQ04    2.
@00009   DFFCcQ4A    2.
@00011   DFFCQ05    2.
@00013   DFFCQ06    2.
@00015   DFFCQ07    1.
@00016   DFFCQ08A    1.
@00017   DFFCQ08B    1.
@00018   DFFCQ08C    1.
@00019   DFFCQ08D    1.
@00020   DFFCQ08E    1.
@00021   DFFCQ08F    1.
@00022   DFFCQ08G    1.
@00023   DFFCQ08H    1.
@00024   DFFCc08I    1.
@00025   DFFCQ08J    1.
@00026   DFFCQ08K    1.
@00027   DFFCQ08L    1.
@00028   DFFCQ08M    1.
@00029   DFFCQ08N    1.
@00030   DFFCQ09    2.
@00032   DFFCQ10    2.
@00034   DFFCQ11    2.
@00036   DFFCQ12    2.
@00038   DFFCd12A    2.
@00040   DFFCbQ13    2.
```

```
@00042   DFFCbQ14    2.
@00044   DFFCbQ15    2.
@00046   DDRCdQ05    1.
@00047   DDRCdQ09    1.
@00048   DFFCd16C    1.
@00049   DFFCc16D    1.
@00050   DFFCc16E    1.
@00051   DFFCc19A    1.
@00052   DFFCc19B    1.
@00053   DFFCc19C    1.
@00054   DFFCc19D    1.
@00055   DFFCc20A    1.
@00056   DFFCc20B    1.
@00057   DFFCc20C    1.
@00058   DSCCQ01    2.
@00060   DSCCQ02    2.
@00062   DSCCcQ03    1.
@00063   DSCCcQ3a    1.
@00064   DSCCcQ3b    1.
@00065   DSCCcQ3C    1.
@00066   DSCCcQ3D    1.
@00067   DSCCcQ3E    1.
@00068   DSCCcQ3F    1.
@00069   DSCCd3G    1.
@00070   DSCCQ05    2.
@00072   DSCCQ06    2.
@00074   DSCCcQ07    2.
@00076   DSCCQ08    2.
@00078   DSCCQ09    2.
@00080   DSCCcQ10    2.
@00082   DSCCbQ18    1.
@00083   DSCCb19A    1.
@00084   DSCCb19B    1.
@00085   DSCCb19C    1.
@00086   DSCCb19D    1.
@00087   DSCCb19E    1.
@00088   DSCCc19F    1.
@00089   DSCCb20A    1.
@00090   DSCCb20B    1.
@00091   DSCCb20C    1.
@00092   DSCCb20E    1.
@00093   DSCCc20H    1.
@00094   DSCCb20G    1.
@00095   DSCCc20I    1.
@00096   DSCCc20J    1.
@00097   DSCCc21A    2.
@00099   DSCCc21B    2.
@00101   DSCCc21C    2.
@00103   DSCCc21D    2.
@00105   DSCCc21E    2.
@00107   DSCCc21F    2.
@00109   DSCCbQ22    1.
@00110   DSCCc22A    1.
@00111   DSCCbQ23    1.
@00112   DSCCcQ24    2.
@00114   DSCCQ11    2.
@00116   DSCCQ12    2.
@00118   DSCCQ13    2.
@00120   DSCCQ14    2.
@00122   DSCCQ15    2.
@00124   DSCCcQ16    2.
@00126   DSCCQ17    2.
@00128   DSCCcQ26    1.
@00129   DSCCcQ27    1.
@00130   DSCCc31A    1.
@00131   DSCCc31B    1.
@00132   DSCCc31C    1.
@00133   DSCCc31D    1.
@00134   DSCCc31E    1.
@00135   DSCCcQ28    1.
@00136   DSCCcQ29    1.
@00137   DSCCcQ30    2.
@00139   DAMCQ01A    2.
@00141   DAMCQ01B    2.
@00143   DAMCQ01C    2.
@00145   DAMCQ01D    2.
@00147   DAMCcQ1E    2.
@00149   DAMCbQ02    1.
@00150   DAMCcQ03    1.
@00151   DAMCcQ4A    1.
@00152   DAMCcQ4B    1.
@00153   DAMCcQ4C    1.
@00154   DAMCcQ4D    1.
@00155   DAMCdQ4F    1.
@00156   DAMCcQ5A    1.
@00157   DAMCcQ5B    1.
@00158   DAMCcQ5C    1.
@00159   DAMCcQ5D    1.
@00160   DAMCcQ6A    1.
@00161   DAMCcQ6B    1.
@00162   DAMCcQ7A    1.
@00163   DAMCcQ7B    1.
@00164   DAMCcQ8A    1.
@00165   DAMCcQ8B    1.
@00166   DAMCdQ09    1.
@00167   DAMCdQ10    1.
@00168   DFBCQ01A    1.
@00169   DFBCc01B    1.
@00170   DFBCQ01C    1.
@00171   DFBCQ01D    1.
@00172   DFBCQ01E    1.
@00173   DFBCQ01F    1.
@00174   DFBCQ01G    1.
@00175   DFBCQ01H    1.
@00176   DFBCQ01I    1.
@00177   DFBCQ01J    1.
@00178   DFBCQ01K    1.
@00179   DFBCQ01L    1.
@00180   DFBCQ01M    1.
@00181   DFBCQ01P    1.
@00182   DFBCQ01Q    1.
@00183   DFBCQ01R    1.
@00184   DFBCQ01S    1.
@00185   DFBCQ01T    1.
@00186   DFBCQ01U    1.
@00187   DFBCc01V    1.
```

```
@00188   DFBCQ01W    1.
@00189   DFBCd01X    1.
@00190   DFBCQ01Z    1.
@00191   DFBCQ1AA    1.
@00192   DFBCQ1BB    1.
@00193   DFBCQ1CC    1.
@00194   DFBCQ1DD    1.
@00195   DFBCQ1FF    1.
@00196   DFBCQ1GG    1.
@00197   DFBCQ1HH    1.
@00198   DFBCQ1JJ    1.
@00199   DFBCQ1LL    1.
@00200   DFBCQ1MM    1.
@00201   DFBCQ1NN    1.
@00202   DFBCQ1OO    1.
@00203   DFBCQ1PP    1.
@00204   DFBCQ1QQ    1.
@00205   DFBCQ1RR    1.
@00206   DFBCQ1SS    1.
@00207   DFBCQ1TT    1.
@00208   DFBCc1UU    1.
@00209   DFBCbQ2A    1.
@00210   DFBCbQ2B    1.
@00211   DFBCbQ2C    1.
@00212   DFBCbQ2D    1.
@00213   DFBCbQ2E    1.
@00214   DFBCbQ2F    1.
@00215   DFBCbQ2G    1.
@00216   DFBCcQ2H    1.
@00217   DFBCbQ2J    1.
@00218   DFBCbQ2K    1.
@00219   DFBCbQ2N    1.
@00220   DFBCbQ2O    1.
@00221   DFBCbQ2P    1.
@00222   DFBCbQ2Q    1.
@00223   DFBCdQ2Z    1.
@00224   DFBCcQ2R    1.
@00225   DFBCcQ2T    1.
@00226   DFBCbQ2V    1.
@00227   DFBCbQ2Y    1.
@00228   DFBCb2AA    1.
@00229   DFBCdQ3B    1.
@00230   DFBCb2BB    1.
@00231   DFBCc2CC    1.
@00232   DFBCb2EE    1.
@00233   DFBCc2FF    1.
@00234   DFBCcQ3A    1.
@00235   DFBCQ03    1.
@00236   DFBCcQ04    1.
@00237   DFBCcQ4A    1.
@00238   DFBCcQ05    1.
