Microdata User Guide

COMMUNITIES SURVEY

2005

Table of Contents

1.0   Introduction

The Communities Survey was conducted by Statistics Canada from February to June 2005 with the cooperation and support of Human Resources and Social Development Canada (formerly Social Development Canada and Human Resources Development Canada). This manual has been produced to facilitate the manipulation of the microdata file of the survey results.

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

Statistics Canada

Client Services
Special Surveys Division
2500 - R, Main Building,
Tunney's Pasture
Ottawa, Ontario K1A 0T6
Telephone: (613) 951-3321 or call toll-free 1 800 461-9050
Fax: (613) 951-4527
E-mail: ssd@statcan.ca

2.0 Background

There is growing recognition that the early years (from 0 to 5 years old) are critical for children's development. The early years shape long-term outcomes related to scholastic achievement, to employment success, to health, to quality of life and to the ability to adapt. In response to this understanding and the Government of Canada's commitment to ensure that all children are ready to learn and participate in society, Human Resources and Social Development Canada is sponsoring Understanding the Early Years (UEY); a national research initiative designed to:

- increase knowledge about what influences healthy child development,

- monitor our progress as a society in terms of improving outcomes for young children, and

- catalyze community action.

The UEY project, which is a joint project between Statistics Canada, Human Resources and Social Development Canada, and the communities themselves, is part of the Understanding the Early Years initiative. The project will assist communities to improve outcomes for the children and families living within them by providing, at the community level, the necessary information to develop and refine local strategies for healthy child development.

The survey population for the UEY project is children living within the communities who attend senior kindergarten (or an equivalent type of learning institution) at the time of the study. The project consists of the following three components:

1. The Early Development Indicator

The Early Development Indicator (EDI) is a questionnaire which is developed and administered by the Canadian Centre for Studies of Children at Risk at McMaster University. This questionnaire measures a variety of factors which have an impact on a child's readiness to learn including physical well being, language and cognitive skills, and social and emotional development. This questionnaire is completed by kindergarten teachers for all kindergarten children in the community project. Where permission from the parent is obtained, results of the EDI are merged with the data from the Communities Survey.

2. The Communities Survey

From among all kindergarten children, a sample of children is selected to participate in the Communities Survey. Data collection is then carried out in two phases: the first consists of interviews conducted over the telephone with the parents of these children, while the second consists of a series of direct measures assessments administered to the children. Statistics Canada's regional offices are responsible for the collection of the Communities Survey.

3. The Ethnographic Study

This component involves community mapping and is done directly by research coordinators within the community. This component of the study maps out various characteristics in the community including the availability of services, parks and play spaces, as well as other community resources.

The UEY project has a five year time frame. In the first year (2000), all three components – the EDI, the Communities Survey and the Ethnographic study – are completed. During years two, three and four (2001 to 2003), only the EDI component is implemented for kindergarten (or equivalent) children. In year five (2004), and in year six (2005) all three components are completed.

In 1999, North York, Ontario was the first community to be involved in this initiative as a pilot project. In the spring of 2000, another five communities (Coquitlam, British Columbia; Prince Albert, Saskatchewan;

Winnipeg, Manitoba; Southwestern Newfoundland and Prince Edward Island) took part in this project. Of the five communities that were surveyed in 2000, only four were followed up in 2004 (Prince Albert, Winnipeg, Southwestern Newfoundland and Prince Edward Island) and the findings were reported in 2004. The seven communities that were selected to participate in 2001 and to follow up in 2005 are: Abbotsford, British Columbia; Saskatoon, Saskatchewan; South Eastman, Manitoba; Mississauga, Ontario; Niagara Falls, Ontario; Montreal, Quebec and Hampton, New Brunswick.

This report will highlight the results of the 2005 community data. Note that the Communities Survey follow-up, implemented in year six (2005), is not administered to the same sample of children as in year two (2001). Instead, the follow-up survey is completed with a new sample of kindergarten (or equivalent) children since it is the community itself rather than the children in the community which is our main interest.

3.0   Objectives

The goal of the Communities Survey in 2001 was to identify and collect information on community influences and other factors that may have an impact on long-term outcomes such as scholastic achievement, employment success, health, quality of life and the ability to adapt for selected communities.

