Table of

Contents

1.0 Introduction 1

2.0 Background 3

3.0 Objectives 5

4.0 Survey Methodology 7
4.1 Definition of the NLSCY Population 7
4.2 NLSCY Sample Design 7
4.2.1 The Main Component 7
4.2.2 The Integrated Component (NPHS) 8
4.2.3 The Territories Component 9
4.2.4 The Child Sample 10
4.3 Sample Allocation 10
4.4 Sample Size 11

5.0 Data Collection 13
5.1 The Household Collection 13
5.2 The School Collection 15
5.3 Computer-Assisted Interviewing 16
5.4 Survey Timing 17
5.5 Proxy Reporting 18
5.6 Interview Length 19
5.7 Interview Training, Supervision and Control 21

6.0 Data Processing 23
6.1 CAI Editing 23
6.2 Data Capture 24
6.3 Minimum Completion Requirements 24
6.4 Head Office Editing 27
6.5 Naming Convention and Coding Structure for NLSCY Variables 29
6.5.1 Naming Convention for Variables 29
6.5.2 Acronym Names for Questionnaire Sections 31
6.5.3 Examples of Variable Names 33
6.5.4 Coding Structure for NLSCY Variables 33
6.6 Coding of Open-ended Questions 35
6.7 Imputation 36
6.8 Creation of Derived Variables 36

7.0 Weighting 39
7.1 Weighting Procedures for the LFS 39

7.2 Weighting Procedures for the Main and Integrated Components 41
7.2.1 Main Component - Old LFS Design (all provinces) 42
7.2.2 Main Component - New LFS Design (All Provinces) 45
7.2.3 Integrated Component (All Provinces Except Québec) 46
7.2.4 Integrated Component - Québec 48
7.2.5 Final Stage of Weighting 50

8.0 NLSCY Concepts and Definitions 51
8.1 Cross-sectional and Longitudinal Estimates 51
8.2 NLSCY Units of Analyses 52
8.3 PMK and Spouse 53
8.4 Family Derived Variables 54
8.5 Socio-Economic Derived Variables 56
8.6 Geographical Indicators 64

9.0 Content and Validation of NLSCY Data 65
9.1 Validation of Scale Data 66
9.1.1 Factor Analyses 67
9.1.2 Calculation of Scores and Item Imputation 68
9.1.3 Reliability Measures for Scales 70
9.2 Demographic Variables 71
9.3 Medical/Biological 73
9.4 Temperament Scale 74
9.5 Education (Child) 75
9.6 Behaviour Scale 76
9.7 Motor and Social Development 82
9.8 Relationships 83
9.9 Parenting Scale 83
9.10 Child Care 87
9.11 Socio-demographic Characteristics 90
9.12 Depression Scale 91
9.13 Family Functioning Scale 92
9.14 Social Support Scale 93
9.15 Education (Parent) 95
9.16 Labour Force 96
9.16.1 Work Duration Derived Variables 96
9.16.2 Comparison with SLID Data 97
9.16.3 Wage Rate Edit 103
9.16.4 Microdata File Suppressions 104
9.17 Income 104
9.17.1 Wage Rate Edit and Income Imputation 105
9.17.2 Comparison to Survey of Consumer Finances . . 107
9.17.3 Microdata File Suppressions 108
9.18 Friends and Family 109
9.19 Feelings and Behaviour 110
9.20 My Parents and Me 111
9.21 PPVT-R 112
9.22 Mathematics Computation Test 115

10.0 Data Quality 117

10.1 Overall Response Rates and Response Bias 117
10.1.1 Component Non-response 122
10.2 Child Questionnaire Response Rates 123
10.3 Parent Questionnaire Response Rates 123
10.4 General Questionnaire Response Rates 124
10.5 PPVT-R Response Rates and Bias 124
10.6 10 to 11 Questionnaire Response Rates and Bias 126
10.7 Mathematics Test Response Rate and Bias 128
10.8 Ceiling Effect for Mathematics Test 134

11.0 Guidelines for Tabulation, Analysis and Release 137
11.1 Rounding Guidelines 137
11.2 Sample Weighting Guidelines for Tabulation 138
11.2.1 Definitions of Types of Estimates: Categorical vs. Quantitative 138
11.2.2 Tabulation of Categorical Estimates 140
11.2.3 Tabulation of Quantitative Estimates 140
11.3 Guidelines for Statistical Analysis 141
11.4 C.V. Release Guidelines 142

12.0 Approximate Sampling Variability Tables 145
12.1 How to Use the C.V. Tables for Categorical Estimates . 147
12.1.1 Examples of using the C.V. tables for Categorical Estimates 149
12.2 How to Use the C.V. Tables to Obtain Confidence Limits 153
12.2.1 Example of Using the C.V. Tables to Obtain Confidence Limits 154
12.3 How to Use the C.V. Tables to Do a T-test 155
12.3.1 Example of Using the C.V. Tables to Do a T-test 156
12.4 Coefficients of Variation for Quantitative Estimates 156
12.5 Release Cut-offs for the NLSCY 157

13.0 Suppression of Confidential Information 159
13.1 Methods Used to Protect Confidentiality 159
13.2 Variables Available on Master File Not Included on Public Use File 161
13.2.1 Geographical Variables 161
13.2.2 Family Demographic variables 161
13.2.3 Ethno-cultural Variables 162
13.2.4 Education Variables 163
13.2.5 Labour Force Variables 164
13.2.6 Income Variables 164
13.2.7 Medical Biological Variables 166
13.3 Remote Access Requests 166

14.0 Record Layout and Univariates 169
14.1 Technical Information 169
14.2 SAS Record Layouts 171
14.2.1 Primary File 171
14.2.2 Secondary File 185
14.3 Univariate Counts 194

Appendix 1 - Content for Release 1 and 2

Appendix 2 - The Labour Force Survey

Appendix 3 - C.V. Tables

Appendix 4 - Univariate Counts - Primary File

Appendix 5 - Univariate Counts - Secondary File

2.0

Introduction

The National Longitudinal Survey of Children and Youth (NLSCY) is a long-term survey designed to measure child development and well-being. The first cycle of the survey was conducted by Statistics Canada in 1994-1995 on behalf of Human Resources Development Canada. This manual has been produced to facilitate the manipulation of the microdata file and to document data quality and other analytical issues regarding the NLSCY.

It should be pointed out that not all the data collected in the first cycle of the NLSCY are included in this first microdata file. The amount of information collected was so extensive a decision was made to have two releases rather than waiting for all of the data to be processed. The second release will be in 1997. The notable sections to be included in this second release are health variables for the child and the parents, the custody history of the child, and data collected from the teacher and the principal. A complete list of the sections included in the first and second release can be found in Appendix 1.

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

At Statistics Canada:

Michael Sivyer
Manager
Dissemination and User Support
Special Surveys Division
Statistics Canada
5(B6) Jean Talon Building
Tunney's Pasture
Ottawa, Ontario K1A 0T6
Telephone:(613) 951-4598
Facsimile:(613) 951-0562
Internet: sivyer@statcan.ca
Toll free #: 1-800-461-9050

Gilles Montigny
Project Manager - NLSCY
Special Surveys Division
Statistics Canada
5(B6) Jean Talon Building
Tunney's Pasture
Ottawa, Ontario K1A 0T6
Telephone:(613) 951-9731
Facsimile:(613) 951-0562
Internet: montgil@statcan.ca

At Human Resources Development Canada:

Susan McKellar
NLSCY Project Coordinator
Applied Research Branch
Human Resources Development Canada
Place du Portage - Phase IV, Mailbox 528
Hull, Quebec K1A 0J9
Telephone:(819) 953-4230
Facsimile:(819) 994-2480
Internet: susan.mckellar@spg.org

4.0

Background

Before the NLSCY was undertaken there were few statistical studies describing a broad range of characteristics of children in Canada. Measures of health, well-being and life opportunities are needed, however, if governments and researchers hope to learn more about the ongoing life conditions of Canadian children and youth, and their developmental experiences. Longitudinal data are central to discovering developmental changes occurring in children over time, and studying the impacts of the social environment of the child and various family-related factors.

Data on the prevalence of, and interaction among, various characteristics and conditions will assist policy makers in understanding the processes that modify risk and protect and encourage the healthy development of children. Such information will enhance the capacity of the various partners in society to develop effective strategies, policies and programs to help children succeed in our changing society.

6.0 Objectives

The primary objective of the NLSCY is to develop a national database on the characteristics and life experiences of children and youth in Canada as they grow from infancy to adulthood. The more specific objectives of the NLSCY are:

- to determine the prevalence of various biological, social and economic characteristics and risk factors of children and youth in Canada;
- to monitor the impact of such risk factors, life events and protective factors on the development of these children; and
- to provide this information to policy and program officials for use in developing effective policies and strategies to help young people live healthy, active and rewarding lives.

Underlying these objectives is the need to:

- fill an existing information gap regarding the characteristics and experiences of children in Canada, particularly in their early years;
- focus on all aspects of the child in a holistic manner (i.e., the child, his/her family, school, and community);
- provide national, and as far as possible, provincial-level data; and
- explore subject areas that are amenable to policy intervention and which affect a significant segment of the population.

8.0

Survey Methodology

The requirement for the NLSCY design was to select a representative sample of children in Canada and to follow and monitor these children over time into adulthood.

8.2 Definition of the NLSCY Population

The target population of the NLSCY for Cycle 1 consisted of Canadian children aged newborn to 11 years of age.[1] There were some exclusions made for operational reasons which are discussed further in this section.

8.4 NLSCY Sample Design

In terms of sampling, the starting point for the NLSCY design was the household. Sampled households actually came from three possible sources which have been labelled as the **Main** Component, the **Integrated** Component and the **Territories** Component.

8.4.2 The Main Component

For Cycle 1 of the NLSCY the requirement was to select households with children, specifically children 0 to 11 years of age. The problem is that the majority of households do not contain children in this age range. In fact approximately only 26% of Canadian households in the 10 provinces contain at least one child in the 0 to 11 age range. A method had to be found to facilitate finding households likely to contain children, otherwise precious dollars would have had to be spent screening households in order to identify those with children. The answer was found with Statistics Canada's Labour Force Survey (LFS).

The reference date for the calculation of age was as of the day of the interview. Collection took place from November 1994 to June 1995.

The Labour Force Survey[2] is conducted on a monthly basis and collects basic demographic information about all household members of a representative sample of Canadian households as well as labour market information about the adults living in these households. For the NLSCY households that were currently or had recently been in the LFS sample were examined to determine which had children. This served as the basis of the household sample for the NLSCY **Main** Component. Approximately 12,900 households were selected for the sample for the NLSCY **Main** Component.

It should be mentioned that the LFS excludes certain populations since they are not part of the LFS sample frame, specifically individuals living in the Yukon or Northwest Territories, individuals living in institutions, and finally individuals living on Indian Reserves. In order to compensate for the first exclusion (i.e., the Yukon and Northwest Territories) the NLSCY introduced the Territories Component as described below. The undercoverage that resulted for the other exclusions (institutions and Indian Reserves) represents approximately 0.5% of children 0 to 11 years of age living in the 10 provinces.

8.4.4 The Integrated Component (NPHS)

At the same time that the NLSCY was being designed there was another national longitudinal survey being launched by Statistics Canada; namely the National Population Health Survey (NPHS). Its purpose is to produce estimates of physical and mental health of Canadians and to identify the factors that determine good and ill health.

Because both the NLSCY and the NPHS needed to collect data on the health of Canadian children, it was decided that a portion of the sample and content of the two surveys would be integrated for the 10 provinces. The children selected by the NPHS were part of the sample for both surveys.

The household sample for the **Integrated** Component was selected in a manner very similar to what was used for the Main Component; it was based on the Labour Force Survey frame.[3] However it was a fresh sample that was selected specifically for the NPHS as opposed to households already participating in the Labour Force Survey. The way that this plan was implemented for Cycle 1 was that for a certain portion of the NPHS household sample, a random selection of one person in the household was

---

Appendix 2 provides a short description of the Labour Force Survey.

In Québec the sample drawn for the integrated sample was a sub-sample of dwellings from a 1993 health survey (Enquête Sociale et de Santé) conducted by Santé Québec.

made, with no restriction on age; i.e., everyone in the household, including children, had an equal chance of being selected. If this selected person was a child aged 0 to 11, then this household was considered to be part of the **Integrated** sample and the NLSCY interview was administered to that household; otherwise the NPHS was conducted. There were approximately 2,700 NLSCY households selected for the Integrated Component.

The exclusions that were discussed above for the Main Component (the Yukon, Northwest Territories, institutions and Indian Reserves) also apply to the NPHS Component.

8.4.6 The Territories Component

The household sample for the **Main** and the **Integrated** Components were both based on the Labour Force sample frame which excludes the Yukon and Northwest Territories (NWT). However there was a requirement to have estimates for the north for both the NLSCY and the NPHS. Therefore the **Territories** Component was introduced which again was an integrated sample for both NLSCY and NPHS. The sample for the **Territories** Component was drawn from the population of private occupied dwellings. The Yukon sample excludes institutions and unorganized areas. The NWT sample has the same exclusions as well as very remote areas and very small communities.

In terms of implementation for the territories sample, if there was a least one child 0 to 11 in any of the selected dwellings, then the NLSCY was conducted for this dwelling. The goal for Cycle 1 was to produce a sample that would yield data for approximately 2,300 children living in each of the Yukon and NWT.

The **Territories** Component is somewhat different than the other components in that it is **fully** integrated with NPHS. For households with children the NLSCY was administered for children living in the household and as well one person was selected at random in each household for the NPHS. For Cycle 1, if that person was 12 years of age or older the NPHS was administered. As such, it was necessary to cut down on the content of both surveys in order to reduce respondent burden. As well, the collection methodology was somewhat different in that it was not computer-assisted interviewing that was used, but rather paper questionnaires.

In this first release of NLSCY data, only data from the 10 provinces have been included. Data for the territories have not yet been processed and will be part of a future release sometime during 1997. When the data for the Yukon and NWT are released, an updated version of this guide will be produced with a full discussion of the sample design and content for the **Territories** Component.

All further discussion in this current version of the microdata guide will be limited to the design and content for the 10 provinces (i.e., the Main the and Integrated Components).

8.4.8 The Child Sample

Once a sample of households was selected for the NLSCY the next step was to select children.

For the **Main** Component one child 0 to 11 years of age who lived the majority of the time in each selected household was selected at random. Then other children in the same economic family[4] as this selected child were selected at random up to a maximum of four children per household.[5]

For the **Integrated** Component a child had already been selected for the health survey as described above. As was done for the Main Component, additional children in the same economic family were selected at random to a maximum of four children aged 0 to 11.

8.6 Sample Allocation

The NLSCY sample for Cycle 1 was constructed taking two important requirements into consideration. A sufficient sample was required in each of the 10 provinces to allow for the production of reliable estimates for all children 0 to 11 years of age. The sample allocation was derived such that the smaller provinces had sufficient sample to meet this requirement.

A second requirement was that it was necessary to have a large enough sample to produce estimates at the Canada level by seven key age groupings or cohorts: 0 to 11 months, 1, 2 to 3, 4 to 5, 6 to 7, 8 to 9, and 10 to 11 years. These groupings will permit analysis every two years by these specific age cohorts while maintaining an overemphasis in the youngest age groups (0 to 11 months and one year olds) which was a requirement for the survey. For the NLSCY Main Component it was possible to oversample households which contained at least one child in

---

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 reasons of response burden, it was decided that a maximum of two children per Cycle 1 household will be followed longitudinally for Cycle 2.

the youngest two age groupings to allow for the sample requirements for these age groups.

8.8 Sample Size

The first cycle of the NLSCY resulted in a responding sample of 13,439 households. In these responding households 22,831 children 0 to 11 years of age were selected to participate in the survey. The following tables provide a breakdown of these children by province and by age.

PROVINCE	RESPONDING SAMPLE SIZE
Newfoundland	1,232
Prince Edward Island	764
Nova Scotia	1,532
New Brunswick	1,426
Québec	4,065
Ontario	6,020
Manitoba	1,789
Saskatchewan	1,878
Alberta	2,185
British Columbia	1,940
TOTAL[1]	22,831

AGE IN YEARS	RESPONDING SAMPLE SIZE
0	2,227
1	2,469
2	1,963
3	1,946
4	1,935

5	1,793
6	1,800
7	1,750
8	1,780
9	1,734
10	1,766
11	1,668
TOTAL	22,831

10.0

Data Collection

Data collection for Cycle 1 of the NLSCY took place between the fall of 1994 and spring of 1995. There were two major forums under which data were collected; namely the household collection and the school collection.

10.2 The Household Collection

For the household collection, data were collected from a variety of respondents using different data collection methods. Below is a description of each type of questionnaire used in the household.

The Household Roster

The first step was to complete a **Household Roster** for each household in the NLSCY sample with a knowledgable household member. This roster asked for basic demographic information for each household member as well as some questions on dwelling conditions. As part of the roster, a "relationship grid" was completed. This grid was used to establish the relationship of everyone in the household to everyone else in the household. Using this relationship information it was possible to derive a series of variables to describe the family situation of the child as discussed in Section 8.4.

Once the household roster was completed, the computer system randomly selected one child 0 to 11 years of age living in the household. A question was asked as to which person in the household was the **Person Most Knowledgeable** about that child. This person was labelled as the **PMK** for this household. In most cases the PMK was the mother of the child.[6] More information about the PMK is presented in Section 8.3.

The PMK was then asked to complete a set of three questionnaires: the General Questionnaire, the Parent Questionnaire, and the Child Questionnaire. Proxy reporting was permitted as discussed in Section 5.5.

The Parent Questionnaire

[6] *Since the PMK was in most cases the mother of the child, in this document the PMK will be referenced as "she".*

The purpose of the Parent Questionnaire was to gather general health information for both the PMK and her spouse/partner and to get some general information on the child's social environment including mental health of the PMK, social support, family functioning and characteristics of the neighbourhood.

The General Questionnaire

The General Questionnaire was completed for both the PMK and the spouse/partner of the PMK. The purpose of the General Questionnaire was to collect socio-economic information for the PMK and spouse/partner. Topic areas included education, labour force and income.

In future cycles of the NLSCY, the Parent and the General Questionnaires will be combined to form one survey instrument, for each of the PMK and spouse/partner.

The Child's Questionnaire

The Child's Questionnaire was completed for selected children in the household, aged newborn to 11 years. Children in the same economic family[7] as the originally selected child (on the household roster) were selected at random to a maximum of four (including the originally selected child). A maximum of four was used for respondent burden reasons. Topic areas on the Child's Questionnaire included among others, health, behaviour, education, literacy, parenting, child care and custody history.

The PPVT-R

The Peabody Picture Vocabulary Test - Revised (PPVT-R) was administered by the interviewer to each selected child aged four to five. Verbal permission was requested of the PMK before the test was administered. The purpose of the test was to measure school readiness of the child. More information about this test is presented in Section 9.21.

Once the entire NLSCY interview had been completed and the Interviewer had left the household she/he completed an assessment questionnaire to assess the conditions under which the test was administered to indicate factors which may have influenced the child's responses and his/her overall reaction to the test.

The 10-11 Questionnaire

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.

This questionnaire was self-completed by each child aged 10 to 11 selected for the NLSCY sample. When the PMK gave permission, the Interviewer provided the questionnaire to the child and encouraged the child to complete the questionnaire in a private setting. Upon completion, the questionnaire was sealed in an envelope to ensure confidentiality of the child. The parent was not permitted to see the child's completed questionnaire. She was informed of this before she gave permission for the child to complete the questionnaire. It was hoped that this procedure would increase the likelihood that the child would provide accurate and honest information.

The objective of this questionnaire was to collect information directly from the child on a variety of aspects of his/her life in order to supplement, and in subsequent analyses, compare with information obtained from the parent and teacher. Some of the topic areas covered were friends and family, school, feelings and behaviours, smoking and drinking and activities.

Neighbourhood Observation by Interviewer

Once the Interviewer left the respondent's home, she/he completed a questionnaire (on the computer) giving her/his perceptions of the neighbourhood in which the respondent resided. This information is intended to supplement the information on the neighbourhood provided by the PMK.

All of the information for the household collection (except for the 10 to 11 Questionnaire) was collected in a face-to-face or telephone interview using computer-assisted interviewing (CAI). Questions were asked to the respondent in the home or by telephone and directly entered into a computer by the interviewer. This made it possible to perform some on the spot edits and basic quality checks to detect errors and to correct errors with the help of the respondent if necessary.

More information about the content of these various questionnaires included in this first release of NLSCY data can be found in Section 9 of this document.

10.4 The School Collection

The school collection was another very important element of the NLSCY. For all children in the Cycle 1 sample who were attending school, the PMK was asked to give written permission to allow for information to be collected from the child's teacher and principal. In cases where the child was in grade 2 or above the PMK was asked to give permission to allow the teacher to administer a short mathematics computation test to the child.

Thus, the school collection involved three questionnaires. These questionnaires were mailed out to teachers and principals, who were asked to complete the questionnaires and mail them back to Statistics Canada in the envelopes provided.

The Teacher's Questionnaire

The goal of the Teacher's Questionnaire was to collect information about the child's academic achievement and behaviour at school, as well as information on characteristics of the class and the teacher's instructional practices.

The Principal's Questionnaire

The goal of the Principal's Questionnaire was to gather information on the school environment in order to assess how this may impact child development. Consequently, the Principal's Questionnaire collected information on school policies, resources and educational climate, rather than data about a specific child.

The Math Computation Test

The math test that the teacher was to administer was a shortened version of the Mathematics Computation Test of the standardized Canadian Achievement Tests, Second Edition (CAT/2). CAT/2 is a series of tests designed to measure achievement in basic academic skills.

10.6 Computer-Assisted Interviewing

Data collection for the NLSCY relied heavily on computer-assisted personal interviewing (CAPI) technology. The CAPI system has two main parts; Case Management and the survey specific part.

The Case Management system controls the case assignment and data transmission for the survey. For the NLSCY, a case refers to a household selected for the NLSCY sample. The Case Management system also automatically records management information for each contact (or attempted contact) with respondents, and provides reports for the management of the collection process.

The Case Management system routes the questionnaire applications and sample file from headquarters to the regional offices, and from the regional offices to the interviewer laptops. The returning data take the reverse

route. All data is encrypted for transmission, and the data are unencrypted only once resident on a separate secure computer with no external access.

The survey-specific part of CAPI includes an introductory component with procedures for contact and selection of households. Once a contact has been made and household composition has been established, the CAPI system generates applicable questionnaire components dependent on the household composition and the outcome of the selection procedures. For Cycle 1 of the NLSCY, some of the specific components that were generated included a Parent and General Questionnaire for the PMK and spouse/partner and Child's Questionnaire for up to four children aged 0 to 11 in each household. These components are discussed in greater detail in Section 5.1.

The use of CAPI technology allowed for high quality collection of complex population-specific content sections. For example, the system facilitated the collection of the relationships of all household members to each other (i.e., the relationship grid). This wealth of information will enable a detailed analysis of family structures, an important concept for analysis of the child information. This type of collection would be very difficult to implement in a paper and pencil environment.

10.8 Survey Timing

The initial plan was to have four collection periods for the household collection. Data for the **Main** Component were to be collected in December 1994 and February 1995 and data for the **Integrated** Component in November 1994 and March 1995, coinciding with NPHS selection periods. The main and integrated samples were split between the two potential collection periods. Each of the four collection periods lasted approximately two weeks.

Once collection actually started it was found that the response rate was not as high as originally hoped. Two back-up procedures were put in place to alleviate this situation. One was to allow for the sample to be **carried forward** to a future collection period in the case of a non-response. For example if in December a household could not be reached because no one was at home for the entire collection period, then this case was sent out again with the February sample and further attempts were made at that time to contact the household.

At the end of the four collection periods it was decided that the response rate could still be improved if more effort were placed on converting non-respondents. In June 1995 all non-responding cases were sent out again to see if these households could be converted to respondents. Statistics Canada's Regional Offices were asked to assign the "best" interviewers to these cases.

The school collection took place from March to June 1995. First, questionnaire packages were mailed to teachers and principals with instructions on how the various instruments should be completed. Approximately one week after the initial mailing a postcard was sent out to thank all respondents and to remind those who had not yet responded to do so. Roughly two weeks later, a second questionnaire package was sent out to teachers and principals who still had not responded. Finally three weeks later non-responding teachers and principals were contacted by telephone and encouraged to participate. It should be noted the school collection was not attempted for households converted in the June follow-up since by this time it was far too late in the school year to allow for this collection to take place.

10.10 Proxy Reporting

For the Child Questionnaire, the Parent Questionnaire and the General Questionnaire,[8] it was intended that the respondent should be the PMK, and indeed in most cases it was the PMK who completed these instruments. However, in some circumstances information was accepted from another household member. For example, if the PMK was away from home for the duration of the interview period, then information was accepted from another household member. For the Parent and General Questionnaires, the spouse/partner sometimes wanted to provide his own information. The following is a summary of who provided the information for each of these questionnaires.

PROXY REPORTING FOR THE CHILD, PARENT AND GENERAL QUESTIONNAIRES

	PMK RESPONDENT	SPOUSE RESPONDENT	OTHER HOUSEHOLD MEMBER RESPONDENT
Child Questionnaire	97.7%	2.2%	0.1%
Parent Questionnaire for the PMK	99.1%	0.8%	0.1%
Parent Questionnaire for the spouse/partner	92.5%	7.5%	-

See Section 5.1 for a description of these questionnaires.

General Questionnaire for the PMK	98.1%	1.6%	0.3%
General Questionnaire for the spouse/partner	86.4%	13.5%	0.1%

For all other questionnaires that were part of the NLSCY (i.e., the 10 to 11 Questionnaire, the PPVT-R, and the Teacher and Principal Questionnaires) proxy reporting was not permitted.

10.12 Interview Length

For the household collection, the interview length for responding NLSCY households was approximately two hours.

The total amount of time that it took to complete the major questionnaires that were part of the NLSCY household collection are presented in the table below. The table gives median interview times (i.e., the time at which 50% of the cases took more time and 50% took less). It should be noted that all extreme times (high and low) were removed before these times were derived.

TOTAL INTERVIEW TIMES IN MINUTES

QUESTIONNAIRE	INTERVIEW TIME FOR RESPONDING HOUSEHOLDS
All questionnaires in the household interview	119
All Child Questionnaires for the household	44
All Parent Questionnaires for the household (for the PMK and spouse/partner)	11
All General Questionnaires for the household (for the PMK and spouse/partner)	14
Total for major components (Child, Parent, General & PPVT)	74

Remaining Components[2]	39

The following table gives the median interview times for various family scenarios. The number of selected children (0 to 11) in the household was the factor that had the strongest impact on interview length. For households for which the PMK had a spouse/partner and four children, the interview length was well over three hours, and in some cases took over four hours. For Cycle 2 of the NLSCY a decision has been made to complete child interviews for a maximum of two children per household in order to reduce response burden.

TOTAL INTERVIEW TIMES BY FAMILY TYPE

FAMILY TYPE	TIME IN MINUTES
PMK, spouse and 1 child	95
PMK, spouse and 2 children	134
PMK, spouse, and 3 children	169
PMK, spouse, and 4 children	200
PMK, no spouse, and 1 child	85
PMK, no spouse, and 2 children	127
PMK, no spouse, and 3 children	161
PMK, no spouse, and 4 children	182

10.14 Interview Training, Supervision and Control

The NLSCY was conducted by Labour Force Survey interviewers. All LFS interviewers are under the supervision of a staff of senior interviewers who are responsible for ensuring that interviewers are familiar with the concepts and procedures involved in the survey, and also for periodically monitoring their interviewers and reviewing their completed documents. 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 the senior and reassigned. The senior interviewers are, in turn, under the supervision of the LFS program managers, located in Statistics Canada regional offices.

For the NLSCY a combination of classroom training and self-study materials were prepared to ensure that interviewers had a proper understanding of survey concepts. The self-study involved the interviewers reading the Interviewer's Manual prepared for the survey and completing home study exercises. During the classroom portion of the training, a program manager or a senior interviewer presented an overview of the survey, went through a mock interview with the participants, gave more specific training on administering the PPVT-R and presented exercises to help interviewers minimize non-response. In total, 14 hours were devoted to these training activities for each interviewer.

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

12.2 CAI Editing

As discussed in Section 5.1, all of the information for the household collection (except for the 10 to 11 Questionnaire) 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 collected information.

Review screens were created for important and complex information. For example, the selection procedures for the PMK, a critical aspect of the survey, were based on the household roster, composed of a demographic record for each household member, and the relationships of each household member to each other household member. As these are critical items for the NLSCY, the collected information was displayed for confirmation with the respondent before continuing the interview.

Range checks were used for continuous variables, to confirm or correct unusual answers during collection. For example, a question was asked about the weight of the child at birth. If the respondent gave a weight that was either significantly high or low, the interviewer was given an instruction to confirm the answer with the respondent.

All flow patterns were automatically built into the CAI system. For example, in the Child Care Section, an opening question was asked if the PMK used daycare or babysitting for the child to allow her (and her spouse/partner) to work or study. If she did, the CAI system continued with a series of questions about the specific care method(s) she used for the child. If not, the CAI system automatically skipped this series of questions.

Some consistency edits were included as part of the CAI system, and interviewers were instructed to "slide back" to the applicable question to correct for inconsistencies. Instructions were displayed to interviewers for handling or correcting problems such as incomplete or incorrect data. For example, in the collection of the Labour Force Section, start and end dates for jobs were critical for continuing the topic. For cases with incomplete date information, the system informed the interviewers that dates were required, and that the rest of the section would be skipped if they did not

enter the required dates. The interviewer was allowed to slide back to the appropriate date field to make a correction.

12.4 Data Capture

There were some questionnaires for the NLSCY that did not make use of computer-assisted interviewing; namely the 10 to 11 Self-complete Questionnaire, the Teacher's Questionnaire and the Principal's Questionnaire. All three of these questionnaires were completed directly by a survey respondent. A brief description of these questionnaires is given in Sections 5.1 and 5.2.

Capture of data for these three questionnaires was accomplished using minicomputers located in each of Statistics Canada's Regional Offices.[9] During this process, any document containing at least one respondent-completed item was captured and an unedited version of the computer record was electronically transmitted to Head Office for further processing. As part of the data capture system there were some quality checks built in to flag unusual entries to warn the capture operator of potentially incorrect entries. As well, the capture systems were built to follow the flow of the various questionnaires and would automatically proceed to the correct question in a skip pattern. The data capture operator however, was permitted to back-up and enter off-path information if that was what appeared on the questionnaire. Capture operators were instructed to use "head-down" keying and enter what they saw (i.e., whatever was on the questionnaires). Mistakes on the questionnaire were then edited at a later stage.

12.6 Minimum Completion Requirements

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

In some cases there was no NLSCY information 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. For these

---

The capture system used was Statistics Canada's generic DC2 system - Data Collection and Capture.

cases where there was no information collected for a household, the household was dropped from the NLSCY file and the sampling weights for responding households were inflated to account for these "dropped" households. This procedure is discussed in detail in Section 7.

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 recontact the respondent.