@00239   DFBCcQ07    1.
@00240   DFBCcQ08    1.
@00241   DFBCd10A    1.
@00242   DFBCd10B    1.
@00243   DFBCd10C    1.
@00244   DFBCd10D    1.
@00245   DFBCd10E    1.
@00246   DFBCd10F    1.
@00247   DFBCd10G    1.
@00248   DFBCd10H    1.
@00249   DFBCd10I    1.
@00250   DFBCd10J    1.
@00251   DFBCd10K    1.
@00252   DFBCd10L    1.
@00253   DPMCdQ04    2.
@00255   DPMCcQ5A    1.
@00256   DPMCcQ5B    1.
@00257   DPMCcQ5C    1.
@00258   DPMCcQ06    1.
@00259   DPMCdQ6A    2.
@00261   DPMCdQ6B    2.
@00263   DPMCdQ6C    2.
@00265   DPMCdQ6D    2.
@00267   DPMCdQ6E    2.
@00269   DPMCdQ6F    2.
@00271   DPMCdQ6G    2.
@00273   DPMCdQ6H    2.
@00275   DPMCdQ6I    2.
@00277   DPMCdQ6J    2.
@00279   DPMCdQ6K    2.
@00281   DPMCdQ6L    2.
@00283   DPMCdQ07    2.
@00285   DPMCcQ8A    1.
@00286   DPMCcQ8B    1.
@00287   DPMCcQ8C    1.
@00288   DPMCcQ09    1.
@00289   DPMCdQ9A    2.
@00291   DPMCdQ9B    2.
@00293   DPMCdQ9C    2.
@00295   DPMCdQ9D    2.
@00297   DPMCdQ9E    2.
@00299   DPMCdQ9F    2.
@00301   DPMCdQ9G    2.
@00303   DPMCdQ9H    2.
@00305   DPMCdQ9I    2.
@00307   DPMCdQ9J    2.
@00309   DPMCdQ9K    2.
@00311   DPMCdQ9L    2.
@00313   DPMCcQ10    1.
@00314   DPMCd11A    2.
@00316   DPMCc11B    2.
@00318   DPMCcQ1A    2.
@00320   DPMCcQ1B    2.
@00322   DPMCcQ1C    2.
@00324   DPMCcQ1D    2.
@00326   DPMCcQ1E    2.
@00328   DPMCcQ1F    2.
@00330   DPMCcQ1G    2.
@00332   DPMCcQ1H    2.
@00334   DPMCcQ1I    2.
@00336   DPMCcQ1J    2.
@00338   DPMCcQ1K    2.
@00340   DPMCcQ1L    2.
@00342   DPMCcQ1M    2.
@00344   DPMCcQ1N    2.
```

```
@00346   DPMCcQ1O    2.
@00348   DPMCcQ1P    2.
@00350   DPMCcQ1Q    2.
@00352   DPMCcQ1R    2.
@00354   DPMCdQ1S    2.
@00356   DPMCdQ1T    2.
@00358   DPMCdQ9U    2.
@00360   DPMCdQ9V    2.
@00362   DPMCbQ3A    1.
@00363   DPMCbQ3C    1.
@00364   DPMCbQ3E    1.
@00365   DPMCd12A    1.
@00366   DPMCd12B    1.
@00367   DPMCd12C    1.
@00368   DPMCd12D    1.
@00369   DPMCd12E    1.
@00370   DPMCd12F    1.
@00371   DPMCd12G    1.
@00372   DPMCd12H    1.
@00373   DPMCd12I    1.
@00374   DPMCd12J    1.
@00375   DPMCd12K    1.
@00376   DPMCd12L    1.
@00377   DPUCQ01    1.
@00378   DPUCQ02    1.
@00379   DPUCQ03    1.
@00380   DPUCcQ3A    3.
@00383   DPUCdQ04    1.
@00384   DPUCQ05    1.
@00385   DDRCdQ01    2.
@00387   DDRCQ03    2.
@00389   DDRCQ04    2.
@00391   DDRCcQ06    1.
@00392   DDRCdQ6A    2.
@00394   DDRCdQ07    2.
@00396   DDRCdQ9A    1.
@00397   DDRCbQ9B    2.
@00399   DDRCdQ9C    2.
@00401   DDRCdQ15    2.
@00403   DDRCQ10    1.
@00404   DDRCQ12    2.
@00406   DDRCd14A    2.
@00408   DDRCc14B    2.
@00410   DDRCc14C    2.
@00412   DDRCd14F    2.
@00414   DDRCc14D    2.
@00416   DDRCdQ17    2.
@00418   DDRCc18A    2.
@00420   DDRCd18B    2.
@00422   DDRCd18C    2.
@00424   DDRCc18D    2.
@00426   DDRCdQ19    1.
@00427   DDRCdQ20    1.
@00428   DATCc1AA    1.
@00429   DATCbQ1A    1.
@00430   DATCc1BB    1.
@00431   DATCbQ1B    1.
@00432   DATCc1CC    1.
@00433   DATCbQ1C    1.
@00434   DATCc1DD    1.
@00435   DATCbQ1D    1.
@00436   DATCbQ1E    1.
@00437   DATCc1EE    1.
@00438   DATCbQ1F    1.
@00439   DATCbQ1G    1.
@00440   DATCc1II    1.
@00441   DATCbQ1I    1.
@00442   DATCdQ07    2.
@00444   DATCQ03    2.
@00446   DATCbQ04    1.
@00447   DATCdQ5A    1.
@00448   DATCbQ5B    1.
@00449   DATCbQ5C    1.
@00450   DATCbQ5D    1.
@00451   DATCbQ5E    1.
@00452   DATCbQ5F    1.
@00453   DATCbQ5G    1.
@00454   DATCdQ05    2.
@00456   DATCdQ10    2.
@00458   DATCbQ11    2.
@00460   DATCdQ12    1.
@00461   DATCdQ13    1.
@00462   DATCc14A    2.
@00464   DATCc14B    2.
@00466   DATCc14C    2.
@00468   DATCc14D    2.
@00470   DHTCcQ03    2.
@00472   DHTCbQ01    4.2
@00476   DHTCbQ02    7.3
@00483   DHTCbQ3A    2.
@00485   DHTCbQ3B    2.
@00487   DHTCbQ3C    2.
@00489   DHTCbQ3G    2.
@00491   DHTCbQ04    2.
@00493   DHTCbQ05    2.
@00495   DHTCdQ06    1.
@00496   DHTCdQ5A    1.
@00497   DHTCdQ5B    1.
@00498   DHTCdQ5C    1.
@00499   DHTCdQ5D    1.
@00500   DHTCdQ5E    1.
@00501   DHTCdQ5F    1.
@00502   DHTCdQ5G    1.
@00503   DHTCdQ5H    1.
@00504   DHTCdQ5I    1.
@00505   DHTCdQ5J    1.
@00506   DWKCcQ1A    2.
@00508   DWKCcQ1B    2.
@00510   DWKCcQ1C    2.
@00512   DWKCcQ1D    2.
@00514   DWKCbQ02    2.
@00516   DWKCcQ4A    1.
@00517   DWKCcQ4B    1.
@00518   DWKCcQ4C    1.
@00519   DWKCcQ4D    1.
@00520   DWKCcQ5A    1.
```

```
@00521   DWKCcQ5B    1.
@00522   DWKCcQ5C    1.
@00523   DWKCdQ5D    1.
@00524   DWKCcQ06    2.
@00526   DWKCcQ7A    2.
@00528   DWKCcQ7B    2.
@00530   DWKCcQ08    3.
@00533   DWKCcQ9A    2.