At that time, the four major objectives were:

- to provide comprehensive data on the determinants of child development, including detailed information about the role of the family, the community's characteristics, and non-parental care arrangements;

- to gain a better understanding of the factors that influence children's readiness to learn at school entry;

- to understand the impact of community-level influences over and above other factors (e.g. individual and family) in the child's life;

- to allow the selected communities to compare the characteristics of their children to the national norms established with the National Longitudinal Survey of Children and Youth.

The objectives of the Communities Survey 2005 remain the same as the Communities Survey 2001. Additionally it allows researchers the ability to measure any changes that might have occurred in the readiness to learn of children in each community as compared to the benchmarks established during the Communities Survey 2001.

4.0   Concepts and Definitions

This chapter outlines concepts and definitions of interest to the users. There are many variables and concepts which are critical to the analysis of the Communities Survey 2005. The following is an explanation of the key concepts. Note that derived variables are those that are not asked directly to the respondents but are calculated using information they have provided.

4.1   Unit of Analysis

For each child, extensive information is gathered on the child's family, parent(s), and neighbourhood. The child should be treated as the unit of analysis, and family and household variables should be treated as characteristics of the child. Thus, the files which have been constructed for all Communities Survey data consist of child records. In order to understand the family situation, estimates such as of the number of children in single-parent families can be produced.

4.2   Person Most Knowledgeable and Spouse

At the time of the telephone interview, the contact person in the household (the person whose name appeared on the sampling frame) was asked to identify who, in the household, was the **p**erson **m**ost **k**nowledgeable (PMK) about the child. The PMK provided the information for the selected child and then gave socio-demographic information about him/herself and the spouse/partner, if present.

The following is the breakdown of the relationship between the PMK and the children for the Communities Survey 2005[1]:

- For 87.9% of the responding children, the PMK was the mother (86.0% the biological mother and 1.9% the step, adoptive or foster mother).

- For 10.8% of the children the PMK was the father (10.5% the biological father and 0.3% the step, adoptive or foster father).

- For 1.3% of the children the PMK was not their parent

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. In most cases this situation occurred when there were language barriers with the PMK but someone else in the household was able to speak the language and provide the necessary information. This situation also occurred where a child had a very young mother living with her own parents, i.e., the child's grandparents, and the grandmother was selected as the PMK.

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

The following is the breakdown of the relationship between the spouse/partner and the Communities Survey 2005 children:

- For 22.4% of the children, the PMK did not have a spouse/partner residing in the household.

---

[1]   These percentage estimates for the PMK and spouse/partner are based on shared weighted data.

- For 67.8% of the children the spouse/partner was the father (64.4% the biological father and 3.4% the step, adoptive or foster father).

- For 9.1% of the children the spouse/partner was the mother (8.8% the biological mother and 0.3% the step, adoptive or foster mother).

- For the remaining 0.7% of the children, the spouse/partner was not their parent.

4.3 Family Derived Variables

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

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

Single-parent Family

Children living with one parent are classified as being members of single-parent families.

In Communities Survey 2005, 77.8% of children were living with two parents, 21.6% with one parent and 0.6% without a parent[2] (GDMCD04).

Intact, Step and Blended Families

Children living with two parents are classified as being members of intact, step and/or blended families based on the relationship of these children to the parents.[3]

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 Communities Survey children, 70.4% were a member of an intact family (GDMCD16).

Step Family

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

---

[2] These estimates for family derived variables are based on shared weighted data.

[3] 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 labeled as a step family. In other Statistics Canada publications children of this type are treated as if they were biological children of both parents.

For the Communities Survey 2005 children, 2.8% were step children themselves (GDMCD03) and 7.4% lived in a step family (GDMCD15).

Blended Family

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

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

- a couple living 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 Communities Survey 2005 children, 5.8% were members of a blended family (GDMCD14).

Economic Family

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

Siblings

Siblings include full, half, step, adopted and foster siblings. Only siblings residing in the household have been included in the calculation of the sibling derived variables. In the case of common-law relationships, if both members have brought their own children into the relationship then these children are considered to be siblings. All siblings living in the household, including adult siblings, are included in the calculation of the sibling derived variables. The sibling derived variables include the number of older siblings, younger siblings and siblings with exactly the same date of birth; i.e., twins (GDMCD08, GDMCD09, GDMCD10 and GDMCD11).

4.4   Socio-economic Derived Variables

In the previous Communities Survey 2001, two derived variables were produced to assist analysts in understanding and explaining the socio-economic situation of the child's family: socio-economic status (SES) and income ratio.