It was necessary to come up with a 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. In order to make this assessment the data collected for each selected child in the household were examined. This was done by looking at certain key questions across the Child Questionnaire. An assessment was made as to whether or not there was an adequate amount of information collected for at least one child in each household. If there was, this household was maintained in the responding sample. All missing variables for this household were set to not-stated or imputed. If there was not adequate information for at least one child then the household was dropped from the responding sample and treated as a non-response.

A **child response code** was formed for each child record on the NLSCY file by looking at key questions across the Child's Questionnaire. The questions that were considered were dependent on age since content varied considerably by age. There were 7 to 8 "key" questions chosen (in a somewhat random fashion) for each age group.

The **child response code** can be used as a measurement of data quality and was used to determine which child records were "good enough to keep".

The **child response code** should be interpreted in the following way:

CHILD RESPONSE CODES

RESPONSE CODE	DESCRIPTION
000	the record has a valid value for all key fields
001	the record has an invalid code (refusal, don't know or not answered) to at least one key field but there is enough information on the record to consider it to be "acceptable"

002	the record has at least one valid value for the key fields but there is **not** enough information to consider the record as "acceptable"
003	the record does not have a valid value on any key fields but the child record was started
004	the record was not started

"Acceptable" and "non-acceptable" were defined as follows.

Calculate:

$$R = \frac{\text{(valid responses to key questions) + (don't know's to key questions)}}{\text{number of key questions}}$$

$$D = \frac{\text{Don't knows to key questions}}{\text{number of key questions}}$$

If R>50% and D<30% the record is acceptable. Otherwise it is not. For a household to be considered a responding NLSCY household there had to be at least one acceptable child record.

The following are the number of child records by response code:

CHILD RESPONSE CODES

RESPONSE CODE	# CHILD RECORDS
000	22,183
001	563
002	140
003	28
004	1,806

In total 22,746 child records were determined to be "good enough to keep" (codes 000 and 001). These children came from 13,439 households, which is the number of households maintained on the NLSCY file. All the appropriate questionnaires were maintained for these responding households. Variables on missing questionnaires for the household were imputed or set to not-stated. There were 22,831 child records for these responding households. Out of these, there were 85 child records that were "not acceptable" but were kept because there was at least one "acceptable" child record for the household.

12.8 Head Office Editing

For the CAI questionnaires for the NLSCY there were two stages of editing conducted.

Pre-edit

The purpose of the Pre-edit was to carry out some basic formatting and preliminary editing. The following are some of the procedures that were carried out:

- Small data base files were created for each section of each questionnaire. A record was created for the section only if the section was applicable. For example, the section on temperament was only applicable for children 3 months to 3 years old. Therefore a temperament record was only created for children in this age group.[10]

- Within several sections, different wording was used for different age groups. For example, in the temperament section, for 0 to 2 year-olds, Question 4 was "How easy is it for you to know what's bothering him/her when he/she cries or fusses?" For 3 year-olds question was "How easy is it for you to know what's

---

For this first release of NLSCY data, for the final output file, complete child records with data for every section of each of the questionnaires were created. If a section is not applicable for a child all of the variables for the section have been set to not-applicable. In subsequent releases however, a series of database files will be released for each section. A record will exist for the section only if the section was applicable. The appropriate software will be provided so that users can easily link variables across files. This will be a far more efficient way to store and manipulate the NLSCY data.

bothering him/her when he/she is irritable?" Initially these questions were stored as separate variables. As part of the pre-edit the different versions were collapsed into one output variable. On the record layout (in Section 14), the various wording for the questions are all given.

- The skip pattens for each section were processed. Codes were set up to distinguish between answers which were valid, not-applicable, refused, don't know, or not-stated. These codes are discussed in detail in Section 6.5.4.

Consistency Editing

After the pre-edit, consistency editing was carried out. The goal of consistency editing is to verify the relationship between two or more variables. For example, in the Socio-Demographic Section, for children who were not born in Canada, there was a question on what year they first immigrated to Canada (ASDCQ2B). 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 it was set to not-stated in the edit. Some of the other consistency editing that was done for the various sections of the questionnaire and any data quality concerns that were noted as a result of this editing are discussed in detail in Section 9 of this document.

For the questionnaires that were collected using a paper version, essentially the same steps of editing were carried out. In the pre-edit, however there was an additional requirement. In some cases a value was captured that was not allowable for a particular item. This was possible due to the fact the data capture operator was given the ability to overwrite the capture edits. These invalid entires were set to a "missing value" in the pre-edit. Another difference is that editing for flow patterns was carried out at the consistency editing stage for the paper questionnaires.

12.10 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 interpretation of the data more straight-forward for the user. These standards are described in this section.

12.10.2 Naming Convention for Variables

In the NLSCY microdata file a naming convention has been used for each variable 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.

The variable names are of the following format:

A SE C Q nnx

where:

A: refers to the NLSCY cycle. "A" means the first cycle, "B" the second, "C" the third etc. Obviously for this first release all variable names will start with an "A".

SE: refers to the **se**ction of the questionnaire where the question was asked or the section from which the variable was derived. The table in Section 6.5.2 gives the acronyms which are used for the sections included in this first release of Cycle 1 data. More information about the content for each of these sections can be found in Section 9.

C: refers to the collection unit or the unit to which the variable refers. There are four possibilities:[11]

C means the variable refers to the child
P means the variable refers to the PMK
S means the variable refers to the spouse/partner
H means the variable refers to the household

---

*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 microdata 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. This issue is discussed further in Section 8.2.*

Q: refers to the variable type. There are four possibilities:

- **Q** means the variable refers to a question that was asked directly on one of the NLSCY questionnaires
- **S** means that the variable refers to a score calculated for one of the scales used on the questionnaire (See Section 9.1)
- **D** means the variable was derived from other questions that were asked on the questionnaire (See Section 6.8)
- **I** means the variable is a flag created to indicate that an item has been imputed (See Section 6.7)

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

12.10.4 Acronym Names for Questionnaire Sections

The following table gives the acronym names that were used for each section of the various NLSCY questionnaires. As explained in Section 6.5.1 this acronym is embedded in the variable name for all variables on the NLSCY microdata file. The acronym is the second and third characters of the variable name.

ACRONYM	SECTION
MM	Variables collected as part of the household roster. Basic demographic variables were collected for each household member. These variables are included on the NLSCY microdata file for the child, the PMK and the spouse/partner.
SD	Socio-demographic variables: - collected for the child on the Child's Questionnaire and for the PMK and spouse/partner on the General Questionnaire.
DM	Demographic variables derived to explain the living arrangements of the child: - derived from information of the household roster and relationship grid.
MD	Medical/biological variables: - asked for children 0 to 3 years of age on the Child's Questionnaire.

TM	Temperament variables. - asked for children aged 3 months to 3 years old, on the Child's Questionnaire.
ED	Education variables. - asked for children 4 to 11 years old on the Child's Questionnaire and about the PMK and spouse/partner on the General Questionnaire.
BE	Behaviour variables: - asked for children 0 to 11 years, on the Child's Questionnaire.
MS	Motor and social development variables: - asked for children 0 to 3 years old, on the Child's Questionnaire.
RL	Social relationship variables: - asked for children 4 to 11 years old, on the Child's Questionnaire.
PR	Parenting style variables: - asked for children 0 to 11, on the Child's Questionnaire.
CR	Child care variables: - collected for children 0 to 11 on the Child's Questionnaire.
A1	Variables from Section A of the 10 to 11 Self-complete Questionnaire: - Section A pertains to friends and family.
D1	Variables from Section D of the 10 to 11 Self-complete Questionnaire: - Section D pertains to behaviour from the child's perspective.
E1	Variables from Section E of the 10 to 11 Self-complete Questionnaire: - Section E pertains to parenting style from the child's perspective.
PP	Variables from the PPVT test: - administered to children in the 4 to 5 age group.
PA	Variables from the PPVT assessment: - answered by the interviewer to describe the conditions under which the PPVT was administered to the child.
DP	Depression scale variables:

- this scale was administered to the PMK, on the Parent Questionnaire.

FN — Family functioning scale variables:
- this scale was administered to the PMK or spouse/partner on the Parent Questionnaire, to measure how family members relate to each other.

SP — Social support scale variables:
- this scale was administered to the PMK or spouse/partner, on the Parent Questionnaire, to determine the availability of social supports.

LF — Labour force variables:
- collected for both the PMK and spouse/partner on the General Questionnaire.

IN — Income variables:
- household income and personal income of the PMK, collected on the General Questionnaire.

MA — Math computation test variables:
- administered to children in grade 2 and over.

GE — Geographic Variables:
- derived from sample information.

12.10.6 Examples of Variable Names

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

ALFSQ2 — This refers to Q2 in the Labour Force Section for the spouse/partner.
The "A" indicates it is a Cycle 1 variable.
The "LF" indicates the Labour Force Section.
The "" indicates it refers to the spouse/partner.
The "Q" indicates it was an item asked directly on the questionnaire.
The "2" is the ID of the item.

APRCS03 — This is a positive interaction score on the parenting scale for a 2 to 11 year-old child.
The "A" indicates it is a Cycle 1 variable.
The "PR" indicates the Parenting Section.
The "C" indicates it refers to the child.
The "S" indicates the variable refers to a score.
The "03" is the ID of the variable.

12.10.8 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 microdata 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: In other cases 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 microdata 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 microdata file to indicate that a question or derived variable is not applicable.

1/ 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 (ACRCQ1G1) is only applicable for children for whom this type of care is used (ACRCQ1G=1). Otherwise there will be a code 996 for this question.

2/ 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 on the microdata file 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.

1/ 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.

2/ In some cases a specific 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.

3/ 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. See Section 6.4 for more information about consistency editing.

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.

12.12 Coding of Open-ended Questions

A few data items on the NLSCY questionnaire were recorded by interviewers in an open-ended format. For example, in the Labour Force Section, if a PMK had worked in the previous 12 months, she was asked to identify a main job. Then there were a series of open-ended questions about this main job:

> Thinking about this main job, what kind of business, service or industry is this?
>
> Again, thinking about this main job, what kind of work were you doing?
>
> In this work, what were your most important duties or activities?

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 his work.

The coding systems used were the 1980 Standard Occupational Classification codes (SOC) and the 1980 Standard Industrial Classification codes (SIC). Grouped versions of these codes are available on the microdata file (ALFPD07 and ALFPD08 for the PMK, and ALFSD07 and ALFSD08 for the spouse/partner).

12.14 Imputation

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. In Section 6.5.4, the various codes that have been used to describe the reason for the item non-response ("refusal", "don't know", "not stated") are described.

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 for some of the scales that were administered. The methods used for imputation for these variables are described in detail in Section 9. 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 microdata file have an "**I**" as the fifth character of the variable name. For example, the name of the imputation flag for household income (AINHQ03) is AINH**I**03A.

12.16 Creation of Derived Variables

A number of data items on the microdata file have been derived by combining items on the questionnaire in order to facilitate data analysis. For example, in the section on child care, the PMK was asked a series of questions about the types of care she used for the child to allow her and her spouse/partner to work or study. For each type of care there was a question on the number of hours per week the child was in that type of care. Using this information, a variable was formed to indicate the primary care arrangement used to allow the PMK and spouse/partner to work or study. It was derived by looking at the number of hours for each care arrangement and setting it to the method for which the number of hours was the greatest.

All derived variables on the NLSCY microdata file have a "**D**" as the fifth character of the variable name. The name of the variable for the primary care arrangement is ACRC**D**01.

14.0

Weighting

The principle behind estimation in a probability sample such as the NLSCY is that each person in the sample "represents", besides himself or herself, several other persons not in the sample. For example, in a 2% simple random sample of the population, each person in the sample represents 50 persons in the population.

The weighting phase is a step which calculates, for each record, what this number is (i.e., the number of individuals in the population represented by this record). This weight appears on the NLSCY microdata file (AWTCW01), and must be used to derive meaningful estimates from the survey. For example, if the number of children living in single parent families is to be estimated, it is done by selecting the records referring to those individuals in the sample with that characteristic and summing the weights found on those records.

Since the NLSCY is based on the Labour Force Survey frame, the derivation of weights for the survey records is clearly tied to the weighting procedure used for the LFS. The LFS weighting procedure is briefly described below.

14.2 Weighting Procedures for the LFS

In the LFS, the subweight attached to each record is the product of the following factors: the basic weight, the cluster sub-weight, the balancing factor for non-response, and the rural-urban factor. These various factors are described below.

Basic Weight

In a probability sample, the sample design itself determines weights which must be used to produce unbiased estimates of the population. Each record must be weighted by the inverse of the probability of selecting the person to whom the record refers. In the example of a 2% simple random sample, this probability would be .02 for each person and the records must be weighted by 1/.02=50. Because all eligible individuals in a dwelling are interviewed for the LFS (directly or by proxy), this probability is essentially the same as the probability with which the dwelling is selected.

Cluster Sub-weight

The cluster delineation is such that the number of dwellings in the sample increases very slightly with moderate growth in the housing stock. Substantial growth can be tolerated in an isolated cluster before the additional sample represents a field collection problem. However, if growth takes place in more than one cluster in an interviewer assignment, the cumulative effect of all increases may create a workload problem. In clusters where substantial growth has taken place, sub-sampling is used as a means of keeping interviewer assignments manageable. The cluster sub-weight represents the inverse of this sub-sampling ratio in clusters where sub-sampling has occurred.

Non-response

Notwithstanding the strict controls of the LFS, some non-response is inevitable, despite all the attempts made by the interviewers. The LFS non-response rate is approximately 5%. For certain types of non-response (eg. household temporarily absent, refusal), data from a previous month's interview with the household if any, is brought forward and used as the current month's data for the household.

In other cases, non-response is compensated for by proportionally increasing the weights of responding households. The weight of each responding record is increased by the ratio of the number of households that should have been interviewed, divided by the number that were actually interviewed. This adjustment is done separately for geographic areas called balancing units. It is based on the assumption that the households that have been interviewed represent the characteristics of those that should have been interviewed. To the extent that this assumption is not true, the estimates will be somewhat biased.

Rural-urban Factor

In NSRUs without sufficient rural and urban population for explicit urban and rural strata to be formed, each primary sampling unit (PSU) is composed of both urban and rural parts. Information concerning the total population in rural and urban areas is available from the 1981 Census for each PSU as well as for each economic region (ER) in which explicit urban/rural stratification is not done. Comparison by ER with the actual 1981 rural or urban census counts indicates whether the selected PSUs over- or under-represent the respective areas. The ratio of actual rural-urban counts is divided by the corresponding estimates. These two factors are computed for each relevant ER at the time of selection of the PSUs and are entered on each sample record according to the appropriate area (rural or urban) of the NSRU. Changes in these factors are incorporated at the time of PSU rotations.

LFS Sub-weight

The product of the previously described weighting factors is called the LFS sub-weight. All members of the same sampled dwelling have the same sub-weight.

14.4 Weighting Procedures for the Main and Integrated Components

The principles behind the calculation of the weights for the NLSCY for the Main and Integrated Components are similar to those for the LFS, since for the most part the NLSCY sample was based on the LFS sampling frame. In Section 4 more information is given about the sampling plans for the Main and Integrated Components.

Households that make up the NLSCY sample were in fact, drawn from four sampling frames.

For the Main Component, the sample was composed of households that were actually part of the LFS sample itself. The households selected for the NLSCY were originally introduced to the LFS sample between April and December of 1994. In October of 1994, a new sampling procedure was introduced for the LFS. Therefore NLSCY households selected from the LFS sample for April-September were part of the **old** LFS design and households selected from the samples for October to December were part of the **new** LFS design. The NLSCY weighting procedures used for the old and new designs differ slightly and were carried out independently as described in Section 7.2.1 and 7.2.2.

For the Integrated sample, (i.e., the sample integrated with the National Population Health Survey - NPHS) the sampling plan was based on the new LFS design for all provinces except for Québec. For these nine provinces a fresh sample of households was selected from the new LFS frame specifically for NPHS and NLSCY. These households had not been previously survey by the LFS. The Québec households included in the NPHS sample came from a frame that was constructed for the Enquête sociale et de santé (ESS) conducted by Santé Québec, in 1992-1993.

The calculation of weights for children is different depending on the sampling frame from which they were selected. The rest of this section is devoted to explaining the adjustments made to the sampling weights for the four sampling frames used the NLSCY, namely:

- Main Component - old LFS design
- Main Component - new LFS design

- Integrated Component - for all provinces except Québec
- Integrated Component - Québec

14.4.2 Main Component - Old LFS Design (all provinces)

In the paragraphs that follow, a number of weight correction factors are discussed. The first five corrections are corrections at the household level. These corrections are the same for all children in a given household. The sixth correction varies for each child selected within a household according to the child's age group and sex. These correction factors, once multiplied by the LFS sub-weight, will give the weights for children in households belonging to the sample for the Main Component - old LFS design. In the final stage, weights are calculated for the four sampling frames put together.

Additional corrections to the LFS sub-weight

All the corrections that follow, made to the LFS sub-weight, are intended to compensate for the particular features of the NLSCY.

Correction 1: Correction for number of rotation groups

The sample for the old LFS survey design is made up of six "rotation groups", each representing one-sixth of the Canadian population. In the NLSCY plan, a number of rotation groups were selected which varied according to the province and the type of household selected. The correction for the number of rotation groups is intended to bring all the estimates obtained into line with the national figure. Consequently, the correction takes the following form:

6 over {number`of ` rotation ` groups ` by ` province ` \and `` type ` of ` household}For all provinces except for Ontario and Alberta, six rotation groups were used. (For Ontario and Alberta, a sufficient number of households were obtained with five rotation groups).

For these eight provinces, households with at least one child aged 0 or 1 were selected from the six rotation groups. The other households, with only children aged 2 to 11 (inclusive), were selected from five rotation groups only. The following table shows the number of rotation groups by province and type of household:

NUMBER OF ROTATION GROUPS SAMPLED

	TYPE OF HOUSEHOLD	PROVINCE
Households with at least one child aged 0 or 1	Households with only children aged 2 to 11	All provinces except Ontario and Alberta
6	5	
Ontario and Alberta	5	4

Correction 2: Corrections for updating of sample frame

Between sample selection and collection, approximately three months elapsed. Thus, some dwellings that were vacant at the time of sampling were inhabited by eligible households at the time of collection. In addition, other households, identified as being outside the target population because they had no children in the target age group, became eligible at the time of collection. In order to take account of households that were eligible at the time of collection but were not selected at the time of sampling, the sub-weight of each of the responding households was adjusted by the following two multiplicative correction factors:

a) Correction for vacant dwellings at the time of sampling:

(`No. ` ` of ` ` vacant ` `dwellings ` `at ` `the ` `time ` ` of ` ` sampling `)+
#
{(`No. ` `of ` ` households ` `\in ` `target` ` population ` `at ` `time ` ` of ` `
collection ` ` \in ` `these ` `vacant ` `dwellings) }
over {No. ` ` of ` `vacant ` ` dwellings ` ` at ` `time ` `of ` `sampling}

b) Correction for household

olds outside the target population at the time of sampling:

(No. ` ` of ` ` households ` `outside ` `target` ` population ` ` at ` `time` ` of ` `sampling)+
{ (Number ` ` of ` ` these ` ` households` ` \in` ` target` ` population ` ` at ` `the ` ` time ` `of ` `collection`)}
over {No. ` `of ` ` households ` ` outside ` `target ` `population ` ` at ` ` the ` `time ` ` of ` `sampling}The correction factor for households occupying vacant dwellings at the time of sampling was 1.0054, while the one for households outside the target population was 1.0049.

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

Sometimes a household included more than one economic family, both with children in the target age group. When this occurred, the child selection procedure required the selection of one of these families at random. To take account of the families that were not selected, the sub-weight associated with such a household was multiplied by the number of economic families present in the household with at least one child in the target age group. This correction affected only four households.

Correction 4: Correction for households with more than four children 0 to 11 years of age

For this survey, a maximum of four children between 0 and 11 years of age were selected per household. If the economic family had more than four children, the children not selected were taken into account by multiplying the sub-weight of the household by a factor equal to the number of children 0 to 11 years of age in the economic family divided by four. This correction affected 42 households.

Correction 5: Correction for household non-response

In surveys such as the NLSCY, some households do not provide responses for a variety of reasons: refusal, special circumstances, language problems, temporary absence. This non-response is usually compensated for by proportionally correcting the sub-weights of the responding households. For the NLSCY, the correction was made by multiplying the sub-weight of the responding households by the following factor:

\sum ` ` of ` `adjusted ` ` weights ` ` of ` ` households ` ` sampled #
{ \in ` ` the ` `replicate ` `within ` `a ` ` stratum ` `of ` `the ` `NLSCY ` }
OVER {\sum ` `of ` `adjusted ` `weights ` `of ` `responding ` ` households } #
\in ` ` the ` ` replicate ` `within ` ` a ` ` stratum ` ` of ` ` the ` ` NLSCYThe adjusted weight is the LFS sub-weight multiplied by the first four correction factors. A different correction was made in each of the strata and replicates specially defined for non-response. The strata and replicates were defined using the following information: self-representing,

non-self-representing or special economic area; urban, rural or mixed area; included or not included in an apartment frame; and by group of rotation group. Each of the strata and replicates retained had to contain at least 10 households and have a response rate of at least 70%.

Correction 6: Correction for post-stratification

Post-stratification was carried out on the sub-weights adjusted by the first five correction factors to ensure that the national and provincial estimates agreed with the January 1995 demographic estimates of the population of children aged 0 to 11.

Post-stratification was done by province, age group (according to the seven main age groups used in the survey), and sex of child, and by census metropolitan area using the "raking ratio" method. Thus, for children belonging to a given domain (formed by their province, age group, sex and census metropolitan area), the ratio of the estimate after post-stratification to the estimate before post-stratification in the domain gave the correction factor for the post-stratification.

14.4.4 Main Component - New LFS Design (All Provinces)

The weight adjustments made for children in households belonging to the **Main Component - new design** were very similar to the ones made for children who were part of the old design as described above. The same six correction factors were applied to the LFS sub-weight.

Correction 1: Correction for number of rotation groups

Like the old design, the sample of the new LFS design is made up of six "rotation groups," each representing one-sixth of the Canadian population. For the NLSCY, three of these rotation groups from the new LFS design were selected. Therefore, the sub-weight of each LFS household was multiplied by 2.

Correction 2: Corrections for updating of sample frame

Again, correction factors were required because of the fact that between sample selection and collection, approximately three months elapsed.

Some dwellings that were vacant at the time of sampling were inhabited by households eligible at the time of collection. For the new design this correction factor was 1.0054.

In addition, other households, identified as being outside the target population because they had no children in the target age group, became eligible at the time of collection. For the new design this correction factor was 1.0049.

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

This correction did not take place for the new design. There were no households in the sample for the new design with more than one economic family, both with children in the target age group.

Correction 4: Correction for households with more than four children 0 to 11 years of age

Again a correction was required for households where there were more than four children 0 to 11 years of age (see Section 7.2.1). This correction affected 27 households for the new design.

Correction 5: Correction for household non-response

A household non-response factor was derived for households in the new plan in exactly the same way as was done for the old design (see Section 7.2.1). The only difference was in the definition of the strata and replicates used for the non-response adjustment, the characteristics used were urban, rural or mixed area; included or not included in an apartment frame; and by group of rotation group.

Correction 6: Correction for post-stratification

For the households in the new LFS design for Québec, post-stratification was carried out on the weights adjusted by the first five correction factors to ensure that estimates agreed with January 1995 demographic estimates of the population of children aged 0 to 11. The post-stratification was done by age group and sex of child, and by census metropolitan area using the "raking ratio" method. Thus, for children belonging to a given domain (formed by their age group, sex and census metropolitan area), the ratio of the estimate after post-stratification to the estimate before post-stratification in the domain gave the correction factor for the post-stratification.

For the other nine provinces, the post-stratification correction factor will be presented in the next section, which describes the NPHS sample frame. Because of certain common strata between these two sample frames, the two were combined and processed together for this final adjustment. This

strategy was necessary for calculating the variance estimate, since the two sample frames were not mutually independent.

14.4.6 Integrated Component (All Provinces Except Québec)

In the paragraphs that follow, the correction factors used for the Integrated Component (in all provinces except Québec) are presented. The Integrated sample selected for NPHS also used the new LFS survey plan, but a fresh sample households was selected i.e., not participating in the LFS.

As an initial weight, the NPHS sample uses the LFS basic weight. The first four corrections are corrections to this basic weight at the household level. These corrections are the same for all children in a given household. The fifth correction varies for each child selected in a household, according to his or her age group and sex. These correction factors, once multiplied by the LFS basic weight, will give the weights of the children in households belonging to the NPHS sample frame, before the final stage of weighting.

Initial corrections made to the LFS basic weight

The initial corrections that follow were made by statisticians of the NPHS project team. Only the list of corrections is presented here. For further details on each of them, the reader may consult the guide to the NPHS "1994-95 Public Use Microdata Files."

- Correction for rotation group
- Correction for cluster growth
- Correction for households not visited
- Correction for empty strata in LFS
- Correction for stabilization
- Correction for multiple dwellings

These corrections made to the LFS basic weight resulted in the NPHS "initial weight."

Additional corrections made to the NPHS initial weight

All the corrections that follow, were made to the NPHS initial weight, and are intended to compensate for the particular features of the NLSCY sample design.

Correction 1: Correction for integration into NPHS

The NPHS sample was constructed such that there were three sub-samples that contained households in the target population for the NLSCY, that is, households with at least one child aged 0 to 11. For one of these sub-samples, only adults were selected i.e., the children in these households had a zero probability of selection. In order to take account of these households not selected for the NLSCY, a multiplicative factor was applied to the households in the other two sub-samples of the NPHS.

Correction 2: Correction for households with multiple economic families

Again it was necessary to make a correction for households with two or more economic families each with children in the target age group (as described in Section 7.2.1). This correction affected four households.

Correction 3: Correction for households with more than four children 0 to 11 years of age

A correction was required for households where there were more than four children 0 to 11 years of age in the household (see Section 7.2.1). This correction affected seven households.

Correction 4: Correction for household non-response

A household non-response factor was derived for households in the NPHS plan in exactly the same way as was done for the new LFS design for the Main Component (see Section 7.2.2).

Correction 5: Correction for post-stratification

As noted in the previous section, the sample frame for the Main Component (new LFS design) and the NPHS sample frame (outside of Québec) are not mutually independent. Some strata were common to the two frames in the sampling. In order to take this situation into account in calculating the variance estimate, the sample from these two frames were combined after the correction for non-response was made. To take account of the relative contribution of each sample frame, a series of multiplicative "alpha" factors were derived by province and child's age group. In each domain representing a given province and age group, the "alpha" factor for the NPHS sample frame was calculated by dividing the sample size for the NPHS sample frame for that domain by the size of the sample for the two sample frames together for the domain. Similarly, "alpha" factors were calculated for the Main Component - new LFS design.

After the adjustments for the "alpha" factors were made, a post-stratification was carried out on the weights established up to this point in order to ensure that the provincial estimates agreed with the January

1995 demographic estimates of the population of children aged 0 to 11. The post-stratification was done by province (except for Québec), age group, and sex of child, and by census metropolitan area using the "raking ratio" method. Thus, for children belonging to a given domain (formed by their province, age group, sex and census metropolitan area), the ratio of the estimate after post-stratification to the estimate before post-stratification in the domain gave the correction factor for the post-stratification.

14.4.8 Integrated Component - Québec

In the paragraphs that follow, the correction factors used for the Integrated Component for Québec are presented. For Quebec, the NPHS used the Enquête sociale et de santé (ESS) conducted by Santé Québec.

Additional corrections made to the ESS weight

The corrections that follow were made by statisticians on the NPHS project team. These corrections were made to the ESS weight provided by Santé Québec. Only the list of corrections is presented here. For further details on each, the reader may consult the guide to the NPHS "1994-95 Public Use Microdata files."

- Correction for cluster growth
- Correction for multiple dwellings

These corrections were applied to the ESS weight to obtain the NPHS "initial weight" for Québec households.

Additional corrections made to the NPHS initial weight

All the corrections that follow, made to the NPHS initial weight, are intended to compensate for the particular features of the NLSCY sample design.

Correction 1: Correction for integration

For the Québec NPHS sample it was necessary to make the same integration adjustment that was made for the other provinces as described in Section 7.2.3.

Correction 2: Correction for households with multiple economic families

Again it was necessary to make a correction for households with two or more economic families each with children in the target age group (as described in Section 7.2.1). This correction affected only one household.

Correction 3: Correction for households with more than four children 0 to 11 years of age

A correction was required for households where there were more than four children 0 to 11 years of age in the household (see Section 7.2.1). This correction affected one household.

Correction 4: Correction for household non-response

A household non-response factor was derived for households in the NPHS design in Québec in exactly the same way as was done for the old LFS design for the Main Component (see Section 7.2.1).

Correction 5: Correction for post-stratification

Post-stratification was carried out on the weights established up to the end of correction factor 4 to ensure that estimates agree with the January 1995 demographic estimates of the population of children aged 0 to 11 in Québec. The post-stratification was done by age group and sex of the child, and by census metropolitan area using the "raking ratio" method. Thus, for children belonging to a given domain (formed by their age group, sex and census metropolitan area), the ratio of the estimate after post-stratification to the estimate before post-stratification in the domain gives the correction factor for post-stratification.

14.4.10 Final Stage of Weighting

Since the NLSCY survey design uses more than one sample frame, it was necessary to take account of the relative contribution of each sample frame to the final estimates in calculating weights for children.

For the nine provinces other than Québec, since the sample frames of the NPHS and the new LFS design of the Main Component were combined into one frame for the post-stratification, it was still necessary to determine the respective contributions of the frame from the old LFS design and this new combined frame. Multiplicative "beta" factors were calculated, according to the same principle as for the "alpha" factors used in the post-stratification, for each sample frame, province and child's age group.

For Québec, three "beta" factors were calculated for each of the three sample frames by child's age group.

By multiplying all the correction factors for each frame by the original weights of the children in responding households, the final weights were obtained. These weights can be found on the microdata file under the name AWTCW01.

There are many variables and concepts which are critical to the analysis of the NLSCY data. In this section there is a brief discussion regarding the types of analyses that are possible with the NLSCY data. This is followed by a description of key variables which have been derived to explain the living arrangements of the child and the socio-economic conditions under which the child lives.

The content areas for each section of the various questionnaires used for the first cycle of the NLSCY are presented in the next section.

16.2 Cross-sectional and Longitudinal Estimates

The NLSCY design and sample has been constructed so that it will be possible to produce both **cross-sectional** and **longitudinal** estimates. For now, with Cycle 1 data, only cross-sectional estimates are possible. Longitudinal information will be available in all subsequent cycles starting from the second cycle.

The allocation of the Cycle 1 sample was such that is will be possible to produce estimates at the national level for the specific age cohorts and at the provincial level for aggregated age groups. This is true for cross-sectional data as well as longitudinal data.