@00535   DWKCcQ9B    2.
@00537   DWKCcQ10    1.
@00538   DWKCc11A    2.
@00540   DWKCc11B    2.
@00542   DWKCc11D    2.
@00544   DWKCc11C    2.
@00546   DWKCd12A    1.
@00547   DWKCd12B    1.
@00548   DWKCd12C    1.
@00549   DWKCd12D    1.
@00550   DWKCd13A    1.
@00551   DWKCd13B    1.
@00552   DWKCd13C    1.
@00553   DWKCd13D    1.
@00554   DWKCd13E    1.
@00555   DWKCd13F    1.
@00556   DWKCd13G    1.
@00557   DWKCd13H    1.
@00558   DWKCd13I    1.
@00559   DWKCdQ14    2.
@00561   DWKCdQ15    2.
@00563   DDACcQ1A    2.
@00565   DDACcQ1B    1.
@00566   DDACcQ1C    1.
@00567   DDACcQ02    2.
@00569   DDACcQ2A    2.
@00571   DDACbQ4C    1.
@00572   DDACbQ4D    1.
@00573   DDACbQ4E    1.
@00574   DDACbQ4F    1.
@00575   DDACcQ05    1.
@00576   DDACdQ19    2.
@00578   DDACdQ23    1.
@00579   DDACdQ24    2.
@00581   DDACdQ25    2.
@00583   DDACd22A    1.
@00584   DDACd22B    1.
@00585   DDACd22C    1.
@00586   DDACd22D    1.
@00587   DDACd22E    1.
@00588   DDACd22F    1.
@00589   DDACd22G    1.
@00590   DDACd22H    1.
@00591   DDACcQ06    2.
@00593   DDACcQ07    2.
@00595   DDACcQ08    1.
@00596   DDACcQ09    1.
@00597   DDECdQ01    2.
@00599   DDECdQ02    2.
@00601   DDECdQ03    2.
@00603   DDECdQ04    2.
@00605   DDECdQ05    2.
@00607   DDECdQ06    2.
@00609   DDECdQ07    2.
@00611   DDECdQ08    2.
@00613   DDECdQ09    2.
@00615   DDECdQ10    2.
@00617   DDECdQ11    2.
@00619   DDECdQ12    2.
@00621   DDECdQ13    2.
@00623   DDECdQ14    2.
@00625   DDECdQ15    2.
@00627   DDECdQ16    2.
@00629   DDECdQ17    2.
@00631   DDECdQ18    2.
@00633   DDECdQ19    2.
@00635   DDECdQ20    2.
@00637   DDECdQ21    2.
@00639   DDECdQ22    2.
@00641   DDECdQ23    2.
@00643   DDECdQ24    2.
@00645   DDECdQ25    2.
@00647   DDECdQ26    2.
@00649   DDECdQ27    2.
@00651   DDECdQ28    2.
@00653   DDECdQ29    2.
@00655   DDECdQ30    2.
@00657   DFFCS01     2.
@00659   DAMCS02     2.
@00661   DFBCS01     2.
@00663   DFBCdS02    2.
@00665   DFBCS03     2.
@00667   DFBCdS04    2.
@00669   DFBCS05     2.
@00671   DFBCS07     2.
@00673   DHTCbS1b    2.
@00675   DPMCcS1     2.
@00677   DPMCbS2B    2.
@00679   DPMCcS3     2.
@00681   DPMCdS4     2.
@00683   DPMCdS5     2.
@00685   DDECdS01    2.
@00687   DDECdS02    2.
@00689   DDECdS03    2.
@00691   DPUCbZQ1    1.
@00692   DPUCbZQ2    1.
@00693   DPUCbZQ3    1.
@00694   DPUCbZQ4    1.
@00695   DPUCbZQ5    1.
@00696   DDRCdZQ3    1.
@00697   DHTCdZQ1    1.
@00698   XSECFLG     1.
@00699   LONGFLG     1.
@00700   FIELDRUK    $12.
@00712   PERSRUK     $14.
@00726   CHILDID     $6.
@00732   DWTCW01C    12.4
@00744   DWTCW01L    12.4
```

```
@00756   DWTCWd1L     12.4
@00768   DMMCQ02    $1.;
```

Appendix II – Changes from previous cycles

1. Revision of weights for previous cycles

The final step in calculating weights for each longitudinal and cross-sectional cohort in each NLSCY cycle is to represent the population of children in Canada's 10 provinces. Cross-sectionally, in Cycle 1, we want to represent children aged 0 to 11 on January 1, 1995. In Cycle 2, we want to represent children aged 0 to 13 on January 1, 1997, and in Cycle 3, children aged 0 to 15 on January 1, 1999. Longitudinally, the children introduced in Cycle 1 must represent children aged 0 to 11 on January 1, 1995. Children aged 0 and 1 on January 1, 1997, are represented by the children introduced in Cycle 2. The weighting step that consists in adjusting the sample weights to obtain target population totals is post-stratification.

During the initial production of final survey weights, we used the preliminary population counts for the appropriate point in time. However, those figures have now been replaced by final counts, which differ from the estimates used. To ensure consistency between our survey estimates and Statistics Canada's population counts, and to comply with Agency policy, we need to revise the sample weights for each cycle.

First, we will examine the magnitude of the changes between the original estimates and the final population count estimates. Second, for each cycle we will determine how the weights were updated. Finally, we will compare selected estimates based on the original weights and the updated weights.

Magnitude of the changes

For each of the three reference years, we compared the original and final population count estimates for the target population. At the level of the total population, the changes were not particularly significant. The total population rose from 4,664,831 to 4,751,437 (+1.86%) in 1995, from 5,456,783 to 5,558,922 (+1.87%) in 1997, and from 6,196,411 to 6,317,254 (+1.95%) in 1999.

The changes were more significant for smaller units such as province or post-stratum. A post-stratum is defined as a unique age-sex-province combination. For each combination, the post-stratification adjustment is the ratio between Statistics Canada's population estimate and the NLSCY's population estimate with adjusted weights for that point in time. Since estimates can be produced for those units, it is important to measure the change in those estimates.

Table 1: Change Between Original and Final Population Estimates, by Province

Province	Increase in 1995	Increase in 1997	Increase in 1999
Newfoundland	0.43%	0.37%	0.40%
Prince Edward Island	0.63%	0.62%	0.60%
Nova Scotia	1.67%	1.60%	1.85%
New Brunswick	1.51%	1.62%	2.04%
Quebec	1.20%	1.16%	1.18%
Ontario	0.82%	0.80%	0.83%
Manitoba	8.58%	9.01%	9.72%
Saskatchewan	7.53%	8.03%	7.75%
Alberta	2.72%	2.86%	2.72%
British Columbia	2.10%	2.03%	2.41%

The changes are fairly minor at the provincial level, except for Manitoba and Saskatchewan, which had increases of over 7% for each reference year.

At the post-stratum level, the increase was 10% for some post-strata in Manitoba and Saskatchewan for each reference year, and 4% for some post-strata in other provinces.

In view of the observed changes, an update is required. The methodology used to adjust the weights in each file produced previously is explained below.

Updating the weights in the master file

Updating the Cycle 1 weights

The methodology used to produce Cycle 1 weights was different from the methodology used in other cycles. First, we have two sets of cross-sectional weights: the cross-sectional weights prior to the cuts made in Cycle 2, the real cross-sectional sample; and the cross-sectional weights following the cuts, the real "longitudinal" sample, on which the longitudinal weights in subsequent cycles will be based. We will refer to the latter set of weights as longitudinal weights.