Socio-economic Status

This derived variable, which was produced in the Communities Survey 2001, will not be available for the Communities Survey 2005. The classification system grouped occupations that appear in Statistics Canada's 1980 Standard Occupational Classification (SOC 1980) into 16 somewhat homogeneous categories, which provides a ranking of occupations according to their social standing or prestige. Since that time, this coding structure has been revised (SOC 1991), however, a new definition of socio-economic status has not yet been developed using this new classification.

Income Ratio

Two derived variables (GINHD04A and GINHD05A) have been created to compare the household income to the low income cut offs (LICO). LICOs are used to distinguish "low income" family units from "other" family units. A family unit is considered "low income" when its income is below the cut off for its family size and its community. A family at or above the low income cut off falls into the "other" category[4]. The variable GINHD03A gives the value of the LICO by geographic area.

4.5   Income Imputation

For the respondents, there exist many cases of partial non-response. In some cases this may be for an entire component or only for certain questions. Imputation is the process whereby missing or inconsistent items are "filled in" with acceptable values. In the Communities Survey 2005, imputation is carried out for certain variables in the adult Income Section.

Imputation flags have been included on the Communities Survey 2005 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 Communities data file have an "I" as the fifth or sixth character of the variable name. For example, the name of the imputation flag for household income (GINHEQ03) is GINHEI03.

Adult Income Imputation

Several income questions were asked during the Communities Survey 2005 household interview. Information on income, broken down into three sources, was asked for the PMK and his/her spouse. Those three income sources are: wages and salary, self-employment net income, and Employment Insurance benefits. Information on income, broken down into four sources was also asked at the household level. Those four income sources were: Child Tax Benefit/National Child Benefit, social assistance, child and spousal support and other sources.

Income is a sensitive topic. As a result, some respondents refused to provide answers to the detailed income questions. Among those, some provided an estimate of their total household income or an estimate of their income using ranges. Finally, for those who responded, amounts declared in the Income Sections were sometimes inconsistent with answers provided in the Labour Force Section (for example, an individual might have reported working in the past 12 months according to answers provided in the Labour Force Section but no wages or self-employment income were reported in the Income Section). Income imputation was carried out to fill out the holes resulting from partial non-response as well as to rectify, when possible, these inconsistencies. Imputation was also done for households whose total reported income was less than $6,000.

Imputation was carried out for each of the income sources. Imputation was done using a nearest neighbour approach. This method first identifies a respondent to the Income Section (a donor) who has similar characteristics as the individual or household with incomplete income data (the recipient). Once the nearest neighbour has been identified, the income amounts reported by the donor are used to impute the missing income amounts for the recipient. Two types of imputation were done. First the three sources of personal income for the PMK and the spouse were imputed. The remaining variables were imputed at the household level.

Household level imputation was done in one of three ways. For households that provided an estimate of household income, this estimate was used to help determine the donor. For households that provided an estimate of household income in ranges, the value of the range was used to help determine the donor. If there was no additional income information, then only other

---

[4] For more information about Statistics Canada's low-income measures, please see Low income cut offs from 1994 - 2003 and low income measures 1992 - 2001, Catalogue no. 75F0002MIE2004002.

household variables, for example province, were used. The imputation flags provide information on how the imputation was done.

Imputation flag = 1   Estimated total was used to determine the donor
Imputation flag = 2   Income range was used to determine the donor
Imputation flag = 3   No additional income information was used to determine the donor

Imputation rates for the income variables can be found in the following table.

Variable	Imputation Rate
GINPc1AA PMK income from wages and salaries	22.7%
GINPc1AB PMK income from self-employment	22.3%
GINPc1AC PMK income from Employment Insurance benefits	21.0%
GINPCD04 Total personal income for PMK	26.5%
GINSc1AA Spouse income from wages and salaries	26.5%*
GINSc1AB Spouse income from self-employment	23.8%*
GINSc1AC Spouse income from Employment Insurance benefits	21.5%*
GINSCD04 Total personal income for spouse	29.6%*
GINHe1AD Household income from child tax benefits	26.0%
GINHe1AE Household income from social assistance	21.0%
GINHe1AF Household income from child and spousal support	20.8%
GINHe1AG Household income from other sources	23.3%
GINHEQ03 Total household income	37.2%

* Households where there was no spouse were not included in the calculation of the imputation rate.