The **longitudinal** sample will be comprised of all children sampled for Cycle 1 of the survey in responding households. The plan is to follow these children over time every two years. Analyses of these children will permit researchers the opportunity to perform in-depth studies of the long term impact of risk factors (such as divorce or the onset of a health condition) and protective factors (such as positive interactions with parents or academic success at school) on these children as they move into adulthood. If a child moves out of the household where he or she was sampled at Cycle 1, that child will be traced to wherever he or she resides at future cycles of the survey. From a longitudinal perspective, the child, not the household, is the statistical unit of analysis.

It should be noted that some children who were participants in Cycle 1 of the NLSCY may not participate in the second or subsequent cycles due to a variety of reasons. This is usually referred to as attrition. The numbers of these children will be carefully monitored and every effort will be made to keep these numbers at a minimum. The Cycle 1 sample and its

allocation was designed with this in mind and as long as future response rates are not lower than expected the sample will still permit longitudinal research by age cohort at the national level.

In the second and subsequent cycles, it is intended that the NLSCY will add children belonging to age groups no longer covered in the longitudinal sample. For example, for Cycle 2 a panel of children 0 and 1 years of age will be added to the Cycle 2 sample. This augmented sample will allow for ongoing **cross-sectional** analyses to supplement the primary longitudinal research. As such, at each cycle it will be possible to get a snap-shot of Canadian children of all ages. At the present time, it is not planned that this augmented component of the sample will be followed longitudinally.

It should be noted the children who immigrate to Canada at any point of time after the Cycle 1 sample was selected and who are in the age cohorts covered in the Cycle 1 sample, will not be included in either cross-sectional or longitudinal estimates. Estimates of the number of children immigrating to Canada will be monitored and a decision may be made in the future to introduce a new sample into the NLSCY to cover these children.

16.4 NLSCY Units of Analyses

The unit of analysis for the NLSCY is intended to be the child and eventually the young adult. For each cycle of the NLSCY, extensive information will be gathered on the child's family, parent(s), and neighbourhood.

It is true that families or households are relatively straightforward units of analysis with cross-sectional data but the situation becomes quite problematic with longitudinal data. Households change composition frequently, due to divorce of parents, or children leaving the parental nest. Attempts have been made in other studies to define "longitudinal households" but the implementation of this concept has never been straightforward. No single definition has been found to be appropriate for most analytic tasks, and many definitions exclude the portion of the population that has undergone the change. Unfortunately, this is often a significant as well as interesting population to study. It has been suggested that a superior alternative is to use the individual as the unit of analysis and present family and household variables as a characteristic of the individual.[12]

For a more complete discussion of units of analyses for longitudinal studies see Duncan, G.D. and Hill M.S. (1985). Conceptions of Longitudinal Households: Fertile or Futile? *Journal of Economic*

Thus the file which has been constructed for this first release of NLSCY data consists of child records. In order to understand the family situation, estimates such as of the number of children in single parent families, or the number of children living in low income households, can be produced.

16.6 PMK and Spouse

In each NLSCY household, one child 0 to 11 years of age was selected at random and a question was asked about who in the household was the **p**erson **m**ost **k**nowledgeable about this child. This person was labelled as the **PMK**. The intention was that the PMK would provide the information for all selected children in the household and then give socio-demographic information about herself and 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.

The following is the breakdown of the relationship of the PMK to the NLSCY children for Cycle 1.

- for 91.3% of responding children, the PMK was the mother (89.9% the biological mother and 1.4% the step, adoptive or foster mother)
- for 8.2% of the children the PMK was the father
- for 0.5% of children the PMK was not a parent.[13]

When the PMK was not a parent, for the majority of cases 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. Only 0.1% of the children did not live with a parent.

If the PMK had a partner residing in the household at the time of the interview, then this person was labelled 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.

---

and Social Measurement, 13:361-375.

These numbers for the PMK and spouse/partner are based on unweighted data.

The following is the breakdown of the relationship of the spouse/partner to the NLSCY children.

- for 14.4% of the children, the PMK did not have a spouse/partner residing in the household
- for 78.1% of children the spouse/partner was the father (73.2% the biological father and 4.9% the step, adoptive or foster father)
- for 7.2% of children the spouse/partner was the mother (biological, step, adoptive or foster)
- for the remaining 0.3% of children, the spouse/partner was not a parent.

16.8 Family Derived Variables

Using NLSCY data, a child's family may be described in several different ways. Many of the family variables that have been used to describe the NLSCY children were derived from what is known as the relationship grid. As part of the household roster 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 for the child that exist on this first microdata file for the NLSCY. The names of the derived variable are given in brackets.

Single-parent family

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

Using the relationship grid, a child's **single-parent status** was derived. There were 84.2% of children living with two parents, 15.7% with one parent and 0.1% without a parent[14] (ADMCD04).

A child's parent status can also be defined in terms of the PMK. There were 84.3% of the NLSCY children living in a household where the PMK had a spouse/partner; and for 15.7% of children the PMK did not have a spouse/partner (ADMPD06A).

These estimates for family derived variables are based on weighted data.

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.

Step, Blended and Intact 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.[15]

Intact family

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

For the NLSCY children, 75.5% were a member of an intact family (ADMCD16).

Step family

A step family consists of a married or common-law couple residing in the same household, with at least one step child living with them who is the biological or adopted child of one parent but not the other parent. 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, 4.6% were step children themselves (ADMCD03) and 8.6% lived in a step family (ADMCD15).

Blended family

Blended families combine children who have different relationships with their parents. 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

---

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 adoptive 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.

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)

- 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, 6.1% were members of a blended family (ADMCD14).

Economic Family

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

For the NLSCY data, 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 included on the microdata file. In the case of common-law relationships, if both members have brought their own children into the relationship then these children are considered as siblings. It should be noted that the classification of siblings was age independent. If an NLSCY child had an adult sibling (for example, 21 years of age) living in the household then this sibling was 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 (ADMCD08, 09, 10 and 11).

16.10 Socio-Economic Derived Variables

There were two derived variables produced from Cycle 1 data to assist analysts in understanding and explaining the socio-economic situation of the child's family.

Socio-economic Status (AINHD08)

Sociologists often use the term "socio-economic status" (SES) to refer to the relative position of a family or individual in an hierarchical social structure, based on their access to, or control over, wealth, prestige and power. In studies of children's academic and social-emotional development, SES is often operationally defined through measures describing the occupational prestige, educational levels, and economic positions of children's parents.

For the first cycle of the NLSCY a measure of SES was derived for each household in the sample and the result assigned to each selected child in that household.[16] It was derived from five sources: the level of education of the PMK, the level of education of the spouse/partner, the prestige of the PMK's occupation, the prestige of the occupation of the spouse/partner, and household income. The method of constructing each component of SES, and the construction of the overall SES measure are described below.

Education - Years of School

The education variable used in the construction of SES was years of schooling. Two such variables were derived independently; one for the PMK and one for the spouse/partner (AED**P**D04 for the **P**MK and AED**S**D04 for the **S**pouse/partner). For the PMK the years of schooling variable was derived based on items AEDPQ01 (years of elementary and high school) and AEDPQ04 (highest level of education attained beyond high school). To create a somewhat continuous interval-level education variable, these two items were recoded to form years of schooling in the following manner:[17]

This particular definition of SES was proposed by Dr. Douglas Willms, Atlantic Centre for Policy Research in Education. University of New Brunswick.

In cases where the PMK had not graduated from high school but had completed a post-secondary degree or certificate, then the post-secondary degree or certificate took precedence. For

AEDPD04	Condition
00	AEDPQ01=1 (no schooling)
03	AEDPQ01=2 (1 to 5 years)
06	AEDPQ01=3 (6 years)
07	AEDPQ01=4 (7 years)
08	AEDPQ01=5 (8 years)
09	AEDPQ01=6 (9 years)
10	AEDPQ01=7 (10 years)
11	AEDPQ01=8 (11 years)
12	AEDPQ01=9 (12 years)
13	AEDPQ01=10 (13 years)
16	AEDPQ04=6 (BA/BSC)
18	AEDPQ04=7 (Masters)
20	AEDPQ04=8 or 9 (MD/PHD)

An extra year was then added to AEDPD04 if the PMK had a diploma from a trade school or community college (i.e., if AESPDQ04= 4 or 5 then AEDPD04 = AEDPD04+1).

The same procedure was used to set up a years of schooling variable for the spouse/partner (AEDSD04).[18]

Occupational Prestige

Occupational status is an important indicator of SES. The occupation variable used in the derivation of SES was a modified version of a scale developed by Pineo, Porter and McRoberts (1977). The classification system groups occupations described in Statistics Canada's *1980 Standard Occupational Classification* into 16 somewhat homogeneous categories, ordered from 1 to 16, where code 1 represents the highest level of occupation and code 16 the lowest. The 16-category scale provides a ranking of occupations according to their social standing or prestige. For the NLSCY, for both the PMK and the spouse/partner, a detailed description was taken of the job considered to be his or her main job during the previous 12 months. The information was used to code occupations into the 1980 classification, and in turn into the 16 prestige categories. For the purposes of deriving SES, the order

---

example, if the PMK had completed only grade 10, but had masters, then AEDPD04 was set to 18.

It was decided that years of schooling was an interesting derived variable itself and therefore this variable has been included on the NLSCY master file for the PMK and spouse/partner (AEDPD04 and AEDSD04). It is not included on the microdata file because of confidentiality concerns. Users can gain access to this variable by way of requests for custom tabulations or by remote access, as discussed in Chapter 13.

of the Pineo-Porter-McRoberts scale was reversed. The final scale used in the derivation of SES had the following values:

01	Farm labourer
02	Unskilled manual
03	Unskilled clerical/sales/service
04	Semi-skilled manual
05	Semi-skilled clerical/sales
06	Farmer
07	Skilled crafts and trade
08	Skilled clerical/sales/service
09	Foreman/forewoman
10	Supervisor
11	Middle manager
12	Technician
13	Semi-professional
14	High-level management
15	Employed professional
16	Self-employed professional
96	Not-applicable - this was assigned for the spouse/partner for cases where the PMK did not have a spouse/partner
99	Not stated

This ordinal scale can be used to rank individuals into the various occupation groups but one cannot assume that the intervals between ranks are equal interval. For example, in this scale a middle manager (code 11) is ranked higher than a supervisor (code 10), which in turn ranked higher than a foreman (code 09). However, this does not imply that the difference in occupation between the middle manager and a supervisor is equivalent to the difference between a supervisor and a foreman. By assuming that the underlying latent construct has a particular distribution, one can assign intervals to the various categories. Mosteller and Tukey (1977) propose a logit transformation to re-express ordinal data on an interval scale. To do this, the percentage of individuals in each occupation group is considered a piece of the logistic distribution. The code assigned to each occupation is the centre of its piece in the logistic distribution. This transformation was employed to scale the 16 occupations.

For each occupation group *x*, the following values were computed:

p	=	the percentage of individuals with an occupation less than occupation *x* (based on the Pineo-Porter-McRoberts category)
pp	=	the percentage of individuals with an occupation less than or equal to occupation *x* (based on the Pineo-Porter-McRoberts category)
phi(p)	=	p*ln(p) + (1-p)*ln (1-p)

$$\text{phi}(pp) = pp*\ln(pp) + (1-pp)*\ln(1-pp)$$

The recoded (logit) value for occupation *x* was assigned to be:

$$\text{PINEOLOG} = \frac{\text{phi}(pp) - \text{phi}(p)}{pp-p}$$

This variable, PINEOLOG (for both the PMK and spouse/partner) was then used in the derivation of SES.

Household Income

The last variable used in the derivation of SES was household income. More detail regarding the collection of household income and data quality issues can be found in Section 9.17. To derive SES, income was coded in $1,000s of dollars, and a few outliers with incomes greater than $150,000 were recoded to $150,000.

Final Derivation of SES

Thus the five variables that were used to derive SES were:

- AEDPD04 (years of schooling for the PMK),
- AEDSD04 (years of schooling for the spouse/partner),
- PINEOLOG-PMK (the pineo occupation code for the PMK transformed to the logit distribution),
- PINEOLOG-SP (the pineo occupation code for the spouse/partner transformed to the logit distribution) and
- HHINC (household income in thousands of dollars)

Each of the five variables were standardized to have a mean of zero and a standard deviation of one. Missing values (i.e., not-stated values) were ignored in the standardization. In the standardization of the spouse/partner variables (AEDSD04 and PINEOLOG-SP), if the PMK did not have a spouse/partner these records were ignored. The SES composite was then calculated by taking the (unweighted) average of the five standardized variables. If one of the five variables had missing data due to non-response (refusal, don't know, etc.) then the average was taken over the remaining non-missing items. If there was no spouse/partner in the household (i.e., the PMK had no spouse/partner) then the average was taken over the three applicable variables (AEDPD04, PINEOLOG-PMK, and HHINC).[19] For two-parent families (i.e., for cases where there

With this procedure, the SES score for single-parent families will tend to be lower because household income, on average, will be lower. However, the SES score will properly reflect the level of education

was a PMK and a spouse/partner), if two or more out of the five input variables were missing, then SES was set to "not-stated" . For single-parent families (i.e., there was no spouse/partner), if one or more out of the three input variables were missing, then SES was set to "not-stated".

On the NLSCY microdata file SES is labelled AINHD08.

Examples of SES

On the microdata file, the value for SES ranges from -2.000 to +1.750. The distribution of SES scores is as follows for children on the file.

SES SCORE RANGE	% CHILDREN WITH SCORE IN RANGE
1.5 or over	2.7%
1.0 to less than 1.5	6.5%
0.5 to less than 1	12.9%
0 to less than 0.5	22.7%
-0.5 to less than 0	27.7%
-1.0 to less than -0.5	16.3%
-1.5 to less than -1.0	7.3%
Less than -1.5	3.2%
Not-stated	0.7%

In order to give a flavour for the types of families associated with various SES scores the following examples are given for illustration purposes. It should be noted that the SES scores given in these examples are approximate and do not correspond to actual records on the NLSCY file. Many more examples are possible for each score involving both one and two parent families.

and the occupational prestige of the single parent. Nevertheless, for most regression analyses where SES is used as a control variable, it would be useful to include a dummy variable denoting whether the family was a single- or two-parent family.

SES SCORE	EXAMPLE
1.5	A family in which: • both the PMK and spouse have a university degree (BA/BSC) • they are both employed professionals • the household income is approximately $77,000
0.5	A family in which: • the PMK has a university degree (BA/BSC) and the spouse has grade 13 • the PMK is employed as a semi-professional and the spouse is employed in a semi-skilled clerical position • household income is approximately $57,000
0.0	A family in which: • the PMK has grade 13 and the spouse grade 12 • the spouse is employed as a semi-professional position and the PMK is not in the labour force • household income is approximately $25,000
-0.5	A family in which: • the PMK and spouse have both completed grade 12 • the PMK is employed in a semi-skilled clerical position and the spouse in a semi-skilled manual position • household income is approximately $16,000
-1.0	A family in which: • neither the PMK nor the spouse have completed high school • the PMK is employed in an unskilled clerical position and the spouse is employed in an unskilled manual position • household income is approximately $20,000
-1.5	A family in which: • neither the PMK nor the spouse have completed high school • neither the PMK nor the spouse are in the labour force • household income is approximately

$12,000

-2.0

A family in which:

- there is no spouse
- the PMK has not completed high school
- the PMK is not in the labour force
- the household income is less than $10,000

Income Ratio (AINHD04 and 05)

NLSCY children can be classified as living in households of various income levels. An income ratio has been derived and assigned to each child record and can be used for analytical purposes to further understand the economic situation of the child. The following is a description of how this ratio was calculated.

Every year Statistics Canada establishes what are known as the low-income cut-offs, which are derived by considering expenditure to income patterns observed in the most recent Family Expenditure Survey. These thresholds or values are calculated for different urban-size and family-size categories and are updated annually using the Consumer Price Index.

The cut-offs that were derived for 1994 were used to calculate the NLSCY **income ratio**. The ratio was simply calculated to be the household income divided by the cut-off value.

Two data quality issues should be raised regarding this income ratio. One is that the cut-offs are based on economic family income. For the NLSCY it was household income that was collected and not economic family income. However in 98.5% of households in the sample the two concepts were equivalent (i.e., there was only one economic family in the household).

Secondly and more importantly, the number of children estimated to live in households with incomes below the cut-off may be overestimated. For the NLSCY household income was collected by asking a single question whereby the PMK was asked to estimate total income from all sources for all household members. The purpose was to get a general indicator of household income. The following is a comparison of the number of households estimated to have a household income below the cut-off by the NLSCY as compared to the number of economic families estimated to be below the cut-off by the Survey of Consumer Finances (SCF). The SCF is an annual survey conducted by Statistics Canada where detailed income information is collected for all household members for all sources.

COMPARISON OF INCOME RATIOS FOR THE NLSCY VS. THE SCF

	COMPARISON OF NLSCY [3] HOUSEHOLDS TO LOW INCOME CUT-OFFS	COMPARISON OF SCF[4] ECONOMIC FAMILIES TO LOW INCOME CUT-OFFS
% below cut-off	24.0	20.6
% over cut-off but less than twice the cut-off	37.3	38.0
% twice the cut-off or more	38.6	41.5

As can be seen in the table, the NLSCY estimates that more households with children are under the cut-off as compared to SCF. In Section 9.17 further data quality issues regarding the income variables collected for the NLSCY are discussed.

16.12 Geographical Indicators

The NLSCY sample was allocated so that provincial analyses will be possible for broad age groupings of children. A variable to indicate province of residence is available on the microdata file (AGEHD03). It was necessary to suppress the province codes on some records on the microdata file due to confidentiality concerns. This is discussed further in Section 9.2.

Sub-provincial analyses may be possible for certain variables, but in order to ensure confidentiality of respondents, sub-provincial indicators have not been included on the microdata file. Census metropolitan area (CMA) is available on the NLSCY master file (AGEHD02) as well as an indicator urban/rural class size (AGEHD01).

User's interested in performing sub-provincial analyses or having the province code available for all records can request custom tabulations or make use of the remote data access service as discussed in Section 13.

18.0

Content and Validation of NLSCY Data

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 which 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 information on the child's parent(s), characteristics of the family and the neighbourhood. This section provides an outline of the content for each section of the questionnaire included in this first release of NLSCY data.

As part of the NLSCY processing system, there were some basic quality checks performed for each section of the questionnaire. Any items for which there was a high level of non-response or which 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.

Statistics Canada has the obligation to ensure the confidentiality of all survey respondents. As a result, it was necessary to suppress or alter some variables on the microdata file. The suppressions made for each section of the questionnaire are discussed here. Users interested in having access to the unscreened data can request custom tabulations or make use of the remote data access service as discussed in Section 13.

Before the section by section discussion of content and validation results, the general validation procedures used for the "scale" data are presented.

18.2 Validation of Scale Data

For some of the concepts that were deemed to be 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 Parenting Scale that was employed was one that was proposed by Dr. M. Boyle at Chedoke-McMaster Hospital, based on work by Dr. Ken Dodge (Vanderbilt University) which was an adaptation of Strayhorn and Weidman's Parent Practices Scale. The scale is intended to measure three different constructs or factors related to parenting; positive interaction, hostile/ineffective parenting and consistent parenting.

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 Scale, 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 made of what specific scales should be used to measure a particular concept, as much as possible, scales were selected that had been used in other studies where the psychometric properties of the measures produced by the scale were available with complete references.

However in many instances the wording of certain questions 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 factorial 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 to exist in other studies were also 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 that were followed for each of these steps are described in detail on the following pages.

The specific details for each scale are discussed later in this section in the appropriate sub-section.

18.2.2 Factor Analyses

The following is a summary of the procedures used in the factor analysis for each scale.

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 factor analysis performed subsequently. 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 also done on each half-sample and the factorial structure and loadings of each factor were compared across the half-samples.

Factor analysis requires that the data have the property of interval or ratio data, that is 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 Always in the respondent's perception. It was felt that this was not necessarily the case for the scales used in the NLSCY.

Therefore 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 which 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.

Initially the factor analysis for each scale to be included in this first release of NLSCY data was carried out using unweighted data. At that point in time the final weights had not yet been calculated. Once the weights were available, work started on repeating the factor analyses using the weighted data. (See Section 7 for a description of the weighting procedures). With the weights, the same factor structure was not always observed.

At this point in time not all of the scale data included in this first release have been re-analysed using the weights. Analyses for the Behaviour Scale and the Parenting Scale on the Child's Questionnaire have been completed using weighted data and scores computed (See Section 9.6 and 9.9 for details). For the Depression Scale, the Family Functioning Scale and the Social Support Scale weights are not available since the unit of analyses for these scales are at the PMK and household level. (The Depression Scale refers to the PMK and the Family Functioning and Social Support Scales refer to the family or household). Therefore the factor structure for these scales was assessed using unweighted data and scores have been computed and included on the microdata file. (See Sections 9.12, 9.13, and 9.14 for details).

18.2.4 Calculation of Scores and Item Imputation

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.

Example:

One of the constructs that emerged in the factor analysis for the Parenting Scale on the Child's Questionnaire was the hostile/ineffective parenting factor. In the factor analysis seven items were found to load into this factor.

APRCQ04	How often do you get annoyed with your child for saying or doing something he/she is not supposed to?
APRCQ08	Of all the times you talk to your child about his/her behaviour, what proportion is praise?
APRCQ09	Of all the times you talk to your child about his/her behaviour, what proportion is disapproval?
APRCQ13	How often do you get angry when you punish your child?
APRCQ14	How often do you think the kind of punishment you give your child depends on your mood?
APRCQ15	How often do you feel you have problems managing your child in general?
APRCQ18	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	1 - never
2 - about once a week or less	2 - less than half the time
3 - a few times a week	3 - about half the time
4 - one or two times a day	4 - more than half the time
5 - many times each day	5 - all the time

In the calculation of the score for this hostile/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 hostile/ineffective parenting score in the range 0 to 28. A score of 0 represents the absence of a problem and a score of 28 is the highest possible score with respect to 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.

Note that the second item that loaded into the hostile/ineffective parenting factor, APRCQ08 (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.

The score for the hostile/ineffective parenting factor is labelled as APRC**S**04 on the record layout for the microdata 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. If less than 10% of the items had a missing value then the items with non-response codes were imputed before the score was computed. The procedure used to impute these missing items is a routine available in SAS in the procedure called PRINQUAL. This procedure indicates, among valid item values, the one that seems the most plausible for a given record. It considers the response profile of the record with the missing item, the response profile of other responding records in the sample as well as the number of factors considered in the analyses.

A flag was created for any item for which values have been imputed to indicate the records for which imputation has taken place. These flags have been included on the microdata file. The flag on the file which corresponds to an item has the same name as the item itself except that the **Q** (question indicator) in the variable name is replaced by **I**. For

example some imputation was carried out for APRCQ04 (How often do you get annoyed with your child for saying or doing something he/she is not supposed to?). The imputation flag for this item is labelled APRCI04.

It should be noted that in addition to the scores, the raw items for each scale are included on the microdata file. This will permit researchers to have the ability 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.

For the scales included in this first release, which have not yet be analysed using the weighted data, scores have not been included on the microdata file. Scores for these scales will be included in the second release of data.

18.2.6 Reliability Measures for Scales

Reliability refers to the accuracy, dependability, consistency or repeatability of score results. In more technical terms, reliability refers to the degree to which the scores are free of measurement errors. There are many ways to measure reliability.

One of the most commonly used reliability coefficients is **Cronbach's alpha** (Cronbach, 1951). Alpha (or α) 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.

Cronbach's α has several interpretations. It can be viewed as the correlation between this 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. In the hostile/ineffective parenting factor, for example, the seven questions actually used for inclusion on the scale can be viewed as a sample from the universe of many 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 α tells 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 α 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 α can be interpreted as a correlation coefficient, it ranges from 0 to 1.

It has been shown that in general, α 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.

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 reliability are below this level. The Cronbach α is given in the documentation for each score which 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 α for each factor score was computed using SAS. Typically the α coefficients calculated using SAS are lower than those calculated using SPSS.

18.4 Demographic Variables

The demographic variables discussed in this section refer to variables collected on the household roster. As part of the household roster some basic demographic information (e.g., age, gender, marital status) was collected for all members of the child's household. The relationship grid was also completed as part of this questionnaire i.e., the relationship of everyone in the household to everyone else. Using this information it was 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 Section 8 (NLSCY Concepts and Definitions).

If was necessary to perform an extensive series of edits on the data that were collected as part of the relationship grid. There were some edits that were carried out as part of the CAI system during collection. However in the data that were received at Head Office there were still inconsistencies.

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

- in all relationships reported, a person could not have more than two parents
- the difference in age between a husband and wife had to be less than 29 years.

In total there were over 30 relationship edits performed. Some of the edits were what is known as "**soft**" edits and some were "**hard**". The first example was a hard edit and the second a soft edit. For all edit failures, the records for the entire household were reviewed manually for obvious mistakes. A correction **had** to be made for the **hard** edit failures. For the **soft** edit failures a correction was made if it was deemed appropriate to do so.

As well there were edits carried out comparing the relationship grid to information collected in the Custody Section. (Variables in the Custody Section will be part if the second release of data). For the most part, in the case of discrepancies, priority was placed on the custody information, since this was collected from the PMK and the information was more detailed. The roster was completed by a knowledgeable household member, not necessarily a parent.

The major source of error for relationship data had to do with step children. There were several cases where a female parent was living with a biological child and a spouse or common-law partner. The relationship of the male partner to the child was coded as "unrelated". For questionnaires completed in French this relationship was often coded as "in-law". In the edit, the relationship code was changed to step child for these cases. As a result of the relationship edits the number of children in a step families increased by close to 40% .

Due to confidentiality concerns it was necessary to suppress some of the demographic variables on the microdata file.

- Detailed age in years for the child has been included, i.e., age for up to four children in the household. As a result of including detailed age, it was necessary to suppress collection date. Collection for the NLSCY took place over an eight month period. By suppressing collection date this casts some doubt on the exact ages of the children.

- For the PMK it was only possible to have age in ranges (15 to 24, 25 to 29, 30 to 34, 35 to 39, and 40+). Age for the spouse/partner has been suppressed entirely. For male PMKs not living with a spouse/partner age group has been set to not-stated. For female PMKs not living with a spouse/partner age group has been set to not-stated for some cases. In total age group of the PMK was set to not-stated for 486 children on the microdata file.

- There were six sets of triplets on the file. For these cases the age of one of the trio has been altered by one.

- There were 27 children on the file who did not live with a parent. The province for all children in these households has been set to not-stated.

- Cases where the PMK was male and there was no spouse/partner caused some concerns with respect to confidentiality. For these cases age group, highest level of education, main source of household income and the province code have all been set to not-stated. There were 165 households and 255 children on the file in this category.

- Cases where the PMK was female and there was no spouse/partner also caused some concern, although not as severe as it was for males. For some of these records the age group of the PMK has been set to not-stated.

It should be noted that the variable "marital status" has not been included on this first release. It was determined that it was important to edit this variable against information in the Custody Section which will only be part of the second release.

18.6 Medical/Biological

The Medical /Biological Section was completed for children in the 0 to 3 age group. The major objective was 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 investigated the biological mother's pregnancy and delivery history, including policy-relevant topics such as the mother's breast-feeding experiences and prenatal lifestyle.

Since birth weight is such an important variable, caution was taking in editing this variable. The records for children with very low birth weights (< 1.5 kilograms) were examined to verify that the response was legitimate. Other variables considered in the edit were the length of the baby at birth, the number of days early of the delivery, the conditions of the delivery (e.g., multiple birth and special medical care) and the health of the child at birth. If there was nothing to collaborate the low birth weight it was set to not-stated.

On the microdata file it was necessary to cap birth weight at the lower end at 1.499 kilograms and less for confidentiality reasons. As well, for multiple births the variable was capped at the upper end at two or more (i.e., twins).

There were a couple of derived variables created for this section that bear note. Two variables were derived to indicate the gestational age of the

child. AMDCD06 gives the gestational age in days and AMDCD07 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). For children in the 0 to 3 age group 9.7% were born early, 89.0% were born in the normal range and 1.2% were born late.

A variable was derived (AMDCD08) 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). For children in the 0 to 3 age group 94.3% were of normal birth weight, 4.9% were of moderately low birth weight and 0.8% were of very low birth weight.

These estimates of premature babies and low birth weight babies are in line with what is found in the literature.

18.8 Temperament Scale

Introduction

Temperament scales are used to measure the temperament of young children (up to and including the age of three) based on the parents' answers to questions about the degree of difficulty their child presents for them. This measure is founded on the assumption that a child's temperament is not solely dependent on biological factors, but is also influenced by the parents' perception of the difficulty of the child.

The temperament scale used in the NLSCY for children 3 to 5 months old was developed by Dr. John Bates of the University of Indiana. This well-established scale, originally known as the Infant Characteristics Questionnaire (ICQ), has been used in large-scale studies and is considered by specialists to be the best available measure for use in household surveys.

The ICQ has been adapted for use in other surveys covering different age groups: 6 to 11 months, 12 to 23 months and two-year-olds. A revised version of the scale, devised by Dr. Jo-Anne Finegan at Toronto's Hospital for Sick Children, is used for three-year-olds.

For children aged 3 to 5 months, the scale made up of questions ATMCQ01 to ATMCQ12, ATMCQ14 to ATMCQ20, ATMCQ23 and ATMCQ33 is intended to measure the extent to which the child is fussy, unadaptable, unpredictable and dull. For children 6 to 11 months old, the foregoing list was expanded to include ATMCQ13 and ATMCQ24 to ATMCQ27. The expanded list of questions measures the same four aspects of temperament as for children 3 to 5 months old.

For children between 1 and 3 years-old, questions ATMCQ1 to ATMCQ15 and ATMCQ17 to ATMCQ33 should theoretically measure the degree to which the child is difficult, irregular, unadaptable, affectively negative and persistent/unstoppable.

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 ATMCQ14, 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 ATMCQ14, the meanings of the ratings are reversed.

Results

Analysis of this scale with weighted data could not be performed in time to be included in this version of the file. It will be completed following this initial release and will be included in the second release scheduled for 1997.

18.10 Education (Child)

The objective of this section was 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 was 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, behaviour problems at school, absenteeism, parental hopes for the child's educational outcomes, number of school changes and residential moves.

For children in grade 1 or higher, additional questions were asked concerning other aspects such as skipping and repeating grades, achievement, special education, parents' perception of school climate and importance of good grades to parents.

The Teacher's Questionnaire which will be included in the second release of NLSCY data next year will provide additional information about the child and his/her school achievement and behaviour.

At the data collection stage, six different questions were asked to determine the child's grade. This was because of the different ways of classifying grade for each province. At the processing stage, these six

questions were collapsed into one variable. On the record layout an indication is given as to what the code means for each province. For example, if the grade code (AEDCD01) is 10, this refers to secondary 1 for Québec and grade 7 for all other provinces. A similar procedure was carried out for grade skipped (AEDCD02) and grade repeated (AEDCD03).

The child's grade was also collected on the Teacher's Questionnaire and on the Math Test administered by the teacher. There was not always consistency across the three data collection units on what the correct grade was. In the edit, priority was placed on what the teacher said in the case of discrepancies.

On the microdata file the variables on language of instruction (AEDCQ12A) and type of school (AEDCQ08) were set to not-stated in some cases because of confidentiality concerns. Only a very small number of records were affected (the variables for 34 children).