The post-stratification domains were different in Cycle 1. Instead of the single-year-of-age - sex - province combinations used in other cycles, the following combinations were used for longitudinal weighting: age 0 - sex - province, age 2-3 - sex - province, age 4-5 - sex - province, age 6-7 - sex - province, age 8-9 - sex - province, and age 10-11 - sex - province. For cross-sectional weighting, the same domains were used with the addition of census metropolitan area (CMA) counts. Then a raking ratio procedure was performed. We will now take a closer look at the methodology used for each set of weights.

Longitudinal weighting

To obtain the final weight, the LFS (old or new design) subweights were multiplied by a series of adjustment factors. The final adjustment was post-stratification. We used the weight obtained before the original post-stratification to compute the new post-stratification adjustment. We also used all the single-year-of-age - sex - province combinations to reconcile the methodology used in Cycle 1 with the methodology used in the other cycles.

Cross-sectional weighting

Originally, we had three sample frames for children outside Quebec and three for children in Quebec. As in the case of longitudinal weighting, the basic weight was multiplied by a series of adjustment factors. The post-stratification adjustment factor was removed, and the basic weight was multiplied by the other adjustment factors. The post-stratification was then recalculated using all the single-year-of-age - sex - province combinations to reconcile the Cycle 1 methodology with the methodology used in subsequent cycles. For the same reasons, the CMA stratum was dropped.

For further details on the Cycle 1 methodology, see the Cycle 1 User's Guide.

Updating the Cycle 2 weights

Longitudinal weighting

In Cycle 2, at the longitudinal level for the cohort introduced in Cycle 1, the weight obtained prior to post-stratification was used as the basic weight. That weight was multiplied by an adjustment factor to correct for non-response, and then post-stratification was carried out in order to represent the population on January 1, 1995. In the updating process, we simply multiplied the initial weight by the non-response adjustment and the new post-stratification adjustment. In contrast to Cycle 1, the age-sex-province post-strata were the same in the updating process as in the original weighting.

Cross-sectional weighting

At the cross-sectional level, for the children introduced in Cycle 1, the weighting was similar to what had been done at the longitudinal level, except that cross-sectional response codes were used. For children introduced in Cycle 2, we used the LFS subweight multiplied by a series of adjustment factors. After the two components were combined to represent the population on January 1, 1997, the final adjustment was post-stratification. In updating the post-stratification, we simply recomputed the post-stratification adjustment and applied it to the cross-sectional weights obtained before the original post-stratification.

Updating the Cycle 3 weights

Longitudinal weighting

In Cycle 3, the methodology was the same as in Cycle 2. At the longitudinal level, the weights obtained prior to the original post-stratification were used during updating to compute the new post-stratification adjustment. The population on January 1, 1995, was represented for children introduced in Cycle 1, and the population on January 1, 1997, was represented for children introduced in Cycle 2.

A second change was made in the longitudinal weights file. Sixty-five children who were longitudinal respondents but cross-sectional non-respondents were left out of the original file. All of them had either died or moved permanently out of the country in a previous cycle. By default, all the variables for those children were set to "null" or "not stated". The 65 respondents were added to the new weights file. The impact of the addition was minimal because of the default values assigned to them.

Cross-sectional weighting

At the cross-sectional level, a cross-sectional sample was added to the two longitudinal cohorts. Once again, we used the weight obtained before the original post-stratification to compute the new post-stratification adjustment. The population on January 1, 1999, was represented.

Updating the bootstrap weight files

The bootstrap weights also have to be updated using the new weights from the master files. The same 1,000 strata/replicate samples were used to minimize the impact of the changes.

Impact of the changes

To measure the impact of the changes, we employed the series of key variables used to produce the approximate CV tables. Twenty-six variables were identified. Estimates were computed for a number of geographic units (Canada, Atlantic region, Prairies, province) and socio-demographic units (three age groups, sex). This provided thousands of estimates for comparison.

Cycle 1

In Cycle 1, the update for changes in the population counts is not the only factor responsible for change. At the longitudinal level (children who survived the cuts and were sampled in Cycle 2), there was also the change in post-strata to reconcile the Cycle 1 methodology with the methodology used in subsequent cycles. Post-stratification uses single years of age instead of two-year cohorts. Nevertheless, the biggest difference between the pre-update and post-update estimates was 0.9% out of a total of 2,914 estimates. The difference between the pre-update and post-update estimates was larger than in other cycles. However, the differences were important only for CVs that were already high. Differences of more than 5% were found only for CVs in excess of 63%. There were only 31 estimates out of 2,914 for which the difference exceeded 5%. In summary, updating had a negligible impact on the estimates.

At the cross-sectional level, with the 22,831 children initially selected, we find the largest differences, all cycles combined. When we updated the weights, we eliminated CMAs from the post-stratification and used single years of age instead of two-year cohorts in order to standardize the methodology used in the various cycles. Moreover, the bootstrap weights were computed without the CMAs, another reason for doing likewise with the sample weights. For provinces such as Quebec and Ontario, which have a number of CMAs and whose post-stratification adjustments were very different before and after post-stratification, the changes may be significant. Of the 2,930 estimates computed, only two had a difference of more than 1%. The proportion of New Brunswick children whose parents expected them to go to university was 1% lower after the update than before. The proportion of spouses in Manitoba who were able to carry on a conversation only in English was 1.3% higher after the update than before. For the coefficients of variation, as in the case of the longitudinal portion, differences of more than 5% occurred only in high CVs (over 36% to be exact). In all, 40 of the 2,930 estimates had CVs with differences in excess of 5%.

At the cross-sectional level, we paid particularly close attention to the variable that indicates the type of community in which the child lives. The values of that variable ranged from “rural” to “urban with a population over 500,000”. With the elimination of CMAs, that variable, which is not on the list of key variables, is one of the variables that should vary the most.

Table 2: Estimates of the Proportion of Children Living in Various Urban/Rural Areas, Before and After Updating, Cross-Sectional Cycle 1, NLSCY, Canada

Rural/urban	Before	After
Urban with over 500,000 population	46.00%	43.97%
Urban with between 100,000 and 500,000 population	17.04%	18.17%
Urban with between 30,000 and 100,000 population	7.55%	7.72%
Urban with between 15,000 and 30,000 population	3.25%	3.31%
Urban with less than 15,000 population	8.27%	8.54%
Rural	17.89%	18.29%

The standard deviation for one of these estimates is about 0.3%. Hence, there is a significant difference between the estimates prior to and after updating for children living in urban areas with a population of more than 500,000.

Cycle 2

For children living in one of Canada's 10 provinces, the largest difference between the 1,848 estimates of key variables before and after updating was 0.4%. The largest difference between CVs was 0.8%. Consequently, updating had a negligible impact.

At the cross-sectional level, out of 2,878 estimates, the largest difference in proportions was 0.5% and 0.7% for the CVs. Consequently, updating had a negligible impact.

Cycle 3

At the longitudinal level, for children introduced in Cycle 1, we computed 2,308 estimates. The largest difference between the proportions observed with the old weights and the new weights was 0.6% Hence, updating had a negligible impact.

For children introduced in Cycle 2, we computed 1,600 estimates. The biggest difference was 0.5%. Thus, the impact was negligible.

At the cross-sectional level, we computed 2,767 estimates. The biggest difference was 0.3%. We also examined the differences in CVs for each estimate. The difference was as high as 8% for some CVs above 90%. The quality does not change in this case, whether we use the old weights or the new ones. Only seven CV estimates had a difference higher than 2%. Consequently, updating had a negligible impact.