4.6   *Early Development Indicator*

The Early Development Indicator (EDI) is a questionnaire which was developed and administered by the Canadian Centre for Children at Risk at McMaster University. This is a teacher-completed checklist that assesses a variety of factors which have an impact on a child's readiness to learn at school. The term "readiness to learn at school" refers to the child's ability to meet the task demands of school, such as being cooperative and sitting quietly and listening to the teacher, and to benefit from the educational activities that are provided by the school. A teacher uses her/his observations after several months of classroom interaction with the child to complete the

questionnaire. Despite being completed for individual children, it is designed only to be interpreted at the group level.

The results of the questionnaire are captured and scored at McMaster University and then sent to Statistics Canada. Where permission has been obtained from the parents, scores from the EDI are linked to data from the Communities Survey 2005 by way of a common identifier. Each of the five domains of school readiness to learn is included in the data file as a scale. They include:

1. physical health and well-being (GEICdS4A);

2. social knowledge and competence (GEICdS5A);

3. emotional health and maturity (GEICdS1A);

4. language and cognitive development (GEICdS3A); and

5. communication skills and general knowledge (GEICdS2A).

5.0   Survey Methodology

The Communities Survey was administered between February and June 2005. Household collection took place in February, March and April by Statistics Canada interviewers who contacted the child's parents and conducted interviews by telephone. In May and June, Statistics Canada interviewers went into the schools to administer the direct measures portion of the survey to those children whose parents had provided consent.

5.1   Population Coverage

The target population for each community is all children enrolled in senior kindergarten in the school boards listed in the following table in the fall of 2004 and are still attending a school within the community in the winter of 2005 (during the household collection period).

Community	School Board or Association
South Eastman, Manitoba	Border Land School Division Division scolaire franco-manitobaine Hanover School Division Red River Valley School Division Seine River School Division
Hampton, New Brunswick	School District # 6
Abbotsford, British Columbia	School District # 34
Mississauga, Ontario	Dufferin-Peel Catholic District School Board Peel District School Board
Niagara Falls, Ontario	District School Board of Niagara Niagara Catholic District School Board
Saskatoon, Saskatchewan	Saskatoon Catholic Schools Saskatoon Public Schools
Montreal, Quebec	Commission scolaire de Montréal

5.2   Sample Design

The sampling unit for the Communities Survey is the child. Children are selected from school files provided to Statistics Canada by the community school boards.

When selecting the sample for the seven participating communities, Statistics Canada was provided with a frame of 9,807 potential children from which to select. A quality assessment of the frame was conducted prior to sample selection which included the removal of duplicate records and the update of missing, incomplete or inconsistent information. A systematic random sample of children was selected from the lists, which were ordered by school and postal code, and information about the sampled individuals was obtained for interviewing. In three of the seven participating communities all the children were selected to be interviewed as the population of children in those communities was under or close to 700. The initial sample allowed for a certain percentage of sampled units to be out-of-scope for the survey (for instance children who were enrolled in kindergarten within the community in the fall of 2004 but who were no longer enrolled during the Communities Survey household interview period). It should also be noted that certain children who were attending a school within the community during the household interview period but who were not enrolled in the school in the fall of 2004 could not be included in the sample and are not accounted for during estimation.

5.3   Sample Size

The following table shows the size of the population and sample for each community.

Community	Population	Sample Size
South Eastman, Manitoba	809	582
Hampton, New Brunswick	697	697
Abbotsford, British Columbia	1,489	582
Mississauga, Ontario	668	668
Niagara Falls, Ontario	721	721
Saskatoon, Saskatchewan	2,307	584
Montreal, Quebec	3,116	575
Total	**9,807**	**4,409**

6.0   Data Collection

Data collection for the surveyed communities took place between February and June of 2005. Household data collection was carried out from February to April by Statistics Canada interviewers using a computer-assisted telephone interview (CATI) application. Interviewers contacted the child's parents and conducted interviews by telephone. Field interviewers then went into the schools in May and June to administer the direct measures portion of the survey to those children whose parents had provided consent. The Early Development Indicator (EDI) was completed in the school by teachers between January and March.

6.1   Process for Informed Consent

In September 2004, school boards in each community received a letter which described the Understanding the Early Years (UEY) project. The letter explained the procedures that would be followed and asked school boards for their cooperation in providing Statistics Canada with kindergarten class list information in order to create a survey frame from which to draw a sample of children.