In the Education Section, there was one question (AEDCQ13) which asked the number of days the child had missed since the beginning of the school year. The answer to this question obviously depends on the collection date which has not been included on the microdata file because of confidentiality concerns. Therefore this variable has been suppressed and a derived variable was created (AEDCD04) to indicate the percent of days missed since the beginning of the school year.

18.12 Behaviour Scale

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

For Cycle 1 of the NLSCY, 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; separation anxiety and opposition were omitted, and indirect aggression and some aspects of conduct disorder were added.

The following indicates the items that were included on the questionnaire to measure these various constructs of behaviour. As discussed in Section 9.1, 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 analyses 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.

Questionnaire Items:

Two and three year-olds:

- **Conduct disorder**
  Items include ABECQ6G from the Ontario Child Health Study (OCHS).

- **Hyperactivity**
  Items include ABECQ6B, Q6I, Q6N, Q6P, Q6S and Q6W from the OCHS and ABECQ6HH from the Montreal Longitudinal Survey.

- **Emotional disorder**
  Items include ABECQ6F, Q6K, Q6Q, Q6V, Q6CC, Q6MM and Q6RR from the OCHS.

- **Anxiety**
  Items include several of the OCHS emotional disorder questions (ABECQ6F, Q6Q, Q6V and Q6CC).

- **Physical aggression**
  Items include ABECQ6X from the Montreal Longitudinal Survey and ABECQ6G from the OCHS.

- **Inattention**
  Items include ABECQ6P from the OCHS and ABECQ6EE, Q6KK and Q6QQ from the Montreal Longitudinal Survey.

- **Prosocial behavior**
  Items include ABECQ6D, Q6U, Q6BB, Q6SS and Q6UU from the Montreal Longitudinal Survey; the last four items are from a scale developed by K. Weir and G. Duveen.

- **Separation anxiety**
  Items include ABEC6DD1, 6LL1, 6PP1 and Q6TT1 from Achenbach's Child Behavior Checklist (CBCL).

- **Opposition**
  Items include ABECQ6E1, Q6J1, Q6R1 and Q6T1 also drawn from Achenbach's CBCL.

Children aged 4 to 11:

- **Conduct disorder**
  Items include ABECQ6C, Q6E, Q6G, Q6L, Q6O (this item is coded "not applicable" for children not in school), Q6T, Q6AA, Q6DD, Q6FF, Q6JJ and Q6PP from the Ontario Child Health Study (OCHS).

- **Hyperactivity**
  Items include ABECQ6B, Q6I, Q6N, Q6P, Q6S and Q6W from the OCHS and ABECQ6HH from the Montreal Longitudinal Survey.

- **Emotional disorder**
  Items include ABECQ6F, Q6K, Q6Q, Q6V, Q6CC, Q6MM and Q6RR from the OCHS.

- **Anxiety**
  Items include ABECQ6Y and Q6II from the Montreal Longitudinal Survey along with several of the OCHS emotional disorder items (ABECQ6F, Q6Q, Q6V and Q6CC).

- **Indirect aggression**
  Items include ABECQ6J, Q6R, Q6Z, Q6LL and Q6TT from Lagerspetz, Bjorngvist and Peltonen of Finland.

- **Physical aggression**
  Items include ABECQ6X from the Montreal Longitudinal Survey and ABECQ6G, Q6AA and Q6NN from the OCHS.

- **Inattention**
  Items include ABECQ6P from the OCHS and ABECQ6EE, Q6KK and Q6QQ from the Montreal Longitudinal Survey.

- **Prosocial behaviour**
  Items include ABECQ6A, Q6H, Q6M, Q6GG and Q6OO from the OCHS and ABECQ6D, 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.

Analysis of the NLSCY Data

To conduct the analysis on the behaviour scale for the NLSCY data, a factor analysis was conducted on the scale for the 2 to 3 age group and the 4 to 11 age group separately. New factor structures emerged which are described in the "Results" Section below.

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 (AWTCW01) by the average weight for all individuals. Thus, the sum of the normalized weights is equal to the sample size.

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.

The score for each factor on the scale was arrived at by totalling 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 an 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.

Results

Two and three year-olds:

There were 3,909 two and three year-olds in the sample. The group was split into two sub-samples of 1,932 and 1,977 individuals, and the analysis for this age group was performed separately for each sub-sample. The non-response rate for most items was about 2.2%. Some individuals were excluded from the analysis that produced the factors. The exclusion criteria were as follows: individuals with eight or more items coded "missing", individuals with one or more refusals, individuals with two or more missing items under hyperactivity and emotional disorder, and individuals with one or more missing items for the other theoretical factors. After the criteria were applied, there were 1,742 and 1,773 individuals left in the sub-samples to be analysed. Data were imputed for only 12 items. The number of imputations ranged between 1 and 8 for those 12 items. A total of 34 values were imputed.

The factor analysis derived five factors for this age group: hyperactivity-inattention (ABECS01), prosocial behaviour (ABECS02), emotional disorder-anxiety (ABECS03), physical aggression-opposition (ABECS04) and separation anxiety (ABECS05). The items making up each factor are listed in the table below.

BEHAVIOUR SCALE FOR 2 AND 3 YEAR-OLDS

FACTOR	SCORE	ITEMS
Hyperactivity - inattention	ABECS01	ABECQ6B, 6I, 6N, 6P, 6S, 6HH, 6QQ
Prosocial behaviour	ABECS02	ABECQ6D, 6U, 6BB, 6SS, 6UU
Emotional disorder - anxiety	ABECS03	ABEQC6F, 6K, 6Q, 6V, 6MM, 6RR

Physical aggression - opposition	ABECS04	ABECQ6G, 6W, 6X, 6E1, 6R1, 6T1, 6Z1, 6NN
Separation anxiety	ABECS05	ABECQ6CC, 6DD1, 6PP1, 6LL1, 6TT1

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). For hyperactivity-inattention (ABECS01), Cronbach's alpha was 0.798. The item that had the greatest effect on this factor was ABECQ6P; removing it lowers Cronbach's alpha to 0.762. The table below shows the Cronbach's alpha for each factor, first including all items, then excluding the item having the greatest effect.

CRONBACH'S ALPHA FOR THE BEHAVIOUR SCALE FOR 2 AND 3 YEAR-OLDS

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 (ABECS01)	0.798	ABECQ6P	0.761
Prosocial behaviour (ABECS02)	0.847	ABECQ6SS	0.795
Emotional disorder-anxiety (ABECS03)	0.593	ABECQ6MM	0.539
Physical aggression-opposition (ABECS04)	0.754	ABECQ6Z1	0.717
Separation anxiety (ABECS05)	0.561	ABECQ6DD1	0.431

Once the factors were identified, the next step was to compute the scores for each factor. The scores for ABECS01, ABECS02, ABECS03, ABECS04 and ABECS05 could not be calculated for 123, 393, 108, 159 and 99 individuals respectively because of unreported values for the items included in the factors.

Children aged 4 to 11:

There were 14,226 children in the 4 to 11 age group. Two sub-samples of 7,073 and 7,153 were created for analysis. The item non-response rate was approximately 2.1% for most of the 47 items involved in the analysis. Individuals were excluded from the analysis on the basis of the following criteria: individuals with eight or more items coded "missing"; individuals with one or more refusals; individuals with two or more missing items under prosocial behaviour, conduct disorder, hyperactivity, anxiety

and emotional disorder; and individuals with one or more missing items for the other factors. After the criteria were applied, 6,620 and 6,683 individuals remained in the sub-samples to be analysed. Data were imputed for 26 items. The number of imputations ranged between 1 and 159 for those 26 items. A total of 363 values were imputed.

Six factors were identified for this age group: hyperactivity-inattention (ABECS06), prosocial behaviour (ABECS07), emotional disorder-anxiety (ABECS08), physical aggression-conduct disorder (ABECS09), indirect aggression (ABECS10) and a new factor, property offence (ABECS11). The items making up each factor are listed in the table below.

BEHAVIOUR SCALE FOR 4 TO 11 YEAR-OLDS

FACTOR	SCORE	ITEMS
Hyperactivity - inattention	ABECS06	ABECQ6B, 6I, 6N, 6P, 6S, 6W, 6HH, 6QQ
Prosocial behaviour	ABECS07	ABECQ6A, 6D, 6H, 6M, 6U, 6BB, 6GG, 6OO, 6SS, 6UU
Emotional disorder - anxiety	ABECS08	ABECQ6F, 6K, 6Q, 6V, 6CC, 6II, 6MM, 6RR
Physical aggression - conduct disorder	ABECS09	ABECQ6G, 6X, 6AA, 6FF, 6JJ, 6NN
Indirect aggression	ABECS10	ABECQ6J, 6R, 6Z, 6LL, 6TT
Property offence	ABECS11	ABECQ6C, 6E, 6L, 6T, 6DD, 6PP

Cronbach's alphas for these factors are given in the table below. Normalized weighted data were used in the computations.

CRONBACH'S ALPHA FOR THE BEHAVIOUR SCALE FOR 4 TO 11 YEAR-OLDS

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 (ABECS06)	0.838	ABECQ6I	0.810
Prosocial behaviour (ABECS07)	0.816	ABECQ6BB	0.789

Emotional disorder - anxiety (ABECS08)	0.794	ABECQ6II	0.756
Physical aggression - conduct disorder (ABECS09)	0.770	ABECQ6AA	0.716
Indirect aggression (ABECS10)	0.781	ABECQ6LL	0.733
Property offence (ABECS11)	0.637	ABECQ6C	0.553

The scores for these factors could not be computed in 338, 647, 324, 358, 814 and 310 cases respectively because of unreported values.

18.14 Motor and Social Development

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 was used to assess these concepts (AMSCQ01 to AMSCQ48).

The Motor and Social Development (MSD) Scale was developed by Dr. Gail Poe of the U.S. National Centre for Health Statistics. The MSD scale consists of a set of 15 questions that measure dimensions of the motor, social and cognitive development of young children from birth through 3 years; the questions vary by age of the child. 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.

A score was calculated for each child by summing the number of "yes" answers to each item in the scale (AMSCS01). 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 old. Therefore a decision was made to produce standardized scores. Each child was assigned a standard score so that the mean MSD score was 100 and the standard deviation was 15 for all age groupings. This standardization was done by 1 month age groups. Therefore children who are 0 months old will have an average MSD score or 100, children who are 1 month old will have an average MSD score of 100, ..., children 47 months old will have an average MSD score of 100. Using this standard score (AMSCS02) it

will be possible to compare the MSD scores of children across the 0 to 3 age group, not controlling for age.

18.16 Relationships

The Relationships Section of the Child's Questionnaire was completed for all children 4 years of age and older. The objective was to provide information about the child's relationships with others. Positive relationships with other children and adults may help to counteract other factors which place a child at risk.

The section collects information about how the child gets along with parents, brothers and/or sisters, teachers, friends, and classmates, with some variation by age of the child. Parents' knowledge of the names of the friends of 8 to 9 and 10 to 11 year olds is also investigated, along with their perception of these other children's behaviour, and whether their own child is shy or outgoing.

The questions on number of days spent doing things with friends, number of friends, and getting along with friends, parents, teachers and siblings (ARLCQ01, Q02, Q06-Q09) are based on those in the Ontario Child Health Study.

18.18 Parenting Scale

The objective of the parenting scale is to measure certain parental behaviours. Specifically, two scales were used; one was designed to measure positive interaction, hostile/ineffective parenting and consistent parenting, and a second scale to measure aversive and non-aversive parental management techniques.

The following indicates the items that were included on the questionnaire to measure these various constructs of parenting. As discussed in Section 9.1, complete factor analyses were carried out for the parenting scales to assess the psychometric properties of these scales for the NLSCY population. As part of these analyses the new factor structures of each construct or factor were compared to what has been found in the past in the literature. The results of these analyses are presented later on in this section.

Questionnaire Items:

For 0 to 11 year-olds:

Questions APRCQ1 to APRCQ18 on positive interaction, hostile/ineffective parenting and consistent parenting were provided by Dr. M. Boyle at Chedoke-McMaster Hospital, based on Dr. Ken Dodge's work (Vanderbilt University) and an adaptation of Strayhorn and Weidman's Parent Practices Scale. (For children 0 to 23 months old only APRCQ1 to Q7 were asked.)

For children 2 to 11 years of age:

Questions APRCQ19 to APRCQ25, which measure aversive and non-aversive parent management techniques, were provided by Dr. M. Boyle.

Analysis of the NLSCY Data

To conduct the analysis on the parenting scales for the NLSCY data, a factor analysis was conducted on the first scale for the 0 to 23 month age group and for the two scales for the 2 to 11 age group separately. New factor structures emerged which are described in the "Results" Section below.

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 (AWTCW01) by the average weight for all individuals. Thus, the sum of the normalized weights is equal to the sample size.

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 the following for the four factors that were found to exist in the parenting scales:

- a lack of positive interactions for the positive interaction factor
- a lack of hostile/ineffective interaction for the hostile/ineffective parenting factor
- a lack of consistency in interactions for the consistency factor
- a lack of punitive/aversive interactions for the aversive factor

Some items were imputed. The imputed values were computed by a procedure (within 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.

The score for each factor on the scale was arrived at by totalling 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 an item 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.

Results

For children 0 to 23 months of age:

In the sample there were 4,696 children in the age group 0 to 23 months. They were divided into two sub-samples of size 2,311 and 2,385, and analyses were done on each sub-sample. The non-response rates for the seven items ranged between 1.9% and 2.5%. Some individuals, namely those with one or more items missing, were excluded from the analysis conducted for the purpose of constructing the factors. After these exclusions, the sub-samples contained 2,245 and 2,307 individuals respectively, for analyses purposes. No imputation took place. As a result of factor analysis, two factors were identified for this age group: the positive interaction factor (APRCS01) and the hostile/ineffective parenting factor (APRCS02). The items that comprise each factor are described in the following table.

PARENTING SCALE FOR CHILDREN 0 to 23 MONTHS OLD

FACTOR	SCORE	ITEMS
Positive interaction	APRCS01	APRCQ1, 2, 3, 6, 7
Hostile/Ineffective	APRCS02	APRCQ4, 5

Cronbach's alpha coefficients (raw values) were calculated with SAS, using the normalized weighted data (in general, Cronbach's alphas calculated with SAS are lower than those produced by the SPSS software package). For the positive interaction factor (APRCS01), the Cronbach's alpha is 0.727. The item that affects the factor the most is APRCQ7. If it were removed from the analysis, the Cronbach's alpha would drop to 0.656. For the hostile/ineffective factor (APRCS02), the Cronbach's alpha is 0.394 (it should be noted only two items make up this factor, and therefore the alpha cannot be calculated if one of the two items is removed). Once the factors were determined, the next step was to calculate the scores for each of the two factors. For the positive interaction factor, scores could not be calculated for 132 individuals, while for the hostile/ineffective factor, scores could not be calculated for 124 individuals, due to missing values for the items comprising these factors.

For children 2 to 11 years of age:

The number of children in the 2 to 11 age group is 18,135. Two sub-samples of 9,090 and 9,045 were created for analysis purposes. The non-response rate per item ranged between 2.1% and 2.7% for the 18 items included in the analysis. Individuals with two or more items coded missing for the positive interaction and hostile/ineffective factors were excluded from the analysis. In addition, individuals with a single missing item for the consistency factor were also excluded from the analysis. After these exclusions, 8,815 and 8,772 individuals respectively were retained for analysis purposes. The data in these samples were imputed for 12 items. For these 12 items, the number of imputations varied between 1 and 16. In all, 91 values were imputed. Three factors were identified for this age group: the positive interaction factor (APRCS03),

the hostile/ineffective factor (APRCS04) and the consistency factor (APRCS05). The items that comprise each factor are described in the following table.

PARENTING SCALE FOR 2 to 11 YEAR-OLDS

FACTOR	SCORE	ITEMS
Positive interaction	APRCS03	APRCQ1, 2, 3, 6, 7
Hostile/ineffective	APRCS04	APRCQ4, 8*, 9, 13, 14, 15, 18
Consistency	APRCS05	APRCQ10, 11, 12*, 16*, 17*

*These Items reversed when calculating score.

The Cronbach's alphas for these factors are described in the following table. Normalized weighted data were used for these calculations.

CRONBACH'S ALPHA FOR THE PARENTING SCALE FOR 2 AND 3 YEAR-OLDS

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 (APRCS03)	0.808	APRCQ2	0.749
Hostile/ineffective (APRCS04)	0.706	APRCQ13	0.654
Consistency (APRCS05)	0.660	APRCQ12	0.569

For each of these factors, the scores could not be calculated for respectively 408, 482 and 534 individuals due to missing values.

For children 2 to 11 years of age - scale on parent management techniques:

The number of children in the 2 to 11 age group is 18,135. Two sub-samples of 9,090 and 9,045 were created for analysis purposes. The non-response rate per item was in the range of 2.5% for the seven items included in the analysis. Individuals with one or more items coded missing for one of the seven items were excluded from the analysis. After these exclusions, 8,848 and 8,801 individuals respectively, were retained for analysis purposes. No imputation took place.

One factor was identified for this age group: the aversive factor (APRCS06). The items that comprise this factor are APRCQ21, 22, 23 and 24. Items 21 and 23 were reversed when calculating the score. Items APRCQ19, 20 and 25 did not present loadings high enough to be included in the factor.

The Cronbach's alpha for this factor is 0.569. The item that contributes the most to lowering the Cronbach's alpha if it is removed from the analysis is APRCQ22. The value that the alpha would take on if this item were removed is 0.377 (normalized weighted data were used for these calculations).

For this factor, the score could not be calculated for 478 individuals, due to missing values.

18.20 Child Care

The objective of the Child Care Section was to provide basic information about the methods of care used for the child while the parents worked or studied, as well as some information on previous care. Concepts measured included both the amount of time spent by the child in child care and the methods of care used for each child. In addition, information was obtained on the number of changes in child care arrangements that the child had experienced in the past 12 months and the reason(s) for changes. The section also identified whether or not a child care centre was profit or non-profit and whether home care was licensed or unlicensed.

A comparison was made of the data collected in this section of the questionnaire to data from the National Child Care Survey (NCCS) which was conducted in 1988 by Statistics Canada. The estimates that were compared are:

- the percentage of children in the various types of care arrangements and
- the average number of hours per week spent in the care arrangement by children who had used it.

The target population for the NCCS included children 0 to 12 years of age. In order to ensure consistency for the comparison, children in the 0 to 11 age range were extracted from the NCCS and estimates were made on this sub-population. The NCCS asked about care arrangements used for any purpose. Again in order to ensure comparability of estimates, only care arrangements where the main activity of the parent was working or studying were included in the NCCS estimates. The table below shows the results of the comparison.

COMPARISON OF CHILD CARE ESTIMATES
NLSCY vs. NCCS

CARE METHOD	% OF CHILDREN USING CARE METHOD		AVERAGE NUMBER OF HOURS USED IN ONE WEEK		NLSCY
NCCS	NLSCY		NCCS		
12.1%	20.5	Someone else's home by a non-relative	20.3	Someone else's home by a relative	6.4%
6.5%	16.5	Own home by brother or sister	18.4	Own home by other relative	1.7%
5.1%	6.3	Own home by non-relative	7.1	Daycare centre	4.8%
5.1%	16.0	Before or after school program	19.7	Own care	6.9%
5.4%	17.5	Other arrangement	18.1		5.5%
3.8%	24.3		31.1		2.7%
1.6%	10.2		12.0		0.4%
5.2%	3.9		5.9		2.4%

The NCCS was conducted approximately seven years earlier than the NLSCY. Therefore it would be expected that changes in the child care field may have taken place in the intervening years. For example, the fact that the NLSCY estimates that more children are being cared for in a daycare centre is not a surprising one. As well, use of before and after school programs has increased, due to the fact that this type of care is now more readily available.

However one area where the estimates do not compare well which is of concern is the percentage of children reported to be in their own care. For the NCCS 5.2% of children were estimated to be in their own care while the parent worked or studied. For the NLSCY the estimate is only 0.4%. While some changes may have taken place over time and parents may now be more reluctant to leave children on their own, it is believed that the NLSCY underestimates the number of children who were in their own care due to a problem on the questionnaire.

On the NLSCY questionnaire, the Child Care Section started out with a filter question:

> Do you currently use child care such as daycare or babysitting while you (and your spouse/partner) are at work or studying?

If the answer was 'yes' to this opening filter question, the interviewer continued on with a set of detailed questions to determine the specific types(s) of care arrangements used; one of these types being the child was in his/her own care. If the answer was 'no' to the filter question these subsequent questions were skipped.

The problem is that if the child was in his or her own care it is likely that the parent would answer 'no' to the filter question and therefore the question on self care would have been skipped.

It may be that the same problem exists for "care in own home by brother or sister". The NCCS estimated that 5.1% of children were cared for by an older brother or sister. For the NLSCY the estimate is 1.7%. Again the problem could be that the parent said 'no' to the initial filter question and therefore the question on sibling care was not asked.

Changes have been made to the questionnaire for Cycle 2 to avoid this problem. The question on self care and care by an older sibling will be asked even if the answer to the filter question is 'no'.

For the Child Care Section there was a series of derived variables created to explain the Primary Care Arrangement used to allow the PMK (and spouse/partner) to work or study. This primary care arrangement was derived by looking at all of the care methods used for the child and selecting the one for which the number of hours used was the greatest.

There were also some derived variables created to link the use of child care to the work arrangements of the parents. For example there is a derived
variable to indicate if child care was used if one of the PMK or spouse/partner currently worked full-time and the other part-time.

Further details regarding these derived variables can be found on the record layout for the Child Care Section. These variables are labelled A**CR**C**D**01 to A**CR**C**D**10. (**CR** indicates the Child Care Section and **D** a derived variable.)

18.22 Socio-demographic Characteristics

The objective of the Socio-demographic Section was to gather information on immigration, ethnic background and the language profile of household members. This will allow analysis for various components of the Canadian population and will permit identification of visible minorities.

As well, there were questions on religious affiliation and frequency of attendance at religious services. Religion, particularly frequency of attendance, is acknowledged as having a positive influence on a child's development.

It was necessary to suppress many of the variables in this section on the microdata file due to confidentiality concerns. The questions on country of birth, ethnicity and religion have all been suppressed while frequency of attendance at religious services has been included.

The questions on mother tongue and language of conversation are included on the microdata file but only with aggregated answer categories:

- English only
- French only
- English and French only
- at least one "other" language indicated.

The aggregated variables for language of conversation are labelled ASD**P**D05B, ASD**S**D05B, and ASD**C**D05B, for the **P**MK, **S**pouse/partner and **C**hild on the microdata file. The mother tongue variables are ASDPD06B, ASDSD06B and ASDCD06B.

For the immigrant population, a derived variable was created to indicate number of years since first immigrating to Canada. It was possible to put a grouped version of this derived variable on the microdata file (ASDPD02B, ASDSD02B, ASDCD02B).

A variable will be created to permit the identification of visible minorities. This variable is based on country of birth, ethnicity and mother tongue. This variable has not yet been derived and will only be available for the second release next year.

Since there are many variables in this section which have been suppressed for the microdata file, researchers who are particularly interested in conducting analyses on socio-demographic variables are encouraged to consider making use of the remote access service described in Section 13.3.

18.24 Depression Scale

Introduction

The Depression scale was administered to the PMK as part of the Parent Questionnaire. Questions for this scale (ADPPQ12A to ADPPQ12L) are a shorter version of the depression rating scale (CES-D), comprising 20 questions, developed by L. S. Radloff of the Epidemiology Study Centre 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.

The depression rating scale includes twelve questions, each of which contains four response categories. In order that the lowest score value 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 (ADPPQ12F, Q12H, and Q12J). The total score (ADPPS01) may therefore vary between 0 and 36, a high score indicating the presence of depression symptoms.

Results

In analysing this scale, unweighted data[20] were used. The sample size was 13,439 PMKs. However, once the observations containing mostly missing values were eliminated, the analysis dealt with only 13,140 PMKs. 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.

In spite of the possibility of extracting more than one factor from the depression rating scale, 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. The Cronbach alpha coefficient (calculated using SAS software) was 0.82. The variable ADPPQ12D showed the highest correlation (0.68) with the total score (once the variable was removed), whereas the variable showing the lowest correlation was ADPPQ12L with a correlation of 0.33. The Cronbach alpha coefficient calculated by omitting one variable was between 0.79 and 0.82 for the 12 variables.

18.26 Family Functioning Scale

Introduction

Questions related to family functioning, i.e. AFNHQ01A to AFNHQ01L, 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 AFNHQ01M, 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

[20] *Weighted data could not be used since the weights developed for the NLSCY are for children only, and not for parents.*

Health Study have shown, for example, that there is an important link between family dysfunction and certain mental conditions in children.

The family functioning scale was administered to either the PMK or spouse/partner as part of the Parent Questionnaire. The unit of analysis for the scale is the family. The scale includes twelve questions, each of which contains four response categories. In order that the lowest score value 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 (AFNHQ01A, Q01C, Q01E, Q01G, Q01I, and Q01K). The total score (AFNHS01) may therefore vary between 0 and 36, a high score indicating family **dysfunction**.

Results

In analysing this scale, unweighted data[21] were used. The sample size for the scale was 13,439 families. However, once the observations containing missing values were eliminated, the analysis dealt only with 13,190 families. 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.

Following single-factor analysis, all 12 variables of the scale were kept since the loading values were well above the established threshold. The Cronbach alpha coefficient (calculated using SAS software) was 0.88. The variable AFNHQ01L showed the highest correlation (0.66) with the total score (once the variable was removed), whereas the variable showing the lowest correlation was AFNHQ01A with a correlation of 0.51. The Cronbach alpha coefficient calculated by omitting one variable was stable at about 0.87 for the 12 variables.

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". Basically this artifact in the scale score is due to the fact than many respondents were consistent in their answering pattern across items.

Weighted data could not be used since the weights developed for the NLSCY are for children only, and not for families.

18.28 Social Support Scale

Introduction

Questions ASPHQ01A to ASPHQ01F are a shorter version of the Social Provisions Scale developed by Dr. Carolyn E. Cutrona and Dr. Daniel W. Russell of Iowa State University. The Social Provisions Scale is used to measure perceived support. The scale was shortened within the framework of an Ontario project called Better Beginnings, Better Future. The shorter version is used to measure the following components of social relationships: guidance, reliable alliances (the assurance that others can be counted upon for practical help) and attachment.

This scale is aimed at determining the level of social support received from friends, family and others.

Other surveys have shown that social support has an effect on self-esteem and on an individual's relationship with children and other members of the family.

The social support scale was administered to either the PMK or spouse/partner as part of the Parent Questionnaire. The unit of analysis for the scale is the family. The scale includes six questions, each of which contains four response categories. In order that the lowest score value 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 (ASPHQ01A, Q01D and Q01E). The total score (ASPHS01) may therefore vary between 0 and 18, a high score indicating the presence of social support.

Results

In analysing this scale, unweighted data[22] were used. The sample size for the scale was 13,439 families. However, once the observations containing missing values were eliminated, the analysis dealt with only 13,253 families. The non-response rate for the different variables was between 1.2 and 1.3%, whereas for the score, a non-response rate of 1.4% was obtained. There was no imputation for the variables in this scale.

Following single-factor analysis, the six items in the scale were kept since loading values were well above the established threshold. The Cronbach alpha coefficient (calculated using SAS software) was 0.82. The variable

---

Weighted data could not be used since the weights developed for the NLSCY are for children only, and not for families.

ASPHQ01C showed the highest correlation (0.65) with the total score (once the variable was removed), whereas the variable showing the lowest correlation was ASPHQ01A with a correlation of 0.45. The Cronbach alpha coefficient calculated by omitting one variable was between 0.78 and 0.82 for the six variables.

When the values for the factor score for the social support scale are examined for the NLSCY children, the distribution that is observed is not a continuous one. In fact the most common scores are 12 and 18. This is a result of the fact that there are six items in the scale and four possible rescaled values (0 to 3).

- Many respondents had a rescaled score of 2 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.

- A second group of respondents has a rescaled score of 3 for every item in the scale and thus an overall score of 18. This means that the respondent answered "strongly agree" to all of the items in the scale which were positive and "strongly disagree" to all of the negative items.

Basically this artifact in the scale score is due to the fact than many respondents were consistent in their answering pattern across items.

18.30 Education (Parent)

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

Research (for example, the Ontario Child Health Study and the National Longitudinal Survey of Youth in the United States) 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.

Due to confidentiality concerns only an aggregated version of the highest level of education attained by the PMK and spouse/partner have been included on the microdata file. These variables (AEDPD02 for the PMK and AEDSD02 for the spouse/partner) have the following values.

- less than secondary
- secondary school graduation

- beyond high school
- college or university degree (including trade).

On the microdata file this variable has been set to not stated for male PMKs who do not live with a spouse/partner.

The other education variable included on the microdata file, is current school status and whether attendance is full-time or part-time. These variables have been included on the file for the PMK, but it was necessary to suppress them for the spouse/partner. If the PMK was a lone parent (i.e., did not live with a spouse/partner), then only the fact as to whether or not she/he is a student has been retained, while the detail about full-time/part-time status has been suppressed.

18.32 Labour Force

Employment stability impacts the home environment, both in terms of income and stress levels. Research, conducted for the Ontario Child Health Study, indicates that parental unemployment can adversely impact child mental health.

The Labour Force Section was 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 included, periods of absence from work, reason for the most recent absence, hours worked, and work arrangements (e.g. shifts) during the previous year. Information was collected on up to six jobs for a one-year period.

Respondents were asked to identify what they considered to be their main job over the previous year (if they had more than one job). A complete description was recorded for this main job and industry and occupation coding was carried out (using 1980 Standard Industrial Classification codes and 1980 Standard Occupational Classification codes).

Data on wages and salaries for this main job were collected. Wage rate data provides an additional source of information on income. This data will be useful in analysing choices which parents, particularly mothers, face in deciding to stay at home or to return to the labour force.

18.32.2 Work Duration Derived Variables

With the data collected in the Labour Force Section it was possible to create a series of derived variables to describe the stability of work for the PMK and spouse/partner over the previous year.

As mentioned above, a series of questions were asked about each of the jobs the PMK and spouse/partner held during the previous year to a maximum of six jobs. For each of these jobs, questions were asked to determine when the job started and ended. As well, in order to address absences within a job the following question was asked as the initial lead-in question to a job:

> Did you have that job one year ago, without a break in employment since then?

The intent was that if there had been a break in employment the respondent was supposed to report this situation as two jobs. The start date for the first job would be when the respondent first started working at the job and the end date would be when the break occurred. The second job would have the point at which the respondent returned to the job as the start date.

In order to measure duration of employment for the PMK over the previous year, an employment vector was set up by superimposing the start and end dates for all jobs held in the previous year. This vector consists of 53 positions. [23] For each position there is an indication if the PMK worked at a job or business that week. If she did not work at a job or business that week the value in the vector is 0, if she worked at one job the value is 1, if she worked at two jobs the value is 2, etc. This employment vector has been included on the microdata file (ALFPD53). As well, there is a derived variable (ALFPD33) based on this vector to indicate the number of weeks the PMK worked at a job or business in the previous year.