Conclusion

Updating the sample weights for the first three cycles was essential to ensure consistency between Statistics Canada's official figures and the NLSCY estimates. We also took the opportunity to standardize the post-stratification methodology by using the same post-strata in all cycles to produce the sample weights, the share-file weights and the bootstrap weights.

The impact of those changes was fairly minor. In Cycles 2 and 3, where the only change involved updating the population counts and the post-stratification, estimates of the proportions indicate that there was virtually no impact. In Cycle, the impact was slightly larger. At the longitudinal level, the differences were slightly larger than for the other cycles, but they were small enough that it is safe to

say the changes had a very minor impact. The largest impact that any update had was on the Cycle 1 cross-sectional weights. Even there, however, it was fairly minor.

Special attention should be paid to Cycle 1 cross-sectional estimates for CMAs or units composed of CMAs. The impact is more serious in those cases. Estimates of totals also merit particular attention. The impact on proportions is negligible, but the differences in the totals are the same as the differences in the counts. Users should therefore exercise greater caution in analyses involving Manitoba and Saskatchewan, the two provinces with the largest count changes.

For the vast majority of studies conducted before the update, no significant change should be observable, and there is no need to review their conclusions. The only exceptions, cases where there may be some doubt, are those mentioned above.

The user should also note that for Cycle 4 cross-sectional weighting, the January 1, 2001, population counts are preliminary and subject to change. A similar revision should be carried out when Statistics Canada finalizes the population counts.

2. Recalculation of the Motor Social Development Score – Cycles 1, 2 and 3

New scores were calculated for the Motor and Social Development (MSD) section of the Child's Questionnaire. This section of the questionnaire was completed by children in the 0 to 3 age group. A scale, made up of 48 questions (xMSCQ01 to xMSCQ48), was used to assess these concepts. According to the age in months, 15 questions will be asked of each child. A raw score and a standardized score were produced for each cycle.

These new scores were calculated in order to correct errors in the scores from previous cycles and to ensure that a standardized approach to calculating the scores is used across all cycles.

A raw score was calculated for each child by summing the number of "yes" answers to each item in the scale (xMSCS01). As well two standardized scores were produced for each cycle, except for Cycle 1. First, a score produced using norms from data from within a given cycle was calculated. Second a score produced using norms from Cycle 1 was calculated. For Cycle 1 these two scores would be the same, so only one score was calculated.

All children, 3 years old or younger, in cycles 1 and 2 and children 3 to 47 months in Cycle 3 old were assigned a standard score using norms based on data from within the cycle. This standardization was done by 1 month age groups. For each month age group the mean and standard deviation of the raw score was found and were used to produce a normalized score with a mean of zero and a standard deviation of one. This score was adjusted such that the mean MSD score was 100 and the standard deviation was 15. Therefore children who are 3 months old have an average MSD score of 100, children who are 4 months old have an average MSD score of 100, and children 47 months old have an average MSD score of 100.

Once these scores were calculated children who were more than 3 standard deviations away from the mean (scores smaller than 55 or greater than 145) were identified, and the norms were recalculated not including these children. These children were considered outliers and are not representative of other children their age. Therefore the average of MSD scores on the data file by age in months may not be exactly 100. Using this standardized score (xMSCS02) makes it possible to compare scores of children across the 0 to 3 age group, not controlling for age.

This score was not calculated for children aged 0-2 months in Cycle 3 as there were not enough respondent children by age in months to establish a norm. In cycles 1 and 2 this is not a problem due to the way that the sampling was done.

A second standardized score (xMSCS02) was calculated for all children 0-47 months in cycles 2 and 3. This score was calculated in the same way as mentioned above; only the norms were derived using the data from Cycle 1, and then applied to the data from the following cycles.

Overall there are no major differences found when comparing the scores found using norms from within a cycle and the scores found using Cycle 1 norms. The score calculated using Cycle 1 norms should be used in order to compare scores over cycles. This score is available for all cycles of data.

3. Health Utility Index

The Health Utility Index is calculated for children for and five year old. In cycles 1 and 2, a provisional measure (HUI2) was used to calculate the index. The index has been recalculated using the HUI3 and included in the re-released files. See chapter 8, Content of the Survey, for more information about the Health Status Index.

4. Revision of the Cycle 3 Maths scores

Changes to CMACS01 & CMACS02 (revision)
For the initial release of this variable for Cycle 3, the raw score and the scaled score were based on 15 of the 20 questions of the test. The normative score was interpolated by inserting the percentile rank obtained with only the 15 questions out of the 20 questions of the short test between the percentiles of the complete test from the CAT/V2. This was done for two reasons:

1° to keep a consistency with the same 15 questions score presented in Cycles 2.

2° because the 5 added questions were selected from a grade level above the current level being test and was not equated for that particular level.

The raw score reflected the outcome of the subset of 15 questions considered for the equating.

Since then, it has been determined that the vertical equating of the classical test would yield consistent measurements of learning gains, whether the added items were equated to a specific grade or not. Item parameters where equated on a normative scale showing relative difficulty levels from all grades and were tested on a moving panel design where children were at different levels throughout the school year. The added gain in precision when measuring the score using the 20 items outweighed the affect on consistency between cycles 2 and 3 as it would imply a systematic loss of information for all subsequent cycles of the survey. This rescaled score has been included on the re-release of the file. The raw score now reflects the outcome as measured over all 20 questions of the test.

Changes to CMACS03 (correction)
For the initial release of this variable for Cycle 3, an error in data transmission resulted in incorrect data for this field. The scaled score using Item Response Theory had been available and was rescaled to reflect consistency between Cycles 3 and 4 releases. The correct values for this score replace the erroneous information that was in the initial release.

Appendix III –Non-response analysis

These two documents are an example of a non-response analysis in a case of examining a sub-domain where the non response is significant. First, an article released in the Daily on June 16th, 2003, followed by a Quality Report on the subject.

National Longitudinal Survey of Children and Youth: Challenges of late adolescence - 2000/01

Note to readers

The National Longitudinal Survey of Children and Youth (NLSCY), developed jointly by Human Resources Development Canada and Statistics Canada, is a comprehensive survey which follows the development of children in Canada and paints a picture of their lives. The survey monitors children's development and measures the incidence of various factors that influence this development, both positively and negatively.

This article highlights findings based on a preliminary analysis of information that has been collected from youth aged 16 and 17 in 2000/01 who were followed from the time they were aged 10 and 11 in 1994/95.

Like all longitudinal surveys, the NLSCY experiences non-response. The level of non-response has gradually increased over the years, in particular for the questionnaire completed by the youth themselves. In 2000/01, the response rate for most of the questions discussed in this article stood between 75 and 80% of those completing other components of the survey. The analysis presented in this article was performed on respondents only. There is therefore, a potential for some degree of non-response bias. A report on data quality relative to this article, providing the coefficients of variation for variables under study and response rates for key questions, is available upon request.

The first cycle of the NLSCY, conducted in late 1994 and early 1995, interviewed parents of about 23,000 children up to the age of 11. The parents gave information, not only about their children, schools and neighbourhoods, but also about themselves and their families. About 3,400 children aged 10 and 11 at that time were asked questions directly about themselves. These data were unique in that they came directly from the child. This release focuses on the self-reported data for these same youth from the fourth cycle of the NLSCY conducted in 2000/01. The youth provided self-assessment of experiences, health status and risky behaviours such as taking money from parents and staying out all night without permission.

The NLSCY asked 16 and 17-year olds to rate how often they had a group of specific emotional problems. These problems ranged from a poor appetite, to a restless sleep, loss of optimism about the future, loneliness, feeling like everything they did was an effort, or difficulty shaking off the blues. If a person responded as having a threshold number and intensity of feelings they were deemed to have symptoms of depression. In this article, youth are identified as having symptoms of depression if they scored in the top 10% of the distribution of the depression scale.