Introductory letters were also mailed to the teachers and principals of the schools involved to solicit their cooperation, as well as to the parents of all kindergarten children in the school board. Parents were informed that their child might be selected to participate in the Communities Survey, and they were asked to sign a permission slip that would allow their child to participate in the direct measures assessment in the event that the child was, in fact, selected.

Statistics Canada then selected a random sample of children from the kindergarten class list. A follow-up letter was mailed to the parents of the selected children explaining when they could expect to receive a call from an interviewer. Telephone interviews began in early February 2005 and parents were asked to confirm verbally whether they gave permission for their child to participate in the direct measures assessment. Statistics Canada interviewers also asked permission of parents to share the survey data with Human Resources and Social Development Canada, and to link the results of the survey to the results from the EDI.

In May and June, Statistics Canada interviewers went to the schools to administer the direct measures assessment to the children whose parents had provided either written or verbal consent. Kindergarten teachers completed the EDI questionnaire for each student in his/her class (with the exception of those for whom there was no parental consent) between the months of January and March.

6.2   Household Data Collection

For the household data collection, Statistics Canada interviewers using a CATI application contacted the person who identified him/herself as the person most knowledgeable (PMK) about the child, usually the mother. Below is a brief description of each section of the computer-assisted interview.

6.2.1   Entry/Exit Component

The first contact with the household was established using the address and telephone number provided by the school board files. Once the sample child was confirmed to be part of the household, the interviewer continued with the interview by preparing a list of all household members, gathering tracing information and recording basic demographic characteristics such as sex, date of birth, marital status, relationships between household members and dwelling information.

6.2.2   The Child Component

The Child questionnaire was completed for the selected kindergarten child in the household. The PMK answered the questions on behalf of the child on such subjects as:

- Health
- Behaviour
- Activities
- Literacy
- Parenting
- Family History
- Childcare
- Sleep Habits
- Positive Behaviour
- Communication
- Use of resources within the community

6.2.3   The Adult (Parent) Component

Questions in this section were asked of the PMK about him/herself and his/her spouse/partner (where applicable). Subjects covered included:

- Education
- Income
- Labour Force Activity
- Health
- Family Functioning
- Neighbourhood Safety
- Socio-demographic characteristics
- Social Support

6.3   Cognitive Measures Administered to the Child

Three direct measures assessments were administered to each sampled child by a Statistics Canada interviewer in the schools: the Peabody Picture Vocabulary Test – Revised (PPVT-R), and shortened versions of the Who Am I? and the Number Knowledge Test. Assessments took approximately 30 minutes in total to administer and were only administered to children for whom consent had been obtained from the parent.

Children who were not able to communicate in English or French, were not given the assessments.

Peabody Picture Vocabulary Test – Revised

The PPVT-R was used to assess the child's level of receptive or hearing vocabulary. It was administered to each selected child. The interviewer says a word as it appears in the test and the child subsequently points to one of four pictures on a card which he believes represents the word. Pictures and words become progressively more difficult as the test proceeds. Results are predictive of school achievement. The test was developed by Lloyd and Leota Dunn at the University of Hawaii and is widely used as a measure of receptive vocabulary for any age group (two and a half years to adulthood). For detailed information on scoring of the PPVT-R, please refer to Chapter 16.0 of the National Longitudinal Survey of Children and Youth, Cycle 5 Microdata User Guide.

Who Am I?

The Who Am I? questionnaire was developed to assess children on entry to school. It assesses the cognitive processes that underlie the learning of early literacy and numeracy skills. The original assessment comprises three scales: copying (circle, cross, square, triangle, and diamond), symbols (printing name, letters, numbers, words, and sentences) and drawing (a picture of self). However, because of time constraints, the drawing task was removed from the Communities Survey. A booklet is used in which the child completes as many tasks as he is able as the assessor turns the pages and gives instructions. The instrument was developed by Dr. Molly de Lemos and colleagues at the Australian Council for Educational Research and can be used with children from 3 to 7 years of age.

Number Knowledge Test

This test assesses the child's understanding of the concept of quantities and of the system of whole numbers. Children who start school with this intuitive knowledge generally do well in math. Children who do not have this understanding, or who are working in a language different from their mother tongue, often have difficulty mastering basic arithmetic and demonstrating number sense. The assessment was developed by Dr. Robbie Case at the Ontario Institute for Studies in Education, University of Toronto. It can be used with children from about 3.5 to 10.5 years of age.