A similar vector and derived variable were computed for the spouse/partner (ALFSD53 and ALFDD33).

In the next release of NLSCY data, next year, additional derived variables will be included to describe the employment picture over the reference year, such as number of weeks worked part-time, number of weeks worked full time, number of gaps in working etc.

18.32.4 Comparison with SLID Data

The reason there are 53 positions in the employment vectors is because employment data were collected over the previous 12 months; which could involve 53 calendar weeks.

In order to assess the quality of the data coming from the Labour Force Section a comparison was made to data coming from the Survey of Labour and Income Dynamics (SLID). SLID, also conducted by Statistics Canada, is a national longitudinal survey of household members aged 16 years of age and over. The goal is to collect detailed income and labour market data.

For comparison purposes SLID employment data were taken for reference year 1993. For the NLSCY, the reference year for employment data was different depending on when the respondent was interviewed. The reference year could have started anywhere from November 1993 to June 1994. It is not expected that these minor differences in the reference year should result in any significant differences when comparing the two data sources. It should be noted that for the SLID comparison all parents of a child 0 to 11 years old were selected from the SLID database to provide a similar target population as the NLSCY.

There were three main measures taken from the SLID data.

1/ **Number of Jobs**

The first measure taken was number of jobs the respondent held during the reference year. Note that for SLID data, if the respondent changed jobs within the same employer without an interruption in work, this counted as one job only. The same is true for the NLSCY.

2/ **Number of weeks employed**

The second measure taken was weeks of employment in the reference year. It should be noted that a person is considered to be employed at a job even if he/she is experiencing an unpaid absence from that job. The one exception to this is if the reason for the absence is a temporary lay-off (which is considered as unemployment or not in the labour force). Basically this is an attempt to measure attachment to an employer even if there is a temporary unpaid absence from that employer.

3/ **Number of weeks working for pay or profit**

The SLID data base also includes information on all absences from a job (including the number of weeks absent, the reason for the absence and whether or not the respondent received pay during the absence). All unpaid absences (such as unpaid absences due to sickness, vacation, maternity leave etc.) were selected and subtracted from the number of weeks employed for the respondent

to produce a third employment measure.[24] This third measure was what the NLSCY was attempting to address - stability of employment income.[25]

These measures were compared to NLSCY measures for the entire population of parents (Tables LFS 1a, LFS 1b, and LFS 1c) and then separately for female parents (Tables LFS 2a, LFS 2b, and LFS 2c) and male parents (Tables LFS 3a, LFS 3b, and LFS 3c). Comparisons were made based on sex, since the notion of PMK does not exist for SLID.

TABLE LFS 1A
SLID - NLSCY COMPARISON OF
NUMBER OF JOBS FOR ALL PARENTS

NUMBER OF JOBS	NLSCY	SLID
0	17.8%	16.8%
1	68.3%	70.1%
2 or more	13.9%	13.1%

TABLE LFS 1B
SLID - NLSCY COMPARISON OF
NUMBER OF WEEKS WORKED FOR ALL PARENTS

	NLSCY WEEKS WORKED	SLID WEEKS EMPLOYED	SLID WEEKS WORKED (employment minus unpaid absences)
% who worked/were employed 0 weeks	17.8%	16.8%	16.8%

Temporary layoffs were not subtracted since they were not considered as employment in the second measure - number of weeks employed.

It should be noted that there may be some small problems with the way this third measure was derived. SLID, for now, only have absence information at the job level. It could be that a person has an absence from one job while working at another job. As a result there could be a ***very slight*** *under-estimate of the number of weeks the respondent worked for pay or profit.*

% who worked/were employed 1 to 52 weeks	17.2%	18.3%	24.7%
% who worked/were employed full year	65.0%	64.9%	58.4%

TABLE LFS 1C
SLID - UNPAID ABSENCES FOR ALL PARENTS

% of parents employed the entire year who had at least one unpaid absence	9.9%
average number of weeks for unpaid absences for these parents	14

TABLE LFS 2A
SLID - NLSCY COMPARISON OF NUMBER OF JOBS FOR FEMALE PARENTS

NUMBER OF JOBS	NLSCY	SLID
0	28.3%	27.6%
1	59.2%	60.2%
2 or more	12.5%	12.2%

TABLE LFS 2B
SLID - NLSCY COMPARISON OF NUMBER OF WEEKS WORKED FOR FEMALE PARENTS

	NLSCY WEEKS WORKED	SLID WEEKS EMPLOYED	SLID WEEKS WORKED (employment minus unpaid absences)
% who worked/were employed 0 weeks	28.3%	27.6%	27.6%

% who worked/were employed 1 to 52 weeks	20.3%	17.8%	27.0%
% who worked/were employed full year	51.4%	54.6%	45.4%

TABLE LFS 2C
SLID - UNPAID ABSENCES FOR FEMALE PARENTS

% of female parents employed the entire year who had at least one unpaid absence	16.7%
average number of weeks for unpaid absences for these parents	16

TABLE LFS 3A
SLID - NLSCY COMPARISON OF
NUMBER OF JOBS FOR MALE PARENTS

NUMBER OF JOBS	NLSCY	SLID
0	5.1%	4.9%
1	79.3%	81.1%
2 or more	15.6%	14.0%

TABLE LFS 3B
SLID - NLSCY COMPARISON OF
NUMBER OF WEEKS WORKED FOR MALE PARENTS

	NLSCY WEEKS WORKED	SLID WEEKS EMPLOYED	SLID WEEKS WORKED (employment minus unpaid absences)
% who worked/were employed 0 weeks	5.1%	4.9%	4.9%
% who worked/were employed 1 to 52 weeks	13.6%	18.7%	22.1%
% who worked/were	81.3%	76.4%	

employed full year

73.0%

TABLE LFS 3C
SLID - UNPAID ABSENCES FOR MALE PARENTS

% of male parents employed the entire year who had at least one unpaid absence	4.5%
average number of weeks for unpaid absences for these parents	7

Discussion

In terms of number of jobs, NLSCY data matches fairly well with SLID data. NLSCY has slightly higher estimates for the number of people who had no jobs during the past year (17.8% vs. 16.8%). One possibility is that people who were only very marginally involved in the labour force (i.e., only for a week or two) may have forgotten to report a job. SLID may have done slightly better here because they were able to feed-back data from the Labour Force Survey which was not done for the NLSCY.

When the number of weeks worked are compared between NLSCY and SLID the following results emerge:

- The NLSCY estimates for **weeks worked** are fairly close to the SLID estimates for **employment**. According to the NLSCY 65.0% of parents are estimated to have worked the entire year. According to SLID 64.9% of parents are estimated to have been **employed** all year. However when unpaid absences (other than temporary layoffs) are subtracted from weeks employed for the SLID data, the percentage of parents who worked for pay all year round reduces to 58.4%. It would seem that for the NLSCY unpaid absences are not being adequately measured.

- In order to explore this matter further, parents on the SLID database who were coded as being employed all year were selected and those who had at least one unpaid absence over the year were identified. Of the parents who were employed all year round close to 10% had at least one unpaid absence and on average this absence lasted 14 weeks. It would seem that the NLSCY is missing many of these unpaid absences.

- The NLSCY estimates that 51.4% of female parents worked all year while SLID estimates that 54.6% were employed all year while 45.4% worked for pay all year round. For females it seems that the NLSCY picks up some unpaid absences but not all.

- The situation seems to be the most serious for males. The NLSCY estimates that 81.3% of male parents worked all year round. SLID estimates that 76.4% were employed all year round and this reduces to 73.0% when unpaid absences are removed. To some extent, the problem could be due to the fact that for the most part employment data for male parents for the NLSCY was provided on a proxy basis by the female parent (i.e., the PMK). It may be that the PMK could not recall unpaid absences or spells of unemployment that her spouse/partner had over the previous year, particularly if they were short spells.

In general, the problem with the NLSCY data is likely due to the fact that there were not specific questions asked about absences from a job and therefore respondents and interviewers may not have handled absences in a consistent way. The lead in question to the first job was:

Did you have that job one year ago, without a break in employment since then?

- In some cases the respondent answered "no" to this question but in the start and end dates, indicated that they had the job all year round. This happened for 1.7% of the parents who worked throughout the year and for approximately 2.5% of female parents who worked throughout the year. For these parents, it would seem that there was an absence but no information was reported about the absence.

- Another problem is that a respondent may have had an unpaid absence from the job during the year but answered yes to the above question because they did not consider the absence to be a break in employment.

For Cycle 2 a new set of Labour Force questions have been developed.

18.32.6 Wage Rate Edit

There was a series of questions in the Labour Force Section, where the PMK and spouse/partner reported earnings for the main job. Respondents were permitted to report earnings by the hour, the week, the year etc. A derived variable was computed so that an hourly wage rate was given for each PMK and spouse/partner (ALFPD12 for the PMK and ALFSD12 for the spouse/partner).

As part of the NLSCY processing system, earning from employment for the main job for the PMK and spouse/partner were calculated and added together and compared to household income. Any cases where there was a large discrepancy were flagged. As well employment earnings for the PMK were compared to the income for the PMK. Again cases where there was a large discrepancy were flagged. All wage rates which were very high (greater than $48.00 per hour) or very low (less than $5.00 per hour) were flagged.

A manual review was carried out of all cases that were flagged. The detailed description of the main job for the PMK and spouse/partner were printed out, along with number of weeks worked at the main and other jobs and all of the information from the Income Section (sources and amount).

In some cases it was determined that there was a problem with the wage rate information and therefore the hourly wage rate was set to not-stated. In other cases the income variables were deemed to be the problem and corrective actions were taken for these variables. This is discussed in Section 9.17.1.

As a result of this edit, the wage rate was set to not-stated for 263 PMK records and for 206 spouse/partner records. In a few cases it was determined that the reporting period was not correct and an appropriate correction was made. For example, it was indicated on the questionnaire that earning per week were reported but it was determined that it was more likely earnings per hour.

18.32.8 Microdata File Suppressions

It was possible to include industry and occupation codes on the microdata file, but only for fairly large aggregate groupings. For occupation, 21 major groups have been included and for industry 13 groups. The Pineo-economic classification code for the main job has also been included on the microdata file. In a few cases industry and occupation codes have been set to not-stated due to confidentiality concerns. The occupation codes for the PMK corresponding to religion and mining have been set to not-stated.

In total:

- the occupation code was set to not-stated for 131 PMKs and for 181 spouse/partners
- the industry code was set to not-stated for 106 PMKs and for six spouse/partners
- the Pineo code was set to not stated for 486 PMKs and for 470 spouse/partners

The hourly wage rate for the PMK and spouse/partner have been included on the microdata file but they have been capped at $24.00 per hour at the upper end and $5.00 per hour at the lower end. The input variables used to calculate the hourly wage rate have been suppressed.

It was possible to include the detailed information on all jobs held by the PMK and spouse/partner in the previous year on the microdata file, except for the start and end dates of the jobs. These dates could potentially give an indication of collection date which was suppressed, as discussed in Section 9.2. However the vectors to indicate the weeks worked over the previous year for the PMK and spouse/partner have been included.

18.34 Income

In the Income Section of the NLSCY, the sources and amount of household income were collected, as well as the personal income of the PMK.

This information will provide an indicator of the family's economic situation, an essential component of the child's environment. Although the PMK's and household's income may not be clear indicators of income distribution among all family members, they act as general indicators of the household's economic situation.

Two approaches were used for collecting amount of income. If the respondent was reluctant or unwilling to provide a specific dollar amount for household and personal income, a "cascade" question was asked. The cascade consisted of income categories.

For household income:

- 81.9% of respondents provided an exact amount
- 14.4% provided an answer to the cascade question
- 3.7% could not or refused to provide any answer.

For PMK income:

- 85.6% of respondents provided an exact amount
- 7.0% provided an answer to the cascade question
- 7.3% could not or refused to provide any answer.

18.34.2 Wage Rate Edit and Income Imputation

As mentioned in the previous section, a wage rate edit was carried out to compare earnings from employment to income. The combined earnings of the PMK and spouse/partner were compared to household income and the earnings of the PMK were compared to the PMK income. Cases where there were large discrepancies were flagged for a manual review. Cases where household income was either very high or very low were also flagged (less than $10,000 or over $150,000). In the manual review, if a decision was made that the household and/or PMK income were not correct they were set to not-stated.

As a result of this edit, household income was set to not-stated for 1,375 records (10.2% of responding households) and the PMK income was set to not-stated for 2,055 records (15.3% of PMKS).

Two observations were made during this edit:

- There was feeling that there was a tendency for household income to be under reported when a comparison was made to earnings from employment. Comparisons with data from the Survey of Consumer Finances presented in the next section substantiated this suspicion.

- There were many cases when the PMK did not correctly report her personal income. Instead of reporting personal income she repeated the figure for household income. This was particularly true in Québec. Changes have been made to the French wording for the PMK income question for Cycle 2.

After the wage rate edit any record with a not-stated value to either household or PMK income was imputed. An exact income value was imputed for any record where only the cascade question had been answered.

The imputation method used was "hot-deck imputation". For any record with a not-stated income value, a "donor" record was found. This donor record was:

- a responding NLSCY record with a valid value for income and

- was deemed to be similar to the record with the not-stated income value.

"Similar" was determined by way of matching variables. If a record had a not-stated value for household income the donor record had to be similar with respect to:

- earnings from employment for the PMK and spouse/partner
- sources of household income
- Pineo occupation code for PMK and spouse/partner
- economic region
- highest level of education for the PMK and spouse/partner.

If there was an answer provided for the income cascade question an exact value was imputed within the appropriate range.

In total household income was imputed for 3,545 (26% of responding households):

For:

- 11.1% of households, a household income was imputed making use of the cascade question - i.e., within the range indicated by the cascade question

- 10.9% of households, a household income was imputed matching on employment earnings of the PMK and spouse/partner

- 4.3% of households, household income was imputed using other matching variables.

The PMK's personal income, was imputed for 3,774 records (28% of responding households):

For:

- 5.3% of PMKs, income was imputed making use of the cascade question - i.e., within the range indicated by the cascade question

- 17.0% of PMKs, income was imputed matching on employment earnings of the PMK

- 5.7% of PMKs, income was imputed using other matching variables.

18.34.4 Comparison to Survey of Consumer Finances

In order to assess the quality of the income data for the NLSCY, comparisons were made to data from the Survey of Consumer Finances (SCF).

SCF is an annual survey conducted by Statistics Canada. For this survey detailed income information is collected individually for each household member. An amount is collected for each specific source of income. These numbers are all added together to get an estimate of total household income, In order to make the comparison, only SCF households with at least one child in the 0 to 11 age group have been included. The reference year for the SCF data is 1994.

The table below compares the distribution of household income from the NLSCY (before the wage rate edit and after imputation) to the SCF distribution.

HOUSEHOLD INCOME DISTRIBUTION

HOUSEHOLD INCOME	NLSCY (before wage rate edit)	NLSCY (after imputation)	SCF
Less than $20,000	18.1%	15.9%	14.1%
$20,000 - $29,999	11.0%	10.9%	10.9%
$30,000 - $39,999	13.9%	13.7%	13.2%
$40,000 - $49,999	14.0%	13.9%	14.2%
$50,000 - $59,999	12.4%	12.7%	13.3%
$60,000 - $69,999	8.8%	9.9%	10.5%
$70,000 - $79,999	7.9%	8.0%	7.7%
$80,000 or more	13.9%	15.1%	16.2%

When the distribution of household income is compared from the NLSCY to SCF it would seem that the NLSCY underestimates household income. The wage rate edit and income imputation resulted in the income distribution coming closer to SCF but there still seems to be an underestimation problem.

One potential reason for the problem may be as a result of the way that the income question was asked. For the NLSCY there was only one overall global question where the respondent had to include income figures for all household member for all sources. Perhaps this is too onerous a task for the respondent and certain portions of income are forgotten and not reported. Various alternatives for the income questions will be considered for future NLSCY cycles.

18.34.6 Microdata File Suppressions

Again it was necessary to make certain suppressions of income data for the microdata file.

The main source for household income has been grouped into three categories:

- wages and salaries, income from self-employment
- worker's compensation, unemployment insurance, social assistance
- other

This variable was suppressed for households where there was a male PMK with no spouse/partner.

A variable was created for household and PMK income (AINHD01A and AINPD02) for all households with the following categories:

- less than $10,000
- $10,000 - $14,999
- $15,000 - $19,999
- $20,000 - $29,999
- $30,000 - $39,999
- $40,000 or more

For households in which there was a couple i.e., the PMK had a spouse/partner it was permissible to have more detail at the upper end. Therefore a second income variable (AINHD01B) was set up with the following categories:

- less than $10,000
- $10,000 - $14,999
- $15,000 - $19,999
- $20,000 - $29,999
- $30,000 - $39,999
- $40,000 - $49,999
- $50,000 - $59,999
- $60,000 - $79,999
- $80,000 or more

This second variable has been set to not-applicable on the microdata file for all households where the PMK does not have a spouse/partner.

18.36 Friends and Family

Friends and Family was one of the sections on the questionnaire completed by children in the 10 to 11 age group. The objective was to determine how well the child felt he/she was getting along with others.

The section collected information on numbers of close friends, time spent with friends, presence of someone the child can confide in, and the quality of relationships with others, such as parents, peers and teachers. This information is important in identifying the extent and quality of the child's social support network. To allow for comparison, the section includes questions which are also included on the Child's Questionnaire completed by the PMK.

There was one group of questions in this section which were part of a scale. Items AA1CQ01, AA1CQ02, AA1CQ03, and AA1CQ04 are intended to measure how well the child gets along with peers. It is part of

the Peer Relations Sub-scale from the Marsh Self-Description Questionnaire, developed by H.W. Marsh. The analysis of the factor structure of this scale has not yet been carried out. The results of this analysis will be included in the second release of NLSCY data.

18.38 Feelings and Behaviour

This section was part of the self-complete questionnaire given to children in the 10 to 11 age group. The objective of this section was to determine the child's perception of his/her general behaviour and the child's engagement in risk-taking behaviours.

This section replicates the behaviour checklist included on the Child's Questionnaire completed by the PMK (Section 9.6) and the one on the Teacher's Questionnaire (which is to be included in the second release of NLSCY data). It is intended to provide indicators of the following behaviours: conduct disorder, hyperactivity, inattention, physical aggression, indirect aggression, emotional disorder, anxiety, and prosocial behaviours. There were also questions which probe for difficult behaviours; these were also included on the Child's Questionnaire completed by the PMK.

The analysis of the factor structure of the behaviour scale has not yet been carried out for the 10 to 11 Questionnaire for NLSCY data. The following indicates the constructs or factors that the behaviour scale is intending to measure, the items that are believed to be included in the factor and the sources for the items.

- **Conduct disorder**:
  Items include AD1CQ01C, E, G, L, O, T, AA, DD, FF, JJ, and PP from the Ontario Child Health Study (OCHS).

- **Hyperactivity**
  Items include AD1CQ01B, I, N, P, S and W from the Ontario Child Health Study and AD1CQ1HH from the Montreal Longitudinal Survey.

- **Emotional disorder**
  Items include AD1CQ01F, K, Q, V, CC, MM, and RR from the Ontario Child Health Study.

- **Anxiety**
  Items include AD1CQ01Y and AD1CQ1II from the Montreal Longitudinal Survey and several of the OCHS emotional disorder items - AD1CQ01F, Q, V and CC.

- **Indirect aggression**

Items include AD1CQ01J, R, Z, LL and TT from Lagerspetz, Bjorngvist and Peltonen of Finland.

- **Physical aggression**
  Items include AD1CQ01X from the Montreal Longitudinal Survey and AD1CQ01G, AA and NN from the Ontario Child Health Study.

- **Inattention**
  Items include AD1CQ01P from the Ontario Child Health Study and AD1CQ1EE, KK, QQ from the Montreal Longitudinal Survey.

- **Prosocial behaviour**
  Items include AD1CQ01A, H, M GG and OO from the Ontario Child Health Study and AD1CQ01D, U, BB, SS, and UU from the Montreal Longitudinal Survey.

In the next release of NLSCY data, a complete analysis will be carried out on the factor structure for the 10 to 11 behaviour scale and scores will be computed and included on the microdata file.

The questions on difficult behaviours of 10 and 11 year-olds are based on the following sources: AD1CQ01A to D are from the National Longitudinal Survey of Youth at Ohio State University; AD1CQ02E is from the Western Australia Child Health Survey, and AD1CQ02F was provided by Dr. R. Tremblay of the University of Montreal.

18.40 My Parents and Me

This section was part of the self-complete questionnaire given to children in the 10 to 11 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 also was 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 interaction, hostile/ineffective parenting, and consistent parenting, aversive and non-aversive parent management techniques.)

The analysis of the factor structure of the parenting scale has not yet been carried out for the 10 to 11 Questionnaire for NLSCY data. It will be carried out by the time of the next release and details will be given at that time.

18.42 PPVT-R

The Peabody Picture Vocabulary Test - Revised (PPVT-R) was administered to the NLSCY children in the 4 to 5 age group.

The PPVT-R was designed to measure receptive or hearing vocabulary and in fact can be used for any 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 4 to 5 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 which 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 5 year-old would be expected to perform better on the PPVT-R than a 4 year-old and thus have a higher score. The standard score takes account of 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 4 to 5 age group. It should be notes 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). Reliability scores for the NLSCY sample will be computed and included in the next release.

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 which may have influenced the child's responses and his/her overall reaction to the test.

Responses to these assessment questions are included on the microdata 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 labelled as APACS01 on the microdata file. It was derived using the following items; APACQ1 (attitude) APACQ2 (rapport), APACQ3 (perseverance/persistence), APACQ4 (cooperation), APACQ5 (motivation/interest), and APACQ9 (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 APACQ1, APACQ2, APACQ3, APACQ4 and APACQ5 (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 labelled as APACS02 on the microdata file. It was derived using the following items; APACQ6 (problems with visual sharpness), APACQ7 (problems with hearing) and APACQ8 (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 APACS03 on the microdata file. It was derived from APACQ13 (light problems) and APACQ14 (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 APACS04 on the microdata file. It was derived from APACQ10 (noise interference), APACQ11 (interruptions), APACQ12 (distractions) and APACQ15 (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.

In order 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

PPVT ASSESSMENT FACTOR	% of children with problem	Average standard score for children with problem[5]	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 [6]	-	
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.

In Section 10.5, response rates for the PPVT-R are presented along with estimates of response bias.

18.44 Mathematics Computation Test

All NLSCY children in grade 2 or above were to complete a Mathematics Computation Test. During the household interview the parent was asked to give consent for the test to be administered by the child's teacher. The test that was administered was a shortened version of the Mathematics Computation Test of the standardized Canadian Achievement Tests, Second Edition (CAT/2). CAT/2 is a test series designed to measure achievement in basic skills. The Mathematics Computation Test measures a student's understanding of the operations of addition, subtraction and multiplication and division of whole numbers. The shortened test that was developed for the NLSCY was a 10 question test for grades 2 and 3 and a 15 question test for children in the higher grades. There were actually three versions of the test administered to the NLSCY children in Cycle 1. Children in grades 2 and 3 were given the level 2 test, children in grades 4 and 5, the level 4 test, and children in grades 6 and 7, the level 6 test.

Every child who completed the mathematics computation test was assigned a raw and a standard score. A raw score was calculated by adding the number of correct responses to the test. Standard scores were developed based on a sample of Canadian children across all ten provinces, referred to as the norm sample. This norm sample was selected by the Canadian Testing Centre. Children in this norm sample in grades 2 and 3 (i.e., who took the level 2 test) were assigned standard scores in the 200 to 400 range (approximately) based on the number of correct responses to the test (i.e., the raw score), children in grades 4 and 5 (i.e., who took the level 4 test) were assigned standard scores in the 264 to 550 range, and children in grades 6 and 7 (i.e., who took the level 6 test) were assigned scores in the 314 to 624 range. Thus children were

essentially assigned a continuous score through equi-percentile equating which is expected to increase over time as the child progresses through school. The CAT/2 scale is designed to have scores ranging from 1 to 999. Equivalence tables were set up to relate the raw score to a standard score by the level of test that was administered based on the results for children in the norm sample.

For children in the NLSCY sample, a standard score was assigned to each child using the equivalence tables. The advantage of using the standard score is that it will possible to track a child's progress over time by comparing his or her standard score to the average standard score for the grade level and by examining growth curves.

It should be noted that there were some problems with the Mathematics Computation Test for Cycle 1 of the NLSCY. For certain grades, the test that was used was not significantly difficult to properly distinguish math computation abilities for children in these grades. The problem was critical for grades 3 and 5 where many children achieved a perfect score on the test. This is often referred to as a "ceiling effect". As a result, great caution should be taken in analysing the data for children in grades 3 and 5. This problem is discussed further in Section 10.8. For Cycle 2 of the NLSCY a more difficult test has been developed for children in these grades.

A second problem is related to response rates. The response rate for the Mathematics Computation Test was quite low; a score is available only for approximately 50% of the responding NLSCY children in grades 2 and over. Reasons for this low response rate are discussed in Section 10.7.

The estimates derived from this survey are based on a sample of children. Somewhat different figures might have been obtained if a complete census had been taken using the same questionnaire, interviewers, supervisors, processing methods, etc. as those actually used. The difference between the estimates obtained from the sample and the results from a complete count taken under similar conditions is called the sampling error of the estimate.

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

In this section some of the non-sampling errors that occurred in the NLSCY are discussed. Non-response to the various components of the NLSCY is discussed in detail. It should be noted that further information regarding data quality in the various sections of the NLSCY questionnaire can be found in Section 9.

20.2 Overall Response Rates and Response Bias

In total, 15,579 households were selected to participate in the NLSCY. Out of these selected households a response was obtained for 13,439 which results in an overall response rate of 86.3%.

In the table that follows the number of households selected in each province is presented as well as the number of responding households and

the response rate. This is followed by a table giving the response rates for the Main and Integrated Components.[26] It should be noted that for the
Main Component only households which were respondents to the Labour Force Survey were included in the NLSCY sample. The impact of not including LFS non-respondents is discussed later on in this section.

NLSCY RESPONSE RATES BY PROVINCE

PROVINCE	SAMPLED[7] HOUSEHOLDS	RESPONDING HOUSEHOLDS	RESPONSE RATE
Newfoundland	889	803	90.3%
Prince Edward Island	481	422	87.7%
Nova Scotia	1,059	926	87.4%
New Brunswick	980	857	87.4%
Quebec	2,867	2,514	87.7%
Ontario	4,268	3,519	82.5%
Manitoba	1,133	1,001	88.3%
Saskatchewan	1,166	1,039	89.1%
Alberta	1,355	1,213	89.5%
British Columbia	1,381	1,145	82.9%
TOTAL[8]	15,579	13,439	86.3%

NLSCY RESPONSE RATES FOR MAIN AND INTEGRATED COMPONENTS

	SAMPLED	RESPONDING	RESPONSE

For the Integrated Component, it was often not even known if non-responding households had children or not. If there were children, the household should have been considered to be a non-respondent to the NLSCY. If there were no children it should have been considered to be out of scope and not included in the response rate calculation. In order to estimate a response rate for the Integrated Component a certain proportion of these households were estimated to have children and therefore labelled as non-responding households and the remainder were considered as out-of-scope.

	HOUSEHOLDS	HOUSEHOLDS	RATE
Main Component	12,904	11,150	86.4%
Integrated Component	2,675	2,289	85.6%
Overall	15,579	13,439	86.3%

As mentioned in Section 5, initially the response rates to the NLSCY were not as high as expected. Therefore in June of 1995 a major follow-up was conducted of all non-responding households. In this follow-up 787 households were converted to respondents thereby increasing the response rate by 5%.

There were many possible reasons why a household chose not to respond to the NLSCY. In some cases 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.

Sometimes 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 recontact the respondent.

It was necessary to come up with a criteria for classifying these "partial" interviews as respondent or non-respondent households. 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. In order to make this assessment the data collected for each selected child in the household were examined. This was done by analysing certain key questions across the Child Questionnaire. An assessment was made as to whether or not there was an adequate amount of information collected for at least one child in each household. If there was, this household was maintained in the responding sample. All missing variables for this household were set to not stated or imputed. If there was not adequate information for at least one child then the household was dropped from the responding sample and treated as a non-response. A more thorough discussion on the procedure for assigning response codes to partial interviews can be found in Section 6.3.

In the table that follows the disposition of the non-responding sample is presented.

THE NLSCY NON-RESPONDING SAMPLE BY REASON FOR NON-RESPONSE

REASON	NON-RESPONDING HOUSEHOLDS	%
Refusal	1,437	67.1
No One at Home	117	5.5
Temporarily Absent	31	1.4
Language Barrier	62	2.9
Special Circumstances (sickness in family, weather conditions etc.)	173	8.1
Partial Response (rejected due to an inadequate amount of information)	283	13.2
Other or Reason Unknown	37	1.7
Total	2,140	

Non-response Bias

Non-response is a type of error that can result in bias in survey estimates. Biased estimates can result if the non-respondents to a survey differ significantly from the respondents.

In order to study the effect of non-response bias, a study was carried out for households included in the sample for the Main Component. Since these households were at one point LFS respondents, certain information was available on both respondents and non-respondents to the Main Component. For the integrated sample there is no prior information available about non-responding household so a non-response bias study was not possible.

There were 12,904 households selected for the Main Component; 11,150 were respondents and 1,754 were non-respondents. Information collected for the LFS was compared for these responding and non-responding groups. The analysis was carried out using weighted data with a correction for the complex sampling design.

The LFS characteristics that were considered in this comparison included:

- Single-parent vs. two-parent family
- CMA/non-CMA

- Age group of parent[27]
- Student status of parent
  - student vs. non-student
  - part-time student vs. full-time student
- Education of parent
- Occupation of parent
- Industry of parent
- Labour force status of parent
- Number of children in the household.

Out of this list there was a significant difference[28] between responding and non-responding households for only four variables.

- Non-responding households were more likely to be in CMAs.

- The parent in non-responding households was more likely to be in the 40+ age group.

- The parent in non-responding households was more likely to have a lower level of education (0 to 8 years of education).

- Households where the parent was a student were more likely to be responding households.

It should be noted that problems associated with the first two variables (CMA and age) are at least partially corrected in the weight adjustments that were carried out. There was a weight adjustment made within CMAs and there was an adjustment made by single years of age for children (See Section 7). Older children are more likely to have older parents so at least some adjustment has been made to compensate for the higher non-response rate for older parents.

The impact of non-responding households where the parent had a lower level of education remains minimal due to the fact that a relatively small proportion of the sample falls into this category.

Other Sources of Bias

For the NLSCY **Main** Component there is another potential source of bias due to the method by which the sample was selected for this component. As discussed in Section 4, the sample for this component was selected from households that had participated in the Canadian Labour Force Survey. Households which had at least one child 0 to 11 years of age at

---

In this analysis the parent referred to the female parent except for families where there was a lone parent and that parent was male.

Only differences that were significant at the 95% confidence level are reported.

the time of the LFS interview were selected for the NLSCY. This sampling methodology results in two problems.