The survey will continue to collect information about these same youth every two years as they move into adulthood.

FRONT-PAGE BULLET

An overwhelming majority (95%) of young people aged 16 and 17 say they are healthy, and that they are optimistic about their future, according to the National Longitudinal Survey of Children and Youth. However, the data also reflect the challenges teens face in dealing with complex decisions about their academic plans, risky behaviours and emotional health.

TEXT

An overwhelming majority (95%) of young people aged 16 and 17 say they are healthy and that they are optimistic about their future, according to the National Longitudinal Survey of Children and Youth (NLSCY).

Changes in family structure between 1994 and 1999 did not make a difference to how children reported their own health in 2000/01. This held true whether children's family arrangement changed during the previous 6 years from living in a two-parent family to living with a single-parent or vice versa

Nonetheless, late adolescence is a time of immense change. Teenagers have to deal with sudden body changes, peer pressure and an emerging sense of self. They are faced with the challenge of dealing with complex decisions regarding future plans, relations with peers and parents, and pressures of everyday life. The NLSCY explored these challenges, and this report releases some preliminary findings.

As they are getting older, some youth are ready to engage in some risky behaviours as might be expected. More than a quarter of 16 and 17 year olds reported staying out all night without permission while 15% admitted they had taken money from parents without permission.

Young people reported more symptoms of depression compared to their reported incidence at younger ages, with 24% of 16 and 17 year olds reporting symptoms of depression versus 9% when they were 12 and 13.

While they are optimistic about their future, a majority of teens also expressed some concern about it, particularly with respect to prospects for fulfilling their academic goals and aspirations. Asked about their academic aspirations, 88% of those aged 16 and 17 reported that they expect to go beyond a secondary or high school graduation.

More than half of 16 and 17 year olds stated that they wished to obtain a university degree. However, given that only 30% of young adults aged 20-24 had gone on to university (2001 Census), some change their plans at least during the first few years following high school.

The information provided by these adolescents adds a new dimension to the understanding of their life, current activities and emotional health.

As they grew older, more youth reported staying out all night and taking money from parents

The persistence of behaviours such as taking money from parents without permission and fighting that caused physical injuries was assessed by comparing responses provided by the same youth when they were aged 12 and 13 in 1996/1997 with their responses when at age 16 and 17 in 2000/01.

This comparison shows that certain behaviours increased, while others tended to remain stable over time. For example, 9% of youth who were aged 12 and 13 in 1996/97 reported staying

out all night without the permission of their parents at least once. By 2000/01, 27% of these youth now aged 16 to 17 reported having stayed out all night without permission at least once, about a third of whom reported doing so three times or more in the previous year.

Looking at another type of risky behaviour in 1996/97, 15% of young people who were aged 12 and 13 reported that they had taken money from their parents without permission at least once. Four years later, among the same group of young people now aged 16 and 17, about 22% admitted they had taken money from their parents without permission once or twice while an additional 7% indicated they had done it three times or more over the previous 12 months.

On the other hand, the proportions of the same young people engaging in other behaviours that might be considered more delinquent in nature, such as fighting that resulted in physical injuries (5%) or damaging property (12%), remained relatively stable during this four-year period.

Some of these risky behaviours are also associated with marijuana smoking. When asked about their drug use during the 12 months prior to the survey, 44% of 16 and 17-year olds reported that they have smoked marijuana. The incidence of marijuana use was 1.8 to 2.6 times higher among youth who reported participating in risky behaviours such as staying out all night without permission, taking money from parents and damaging others' property.

Among 16 and 17-year olds who reported staying out all night without permission, 72% reported that they have tried marijuana smoking, while 28% reported that they have not smoked marijuana. Moreover, for 16 and 17-year olds who reported that they had taken money from their parents, 64% reported that they smoked marijuana. For 16 and 17-year olds who reported that they have damaged others' property, 69% reported that they have smoked marijuana.

Symptoms of depression increased with age

Data from the NLSCY show that reports of symptoms of depression, anxiety and distress among adolescents increased as they grew older between 1996/97 and 2000/01.

The NLSCY asked 16 and 17-year olds to rate how often they had a group of specific emotional problems. These problems ranged from a poor appetite, to a restless sleep, loss of optimism about the future, loneliness, feelings of being disliked by other people, crying spells, feeling depressed, feeling like everything they did was an effort, or difficulty shaking off the blues. If a person responded as having a threshold number and intensity of feelings they were deemed to have symptoms of depression.

In 1996/97, 9% of youth aged 12 and 13 reported having symptoms of depression. By 2000/01, when these same young people were aged 16 and 17, the proportion who reported having symptoms of depression had more than doubled to 24%.

A more in-depth analytical paper on late adolescence will be available in Fall 2003, and will explore the relationship between risky behaviours as well as depression and school performance, parenting style, socioeconomic background and family arrangement. For more information, about the data collected during the first four cycles of the NLSCY or to enquire about the concepts, methods or data quality of this release, contact Client Services, (1-888-297-7355; 613-951-7355; ssd@statcan.ca) or Teresa Abada (613-951-3017; teresa.abada@statcan.ca), Statistics Canada, or Satya Brink (613 – 953-6322; satya.brink@hrdc-drhc.gc.ca), Human Resources Development Canada.

Quality report regarding the article in the Daily pertaining to adolescents in the NLSCY

Like all surveys, the NLSCY must deal with non-response. There are two large categories of non-response: total non-response and partial non-response. Total non-response is the complete absence of data (or too little data to be considered a response) for a sampled unit. To compensate for the total non-response, the survey weighting is adjusted. Partial non-response is the absence of information for certain questions only, with the person selected having nonetheless adequately answered enough questions to be considered a respondent. The purpose of this report is to study the partial non-response to the self-administered component of the NLSCY which was largely used for the article in the Daily.

This self-administered component is intended for children 10 years of age and older. It covers many areas: family and friends, health, the child's behaviour, tobacco and other drugs, love relationships and decision-making.

In the article in the Daily, the analysis pertains only to the 1800 adolescents (approximately) who were 16-17 in 2000/01 and considered respondents for the main part of the questionnaire from the beginning of the NLSCY (they were 10-11 years old at the time). The self-administered component shows non-response. Some teens, in fact, did not reply to this part but their parents and/or they nonetheless gave the information in the main questionnaire. Erosion is, moreover, a classic phenomenon in longitudinal surveys: it is there regarding the self-administered component. During the course of the various cycles of the survey and as they get older, these adolescents seem to « tire » of the self-administered component.

75-80% Response rate (approximately) in 2000/01

The older these adolescents get, the less they respond.

In the article that interests us, the behaviour of adolescents evolves since the first cycle in 1994/95 by their response but also by their lack of response. We will not present here an overall response rate for the self-administered component for the four cycles of the survey but response rates for one or two questions pertaining to the theme addressed in the article.

For questions pertaining to delinquency, the response rate was approximately 89% in cycle 2 (the children were 12-13 years old); the rate is about 77% in Cycle 4. The same for questions pertaining to depression; the response rate is roughly 80% in cycle 2 and 70% in Cycle 4.

From one theme to another, the response rates are comparable.

Across the various subjects dealt with in the article, the response rates are generally comparable. There is no theme to which the adolescents did not reply en masse. The table below presents the response rates by subject addressed in the article.

We bear in mind here that the response rates are calculated for one or two questions pertaining to a particular section. It gives an approximation of what the response rate could be in each section.