Dr. Case and his colleague Dr. Yukari Okamoto at the University of California developed a shortened version of this assessment for the National Longitudinal Survey of Children and Youth. The test is administered orally and the questions are asked until the child fails to answer more than half the problems in a level.

6.4   Interview Length

For the household data collection, the interview length for responding Communities Survey households was approximately one hour and forty minutes. Administration of the child and adult questionnaires took an average of 70 minutes to complete. Additional time was required for the interviewer to introduce the survey, complete the household roster, etc.

Administration of the cognitive tests in the schools required approximately 10 minutes per test or 30 minutes in total for each child.

6.5   Interviewer Training, Supervision and Control

The Communities Survey 2005 was conducted by the Labour Force Survey (LFS) 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 of 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 interviewer and reassigned. The senior interviewers are, in turn, under the supervision of the LFS program managers, located in Statistics Canada regional offices.

For the Communities Survey 2005, a combination of classroom training and self-study materials ensured that interviewers had a thorough understanding of survey concepts. As part of the self-study portion of the training, interviewers read the Interviewer's Manual and completed home study exercises before a one day training session took place. Classroom training for the household portion of the survey was given in the Regional Offices by Head Office staff. During this training, a program manager or a senior interviewer presented an overview of the survey,

went through a mock interview with the participants and presented exercises to help interviewers minimize non-response.

Training for the direct measures portion of the survey (which took place in the schools) was given by a senior interviewer during a one day training session. Interviewers were again required to read the Interviewer's Manual and familiarize themselves with the appropriate materials in advance of training. A video was provided to each interviewer for viewing prior to the direct measures training session.

7.0 Data Processing

The main output of the Communities Survey 2005 is a "clean" data file. This chapter presents a brief summary of some of the processing steps involved in producing this file.

The Communities Survey 2005 relies mainly on computer-assisted telephone interviews, which presents many advantages when it comes to the processing of the data, not the least of which is a reduction in the need for data capture. On-line flow and consistency edits which are programmed into the application mean that the data are available more quickly for Head Office processing and the quantity of processing is reduced.

Once the data file is received at Head Office, processing of the Communities Survey 2005 data is done using the Generalized Processing Environment. This is a generic system that follows a series of steps to "clean" a file from beginning to end. The main steps include:

- Clean up
- Relationship edits
- Pre-edit
- Flow edits
- Coding
- Consistency edits
- Derived variables
- Final processing file

7.1 Computer Generated Edits

As mentioned above, all of the information for the household was collected during a telephone interview using a computer-assisted interviewing (CAI) application. As such, it was possible to build various edits and checks into the questionnaire for the various household CAI components in order to ensure a high quality of the information collected. Below are specific examples of the types of edits used in the application:

Review Screens
Review screens were created for important and complex information. For example, the selection procedures for the person most knowledgeable (PMK), a critical element of the survey, were based on the household roster. The household roster screen showed the demographic information for each household member and his/her relationship to every other household member. The collected information was displayed on the screen for the interviewer to confirm with the respondent before continuing the interview.

Range Edits
Range edits were built into the CAI system for questions which ask for numeric values. If values entered were outside the range, the system generated a pop-up window which stated the error and instructed the interviewer to make corrections to the particular question. For example, one question asks about the weight of the child at birth. If the weight entered into the computer was either significantly high or low, a pop-up message would appear asking the interviewer to confirm the answer with the respondent.

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

General Consistency Edits
Some consistency edits were included as part of the CAI system which allowed interviewers to "slide back" to previous questions to correct for inconsistencies. Instructions were displayed to interviewers for handling or correcting problems such as incomplete or incorrect data. For example, in the collection of the Labour Force Section, the number of weeks worked, not working, and looking for work should not total more than 53 weeks. If this was the case, the system generated a pop-up window that stated the error and instructed the interviewer to slide back to the appropriate question to confirm the data and make corrections as required.

Consistency Edits
Edits were also performed to ensure consistency. When inconsistencies were identified, the interviewer was asked by the system to confirm the data with the respondent through a series of questions. For example, in the Socio-demographic Section if the respondent answered Canada as the country in which they were born, the consistency edit program will automatically select the answer of "Canada, citizen by birth" when asked "Of what country are you a citizen".

7.2   Data Capture

Paper Questionnaires
Some questionnaires did not make use of computer-assisted interviewing. One of the direct assessment measures for example, the "Who am I?", was completed by the child using a paper questionnaire.