One problem is that only respondents to the LFS were considered for the NLSCY sample for the Main Component. It could be that some of the LFS non-respondents had children in the appropriate target age group; but these households were not included in the NLSCY sample. The response rate to the LFS was approximately 95% in the time period in which the NLSCY sample was selected. It is estimated that approximately 700 households with children were missed due to the fact that no attempt was made to make contact with non-responding LFS households.

A second complication was due to the fact that only households for which there were children when the LFS was conducted were included in the NLSCY sample. It is possible that households were not included in the sample since they were vacant or only contained members 12 years of age or older at the time of the LFS. Some of these household may have had children (0 to 11) living with them a few months later when the NLSCY interview took place. It is estimated that approximately 240 households with children were missed in the NLSCY sample for this reason. It is likely that a large portion of these 240 household would represent households with a newborn, since the newborn came into the family after the time of the LFS. The weighting procedures carried out (see Section 7) compensates for the under representation of 0 year-olds at the global level, but there is likely an under representation of children 0 to 3 months old.

In total 3080 households were missed due to non-response to the NLSCY interview (2140) or due to the two problems discussed above (940). A complete interview was obtained for 13,439 households which represents 81.4% of the total households estimated to have children in the 0 to 11 age group.

20.2.2 Component Non-response

As discussed in Section 5, there were several respondents or components to the NLSCY interview. The PMK provided detailed information about each selected child. In the parent and the general interview, the PMK provided information about herself and her spouse/partner. The PPVT-R test was administered to children in the 4 to 5 age group. Children in the 10 to 11 age group completed a questionnaire on their own. For school-aged children the teacher completed a questionnaire about the child and if the child was in grade 2 or above a Math Test was administered. There was a potential for non-response for each of these individual components.

It should be noted however, that if a household was deemed to be a responding household, then all required components were created for that household; even if there were no data provided for a particular

component. For example, if there was a 10 year-old in a responding household who would not complete the 10 to 11 Questionnaire, then this component still exists for the child, with all variables set to not-stated. Likewise if a parent completed a Child Questionnaire for one child in the household but refused to provide information for a second child then there is a record for this second child with not-stated values for all variables.

The following sections provide a summary of the degree to which there is complete data for the various NLSCY components. A brief summary of the content for each of these components can be found in Section 5. As it can be seen in the sections that follow, the impact of partial non-response on data quality is minimal. The one exception to this is the Mathematics Computation Test (Section 10.7).

20.4 Child Questionnaire Response Rates

In order to assess the completeness of the child data (i.e., the information provided by the PMK about the child) several key questions were identified across the Child Questionnaire. One item was selected from most of the sections on this questionnaire in a somewhat random fashion to assess data quality. Of the responding sample of 22,831 children:

- there were answers to all key questions for 97.1% of the children,
- there was partial information for 1.3% and,
- for 1.6% of these children there was a non-response to all key items.

20.6 Parent Questionnaire Response Rates

This questionnaire was administered for the both the PMK and spouse/partner. Again key questions were identified to assess completeness. Out of the 24,692 PMKs and spouse/partners:

- there were answers to all key questions for 91.8% cases,
- there was partial information for 6.3%[29] and,

The reason for the high number of partials for the parent questionnaire was because of one of the questions that was chosen as a key item. This item was question 6A from the Neighbourhood Section - "If there is a problem around here, the neighbours get together to deal with it." A fairly high number of parents answered

- for 1.9% of cases there was a non-response to all key items.

20.8 General Questionnaire Response Rates

This questionnaire was also administered to the PMK and spouse/partner. Out of the 24,692 PMKs and spouse/partners:

- there were answers to all key questions for 98.5% of cases,
- there was partial information for 0.1% and,
- for 1.4% of cases there was a non-response to all key items.

20.10 PPVT-R Response Rates and Bias

For the 3,728 children in the 4 to 5 age group:

- a PPVT-R score was calculated for 88.8% of these children,
- 1.0% of children started the test but did not finish so a score could not be calculated,
- for the remaining 10.2% of children the test was not started.

In order to assess non-response bias for the PPVT-R, characteristics of the children who completed the test (88.8%) were compared with those who did not (11.2%).

The following table presents the variables included in this non-response bias study and the results. An explanation is given for differences significant at the 95% confidence level.

NON-RESPONSE BIAS FOR PPVT-R

VARIABLE	RESULT
Sex of the child (AMMCQ02)	Girls were more likely to respond to the PPVT-R than boys. The response rate for girls was 90.5%

don't know to this question.

and the response rate for boys was 87.3%.

Parent Status (ADMCD04)
- child lives with
    - two parents
    - one parent
    - no parents

No effect

Score on the hyperactivity factor on the behaviour scale on the Child's Questionnaire (ABECS06)

Children who were more hyperactive were more likely to be respondents.

Average hyperactivity score
- Respondents - 5.0
- Non-respondents - 4.6

Score on the prosocial factor on the behaviour scale on the Child's Questionnaire (ABECS07)

Children who were more prosocial were more likely to be respondents.

Average prosocial score
- Respondents - 11.5
- Non-respondents - 10.5

Score on the emotional disorder factor on the behaviour scale on the Child's Questionnaire (ABECS08)

No effect

Score on the conduct disorder factor on the behaviour scale on the Child's Questionnaire (ABECS09)

Children who had higher conduct disorder scores were more likely to be respondents.

Average conduct disorder score
- Respondents - 1.7
- Non-respondents - 1.3

Score on the indirect aggression factor on the behaviour scale on the Child's Questionnaire (ABECS10)

Children who scored higher on the indirect aggression scale were more likely to be respondents.

Average indirect aggression score
- Respondents - 0.8
- Non-respondents - 0.6

Household income (AINHD01)

No effect	Current working status of PMK (ALFPD28) - full-time, part-time or not working
No effect	Highest level of education of PMK
No effect	

Province	The response rate was lower in Manitoba, Saskatchewan and Alberta, (84.1%, 80.6% and 83.1%)[9]

20.12 10 to 11 Questionnaire Response Rates and Bias

Again key questions (nine in total) were identified on the 10 to 11 Questionnaire in order to assess completeness. Out of the 3,434 children in the 10 to 11 age group selected in responding households:

- there were answers to all key questions for 75.1% of these children,
- at least half but not all key questions were answered for 11.6% of the children,
- at least one but less than half of the key questions were answered for 3.0% of the children and,
- for 10.3% there was a non-response to all key items.

In order to assess non-response bias for the 10 to 11 Questionnaire, characteristics of the children who answered at least half of the key questions (86.7%) were compared with those who did not (13.3%).

The following table presents the variables included in this non-response bias study and the results. Only results significant at the 95% confidence level are presented. Children who answered at least half of the key questions are labelled as respondents in this table.

NON-RESPONSE BIAS FOR 10 TO 11 QUESTIONNAIRE

VARIABLE	RESULT
Sex of the child (AMMCQ02)	Girls were more likely to respond to the 10 to 11 Questionnaire than boys. The response rate for girls was 87.8% and the response rate for boys was 85.5%
Parent Status (ADMCD04) - child lives with - two parents - one parent - no parents	No effect
Score on the hyperactivity factor on the behaviour scale on the Child's Questionnaire (ABECS06)	No effect
Score on the prosocial factor on the behaviour scale on the Child's Questionnaire (ABECS07)	Children who were more prosocial were more likely to be respondents. Average prosocial score • Respondents - 13.0 • Non-respondents - 12.4
Score on the emotional disorder factor on the behaviour scale on the Child's Questionnaire (ABECS08)	No effect
Score on the conduct disorder factor on the behaviour scale on the Child's Questionnaire (ABECS09)	No effect
Score on the indirect aggression factor on the behaviour scale on the Child's Questionnaire (ABECS10)	No effect
How well the child is doing at school in reading based on information from the parent on the Child's Questionnaire (AEDCQ14A)	Children who were doing poorly or very poorly in reading were more likely to be non-respondents. For children who had poor or very poor reading skills the response rate was 67.3%. For children who were reported to have average or above average skills the response rate was 88.4%

How well the child is doing at school "overall" based on information from the parent on the Child's Questionnaire (AEDCQ14D)

Children who were doing poorly or very poorly in school were more likely to be non-respondents. For children who very doing poorly or very poorly in school the response rate was 59.3%. For children who were reported to have average or above average skills the response rate was 88.2%	
Household income (AINHD01)	Children living in households with lower incomes were more likely to be non-respondents. Average household income • Respondents - $50,466 • Non-respondents - $43,633
Highest level of education of PMK Children for which the PMK had a higher level of education were more likely to be respondents	
Province	The response rate was lower in Manitoba, Saskatchewan and Alberta, (82.5%, 76.8% and 84.7%)[10]

20.14 Mathematics Test Response Rate and Bias

As mentioned earlier, one component of the NLSCY was administered at school. Wherever parents and school boards consented, the child's teacher and principal were contacted and asked a number of questions about the child and his/her school environment. Children in grade 2 and above were also given a short mathematical skills test. Mathematics tests were completed for only 50.5% of the children in grade 2 and over who were part of the NLSCY responding sample. The table that follows shows the distribution by province.

DISTRIBUTION OF MATHEMATICS TESTS BY PROVINCE

	Number of children	Number of mathematics tests completed

	"eligible" for the mathematics test	
Newfoundland	541	378
Prince Edward Island	281	153
Nova Scotia	549	326
New Brunswick	534	316
Québec	1,372	505
Ontario	2,208	1,160
Manitoba	646	327
Saskatchewan	699	304
Alberta	835	436
British Columbia	702	334
Total	8,367	
	4,239	

The number of math tests completed divided by the number of children "eligible" for the test represents the response rate for the math test. A lower response rate has two potential consequences. First, it reduces the actual sample size for which users will have data. Second, non-respondents may have different characteristics from respondents, which would produce a bias in the results.

The math test response rate was lower than originally hoped. Various factors affected the response rate. Although no one factor was particularly detrimental, a combination of factors had an impact on the response rate. Nevertheless, it is unlikely that all these factors had the same effect on potential bias. For example, to boost the response rate in households, follow-up collection was carried out in June of 1996. For

operational reasons, no math tests were administered at that time. In addition, a number of consent forms in Québec were processed too late for the test to be administered. While these factors did contribute to non-response to the math test, they probably had less effect on potential bias than cases where parents refused permission for their children to take the test. The various components of test non-response are shown in the table below.

NON-RESPONSE FACTORS

Component of non-response	Portion of non-response (%)
June follow-up	5.9
Consent form not received	13.7
Parent refusal	3.4
School board refusal	9.6
Teacher non-response	17.4
Other	1.8
Total	51.8

A study was done to assess the impact that the low response rate had on the results. It is difficult to quantify the actual impact; however it is possible to examine some of the characteristics observed in household interviews and compare distributions for cases where there was a response to the mathematics test vs. a non-response. If those characteristics are related to the test results, and if a difference in behaviour is noted between respondents and non-respondents, it can be assumed that there may be some bias in the data.

The table that follows provides an example of this phenomenon for grade 2 students.

DISTRIBUTION OF MATHEMATICS TEST RESPONDENTS AND NON-RESPONDENTS BY HOUSEHOLD INCOME CATEGORY

	0-$14k	$15k-$29k	$30k-$49k	$50k-$59k	$60k +	Tot.
Respondents Math. (%)	14.3	28.5	29.6	15.2	12.4	
100	Non-respondents Math. (%)	16.9	28.6	26.6	16.3	11.5

As the table shows, there is a difference between the distributions. The next table shows the average score to the math test by household income class.

AVERAGE RAW SCORES OF MATHEMATICS TEST RESPONDENTS (GRADE 2) BY HOUSEHOLD INCOME CATEGORY

	0-$14k	$15k-$29k	$30k-$44k	$45k-$59k	$60k +
Score	5.24	5.68	6.08	6.32	6.55

There is a higher percentage of children that are non-respondents to the math test for the lower income class and a lower percentage in the higher income class. If it is assumed that the average math score is the same for respondents and non-respondents within an income class, the results from the responding sample could be expected to be slightly higher than the results that would have been obtained in the whole population.

The following table presents the list of variables that were compared and shows the ones that had significant differences between respondents and non-respondents to the mathematics test.

COMPARISONS BETWEEN RESPONDENTS AND NON-RESPONDENTS TO THE MATH TEST FOR VARIOUS CHARACTERISTICS OF THE NLSCY

VARIABLE	DIFFERENCES	GROUP FOR WHICH RESPONSE RATE IS LOWER
Failed a grade	x	

failed a grade	How well child is doing in reading (according to PMK)	x
poorly, very poorly	How well the child is doing in Math (according to PMK)	x
average, poorly, very poorly	How well the child is doing in composition (according to PMK)	x
poorly, very poorly, not-applicable	How well the child is doing in general (according to PMK)	x
poorly, very poorly	Received help outside the school	x
received help	Contacted by school regarding behaviour	x
	Looks forward to go to school (according to PMK)	

twice or more			
no			Important that child has good grades (according to PMK)
	x	not important/ refusal	How far it is hoped that the child will go in school (according to PMK)
	x	primary/secondary/ other/CEGEP/trade	Progress important at the school
	x	refusal	Child enjoys being at school (according to PMK)
	x	refusal/disagree	Parents welcome at school
	x	refusal/disagree	School spirit high
	x	refusal	Child receives special education
	x	yes/refusal	PMK has a high school diploma
	x	no	PMK went to a post-secondary establishment
	x	no	Household income[11]
	x	0-$15k, refusal	PMK income
	x		Number of children in the household

no		CMA
x	certain CMAs in Quebec, as well as Kitchener and Regina	

Child's health

no
Reads well with glasses

no

Reads well without glasses	no
Needs a hearing aid	no

The observed differences in the various characteristics appear to give evidence that there may be a tendency to overestimate the average scores of the mathematics test of the responding sample compared to the results that would have been obtained if everybody in the sample had been a respondent.

20.16 Ceiling Effect for Mathematics Test

The mathematics tests administered in Cycle 1 were fairly short. There were 10 questions in the test for grade 2 and 3 students, and 15 questions in the test given to students in higher grades. Furthermore, in order to streamline administrative procedures, tests with the same level of difficulty were used for two grades (e.g. second and third graders took the same test, as did grade 4 and 5 students and grade 6 and 7 students). Although the problem did not show up in the initial testing, a ceiling effect was noted, especially among third and fifth graders (the ceiling is the highest possible mark on the test, and a high ceiling effect indicates that "too many" children had perfect marks).

DISTRIBUTION OF CHILDREN WITH PERFECT MARKS BY GRADE

	Percentage of children with perfect marks
Grade 2	10.6 %
Grade 3	38.0 %
Grade 4	3.2 %
Grade 5	14.7 %
Grade 6	4.5 %

Comparisons at the provincial level reveal even more pronounced differences. Québec in particular had a more serious ceiling effect. Consequently, even though the mathematics test scores are available for all grades, it is recommended that the data for grades 3 and 5 not be used, or that they be used with extreme caution!

For the next cycle, a number of steps have been taken to improve the results by increasing the response rate and reducing the ceiling effect. First, there will be a different test for each grade. All mathematics tests will consist of 15 questions. An aptitude indicator will be administered during the home interview to help identify the child's grade and to assist in the imputation of missing responses where necessary. In addition, in order to improve response rates, more effort has been put in encouraging school boards to co-operate and a better tracking system for consent forms has been implemented. These measures should help eliminate most of the problems encountered in Cycle 1.

22.0

Guidelines for Tabulation, Analysis and

Release

This section of the documentation outlines the guidelines to be adhered to by users tabulating, analysing, 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.

22.2 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 corresponding unrounded components and then are to be rounded themselves to the 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.

22.4 Sample Weighting Guidelines for Tabulation

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 sampling 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.

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.

22.4.2 Definitions of Types of Estimates: Categorical vs. Quantitative

It should be pointed out that 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 (AWTCW01) is a "child" weight. Estimates of parents or families cannot be made from the NLSCY microdata file. A further discussion of units of analyses can be found in Section 8.1 of this document.

Before discussing how the NLSCY data can be tabulated and analysed, it is useful to describe the two main types of point estimates of population characteristics which can be generated from the microdata file for the NLSCY.

Categorical Estimates

Categorical estimates are estimates of the number, or percentage of the surveyed population possessing certain characteristics or falling into some defined category. The number of children who were born before the due date or the proportion of children who were in excellent health at birth are examples of such estimates. 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

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 formX hat over Y hat where X hat is an estimate of the surveyed population quantity total and Y hat is an estimate of the number of persons 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.

Examples of Quantitative Questions:

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

22.4.4 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. Proportions and ratios of the form X hat over Y hat are obtained by:

(a) summing the final weights of records having the characteristic of interest for the numerator (X hat),
(b) summing the final weights of records having the characteristic of interest for the denominator(Y hat), then

(c) dividing the numerator estimate by the denominator estimate.

22.4.6 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,[30]

Do not include: "don't know", "refusal" and "not-stated" codes in this tabulation (i.e., records for which the number of days of special care is code 997, 998 or 999). For cases where the number of days is not-applicable (i.e., 996) because no care was received recode the number of days to 0 to perform the calculation.

- then sum this value over all records for which the child was born prematurely

To obtain a weighted average of the form X hat over Y hat, the numerator(X hat),is calculated as for a quantitative estimate and the denominator(Y hat),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[31], then
(c) divide estimate (a) by estimate (b).

22.6 Guidelines for Statistical Analysis

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.

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.

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.

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
- Calculate the AVERAGE weight for these records by summing the original person weights from the microdata file for these

---

Do not include premature babies for which the number of days was don't know, refusal, or not-stated in this calculation (i.e., 997, 998 or 999).

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.

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.

22.8 C.V. 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*. Data quality is affected by both sampling and non-sampling errors as discussed in Section 10. However for this purpose, the quality level of an estimate will be determined only on the basis of sampling error as reflected by the coefficient of variation as shown in the table below. Nonetheless users should be sure to read Section 10 to be more fully aware of the quality characteristics of these data.

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.

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."

24.0

Approximate Sampling Variability 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, which can be found in Appendix 3, allow the user to obtain an approximate coefficient of variation based on the size of the estimate calculated from the survey data.

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.

For the NLSCY the sample was constructed taking account of two important requirements.

- a sufficient sample was required in each of the 10 provinces to allow for the production of reliable estimates for all children 0 to 11 years of age.

- a second requirement was that it was necessary to have a large enough sample to produce estimates at the Canada level by seven key age groupings or cohorts: 0 to 11 months, 1, 2 to 3, 4 to 5, 6 to 7, 8 to 9, and 10 to 11 years.

The tables that follow show the design effects, sample sizes and population counts first by province and then by age groupings which were used to produce the Approximate Sampling Variability Tables.

PROVINCE	DESIGN EFFECT	SAMPLE SIZE	POPULATION
Newfoundland	2.8	1,232	89,533
Prince Edward Island	2.4	764	23,161
Nova Scotia	3.4	1,532	144,744
New Brunswick	3.1	1,426	115,913
Québec	2.5	4,065	1,099,033
Ontario	4.1	6,020	1,777,525
Manitoba	3.4	1,789	183,268
Saskatchewan	2.4	1,878	176,449

Alberta	2.3	2,185	489,604
British Columbia	3.5	1,940	574,160
Atlantic provinces	3.3	4,954	373,351
Prairies	2.4	5,852	849,321
Total[12]	3.6	22,831	4,673,390

AGE GROUP	**DESIGN EFFECT**	**SAMPLE SIZE**	**POPULATION**
0 to 11 months	3.3	2,227	370,887
1 year	3.4	2,469	381,711
2 to 3 years	2.3	3,909	791,754
4 to 5 years	1.9	3,728	800,064
6 to 7 years	2.5	3,550	763,632
8 to 9 years	2.2	3,514	783,049
10 to 11 years	2.1	3,434	782,293
0 to 3 years	2.8	8,605	1,544,352
4 to 11 years	2.2	14,226	3,129,038
4 to 7 years	2.2	7,278	1,563,696
8 to 11 years	2.2	6,948	1,565,342
Total (0 to 11)	3.6	22,831	4,673,390

All coefficients of variation in the Approximate Sampling Variability Tables are approximate and, therefore, unofficial. Estimates of actual variance for specific variables may be obtained from Statistics Canada on a cost-recovery basis. The use of actual variance estimates would likely result in estimates with lower variances; for example it could be that estimates in the "unacceptable" category according to the Approximate Sampling Variability Tables may move up to the "marginal" category. See Section 11.4 for more information on c.v. release guidelines.

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.

24.2 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. See Appendix 3 for the actual "Approximate Sampling Variability Tables".

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 (i.e., less than 2,500 grams) 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 (e.g. those in a particular sex or age group within province), 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 differenced hat = X hat SUB 1 - X hat SUB 2is:

sigma_d hat ~=~ sqrt{ {(X hat_1`alpha_1)}^2 ~+~ {(X hat_2`alpha_2)}^2} where X hat SUB 1is estimate 1,X hat SUB 2is estimate 2, and α_1 and α_2 are the coefficients of variation of X hat SUB 1and X hat SUB 2respectively. The

coefficient of variation of dhat is given by σ SUB {dhat} ` / ` {dhat}. This formula is accurate for the difference between separate and uncorrelated characteristic

s, but is only approximate otherwise.

Rule 4: Estimates of Ratios

In the case where the numerator is a subset of the denominator, the ratio should be treated as a percentage and Rule 2 applied. This would apply, for example, to the case where the denominator is the number of low birth weight babies and the numerator is the number of low birth weight babies who were born prematurely (gestational age 258 days or less).

In the case where the numerator is not a subset of the denominator, 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. For example, this would apply to an estimate such as, the ratio of the number of female babies who were of low birth weight as compared to the number of male babies who were of low birth weight. The standard error of such a ratio(R hat = X hat SUB 1 ` / ` X hat SUB 2)is:

sigma_R hat ~=~ R hat `sqrt{{alpha_1}^2 ~+~ {alpha_2}^2} where $\alpha_1$ and $\alpha_2$ are the coefficients of variation of X hat SUB 1(the number of low birth weight female babies) andX hat SUB 2 (the number of low birth weight male babies) respectively. The coefficient of variation of R hat is given by σ SUB R hat ` / ` R hat. The formula will tend to overstate the error, if X hat SUB 1and X hat SUB 2are positively correlated and understate the

error if X hat SUB 1 and X hat SUB 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.

24.2.2 Examples of using the 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, 84,085 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 3 age group. Note that the question on birth weight was applicable only to children in the 0 to 3 age group and therefore this is the table that should be used to determine the c.v. for this estimate.

(2) The estimated aggregate (84,085) 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 85,000.

(3) The coefficient of variation for an estimated aggregate is found by referring to the first non-asterisk entry on that row, namely, 7.3%.

(4) The approximate coefficient of variation of the number of low birth weight babies is estimated to be 7.3%. The finding that there were 84,085 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.

Example 2 : Estimates of Proportions or Percentages Possessing a Characteristic

Using NLSCY data, it is estimated that 70.8% (59,567/84,085) 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 3 age group. Note that the questions on birth weight and delivery time were applicable only to children in the 0 to 3 age group and therefore this is the table that 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 (70.8%) and the numerator portion of the percentage (59,567) in determining the coefficient of variation.

(3) The numerator, 59,567, does not appear in the left-hand column (the 'Numerator of Percentage' column) so it is necessary to use the figure closet to it, namely 60,000. 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 5.0% is the coefficient of variation to be used.

(5) The approximate coefficient of variation of the percentage of low birth weight babies who were prematurely is estimated to be 5.0%. The finding that 70.8% of low birth weight babies were born prematurely 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.

Example 3 : Estimates of Differences Between Aggregates or Percentages

Using NLSCY data, it is estimated that 6.1% (45,690/753,203) of female babies were born prematurely, while 4.9% (38,395/791,149) 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 3 age group in the same manner as described in example 2 gives the c.v. of the estimate for female babies as 10.3%, and the c.v. of the estimate for male babies as 10.9%.

(2) Using Rule 3, the standard error of a difference(d hat = X hat SUB 1 - X hat SUB 2) is:

sigma_d hat ~=~ sqrt{ {(X hat_1`alpha_1)}^2 ~+~ {(X hat_2`alpha_2)}^2}
whereX hat SUB 1is estimate 1 (the percent of low birth weight female babies),X hat SUB 2is estimate 2 (the percent of low birth weight male babies) , and α_1 and α_2 are the coefficients of variation of X hat SUB 1and X hat SUB 2respectively.

That is, the standard error of the difference d hat = (.061-.049) = .012 is:

stackalign{sigma_d hat ~=&~ sqrt{ {[(.061)(.103)]}^2 ~+~ {[(.049)(.109)]}^2}#~&~#=&~ sqrt{~(.000039476) ~+~(.000028526)}#~&~#=&~ .008} (3) The coefficient of variation ofd hatis given by

STACKALIGN { σ_d hat ` / ` d hat ~ = & ~ .008 ` / ` .012#
~ = & ~ .667 }

(4) So the approximate coefficient of variation of the difference between the estimates is 66.7%. This estimate is "unacceptable" since the coefficient of variation is over 33.3%. Statistics Canada recommends not to release estimates of unacceptable quality.

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(=X hat SUB 1)is the number of low birth weight female babies and denominator (=X hat SUB 2)of the estimate is the number of low birth weight male babies.

(2) Refer to the table for the 0 to 3 age group. The questions on birth weight were applicable only to children in the 0 to 3 age group.

(3) The numerator of this ratio estimate is 45,690. The figure closest to it is 45,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 10.3%.

(4) The denominator of this ratio estimate is 38,395. The figure closest to it is 40,000. The coefficient of variation for this estimate is found by referring to the first non-asterisk entry on that row, namely, 10.9%.

(5) So the approximate coefficient of variation of the ratio estimate is given by Rule 4, which is,

alpha_R hat ~=~ sqrt{{alpha_1}^2 ~+~ {alpha_2}^2} where α_1 and α_2 are the coefficients of variation ofX hat SUB 1 andX hat SUB 2 respectively .

That is ,

stackalign{alpha_R hat ~=&~ sqrt{(.103)^2 ~+~ (.109)^2}#~&~#=&~0.150} The obtained ratio of female babies who were of low birth weight versus male babies who were of low birth wei

ght is 45,690/38,395 which is 1.19 : 1. The coefficient of variation of this estimate is 15.0%, which is "acceptable" and no warning message is required to produc

e this estimate since the c.v. for the estimate is in the 0.0% to 16.5% range.

24.4 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,X hat,are generally expressed as two numbers, one below the estimate and one above the estimate, as (X hat ` - ` k, ` ~ X hat ` + ` k)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 estimateX hat,and then using the following formula to convert to a confidence interval CI:

{CI}_X ~=~ [`X hat ~-~ t `` X hat `` alpha_{X hat} ~,~ X hat ~+~ t `` X hat `` alpha_{X hat} `] where α_X hatis the determined coefficient of variation of X hat,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: Release guidelines which apply to the estimate also apply to the confidence interval. 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 the high levels of error.

24.4.2 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.

STACKALIGN { X hat ~ = & ~ 5.5% ~ (0.055 ~ expressed ~ as ~ a ~ proportion) #
~ & ~ #
t ~ = & ~ 2 #
~ & ~ #
α_X hat ~= & ~7.3% ~ (0.073~expressed ~as~a~proportion) #
~ & ~is ~the ~coefficient ~of ~ variation ~ of ~ this ~ estimate #
~ & ~ as ~determined ~ by ~ the ~ tables #
~ & ~ #
CI_X ~ = & ~ {0.055 ` - ` (2)(0.055)(0.073), ~ ~0.055 `~+ ` (2)(0.055)(0.073)} #
~ & ~ #
CI_X ~ = & ~ {0.055 ` - ` 0.008, ~ ~ 0.055 ` + ` 0.008} #
~ & ~ #
CI_X ~ = & ~{0.047, ~ 0.063}}

With 95% confidence it can be said that between 4.7% and 6.3% of babies who were 0 to 3 years old at the time of the survey were of low birth weight.

24.6 How to Use the C.V. Tables to Do a T-test

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.

> LetX hat_1andX hat_2be sample estimates for two characteristics of interest. Let the standard error on the difference X hat_1 ` - ` X hat_2beσ_d hat.
>
> If ~ t ~=~ {{X hat}_1 ~-~ {X hat}_2 } over sigma_{d hat} ~ 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.

24.6.2 Example of Using the 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 female babies who were of low birth weight and the proportion of male babies who were of low birth weight. From example 3 (Section 12.1.1), the standard error of the difference between these two estimates was found to be = .008. Hence,
t~=~{{X hat}_1 ~-~ {X hat}_2} over sigma_{d hat} ~=~ {.061 ~-~ .049} over {.008} ~=~ {{.012} over {.008}} ~=~ {1.5}.Since t = 1.5 is between -2 and 2, no conclusion at the 0.05 level of significance can be made regarding the difference in proportions of male of female babies who were of low birth weight.

24.8 Coefficients of Variation for 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 (i.e., the estimate of the number of persons contributing to the quantitative 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.

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.

24.10 Release 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 warning note attached to caution subsequent users about the high level of error associated with the estimate.

GEOGRAPHICAL RELEASE CUT-OFFS

Province	Acceptable - estimates at or above	Marginal - estimates between	Unacceptable - estimates at or below
Newfoundland	7,000	2,000 & 7,000	2,000
Prince Edward Island	2,500	500 & 2,500	500
Nova Scotia	11,000	3,000 & 11,000	3,000
New Brunswick	8,500	2,000 & 8,500	2,000
Québec	24,500	6,000 & 24,500	6,000
Ontario	43,500	11,000 & 43,500	11,000
Manitoba	12,000	3,000 & 12,000	3,000
Saskatchewan	8,000	2,000 & 8,000	2,000
Alberta	18,000	4,500 & 18,000	4,500
British Columbia	35,500	9,000 & 35,500	9,000
Atlantic provinces	9,000	2,000 & 9,000	2,000
Prairie provinces	12,500	3,000 & 12,500	3,000
Total[13]	27,000	6,500 & 27,000	6,500

AGE GROUP RELEASE CUT-OFFS

Age Group	Acceptable - estimates at or above	Marginal - estimates between	Unacceptable - estimates at or below
0 to 11 months	19,000	5,000 & 19,000	5,000
1 year	18,500	4,500 & 18,500	4,500
2 to 3 years	16,500	4,000 & 16,500	4,000
4 to 5 years	14,500	3,500 & 14,500	3,500
6 to 7 years	19,500	5,000 & 19,500	5,000

8 to 9 years	17,500	4,500 & 17,500	4,500
10 to 11 years	17,000	4,500 & 17,000	4,500
0 to 3 years	18,000	4,500 & 18,000	4,500
4 to 11 years	17,500	4,500 & 17,500	4,500
4 to 7 years	17,000	4,500 & 17,000	4,500
8 to 11 years	18,000	4,500 & 18,000	4,500
TOTAL	27,000	6,500 & 27,000	6,500

It should be noted that the 'Public Use' NLSCY microdata file differs in a number of important respects from the survey 'master' file held by Statistics Canada. These differences are the result of actions taken to protect the anonymity of individual survey respondents. Actions taken to ensure confidentiality for survey respondents are discussed in Section 13.2. The methods used to detect confidentiality problems are discussed in Section 13.1.