Variables: all the variables pertain to Cycle 4 unless otherwise mentioned.	Response rate
Health	80%
How often do your parents encourage you?	71%
Close friends push me to succeed	83%
How often hang with kids in trouble?	86%
Depression score in cycle2	80%
Depression score in cycle4	71%
Happy with life now, The next five years look good to me	77%
Important to have good grades?	87%
Risky behaviours Stayed all night out without permission in cycle 2	89%
Cycle 4	77%
Have you seriously thought about suicide?	77%
Have you ever smoked cigarettes?	77%
I know my major in college, in university	71%
Stressful events : one example Break-up with boyfriend/girlfriend	76%
Self esteem	77%

The importance of analyzing the respondents: for example, girls respond more than boys

In the case of this article, the analysis is based on a sub-population of the sample, as it is only interested in **respondents**. When interpreting the results, care must be taken to take this behaviour into consideration. When a number of adolescents do not respond, how should their non-response be interpreted? Is their behaviour comparable to the respondents' behaviour in other ways? For example, do the girls who do not respond behave like the girls who do respond in other ways? Or on the contrary, does the behaviour of these teens who do not respond translate into very different, even risky, behaviour for certain questions?

No detailed analysis of non-response has been done for the subjects addressed here. It should be noted that most of the subjects correspond to questions in the self-administered component but some appear as well in the parental component.

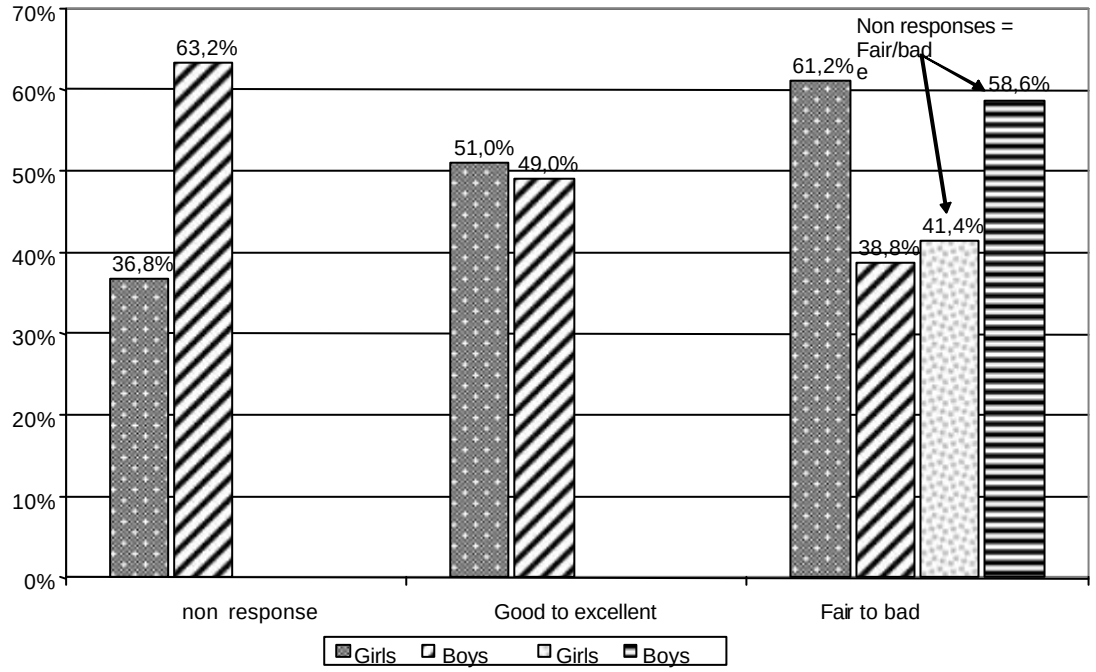
Let us look at sex only as a non-response characteristic. Girls consistently respond more than boys in the NLSCY sample. This phenomenon is amplified in Cycle 4.

Response rate by section and sex in % - weighted

Variables: all variables pertain to Cycle 4 unless otherwise stated.	Response rate	Girls	Boys
Health	80%	85%	76%
How often do your parents encourage you?	71%	75%	67%
Close friends push me to succeed	83%	86%	81%
How often hang with kids in trouble?	87%	88%	86%
Depression score in cycle2	80%	84%	78%
Depression score in cycle4	71%	76%	67%
Happy with life now, The next five years look good to me	77%	80%	75%
Important to have good grades?	87%	87%	86%
Risky behaviours : one example Stayed all night out without permission in cycle 2	89%	92%	88%
Cycle 4	77%	80%	75%
Have you seriously thought of suicide	77%	79%	74%
Have you ever smoked cigarettes?	77%	79%	74%
I know my major in college, in university	71%	76%	68%
Stressful events : one example Break-up with boyfriend/girlfriend	76%	79%	73%
Self esteem	77%	80%	75%

Let's take the question regarding health as an example: In general, would you say that the « adolescent's » health is excellent…to poor?

NLSCY Cycle 4 – In general, would you say that the "adolescent's" health is excellent…to poor?



The distribution of responses to this question is the following: 20% non-response, 75% answered « good to excellent » and the remaining 5% « fair or poor ».

If we only concentrate on the respondents, then 95% of the answers belong to the “good to excellent” category. Furthermore, we can note that the majority of teens whose health seems poor are girls.
In an **extreme** case, in which those who did not respond to the question had « poor or fair » health, the result would be that 76% of the children were in « good or excellent » health and that the 24% in poor health would be mostly boys!
This example certainly does not reflect reality. However, it highlights the importance of considering non-response in all analyses of the population groups of interest.
It is important to establish certain hypotheses regarding the behaviour of the non-respondents if we wish to extrapolate the results to a larger population.

Acceptable coefficients of variation
For each estimate presented in the study, a coefficient of variation (CV) is calculated using bootstrap weights.
The sample used for the study is the longitudinal sample also known as the « funnel » as it is only interested in the children replying to the main questionnaire over the course of the four survey cycles. One thousand bootstrap weights were generated for each child responding. All CV corresponding to the figures presented in the article allow us to say that the estimates are acceptable. CVs [are] mostly below 16.5% and below 33%.