The Early Development Instrument was administered by McMaster University, who processed the results and delivered the data file to Human Resources and Social Development Canada, which then sent the file to Statistics Canada for linkage to the Communities Survey.

Data Capture for the Paper Questionnaire
The "Who Am I?" questionnaires were scored at Statistics Canada's Head Office. The scored results were provided to Head Office processing staff for further processing. As part of the capture system, some quality checks were built in to flag unusual entries.

7.3   Clean Up

Defining Requirements
The purpose of this step is to drop full-duplicate records and split-off records with duplicate identification numbers for examination. Then the data is split between response and non-response based on pre-determined criteria.

A review was done of the responding and non-responding questionnaires and specifications were created based on this analysis to determine which records would be dropped due to non-response. Essentially, if a record was missing key information or had more than half the questions unanswered, they were dropped from the file.

At the end of this step, records are processed by questionnaire type; that is the Adult Questionnaire, the Child Questionnaire, and the Household Questionnaire.

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

7.4   Relationship Edits

The relationship edit establishes the relationship between the members of the household and creates the family derived variables. This step performs a standard set of edits against the relationship information entered for all members of a given household; some inconsistencies are corrected automatically by an application using a set of rules, while others are flagged for manual review and recoding. A related set of derived variables is produced through the relationship edits.

7.5   Pre-edits

The purpose of the pre-edit is to carry out some basic formatting and preliminary editing. As an example, for all records where values were missing (blank) from the collection, the value of "9", "99", "999"… was inserted to indicate that no information was collected. The "Don't know" values returned by the CAI application as code "9" are changed to "7" in the pre-edits. As well, the "Mark all that apply" questions were de-strung and values converted to Yes (1) or No (2) responses. Finally, all text answers were removed from the processing file and set aside to be handled separately.

7.6   Flow edits

The flow edits replicate the flow patterns from the questionnaire. Variables which are skipped based on flows are converted from "Not stated" to "Valid skip" codes (6, 96, 996…).

For skips based on the answer to certain questions, all skipped questions are set to "Valid skip". For skips based on "Don't know" and "Refusal", all skipped questions are set to "Not stated".

7.7   Coding of Open-ended Questions

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

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

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

How they are coded
The open-ended questions were coded using various standard classifications. Occupation questions were coded using the 1991 Standard Occupational Classification codes (SOC) and the industry questions were coded using the 1997 North American Industry Classification System (NAICS). Grouped versions of these codes are available on the data file (GLFPcD7A and GLFPcD8A for the PMK, and GLFSCD7A and GLFSCD8A for the spouse/partner).

7.8   Consistency Editing

After the flow edits were completed, consistency editing was carried out to verify the relationship between two or more variables. Decision tables are used to specify the consistency edits. The LogiPlus software was used to input the decision tables and generate the SAS code. A report

with the "Before" and "After" counts of the variables is generated. Additionally, a report is generated providing the rule counts for each decision table.

For example, in the Socio-demographic Section, for children who were not born in Canada, question GSDCQ3 asks what year they first immigrated to Canada. There was a consistency edit which compared the year of immigration to the child's year of birth. If the year of immigration was before the year of birth then the year of immigration was set to "Not stated" in the edit.

7.9   Imputation Flags

For various reasons, certain variables may be missing for responding households on the Communities Survey 2005 file. This is usually referred to as item non-response or partial non-response.

For a few variables on the file, 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 Communities Survey 2005, imputation was carried out for household income and PMK income. See Section 4.5 for more details on imputation.

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

All imputation flag variables on the data file have an "I" as the fifth or sixth character of the variable name. For example, the imputation flag variable for the income of the PMK would be named GINPEI03.

Please note that if one or several of the variables of an answered component presented a non-response, this or these variables were not imputed except for the income variables mentioned above.

7.10 Derived Variables

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

Derived Variable Name
All derived variables on the data file have a "D" as the fifth character of the variable name. For example, the name of the variable for the "Number of hours per week spent in the primary care arrangement" is GCRCD02.

7.11 Standard Coding Structures

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

Refusals

During a CAI interview, the respondent may 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 data file, an item which was refused is indicated by a code "8". For a variable that is one digit long the code is "8", for a two-digit variable "98", for a three-digit variable "998", etc.

Don't know

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

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

Not applicable

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

In some cases a single question or series of questions was not applicable. For example, the question on number of hours per week the child is cared for in a daycare centre (GCRCQ1GA) is only applicable for children for whom this type of care is used (GCRCQ1GA = 1). Otherwise there will be a code 996 for this question.