Users requiring access to information excluded on the microdata file may purchase custom tabulations or make use of the Remote Access service described in Section 13.3.

26.2 Methods Used to Protect Confidentiality

Several measures were taken to assess disclosure risk for the NLSCY public use microdata file. Principal among these was an extensive review of all variables proposed for the public use microdata file to identify those variables considered to be "key" or "indirect identifiers". These variables are ones which may not spontaneously lead to the identification of an individual on their own but when considered in conjunction with other variables on the file could lead to disclosure.[32] For example, a child with a mother tongue of French would not be considered to be a problem with respect to confidentiality. However if that child has parents with a mother tongue of Chinese and it is known that the child lives in rural Alberta, then the risk of disclosure increases. An assessment of risk was made based on the variables considered to be indirect identifiers.

Due to the hierarchical nature of the file, all analyses to assess risk of disclosure was carried out at the family level. For example, when the variables related to language (e.g., mother tongue) were checked for risk of disclosure, a new variable was created that comprised language information for all children in the family (up to four) and language for the parents. When occupation of the parents was considered, the occupation of both parents was considered simultaneously.

There were essentially three procedures used for these variables to analyse risk of disclosure.

It should be noted that any variable considered to be a direct identifier such as the name, address or telephone number of a respondent has been suppressed on the microdata file.

1/ For cases where similar variables existed for the Census, Census data were retrieved to see if these variables (or combination of variables) were unique in the Census.

2/ For other variables, in order to assess risk systematically, an approach developed for the Census was adapted for the NLSCY. This general approach uses Census software to look at three-way combinations of variables designated to be "indirect identifers". Unusual combinations of these variables could in theory lead to spontaneous recognition an individual on the microdata file. There were two objectives:

- to identify combinations of variables that result in a high proportion of uniques i.e., what variables are "causing" an abnormal number of unique combinations.

- to determine what individual records emerge as uniques in many three-way combinations -- an indicator that the record in question is quite unusual and perhaps identifiable.

Because the NLSCY sample consists of approximately 0.5% of Canadian children, one should expect high proportions of unique combinations when several desegregated variables are combined. Therefore the goal was not to ensure that there were no unique combinations on the microdata file. This would involve making suppressions or recodes on virtually every record on the file. Instead the approach taken was to systematically identify variables and records causing the most problems and focus attention on them.

3/ Finally, all univariate counts were reviewed in isolation to assess any potential confidentiality problems. Top and bottom capping of values or regrouping of values was sometimes carried out.

Changes and suppressions made on the microdata file as a result of this analysis are presented in the next section.

26.4 Variables Available on Master File Not Included on Public Use File

The following is a summary of the actions which have been taken on the microdata file to reduce the risk of disclosure for individual respondents. Most of these suppressions are described in Section 9 by content area. A summary is provided here for convenience.

It should be noted that in the univariate counts given in Section 14, counts from the master file are presented. This way users can be aware of what is available on the master file. There is a note on the record layout explaining the nature of the

suppression that was made on the microdata file for the variable. Most suppressions involved setting a variable to not stated.

26.4.2 Geographical Variables

- It was necessary to suppress the province code (AGEHD03) on some records on the microdata file. This was done for children who did not live with a parent and children who lived in a family with a male PMK with no spouse/partner. As a result the province code was set to Not-Stated for 281 children.

- Sub-provincial indicators have not been included on the microdata file. Census Metropolitan Area (CMA) is available on the NLSCY master file (AGEHD02) as well as an indicator of urban/rural class size (AGEHD01).

26.4.4 Family Demographic variables

- Detailed age in years for the child (AMMCQ01) has been included on the microdata file (i.e., age for up to four children in the household). As a result of including detailed age, it was necessary to suppress collection date. Collection for the NLSCY took place over an eight month period. By suppressing collection date this casts some doubt on the exact ages of the children.

- It was only possible to have age in ranges for the PMK (ADMPD06D with ranges 15 to 24, 25 to 29, 30 to 34, 35 to 39, and 40+). Age for the spouse/partner has been suppressed entirely. The age group for male PMKs not living with a spouse/partner has been set to not-stated. For female PMKs not living with a spouse/partner age group has been set to not-stated for a few cases. In total, age group of the PMK was set to not-stated for 486 children on the microdata file.

- There were six sets of triplets on the file. The age of one of the trio has been altered by one and flow patterns have been adjusted accordingly.

- There is a variable on the microdata file indicating the number of people living in the household (ADMHD02). It has been capped at 6. The variable indicating the total number of persons in the economic family has been suppressed.

- On the microdata file, the sibling variables (total number of siblings (ADMCD08), number of older siblings (ADMCD09), number of younger siblings (ADMCD10) and number of siblings of exactly the same age (ADMCD11)) have all in effect been made into dichotomous variables. A code "0" means there is no such sibling and a code "1" means there is one or more of such a sibling.

- The variables on age of biological mother at birth of child (ADMCD18 and D18B) and age of biological father at birth of child (ADMCD19 and D19B) have been suppressed.

26.4.6 Ethno-cultural Variables

- It was necessary to suppress many of the variables in this section on the microdata file due to confidentiality concerns. The questions on country of birth, ethnicity and religion have all been suppressed while frequency of attendance at religious services has been retained.

- The questions on mother tongue and language of conversation are included on the microdata file but only with aggregated answer categories:

    - English only
    - French only
    - English and French only
    - At least one "other" language indicated.

    - Language of conversation:

      - the aggregated variables are labelled as ASD**P**D05B, ASD**S**D05B, and ASD**C**D05B, for the **P**MK, **S**pouse/partner and **C**hild on the microdata file. There were a few suppressions for this variable.

    - Mother tongue:

      - the aggregated variables are ASDPD06B, ASDSD06B and ASDCD06B.

- Immigrant population:

  - a derived variable was created to indicate number of years since first immigrating to Canada. It was possible to put a grouped

version (0 to 4 years, 5 to 9 years, 10 or more years) of this derived variable on the microdata file (ASDPD02B, ASDSD02B, ASDPC02B).

26.4.8 Education Variables

- Due to confidentiality concerns only an aggregated version of the highest level of education attained by the PMK and spouse/partner have been included on the microdata file. These variables (AEDPD02 for the PMK and AEDSD02 for the spouse/partner) have the following values: less than secondary, secondary school graduation, beyond high school, college or university degree (including trade).

  On the microdata file this variable has been set to not-stated for male PMKs who do not live with a spouse/partner.

- The other education variable included on the microdata file for parents, is current school status and whether attendance is full-time or part-time. These variables have been included on the file for the PMK (AEDPQ05 and Q06), but it was necessary to suppress them for the spouse/partner. If the PMK was a lone parent (i.e., did not live with a spouse/partner), then only the fact as to whether or not she/he is a student has been retained, while the detail about full-time/part-time status has been suppressed.

- For the education variables on the microdata for children, the variables on language of instruction (AEDCQ12A) and type of school (AEDCQ08) were set to not-stated in some cases because of confidentiality concerns. Only a very small number of records were affected (the variables for 34 children).

26.4.10 Labour Force Variables

- It was possible to include industry and occupation codes for the main job for the PMK and spouse/partner on the microdata file, but only for fairly large aggregate groupings. There are 21 major groups for occupation and 13 groups for industry.

  The Pineo-economic classification code for the main job has also been included on the microdata file.

  In a few cases industry and occupation codes have been set to not-stated due to confidentiality concerns. For the PMK, the occupation

codes corresponding to religion and mining have been set to not-stated.

In total:

- the occupation code was set to not-stated for 131 PMKs and for 181 spouse/partners.
- the industry code was set to not-stated for 106 PMKs and for 6 spouse/partners.
- the Pineo code was set to not-stated for 486 PMKs and for 470 spouse/partners.

- The hourly wage rate for the PMK and spouse/partner have been included on the microdata file but they have been capped at $24.00 per hour at the upper end and $5.00 per hour at the lower end. The input variables used to calculate the hourly wage rate have been suppressed.

- It was possible to include the detailed information on all jobs held by the PMK and spouse/partner in the previous year on the microdata file, except for the start and end dates of the jobs. These dates could potentially give an indication of collection date which was suppressed. However the vectors to indicate the weeks worked over the previous year for the PMK and spouse/partner have been included.

26.4.12 Income Variables

- The only variable that was allowed to go on the microdata file for sources of income was the main source of household income (AINHD02B) in three major categories:

  - wages and salaries, income from self-employment
  - worker's compensation, unemployment insurance, social assistance
  - other

  This variable was suppressed for households where there was a lone male PMK with no spouse/partner.

- A variable was created for household and PMK income (AINHD01A and AINPD02) for all households with the following categories:

  - less than $10,000
  - $10,000 - $14,999
  - $15,000 - $19,999

- $20,000 - $29,999
- $30,000 - $39,999
- $40,000 or more

- For households in which there was a couple i.e., the PMK had a spouse/partner it was permissible to have more detail at the upper end. Therefore a second income variable (AINHD01B) was set up with the following categories:

    - less than $10,000
    - $10,000 - $14,999
    - $15,000 - $19,999
    - $20,000 - $29,999
    - $30,000 - $39,999
    - $40,000 - $49,999
    - $50,000 - $59,999
    - $60,000 - $79,999
    - $80,000 or more

  This second variable has been set to not-applicable on the microdata file for all households where the PMK does not have a spouse/partner.

- The microdata file includes the ratio of household income to the low income cut-off for the economic family (i.e., the LICO) in ranges (<0.75, ≥0.75-<0.9, ≥0.9-<1.0, ≥1.0-<1.1, ≥1.1-<1.25, ≥1.25). Again it was not possible to give the exact ratio.

- The Socio-economic status variable discussed in Section 8.5 has been included on the microdata file. It was necessary to cap this variable at -2.0 at the lower end and +1.75 at the upper end.

26.4.14 Medical Biological Variables

- On the microdata file it was necessary to cap birth weight (AMDCQ13B) at the lower end at 1.499 kilograms and less.

- For multiple births the variable (AMDCQ15) was capped at the upper end at 2 or more (i.e., twins).

26.6 Remote Access Requests

During the past few years, as the surveys conducted by Statistics Canada have grown in scope and the number of variables collected increased substantially, suppression and collapsing of confidential variables has become a source of concern for many users of the data. This is particularly true for users of longitudinal data sets such as the NLSCY. As the number of variables collected about NLSCY survey respondents grows over time, more and more is known about these individuals and the protection of the person's confidentiality becomes a difficult task. At the same time, if the variables collected cannot be made readily available to users, it becomes difficult to justify the expense required to collect these variables. Often these are the very variables that are critical to a complete and comprehensive analyses of the survey data.

Statistics Canada has been striving to find a solution to this problem. In order to somewhat alleviate the problem, for the NLSCY, a procedure has been implemented whereby "place holders" for all confidential variables have been introduced on microdata file and the metadata associated with these variables have been provided (i.e., univariate distributions of the confidential variables at the aggregate level have been included in Section 14). In this way users of NLSCY data can be aware of the confidential variables which are available on the master file and can contact Statistics Canada to request special "custom" tabulations on these variables if so desired.

Unfortunately custom tabulations can be a somewhat iterative and potentially costly procedure.

For this first release of NLSCY an alternative approach is being investigated. A service which has been labelled as "Remote Access" is being proposed as a way to reduce direct involvement by Statistics Canada personnel in dealing with custom requests. With Remote Access, researchers gain access to an "enhanced" public use microdata file and **directly** formulate and test retrieval code, including queries on confidential variables. The code for these custom tabulations are transmitted electronically to Statistics Canada via the INTERNET, and moved into the Department's internal, secure environment. Next, the code would be processed on a PC, the results vetted for confidentiality, and shipped back to the client.

It should be noted that the onus is with the user to submit retrieval programs which are correct and tested. Statistics Canada will review results only for confidentiality concerns and will not make any assessment whatsoever as to whether or not the submitted program has worked properly. Initially, there should be some discussion to ensure that Statistics Canada has a copy of the software used in the submitted program.

This service will be offered to researchers who have purchased a NLSCY microdata file. The microdata file for the NLSCY has been structured as described above so there are "place holders" for all confidential variables and the metadata for these variables is available (in Section 14). The Remote Access Service is still at the pilot stage, so initially the service will be offered free of change (until the end of 1997).

At that time the project will be evaluated and an appropriate costing algorithm will be developed. The goal will be to keep the cost at a minimum since the burden of the work will rely with the user and not Statistics Canada personnel. Users interested in making use of this service can get in touch with Statistics Canada contacts mentioned in Section 1.

28.2 Technical Information

The microdata file for the NLSCY is stored as an ASCII text file. There are actually two microdata files; a primary file (NLSCYA\PRIMARY.TXT) and a secondary file (NLSCYA\SECONDRY.TXT). The primary file contains the variables that will be of most interest to users. The secondary file contains variables such as imputation flags, variables that were for the most part suppressed, and detailed information not likely to be of interest in primary analyses (for example details of the six jobs collected in the Labour Force Section of the questionnaire). The main reason for making the split into the two files was due to space. The primary file takes up 24 megabytes and the secondary file 14 megabytes.

SAS and SPSS record layouts have been included for reading the microdata files. The SAS record layouts are included in Section 14.2. Format and label statements are also included.

The electronic metadata (i.e., this document, record layouts, univariate counts etc.) is a Windows Help file, which facilitates bookmaking and annotation. Hypertext links provide navigation through the metadata based on links from the table of contents to each section or sub-section. A README file is included with a list and description of each file on the CD-ROM.

For this first release of NLSCY data, the files consist of complete child records with data for all of the sections of the various questionnaires. The data collected for the PMK and spouse/partner have been replicated for each child in the household. For example if there were three children in the family, the education variables for the PMK and the spouse have been written to all three child records. If a section of a questionnaire (or a complete questionnaire) was not applicable for a child, all of the variables for that section (or questionnaire) have been set to not-applicable. For example if the PMK did not have a spouse, the spouse variables

have been set to not-applicable for all children in the family.[33] All variables on the microdata file which are longer than one byte have been zero-filled. The record layouts included with the release package declare all variables as numeric, except those for gender (character M or F).

There are a total of 22,831 child records included on the NLSCY microdata file; i.e.; on both the primary and secondary file. The unit of analysis for all estimates made from the NLSCY file should refer to the child. The sample design used for the NLSCY was not self-weighting. When producing estimates for children, including the production of ordinary statistical tables, users must apply the proper sampling 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 has been included on both the primary and secondary files and is labelled AWTCW01. This is a cross-sectional child weight to be used for analysing Cycle 1 data. It represents the total number of children represented by the child record for Cycle 1 data. In subsequent releases both cross-sectional and longitudinal child weights will be provided.

There are two identification variables on the microdata file(s). One identifies children (CHILDID) and one identifies households (AIDHD01). The child ID (CHILDID) is unique for each child on the file and can be used to link records between the primary and secondary files. In subsequent releases of NLSCY data in the years to come, the ID for each child (CHILDID) will remain the same. There will be additional children added to the file, but once a child is included on the file, the ID for that child will remain the same. Children on this first microdata file have IDs in the range 100001 to 122831. The first digit (i.e., the '1') indicates that the child came into the sample in Cycle 1. For children introduced at Cycle 2 the first digit of CHILDID will be '2' etc.

All children in the same household will have the same household ID (AIDHD01). Over the ensuing years children will not continue to live in the same households and will move out, or families may divide. Therefore the household ID will not remain the same over time. For each cycle there will be a household ID which can be used to determine which children live in similar households for that particular cycle.

In Appendix 4, counts are given on all of the categorical variables included on the microdata file. Both unweighted and weighted counts are given for each variable. For continuous variables only ranges of values are given. It should be noted that the counts that have been provided are actually counts from the NLSCY 'master' file held by Statistics Canada. Certain suppressions and grouping of values for certain variables were required for the public use file in order to protect the anonymity of survey respondents. Actions taken to ensure the confidentiality of survey respondents are summarized in Section 13.2. Counts from the master file have been provided so that users are aware of the confidential variables that are available on

---

In subsequent releases of the NLSCY data, a series of data base files will be released for each section of the questionnaire. A record will exist for the section only if the section was applicable. The appropriate software will be provided so that users can easily link variables across files. This will be a far more efficient way to store and manipulate NLSCY data.

the master file and can contact Statistics Canada to request special "custom" tabulations on these variables if desired or make use of the Remote Access service discussed in Section 13.3. For variables which were suppressed or altered for the public use file, there is an indication on the record layout explaining what was done.

The NLSCY microdata file documentation system has employed certain standards to label variable names and values. The intent is to make interpretation of the data more straight-forward for the user. These standards and examples are provided in Section 6.5 of this guide.

28.4 SAS Record Layouts

28.4.2 Primary File

```
/*********************************************************/
/* NLSCY  PRIMARY MICRO DATA FILE                        */
/*********************************************************/

DATA PRIM;
INFILE 'NLSCYA\PRIMARY.TXT' LRECL=1077 MISSOVER PAD;

INPUT
@1   AGEHD01   2.
@3   AGEHD02   2.
@5   AGEHD03   2.
@7   AMMPQ01   3.
@10   AMMPQ02   $1.
@11   AMMPQ03A   4.
@15   AMMPQ03B   2.
@17   AMMPQ03C   2.
@19   AMMSQ01   3.
@22   AMMSQ02   $1.
@23   AMMSQ03A   4.
@27   AMMSQ03B   2.
@29   AMMSQ03C   2.
@31   AMMCQ01   3.
@34   AMMCQ02   $1.
@35   AMMCQ03A   4.
@39   AMMCQ03B   2.
@41   AMMCQ03C   2.
```

@43 ADMCD01 2.
@45 ADMHD02 2.
@47 ADMCD03 2.
@49 ADMCD04 1.
@50 ADMCD05 1.
@51 ADMCD06 2.
@53 ADMPD06A 1.
@54 ADMCD06B 2.
@56 ADMCD06C 1.
@57 ADMPD06D 2.
@59 ADMSD06E 2.
@61 ADMHD06F 2.
@63 ADMHD07 2.
@65 ADMCD08 2.
@67 ADMCD09 2.
@69 ADMCD10 2.
@71 ADMCD11 2.
@73 ADMCD12 2.
@75 ADMCD13 2.
@77 ADMCD14 1.
@78 ADMCD15 1.
@79 ADMCD16 1.
@80 ADMCD17 2.
@82 ADMCD18 2.
@84 ADMCD18B 2.
@86 ADMCD19 2.
@88 ADMCD19B 2.
@90 ADMCD20 2.
@92 AEDPQ01 2.
@94 AEDPQ02 1.
@95 AEDPQ03 1.
@96 AEDPQ04 2.
@98 AEDPQ05 1.
@99 AEDPQ06 1.
@100 AEDPD01 2.
@102 AEDPD02 1.
@103 AEDPD04 2.
@105 AEDSQ01 2.
@107 AEDSQ02 1.
@108 AEDSQ03 1.
@109 AEDSQ04 2.
@111 AEDSQ05 1.
@112 AEDSQ06 1.
@113 AEDSD01 2.
@115 AEDSD02 1.
@116 AEDSD04 2.
@118 ALFPD02 2.
@120 ALFPD03 2.

@122 ALFPD04 1.
@123 ALFPD04B 1.
@124 ALFPD05 4.
@128 ALFPD06 4.
@132 ALFPD07 2.
@134 ALFPD08 2.
@136 ALFPD09 2.
@138 ALFPD10 1.
@139 ALFPD12 5.
@144 ALFPD19 1.
@145 ALFPD20 1.
@146 ALFPD21 1.
@147 ALFPD22 1.
@148 ALFPD23 1.
@149 ALFPD24 1.
@150 ALFPD25 1.
@151 ALFPD26 2.
@153 ALFPD27 3.
@156 ALFPD28 1.
@157 ALFPD30A 1.
@158 ALFPD30B 1.
@159 ALFPD30C 1.
@160 ALFPD30D 1.
@161 ALFPD30E 1.
@162 ALFPD30F 1.
@163 ALFPD30G 1.
@164 ALFPD30H 2.
@166 ALFPD31A 1.
@167 ALFPD31B 1.
@168 ALFPD31C 1.
@169 ALFPD32 2.
@171 ALFPD33 2.
@173 ALFPD53 53.
@226 ALFSD02 2.
@228 ALFSD03 2.
@230 ALFSD04 1.
@231 ALFSD04B 1.
@232 ALFSD05 4.
@236 ALFSD06 4.
@240 ALFSD07 2.
@242 ALFSD08 2.
@244 ALFSD09 2.
@246 ALFSD10 1.
@247 ALFSD12 5.
@252 ALFSD19 1.
@253 ALFSD20 1.
@254 ALFSD21 1.
@255 ALFSD22 1.

@256 ALFSD23 1.
@257 ALFSD24 1.
@258 ALFSD25 1.
@259 ALFSD26 2.
@261 ALFSD27 3.
@264 ALFSD28 1.
@265 ALFSD30A 1.
@266 ALFSD30B 1.
@267 ALFSD30C 1.
@268 ALFSD30D 1.
@269 ALFSD30E 1.
@270 ALFSD30F 1.
@271 ALFSD30G 1.
@272 ALFSD30H 2.
@274 ALFSD31A 1.
@275 ALFSD31B 1.
@276 ALFSD31C 1.
@277 ALFSD32 2.
@279 ALFSD33 2.
@281 ALFSD53 53.
@334 AINHD01A 2.
@336 AINHD01B 2.
@338 AINPD02 2.
@340 AINHD03 5.
@345 AINHD04 5.
@350 AINHD05 2.
@352 AINHD07 2.
@354 AINHD08 7.
@361 ADPPQ12A 1.
@362 ADPPQ12B 1.
@363 ADPPQ12C 1.
@364 ADPPQ12D 1.
@365 ADPPQ12E 1.
@366 ADPPQ12F 1.
@367 ADPPQ12G 1.
@368 ADPPQ12H 1.
@369 ADPPQ12I 1.
@370 ADPPQ12J 1.
@371 ADPPQ12K 1.
@372 ADPPQ12L 1.
@373 ADPPS01 2.
@375 AFNHQ01A 1.
@376 AFNHQ01B 1.
@377 AFNHQ01C 1.
@378 AFNHQ01D 1.
@379 AFNHQ01E 1.
@380 AFNHQ01F 1.
@381 AFNHQ01G 1.

@382 AFNHQ01H 1.
@383 AFNHQ01I 1.
@384 AFNHQ01J 1.
@385 AFNHQ01K 1.
@386 AFNHQ01L 1.
@387 AFNHQ01M 1.
@388 AFNHQ02 2.
@390 AFNHS01 2.
@392 ASPHQ01A 1.
@393 ASPHQ01B 1.
@394 ASPHQ01C 1.
@395 ASPHQ01D 1.
@396 ASPHQ01E 1.
@397 ASPHQ01F 1.
@398 ASPHQ02A 1.
@399 ASPHQ02B 1.
@400 ASPHQ02C 1.
@401 ASPHQ02D 1.
@402 ASPHS01 2.
@404 ASDPD01 3.
@407 ASDPD02 3.
@410 ASDPD02B 1.
@411 ASDPD03 2.
@413 ASDPD04 2.
@415 ASDPD05 2.
@417 ASDPD05B 1.
@418 ASDPD06 2.
@420 ASDPD06B 1.
@421 ASDSD01 3.
@424 ASDSD02 3.
@427 ASDSD02B 1.
@428 ASDSD03 2.
@430 ASDSD04 2.
@432 ASDSD05 2.
@434 ASDSD05B 1.
@435 ASDSD06 2.
@437 ASDSD06B 1.
@438 ASDCD01 3.
@441 ASDCD02 3.
@444 ASDCD02B 1.
@445 ASDCD03 2.
@447 ASDCD04 2.
@449 ASDCD05 2.
@451 ASDCD05B 1.
@452 ASDCD06 2.
@454 ASDCD06B 1.
@455 AMDCQ01A 1.
@456 AMDCQ01B 1.

@457 AMDCQ01C 1.
@458 AMDCQ02 2.
@460 AMDCQ03 1.
@461 AMDCQ04 2.
@463 AMDCQ05A 1.
@464 AMDCQ05B 1.
@465 AMDCQ05C 1.
@466 AMDCQ06 2.
@468 AMDCQ07 1.
@469 AMDCQ08A 1.
@470 AMDCQ08B 1.
@471 AMDCQ08C 1.
@472 AMDCQ09A 1.
@473 AMDCQ09B 1.
@474 AMDCQ09C 1.
@475 AMDCQ09D 1.
@476 AMDCQ10A 1.
@477 AMDCQ10B 1.
@478 AMDCQ10C 1.
@479 AMDCQ10D 1.
@480 AMDCQ12A 1.
@481 AMDCQ12C 3.
@484 AMDCQ13B 5.
@489 AMDCQ14B 2.
@491 AMDCQ15 1.
@492 AMDCQ16 1.
@493 AMDCQ17 1.
@494 AMDCQ18 1.
@495 AMDCQ21A 1.
@496 AMDCQ21B 1.
@497 AMDCQ21C 1.
@498 AMDCQ21D 1.
@499 AMDCQ21E 1.
@500 AMDCQ21F 3.
@503 AMDCQ22 2.
@505 AMDCQ23A 1.
@506 AMDCQ23B 1.
@507 AMDCQ23C 1.
@508 AMDCQ23D 3.
@511 AMDCQ23F 1.
@512 AMDCQ24A 1.
@513 AMDCQ24B 3.
@516 AMDCQ25 1.
@517 AMDCQ26 1.
@518 AMDCQ27 2.
@520 AMDCQ28A 1.
@521 AMDCQ28B 1.
@522 AMDCQ28C 1.

@523 AMDCQ28D 1.
@524 AMDCQ28E 1.
@525 AMDCQ28F 1.
@526 AMDCQ28G 1.
@527 AMDCQ28H 1.
@528 AMDCQ28I 1.
@529 AMDCQ28J 1.
@530 AMDCQ28K 1.
@531 AMDCQ28L 1.
@532 AMDCQ28M 1.
@533 AMDCD01 1.
@534 AMDCD02 2.
@536 AMDCD03 2.
@538 AMDCD04 2.
@540 AMDCD05 2.
@542 AMDCD06 3.
@545 AMDCD07 1.
@546 AMDCD08 1.
@547 AMDCD09 1.
@548 AMDCD10 1.
@549 ATMCQ01 2.
@551 ATMCQ02 2.
@553 ATMCQ03 2.
@555 ATMCQ04 2.
@557 ATMCQ05 2.
@559 ATMCQ06 2.
@561 ATMCQ07 2.
@563 ATMCQ08 2.
@565 ATMCQ09 2.
@567 ATMCQ10 2.
@569 ATMCQ11 2.
@571 ATMCQ12 2.
@573 ATMCQ13 2.
@575 ATMCQ14 2.
@577 ATMCQ15 2.
@579 ATMCQ16 2.
@581 ATMCQ17 2.
@583 ATMCQ18 2.
@585 ATMCQ19 2.
@587 ATMCQ20 2.
@589 ATMCQ21 2.
@591 ATMCQ22 2.
@593 ATMCQ23 2.
@595 ATMCQ24 2.
@597 ATMCQ25 2.
@599 ATMCQ26 2.
@601 ATMCQ27 2.
@603 ATMCQ28 2.

@605 ATMCQ29 2.
@607 ATMCQ30 2.
@609 ATMCQ31 2.
@611 ATMCQ32 2.
@613 ATMCQ33 2.
@615 AEDCD01 2.
@617 AEDCQ02 1.
@618 AEDCQ03 1.
@619 AEDCQ04 1.
@620 AEDCD02 2.
@622 AEDCQ06 1.
@623 AEDCD03 2.
@625 AEDCQ08 1.
@626 AEDCQ09A 1.
@627 AEDCQ09B 2.
@629 AEDCQ10 2.
@631 AEDCQ11 2.
@633 AEDCQ12A 1.
@634 AEDCQ12B 1.
@635 AEDCQ12C 1.
@636 AEDCQ12D 1.
@637 AEDCQ12E 1.
@638 AEDCQ13 3.
@641 AEDCQ14A 2.
@643 AEDCQ14B 2.
@645 AEDCQ14C 2.
@647 AEDCQ14D 2.
@649 AEDCQ15A 1.
@650 AEDCQ15B 1.
@651 AEDCQ16 1.
@652 AEDCQ17 2.
@654 AEDCQ18A 1.
@655 AEDCQ18B 2.
@657 AEDCQ19A 1.
@658 AEDCQ19B 1.
@659 AEDCQ19C 1.
@660 AEDCQ19D 1.
@661 AEDCQ20 1.
@662 AEDCD04 3.
@665 AMACS01 2.
@667 AMACS02 3.
@670 ABECQ1 2.
@672 ABECQ2 2.
@674 ABECQ3 2.
@676 ABECQ4 2.
@678 ABECQ5 1.
@679 ABECQ5A 2.
@681 ABECQ6A 1.

@682 ABECQ6B 1.
@683 ABECQ6C 1.
@684 ABECQ6D 1.
@685 ABECQ6E 1.
@686 ABECQ6E1 1.
@687 ABECQ6F 1.
@688 ABECQ6G 1.
@689 ABECQ6H 1.
@690 ABECQ6I 1.
@691 ABECQ6J 1.
@692 ABECQ6J1 1.
@693 ABECQ6K 1.
@694 ABECQ6L 1.
@695 ABECQ6M 1.
@696 ABECQ6N 1.
@697 ABECQ6O 1.
@698 ABECQ6P 1.
@699 ABECQ6Q 1.
@700 ABECQ6R 1.
@701 ABECQ6R1 1.
@702 ABECQ6S 1.
@703 ABECQ6T 1.
@704 ABECQ6T1 1.
@705 ABECQ6U 1.
@706 ABECQ6V 1.
@707 ABECQ6W 1.
@708 ABECQ6X 1.
@709 ABECQ6Y 1.
@710 ABECQ6Z 1.
@711 ABECQ6Z1 1.
@712 ABECQ6AA 1.
@713 ABECQ6BB 1.
@714 ABECQ6CC 1.
@715 ABECQ6DD 1.
@716 ABEC6DD1 1.
@717 ABECQ6EE 1.
@718 ABECQ6FF 1.
@719 ABECQ6GG 1.
@720 ABECQ6HH 1.
@721 ABECQ6II 1.
@722 ABECQ6JJ 1.
@723 ABECQ6KK 1.
@724 ABECQ6LL 1.
@725 ABEC6LL1 1.
@726 ABECQ6MM 1.
@727 ABECQ6NN 1.
@728 ABECQ6OO 1.
@729 ABECQ6PP 1.