CV Calculation with the Funnel Bootstrap Weights

Variable of interest	Numerator	Denominator	Estimated Ratio	CV bootstrap	CI95 -low	CI95 - upper
Health	in good health	RTQ*	94.4%	1.08%	92.4%	96.4%
Headache	never headache	RTQ	41.1%	4.22%	37.7%	44.5%
	More than once a week	RTQ	14.4%	9.42%	11.8%	17.1%
Stomach ache	Never stomach ache	RTQ	60.7%	2.91%	57.2%	64.1%
	More than once a week	RTQ	6.2%	15.1%	4.3%	8.0%
Backache	Never backache	RTQ	53.2%	3.6%	49.4%	56.9%
Difficult to sleep	Never	RTQ	53.9%	3.83%	49.8%	58.0%
	More than once a week	RTQ	15.3%	9.06%	12.6%	18.0%
Friends push to succeed	Most True	RTQ	67.9%	2.66%	64.3%	89.7%
Hang with kids in trouble	Never seldom	RTQ	87.5%	1.32%	85.2%	89.7%
	Sometimes	RTQ	10.2%	10.3%	8.2%	12.3%
Hang * Expect you will go in school	Expect will go in school more than technical	Never or seldom to the question hang kids in trouble	86.7%	1.53%	84.1%	89.3%
Happy with life	Agree	RTQ	84.9%	1.7%	82.1%	87.7%
	Disagree	RTQ	15.1%	9.55%	12.3%	17.9%
Next five years look good	Agree	RTQ	90.0%	1.24%	87.8%	92.2%
	Disagree	RTQ	10%	11.22%	7.8%	12.1%
Important to have good grades	Very	RTQ	69.1%	2.29%	66.0%	72.2%
	Somewhat	RTQ	27.9%	5.3%	25%	30.8%
	Not	RTQ	3%	21.78%	1.7%	4.3%
Out all night without permission cycle 2	Once or twice	RTQ	7.1% (1)	12.46%	5.4%	8.9%
Out all night without permission Cycle 4	Once or twice	RTQ	16.8% (2)	9.35%	13.7%	19.9%
Comparison cycle2 – Cycle 4 between	(1) and (2) above	Significantly different	-9.7%	16.05%	-12.8%	-6.7%
Out all night without permission cycle 2	Once or more	RTQ	8.52	11.64%	6.6%	10.5%
Out all night without permission cycle 1	Once or more	RTQ	26.4%	6.71%	23%	29.9%

Variable of interest	Numerator	Denominator	Estimated Ratio	CV bootstrap	CI95 -low	CI95 - upper
Comparison cycle2 – Cycle 4	Once or more	Significantly different	-11.9%	11.9%	-22.1%	-13.7%
Out all night without permission Cycle 4	3 times or more	RTQ	9.6%	13.08%	7.1%	12.1%
Take money from parents cycle 2	Once	RTQ	12.0%(1)	9.78%	9.7%	14.3%
	Never seldom	RTQ	86.5%	1.42%	84.1%	88.9%
Take money from parents Cycle 4	Once	RTQ	21.8%(2)	8.11%	18.4%	25.3%
	Never	RTQ	71.2%	2.54%	67.6%	74.7%
	3 and More	RTQ	7%	13.04%	5.2%	8.8%
Comparison cycle 2 – cycle4 between	(1) and (2) above	Significantly different	-9.8%	15.13%	-12.77%	-6.93%
Physical injuries Cycle2	Never	RTQ	94.6%	0.96%	92.8%	96.3%
	Yes	RTQ	5.4% (1)	16.61%	3.7%	7.2%
Physical injuries Cycle4	Never	RTQ	94.7%	0.91%	93.04%	96.41%
	Yes	RTQ	5.3% (2)	16.27%	3.6%	7.0%
Comparison between cycle2-cycle4	(1) and (2) above	NO DIFFERENCE	0.17%	50.4%		
Damage other property cycle 2	Never seldom	RTQ	88.3%	1.31%	86%	90.6%
Depression in cycle 2	High score	RTQ	9.4% (1)	11.36%	7.3%	11.5%
Depression in Cycle 4	High score	RTQ	23.8% (2)	7.42%	20.3%	27.3%
Comparison between cycle 2 Cycle 4	(1) and (2)	Significantly different	-14.4%	9%	-16.9%	-11.9%
High score in depression in Cycle 4	Girls	RTQ	61.8% (1)	6.43%	54%	69.6%
High score in depression in Cycle 4	Boys	RTQ	38.2% (2)	10.41%	30.4%	46%
Comparison between	Girls (1) and boys (2)	Difference	23.6%	25.15%	12%	35.3%
Depression and self esteem Cycle 4	Low depression And high self esteem score	High self esteem score	80.4%	2.24%	76.9%	83.9%
Considered suicide	Respond Yes	RTQ	12.3%	10.85%	9.7%	14.9%

Variable of interest	Numerator	Denominator	Estimated Ratio	CV bootstrap	CI95 -low	CI95 - upper
Considered suicide by gender	Girls who considered suicide	Kids who considered suicide	60.6% (1)	9.14%	49.8%	71.5%
Low self esteem score and gender Cycle 4	Girls and low self esteem score	Low self esteem score	59.1% (1)	9.97%	47.5%	70.6%
	Boys and low self esteem score	Low self esteem score	40.9% (2)	14.39%	29.4%	52.5%
Comparison between	Girls (1) and boys (2)	NOT significative difference	18.1%	82.68	-11.2%	47.5%
Risky behaviours and low depression score in Cycle 4						
Break-up and depression	Low depression score and break-up	Low depression score	33.8% (2)	6.88%	29.1%	38.1%
	Low depression score and NO break-up	Low depression score	65.8	3.5%	61.7%	70.7%
Comparison between	Break up and no break up for Low depression score	DIFFERENT				
	High depression score and break up	High depression score	50.1% (1)	8.07%	42.1%	58.0%
Comparison between	High (1) and low (2) depression score for Break up	DIFFERENT	-16.3%	29.3%		
Problem in school and depression	Low depression score And problem	Low depression score	23% (1)	8.92%	19%	27%
	Low depression score And NO problem	Low depression score	77% (2)	2.66%	73%	81%
	High depression score And problem	High depression score	53.3% (3)	8.99%	43.9%	62.7%

Variable of interest	Numerator	Denominator	Estimated Ratio	CV bootstrap	CI95 -low	CI95 - upper
Comparison between	Problem (1) and no problem (2) For low score	DIFFERENT				
	Problem in low score (1) and in high score (3)	DIFFERENT				
Death of someone close and depression	Low depression score And death	Low depression score	32.3% (1)	6.64%	28.1%	36.5%
	Low depression score And NO death	Low depression score	67.7% (2)	3.16%	63.5%	71.9%
	High depression score and death	High depression score	47.9% (3)	10.2%	38.3%	57.5%
Comparison between	Low depression score And death (1) /no death (2)	DIFFERENT				
Smoking	Death - low score (2) and death-high score (3)	DIFFERENT				
Have you ever smoked	Never	RTQ	35.7%	5.92%	31.5%	39.8%
	A few	RTQ	31.0%	6.04%	27.4%	34.7%
	Anymore	RTQ	9.5%	16.46%	7%	12%
	1 to 5 days/week	RTQ	5.9%	17.1%	3.9%	7.8%
	6-7 days /week	RTQ	17.9%	8.19%	15%	20.8%
Know my major in college	Not like me	RTQ	28.7%	6.73%	24.9%	32.5%
	Neutral	RTQ	18.7%	9.76%	15.1%	22.3%
	Like me	RTQ	52.6%	4.04%	48.4%	56.7%
Smoking marijuana	Never smoke or smoke few times	RTQ	84.7%			
	A few times or 1-2/month	RTQ	29.43%	6.8%	25.5%	33.4%
	Smoke 1 to 7 days a week	RTQ	15.3%	9.51%	12.5%	18.2%
Marijuana and risky behaviours						
Night out without permission	Don't smoke	RTQ	27.6%	14.21%	19.9%	35.2%

Variable of interest	Numerator	Denominator	Estimated Ratio	CV bootstrap	CI95 -low	CI95 - upper
Night out without permission	Smoke	RTQ	71.53%	5.6%	63.7%	79.4%
Operate vehicles under drugs	Never	RTQ	89%	1.35%	86.6%	91.4%
	At least once	RTQ	11%	11%	8.6%	13.4%
Passager dans un véhicule où le conducteur avait consommé des drogues	Never	RTQ	41.4%	5.26%	37.1%	45.6%
	At least once	RTQ	58.6%	3.71%	54.4%	62.9%
Marijuana et conduite	Smoke 1-7 days a week and never operate a vehicle under drugs	Smoke 1-7 days a week marijuana	64.3%	7.28%	55.1%	73.5%
	Smoke 1-7 days a week and operate at least once a vehicle under drugs	Smoke 1-7 days a week marijuana	33.7%	13.74%	24.6%	42.7%

* RTQ (respondents to question mentioned in first column)