@730 ABEC6PP1 1.
@731 ABECQ6QQ 1.
@732 ABECQ6RR 1.
@733 ABECQ6SS 1.
@734 ABECQ6TT 1.
@735 ABEC6TT1 1.
@736 ABECQ6UU 1.
@737 ABECQ7A 1.
@738 ABECQ7B 1.
@739 ABECQ7C 1.
@740 ABECQ7D 1.
@741 ABECQ7E 1.
@742 ABECQ7F 1.
@743 ABECS01 2.
@745 ABECS02 2.
@747 ABECS03 2.
@749 ABECS04 2.
@751 ABECS05 2.
@753 ABECS06 2.
@755 ABECS07 2.
@757 ABECS08 2.
@759 ABECS09 2.
@761 ABECS10 2.
@763 ABECS11 2.
@765 AMSCS01 2.
@767 AMSCS02 3.
@770 ARLCQ01 2.
@772 ARLCQ02 2.
@774 ARLCQ03 2.
@776 ARLCQ04 1.
@777 ARLCQ05 1.
@778 ARLCQ06 2.
@780 ARLCQ07 2.
@782 ARLCQ08 2.
@784 ARLCQ09 2.
@786 APRCQ01 2.
@788 APRCQ02 2.
@790 APRCQ03 2.
@792 APRCQ04 2.
@794 APRCQ05 2.
@796 APRCQ06 2.
@798 APRCQ07 2.
@800 APRCQ08 2.
@802 APRCQ09 2.
@804 APRCQ10 2.
@806 APRCQ11 2.
@808 APRCQ12 2.
@810 APRCQ13 2.

@812 APRCQ14 2.
@814 APRCQ15 2.
@816 APRCQ16 2.
@818 APRCQ17 2.
@820 APRCQ18 2.
@822 APRCQ19 2.
@824 APRCQ20 2.
@826 APRCQ21 2.
@828 APRCQ22 2.
@830 APRCQ23 2.
@832 APRCQ24 2.
@834 APRCQ25 2.
@836 APRCQ26A 1.
@837 APRCQ26B 1.
@838 APRCQ26C 1.
@839 APRCQ26D 1.
@840 APRCQ26E 1.
@841 APRCQ26F 1.
@842 APRCQ26G 1.
@843 APRCQ26H 1.
@844 APRCQ26I 1.
@845 APRCQ26J 1.
@846 APRCQ26K 1.
@847 APRCQ27 1.
@848 APRCQ28 1.
@849 APRCS01 2.
@851 APRCS02 2.
@853 APRCS03 2.
@855 APRCS04 2.
@857 APRCS05 2.
@859 APRCS06 2.
@861 ACRCQ1A 1.
@862 ACRCQ1B 1.
@863 ACRCQ1BA 3.
@866 ACRCQ1BB 1.
@867 ACRCQ1C 1.
@868 ACRCQ1CA 3.
@871 ACRCQ1CB 1.
@872 ACRCQ1D 1.
@873 ACRCQ1DA 3.
@876 ACRCQ1E 1.
@877 ACRCQ1EA 3.
@880 ACRCQ1F 1.
@881 ACRCQ1FA 3.
@884 ACRCQ1G 1.
@885 ACRCQ1GA 3.
@888 ACRCQ1GB 1.
@889 ACRCQ1H 1.

@890 ACRCQ1HA 2.
@892 ACRCQ1I 1.
@893 ACRCQ1IA 3.
@896 ACRCQ1J 1.
@897 ACRCQ1JA 3.
@900 ACRCQ03 2.
@902 ACRCQ04 2.
@904 ACRCQ05A 1.
@905 ACRCQ05B 1.
@906 ACRCQ05C 1.
@907 ACRCQ05D 1.
@908 ACRCQ05E 1.
@909 ACRCQ05F 1.
@910 ACRCQ05G 1.
@911 ACRCQ06 1.
@912 ACRCQ07 2.
@914 ACRCQ08A 1.
@915 ACRCQ08B 1.
@916 ACRCQ08C 1.
@917 ACRCQ08D 1.
@918 ACRCQ08E 1.
@919 ACRCQ08F 1.
@920 ACRCQ08G 1.
@921 ACRCQ08H 1.
@922 ACRCQ08I 1.
@923 ACRCD01 2.
@925 ACRCD02 3.
@928 ACRCD03 2.
@930 ACRCD04 3.
@933 ACRCD05 1.
@934 ACRCD06 3.
@937 ACRCD07 1.
@938 ACRCD08 1.
@939 ACRCD09 1.
@940 ACRCD10 1.
@941 APPCS01 3.
@944 APPCS02 3.
@947 APPCD01 2.
@949 APPCD02 1.
@950 AA1CQ01 2.
@952 AA1CQ02 2.
@954 AA1CQ03 2.
@956 AA1CQ04 2.
@958 AA1CQ05 2.
@960 AA1CQ06 2.
@962 AA1CQ07 1.
@963 AA1CQ08A 1.
@964 AA1CQ08B 1.

@965 AA1CQ08C 1.
@966 AA1CQ08D 1.
@967 AA1CQ08E 1.
@968 AA1CQ08F 1.
@969 AA1CQ08G 1.
@970 AA1CQ08H 1.
@971 AA1CQ08I 1.
@972 AA1CQ08J 1.
@973 AA1CQ08K 1.
@974 AA1CQ08L 1.
@975 AA1CQ08M 1.
@976 AA1CQ08N 1.
@977 AA1CQ09 2.
@979 AA1CQ10 2.
@981 AA1CQ11 2.
@983 AA1CQ12 2.
@985 AD1CQ01A 1.
@986 AD1CQ01B 1.
@987 AD1CQ01C 1.
@988 AD1CQ01D 1.
@989 AD1CQ01E 1.
@990 AD1CQ01F 1.
@991 AD1CQ01G 1.
@992 AD1CQ01H 1.
@993 AD1CQ01I 1.
@994 AD1CQ01J 1.
@995 AD1CQ01K 1.
@996 AD1CQ01L 1.
@997 AD1CQ01M 1.
@998 AD1CQ01N 1.
@999 AD1CQ01O 1.
@1000 AD1CQ01P 1.
@1001 AD1CQ01Q 1.
@1002 AD1CQ01R 1.
@1003 AD1CQ01S 1.
@1004 AD1CQ01T 1.
@1005 AD1CQ01U 1.
@1006 AD1CQ01V 1.
@1007 AD1CQ01W 1.
@1008 AD1CQ01X 1.
@1009 AD1CQ01Y 1.
@1010 AD1CQ01Z 1.
@1011 AD1CQ1AA 1.
@1012 AD1CQ1BB 1.
@1013 AD1CQ1CC 1.
@1014 AD1CQ1DD 1.
@1015 AD1CQ1EE 1.
@1016 AD1CQ1FF 1.

```
@1017   AD1CQ1GG   1.
@1018   AD1CQ1HH   1.
@1019   AD1CQ1II   1.
@1020   AD1CQ1JJ   1.
@1021   AD1CQ1KK   1.
@1022   AD1CQ1LL   1.
@1023   AD1CQ1MM   1.
@1024   AD1CQ1NN   1.
@1025   AD1CQ1OO   1.
@1026   AD1CQ1PP   1.
@1027   AD1CQ1QQ   1.
@1028   AD1CQ1RR   1.
@1029   AD1CQ1SS   1.
@1030   AD1CQ1TT   1.
@1031   AD1CQ1UU   1.
@1032   AD1CQ02A   1.
@1033   AD1CQ02B   1.
@1034   AD1CQ02C   1.
@1035   AD1CQ02D   1.
@1036   AD1CQ02E   1.
@1037   AD1CQ02F   1.
@1038   AD1CQ03   1.
@1039   AE1CQ01A   1.
@1040   AE1CQ01B   1.
@1041   AE1CQ01C   1.
@1042   AE1CQ01D   1.
@1043   AE1CQ01E   1.
@1044   AE1CQ01F   1.
@1045   AE1CQ01G   1.
@1046   AE1CQ01H   1.
@1047   AE1CQ01I   1.
@1048   AE1CQ01J   1.
@1049   AE1CQ01K   1.
@1050   AE1CQ01L   1.
@1051   AE1CQ01M   1.
@1052   AE1CQ01N   1.
@1053   AE1CQ01O   1.
@1054   AE1CQ01P   1.
@1055   AE1CQ01Q   1.
@1056   AWTCW01   9.
@1065   AIDHD01   7.
@1072   CHILDID   6.;
```

28.4.4 Secondary File

```
/*********************************************************/
/* NLSCY  SECONDARY MICRO DATA FILE                      */
/*********************************************************/

DATA SEC;
 INFILE  '\NLSCYA\SECONDRY.TXT'  LRECL=575  MISSOVER PAD;

INPUT
@1   ALFPQ01   2.
@3   ALFPD01A   2.
@5   ALFPQ02   1.
@6   ALFPQ04A   1.
@7   ALFPQ06A   1.
@8   ALFPQ08A   2.
@10   ALFPQ09A   2.
@12   ALFPQ10A   1.
@13   ALFPQ11A   1.
@14   ALFPQ04B   1.
@15   ALFPQ06B   1.
@16   ALFPQ08B   2.
@18   ALFPQ09B   2.
@20   ALFPQ10B   1.
@21   ALFPQ11B   1.
@22   ALFPQ04C   1.
@23   ALFPQ06C   1.
@24   ALFPQ08C   2.
@26   ALFPQ09C   2.
@28   ALFPQ10C   1.
@29   ALFPQ11C   1.
@30   ALFPQ04D   1.
@31   ALFPQ06D   1.
@32   ALFPQ08D   2.
@34   ALFPQ09D   2.
@36   ALFPQ10D   1.
@37   ALFPQ11D   1.
@38   ALFPQ04E   1.
@39   ALFPQ06E   1.
@40   ALFPQ08E   2.
@42   ALFPQ09E   2.
@44   ALFPQ10E   1.
@45   ALFPQ11E   1.
@46   ALFPQ04F   1.
```

@47 ALFPQ06F 1.
@48 ALFPQ08F 2.
@50 ALFPQ09F 2.
@52 ALFPQ10F 1.
@53 ALFPMNJB 2.
@55 ALFPQ16 1.
@56 ALFPQ16A 2.
@58 ALFPQ16B 1.
@59 ALFPQ16C 9.
@68 ALFPQ16D 2.
@70 ALFPQ16E 6.
@76 ALFSQ01 2.
@78 ALFSD01A 2.
@80 ALFSQ02 1.
@81 ALFSQ04A 1.
@82 ALFSQ06A 1.
@83 ALFSQ08A 2.
@85 ALFSQ09A 2.
@87 ALFSQ10A 1.
@88 ALFSQ11A 1.
@89 ALFSQ04B 1.
@90 ALFSQ06B 1.
@91 ALFSQ08B 2.
@93 ALFSQ09B 2.
@95 ALFSQ10B 1.
@96 ALFSQ11B 1.
@97 ALFSQ04C 1.
@98 ALFSQ06C 1.
@99 ALFSQ08C 2.
@101 ALFSQ09C 2.
@103 ALFSQ10C 1.
@104 ALFSQ11C 1.
@105 ALFSQ04D 1.
@106 ALFSQ06D 1.
@107 ALFSQ08D 2.
@109 ALFSQ09D 2.
@111 ALFSQ10D 1.
@112 ALFSQ11D 1.
@113 ALFSQ04E 1.
@114 ALFSQ06E 1.
@115 ALFSQ08E 2.
@117 ALFSQ09E 2.
@119 ALFSQ10E 1.
@120 ALFSQ11E 1.
@121 ALFSQ04F 1.
@122 ALFSQ06F 1.
@123 ALFSQ08F 2.
@125 ALFSQ09F 2.

@127 ALFSQ10F 1.
@128 ALFSMNJB 2.
@130 ALFSQ16 1.
@131 ALFSQ16A 2.
@133 ALFSQ16B 1.
@134 ALFSQ16C 9.
@143 ALFSQ16D 2.
@145 ALFSQ16E 6.
@151 AINHQ01A 1.
@152 AINHQ01B 1.
@153 AINHQ01C 1.
@154 AINHQ01D 1.
@155 AINHQ01E 1.
@156 AINHQ01F 1.
@157 AINHQ01G 1.
@158 AINHQ01H 1.
@159 AINHQ01I 1.
@160 AINHQ01J 1.
@161 AINHQ01K 1.
@162 AINHQ01L 1.
@163 AINHQ01M 1.
@164 AINHQ02A 2.
@166 AINHD02B 1.
@167 AINHQ03 6.
@173 AINPQ04 6.
@179 AINHI03A 2.
@181 AINHI03B 1.
@182 AINPI04A 2.
@184 AINPI04B 1.
@185 ASDPQ1 2.
@187 ASDPQ2AA 1.
@188 ASDPQ2AB 1.
@189 ASDPQ2AC 1.
@190 ASDPQ2AD 1.
@191 ASDPQ2B 1.
@192 ASDPQ3 4.
@196 ASDPQ4A 1.
@197 ASDPQ4B 1.
@198 ASDPQ4C 1.
@199 ASDPQ4D 1.
@200 ASDPQ4E 1.
@201 ASDPQ4F 1.
@202 ASDPQ4G 1.
@203 ASDPQ4H 1.
@204 ASDPQ4I 1.
@205 ASDPQ4J 1.
@206 ASDPQ4K 1.
@207 ASDPQ4L 1.

@208 ASDPQ4M 1.
@209 ASDPQ4N 1.
@210 ASDPQ4O 1.
@211 ASDPQ4P 1.
@212 ASDPQ4Q 1.
@213 ASDPQ4R 1.
@214 ASDPQ4S 1.
@215 ASDPQ5A 1.
@216 ASDPQ5B 1.
@217 ASDPQ5C 1.
@218 ASDPQ5D 1.
@219 ASDPQ5E 1.
@220 ASDPQ5F 1.
@221 ASDPQ5G 1.
@222 ASDPQ5H 1.
@223 ASDPQ5I 1.
@224 ASDPQ5J 1.
@225 ASDPQ5K 1.
@226 ASDPQ5L 1.
@227 ASDPQ5M 1.
@228 ASDPQ5N 1.
@229 ASDPQ5O 1.
@230 ASDPQ5P 1.
@231 ASDPQ5Q 1.
@232 ASDPQ5R 1.
@233 ASDPQ5S 1.
@234 ASDPQ6A 1.
@235 ASDPQ6B 1.
@236 ASDPQ6C 1.
@237 ASDPQ6D 1.
@238 ASDPQ6E 1.
@239 ASDPQ6F 1.
@240 ASDPQ6G 1.
@241 ASDPQ6H 1.
@242 ASDPQ6I 1.
@243 ASDPQ6J 1.
@244 ASDPQ6K 1.
@245 ASDPQ6L 1.
@246 ASDPQ6M 1.
@247 ASDPQ6N 1.
@248 ASDPQ6O 1.
@249 ASDPQ6P 1.
@250 ASDPQ6Q 1.
@251 ASDPQ6R 1.
@252 ASDPQ6S 1.
@253 ASDPQ8 2.
@255 ASDPQ9 2.
@257 ASDSQ1 2.

```
@259  ASDSQ2AA  1.
@260  ASDSQ2AB  1.
@261  ASDSQ2AC  1.
@262  ASDSQ2AD  1.
@263  ASDSQ2B  1.
@264  ASDSQ3  4.
@268  ASDSQ4A  1.
@269  ASDSQ4B  1.
@270  ASDSQ4C  1.
@271  ASDSQ4D  1.
@272  ASDSQ4E  1.
@273  ASDSQ4F  1.
@274  ASDSQ4G  1.
@275  ASDSQ4H  1.
@276  ASDSQ4I  1.
@277  ASDSQ4J  1.
@278  ASDSQ4K  1.
@279  ASDSQ4L  1.
@280  ASDSQ4M  1.
@281  ASDSQ4N  1.
@282  ASDSQ4O  1.
@283  ASDSQ4P  1.
@284  ASDSQ4Q  1.
@285  ASDSQ4R  1.
@286  ASDSQ4S  1.
@287  ASDSQ5A  1.
@288  ASDSQ5B  1.
@289  ASDSQ5C  1.
@290  ASDSQ5D  1.
@291  ASDSQ5E  1.
@292  ASDSQ5F  1.
@293  ASDSQ5G  1.
@294  ASDSQ5H  1.
@295  ASDSQ5I  1.
@296  ASDSQ5J  1.
@297  ASDSQ5K  1.
@298  ASDSQ5L  1.
@299  ASDSQ5M  1.
@300  ASDSQ5N  1.
@301  ASDSQ5O  1.
@302  ASDSQ5P  1.
@303  ASDSQ5Q  1.
@304  ASDSQ5R  1.
@305  ASDSQ5S  1.
@306  ASDSQ6A  1.
@307  ASDSQ6B  1.
@308  ASDSQ6C  1.
@309  ASDSQ6D  1.
```

@310 ASDSQ6E 1.
@311 ASDSQ6F 1.
@312 ASDSQ6G 1.
@313 ASDSQ6H 1.
@314 ASDSQ6I 1.
@315 ASDSQ6J 1.
@316 ASDSQ6K 1.
@317 ASDSQ6L 1.
@318 ASDSQ6M 1.
@319 ASDSQ6N 1.
@320 ASDSQ6O 1.
@321 ASDSQ6P 1.
@322 ASDSQ6Q 1.
@323 ASDSQ6R 1.
@324 ASDSQ6S 1.
@325 ASDSQ8 2.
@327 ASDSQ9 2.
@329 ASDCQ1 2.
@331 ASDCQ2AA 1.
@332 ASDCQ2AB 1.
@333 ASDCQ2AC 1.
@334 ASDCQ2AD 1.
@335 ASDCQ2B 1.
@336 ASDCQ3 4.
@340 ASDCQ4A 1.
@341 ASDCQ4B 1.
@342 ASDCQ4C 1.
@343 ASDCQ4D 1.
@344 ASDCQ4E 1.
@345 ASDCQ4F 1.
@346 ASDCQ4G 1.
@347 ASDCQ4H 1.
@348 ASDCQ4I 1.
@349 ASDCQ4J 1.
@350 ASDCQ4K 1.
@351 ASDCQ4L 1.
@352 ASDCQ4M 1.
@353 ASDCQ4N 1.
@354 ASDCQ4O 1.
@355 ASDCQ4P 1.
@356 ASDCQ4Q 1.
@357 ASDCQ4R 1.
@358 ASDCQ4S 1.
@359 ASDCQ5A 1.
@360 ASDCQ5B 1.
@361 ASDCQ5C 1.
@362 ASDCQ5D 1.
@363 ASDCQ5E 1.

@364 ASDCQ5F 1.
@365 ASDCQ5G 1.
@366 ASDCQ5H 1.
@367 ASDCQ5I 1.
@368 ASDCQ5J 1.
@369 ASDCQ5K 1.
@370 ASDCQ5L 1.
@371 ASDCQ5M 1.
@372 ASDCQ5N 1.
@373 ASDCQ5O 1.
@374 ASDCQ5P 1.
@375 ASDCQ5Q 1.
@376 ASDCQ5R 1.
@377 ASDCQ5S 1.
@378 ASDCQ6A 1.
@379 ASDCQ6B 1.
@380 ASDCQ6C 1.
@381 ASDCQ6D 1.
@382 ASDCQ6E 1.
@383 ASDCQ6F 1.
@384 ASDCQ6G 1.
@385 ASDCQ6H 1.
@386 ASDCQ6I 1.
@387 ASDCQ6J 1.
@388 ASDCQ6K 1.
@389 ASDCQ6L 1.
@390 ASDCQ6M 1.
@391 ASDCQ6N 1.
@392 ASDCQ6O 1.
@393 ASDCQ6P 1.
@394 ASDCQ6Q 1.
@395 ASDCQ6R 1.
@396 ASDCQ6S 1.
@397 ASDCQ8 2.
@399 ASDCQ9 2.
@401 ABECI6A 1.
@402 ABECI6B 1.
@403 ABECI6C 1.
@404 ABECI6D 1.
@405 ABECI6E 1.
@406 ABECI6E1 1.
@407 ABECI6F 1.
@408 ABECI6G 1.
@409 ABECI6H 1.
@410 ABECI6I 1.
@411 ABECI6J 1.
@412 ABECI6J1 1.
@413 ABECI6K 1.

@414 ABECI6L 1.
@415 ABECI6M 1.
@416 ABECI6N 1.
@417 ABECI6O 1.
@418 ABECI6P 1.
@419 ABECI6Q 1.
@420 ABECI6R 1.
@421 ABECI6R1 1.
@422 ABECI6S 1.
@423 ABECI6T 1.
@424 ABECI6T1 1.
@425 ABECI6U 1.
@426 ABECI6V 1.
@427 ABECI6W 1.
@428 ABECI6X 1.
@429 ABECI6Y 1.
@430 ABECI6Z 1.
@431 ABECI6Z1 1.
@432 ABECI6AA 1.
@433 ABECI6BB 1.
@434 ABECI6CC 1.
@435 ABECI6DD 1.
@436 ABECI6EE 1.
@437 ABECI6FF 1.
@438 ABECI6GG 1.
@439 ABECI6HH 1.
@440 ABECI6II 1.
@441 ABECI6JJ 1.
@442 ABECI6KK 1.
@443 ABECI6LL 1.
@444 ABECI6MM 1.
@445 ABECI6NN 1.
@446 ABECI6OO 1.
@447 ABECI6PP 1.
@448 ABECI6QQ 1.
@449 ABECI6RR 1.
@450 ABECI6SS 1.
@451 ABECI6TT 1.
@452 ABECI6UU 1.
@453 AMSCQ01 1.
@454 AMSCQ02 1.
@455 AMSCQ03 1.
@456 AMSCQ04 1.
@457 AMSCQ05 1.
@458 AMSCQ06 1.
@459 AMSCQ07 1.
@460 AMSCQ08 1.
@461 AMSCQ09 1.

@462 AMSCQ10 1.
@463 AMSCQ11 1.
@464 AMSCQ12 1.
@465 AMSCQ13 1.
@466 AMSCQ14 1.
@467 AMSCQ15 1.
@468 AMSCQ16 1.
@469 AMSCQ17 1.
@470 AMSCQ18 1.
@471 AMSCQ19 1.
@472 AMSCQ20 1.
@473 AMSCQ21 1.
@474 AMSCQ22 1.
@475 AMSCQ23 1.
@476 AMSCQ24 1.
@477 AMSCQ25 1.
@478 AMSCQ26 1.
@479 AMSCQ27 1.
@480 AMSCQ28 1.
@481 AMSCQ29 1.
@482 AMSCQ30 1.
@483 AMSCQ31 1.
@484 AMSCQ32 1.
@485 AMSCQ33 1.
@486 AMSCQ34 1.
@487 AMSCQ35 1.
@488 AMSCQ36 1.
@489 AMSCQ37 1.
@490 AMSCQ38 1.
@491 AMSCQ39 1.
@492 AMSCQ40 1.
@493 AMSCQ41 1.
@494 AMSCQ42 1.
@495 AMSCQ43 1.
@496 AMSCQ44 1.
@497 AMSCQ45 1.
@498 AMSCQ46 1.
@499 AMSCQ47 1.
@500 AMSCQ48 1.
@501 APRCI01 1.
@502 APRCI02 1.
@503 APRCI03 1.
@504 APRCI04 1.
@505 APRCI05 1.
@506 APRCI06 1.
@507 APRCI07 1.
@508 APRCI08 1.
@509 APRCI09 1.

```
@510   APRCI10   1.
@511   APRCI11   1.
@512   APRCI12   1.
@513   APRCI13   1.
@514   APRCI14   1.
@515   APRCI15   1.
@516   APRCI16   1.
@517   APRCI17   1.
@518   APRCI18   1.
@519   APACQ01   2.
@521   APACQ02   2.
@523   APACQ03   2.
@525   APACQ04   2.
@527   APACQ05   2.
@529   APACQ06   1.
@530   APACQ07   1.
@531   APACQ08   1.
@532   APACQ09   2.
@534   APACQ10   2.
@536   APACQ11   2.
@538   APACQ12   2.
@540   APACQ13   2.
@542   APACQ14   2.
@544   APACQ15   2.
@546   APACS01   2.
@548   APACS02   2.
@550   APACS03   2.
@552   APACS04   2.
@554   AWTCW01   9.
@563   AIDHD01   7.
@570   CHILDID   6.;
```

28.6 Univariate Counts

The univariate counts for the primary and secondary files can be found in Appendix 4 (primary file) and Appendix 5 (secondary file).

Appendix 1 - Content for Release 1 and 2

Not all the information collected for the first cycle of the NLSCY are included in this first microdata file. The amount of information collected was so extensive a decision was made to have two releases rather than waiting for all of the data to be processed. The second release will be in 1997. The notable sections to be included in the second release are health variables for the child and the parents, the custody history of the child, and data collected from the teacher and the principal. A complete list of the sections included in the first and second release can be found below.

RELEASE 1 (i.e., current release)

Content: **Child Information**

Medical/Biological - pre-natal, delivery, post-natal health etc.
Temperament
Education
Behaviour
Motor and social development
Relationships
Parenting
Child care
Socio-demographic characteristics

Parent Information
Depression scale (for PMK)
Family functioning
Social support

General Questionnaire data (PMK and Spouse)
Socio-demographic characteristics
Education
Labour force
Income

10 - 11 Questionnaire
Friends and family
Feelings and behaviour
My parents and me

PPVT

Math Test

Demographic Data - (i.e., Family Composition variables)

RELEASE 2

Date: 1997

Content: **Child Information**
Health
Literacy
Activities
Family and custody history

Parent Information
Adult health
Neighbourhood safety and observation

General Questionnaire data (PMK and Spouse)
Restrictions of activities
Chronic conditions

10 - 11 questionnaire
School
About me
Puberty
Smoking, drinking and drugs
Activities

Teacher Questionnaire

Principal Questionnaire

Appendix 2 - The Labour Force Survey

Survey Coverage

The Labour Force Survey (LFS) is a monthly household survey carried out by Statistics Canada in approximately 59,000 households throughout the country. The LFS is used to produce monthly estimates of employment, self-employment and unemployment. Information on variables such as industry and occupation of employment, educational attainment, ethnic origin, and country of birth is obtained. Approximately 97% of the population 15 years of age and over is covered in the survey. Excluded from the LFS are the populations in the Yukon and the Northwest Territories, residents of Indian Reserves, full-time members of the Canadian Armed Forces, and residents of institutions, such as chronic care hospitals, prisons and child residential treatment facilities. Civilian members of the Armed Forces' households and native people living "off-reserve" are captured by the survey.

Sample Design

The Labour Force Survey employs a stratified, multistage probability sample[34] design based on an area frame in which dwellings (residences) are the sampling units. All eligible individuals who occupy one of the selected dwellings are part of the LFS sample. For design purposes, each province of Canada constitutes an independent sample and is divided into two parts composed of large cities and rural areas plus small urban centres. Through stratification, these parts are broken down into clusters of dwellings, e.g., city blocks, from which dwellings are selected.

It should be noted that at the time of sample selection, no information is known about the persons living within a selected dwelling, who are collectively known as a household. It is the dwelling, not the household, that is chosen for the sample. If the household moves, whoever is living in the dwelling at the time of the interview is included in the sample.

Each dwelling is retained in the LFS sample for six consecutive months and no substitution of dwellings takes place in the event that information cannot be obtained from a dwelling. The entire sample is divided into six representative parts or rotation groups. Each rotation group contains some 10,000 households, representing about 20,000 individuals. The rotation of dwellings in the sample is carried out so that one-sixth of the sample is changed each month. In other words, each month one-sixth of

---

*Refer to **Methodology of the Canadian Labour Force Survey: 1984-1990**, Statistics Canada, Catalogue 71-526 for more details.*

the dwellings, having completed the six month stay in the sample, are replaced by new dwellings in the same or a similar area.

Dwellings which are currently in the sample are referred to as the active sample. Dwellings which are no longer part of the sample are called rotates out.

The LFS sample frame has been recently redesigned to incorporate new elements. This new frame was phased in as of October 1994.

LFS Collection Methodology

Data collection for the LFS is carried out during the week following the LFS reference week which is normally the week containing the 15th day of the month; thus collection is usually the third week of the month. Statistics Canada interviewers, who are part-time employees hired and trained specifically to carry out the survey, contact each of the dwellings in the sample, through personal or telephone interviews, to obtain the required information. The interviews are carried out using Computer-Assisted Personal Interviewing (CAPI).

Each interviewer contacts approximately 65 designated dwellings per month, one-sixth of which will be "new" dwellings. Each of these "new" dwellings is visited personally by the interviewer who collects information for all household members from one knowledgeable and responsible member. Subsequent interviews may be conducted by telephone provided the knowledgeable and responsible member agrees to this procedure. Currently, approximately 85% of the LFS interviews after the first month are conducted by telephone.

Using the LFS Frame for the NLSCY

One advantage of using the LFS survey frame for other surveys is that each rotation group of the LFS provides a sample capable of producing representative statistics for Canada and each province. In addition, the household composition information collected for the LFS is available to select a sample. Furthermore, LFS interviewers are available to do surveys when they are not working on the LFS and are familiar with the CAPI collection methodology. Because of these factors, the LFS frame was chosen for the NLSCY.

Depending on the level of reliability required, the budget and the available collection capacity, from one to six active rotation groups can be surveyed in a non-LFS collection week. This capacity can be expanded by the addition of dwellings which have rotated out prior to the survey reference month. In theory, this approach can be used to augment a survey's sample infinitely. In practice, however, a combination of cost and statistical reliability limit the additional "take" to roughly three times the regular LFS sample, that is about 15 rotation groups. With regard to the NLSCY, nine rotation groups were sufficient to cover all age groups; a combination of active rotation groups and rotates out were used.

The LFS Household Record collects basic demographic information such as age, sex, marital status, educational attainment, economic family association and relationship to head of economic family for all members of all households identified in selected dwellings. The age data from this record were used to facilitate the selection of dwellings with children for the NLSCY. This alleviated much of the need to screen dwellings to determine if children under 12 resided in them.

Appendix 3 - C.V. Tables

Appendix 4 - Univariate Counts - Primary File

Appendix 5 - Univariate Counts - Secondary File

Excludes Yukon and Northwest Territories.

This is the difference between the total time and the time required for the major components. This would include time for the interviewer to introduce the survey, complete the household roster, the relationships, the neighbourhood observation, the informed consent for the school collection, set-up time for the 10 to 11 Questionnaire, time for the computer to generate the various questionnaires, etc.

For the NLSCY estimate the weight factor used to produce the estimate is the sampling weight adjusted for non-response. Post-stratification, however, was not carried out for the household weight.

For the purposes of this comparison, only economic families with a least one child in the 0 to 11 age group in the SCF sample were included.

The differences in scores for children with attitude problems, physical problems and distraction problems are all significant at the 95% confidence level.

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.

This includes households with at least one child 0 to 11 years of age at the time of the NLSCY interview.

Excludes Yukon and Northwest Territories.

One reason for this, is that when the June follow-up for non-respondents was carried out, the response rate for these provinces was already quite high. Therefore it was agreed that for these provinces only, the follow-up could be done completely by telephone. This precluded administering the PPVT-R since it had to be administered in person.

One reason for this, it that when the June follow-up for non-respondents was carried out, the response rate for these provinces was already quite high. Therefore it was agreed that for these provinces only, the follow-up could be done completely by telephone. This precluded administering the 10 to 11 Self-complete Questionnaire.

Household income has been imputed. For this reason the previous tables with income data did not show missing values. However there was more non-response to the math test for the children where the PMK refused to give a household income.

Excludes the Yukon and Northwest Territories.

Excludes the Yukon and Northwest Territories.