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 the following reasons.

➢ As part of the CAI interview, the interviewer was permitted to enter a "Refusal" or "Don't know" code, as described above. When this happened, the CAI system was often programmed to skip out of this particular section of the questionnaire. In the case of a "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 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 of the respondent. In some cases it is not even known if the question was applicable to the respondent.

➢ In some cases a questionnaire was not started or it was started but ended prematurely. For example, there may have been some kind of an interruption, or the respondent decided that he/she wished to terminate the interview. If there was enough information collected to establish the household as a responding household, then all remaining unanswered questions 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".

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

For derived variables, if one or more of the input variables (to the derived variable) had a "Refusal", "Don't know" or "Not stated" code, then the derived variable was set to "Not stated".

An item which was coded as "Not stated" is indicated by a code "9". For a variable that is one digit long the code is "9", for a two-digit variable "99", for a three-digit variable "999", etc.

7.12 Naming Convention

The Communities Survey 2005 data file documentation system has employed certain standards to label variable names and values. The intent is to make data interpretation more straight-forward for the user.

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

Format for Variable Names

The **first** character of the variable name refers to the Communities Survey reference year:

"C" indicates the 1999 reference year,
"D" indicates the 2000 reference year,
"E" indicates the 2001 reference year,
"F" indicates the 2004 reference year and
"G" indicates the 2005 reference year.

The **second and third** characters refer to the section of the questionnaire where the question was asked or the section from which the variable was derived. Refer to Section 7.13 for the acronym names used for each section of the various questionnaires.

The **fourth** character refers to the collection unit or the unit to which the variable refers. There are five possibilities[5].

"C" if the variable refers to the child,
"P" if the variable refers to the PMK,
"S" if the variable refers to the spouse/partner,
"H" if the variable refers to the household and
"W" if the variable refers to a weight.

The **fifth, sixth, seventh and eighth** characters of the variable name (for example GSDCQ2B and GLFPbD38) could refer to the following:

Q   refers to the variable for a question that was asked directly on the questionnaire.

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

---

[5] 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 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. However it will **not** be possible to produce estimates of the number of low income households or depressed PMKs.

D   means the variable was derived from two or more questions that were asked on the questionnaire or coded variables.

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

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

7.13 Acronyms for the Questionnaire Sections

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

Acronym	Section	Collected or Derived From:
GE	Geographic	Sample information.
HH	Household	Dwelling characteristics.
MM	Variables collected as part of the household roster.	Basic demographic variables for each household member. These variables are included on the data file for the child, the PMK and the spouse/partner.
DM	Demographic - derived to explain the living arrangements of the child.	Information from the household roster and relationship grid.
SD	Socio-demographic	Child on the Child Questionnaire and for the PMK and spouse/partner on the Adult Questionnaire.
HL	Health	PMK and spouse/partner on the Adult Questionnaire and for the child on the Child Questionnaire.
CH	Adult Chronic Conditions	PMK and spouse/partner in the Health Section of the Adult Questionnaire.
RS	Restriction of Activities	PMK and spouse/partner in the Health Section of the Adult Questionnaire.
DP	Depression Scale	Parent Questionnaire (this scale was administered to the PMK).
ED	Education	PMK and spouse/partner on the Adult Questionnaire.
LF	Labour Force	PMK and spouse/partner on the Adult Questionnaire.
IN	Income	Household income and personal income of the PMK collected on the Adult Questionnaire.
FN	Family Functioning	Adult Questionnaire (section asked to the PMK or spouse/partner).
AC	Activities	Child on the Child Questionnaire.
BE	Behaviour	Child on the Child Questionnaire.
PR	Parenting Style	Child on the Child Questionnaire.
CR	Child Care	Child on the Child Questionnaire.

Acronym	Section	Collected or Derived From:
PP	PPVT Test	Child on the Child Questionnaire.
SF	Neighbourhood Safety	Adult Questionnaire (section asked to the PMK or spouse/partner).
SP	Social Support	Adult Questionnaire (section asked to the PMK or spouse/partner).
SL	Sleep	Child on the Child Questionnaire.
WM	Who Am I?	Direct Measure.
KN	Number Knowledge	Direct Measure.
CC	Communities Component	Child on the Child Questionnaire.
WT	Weight	Weight as part of the sample design.

7.14 Examples of Variables Names

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

Variable GLFSQ02	Refers to:
G	Indicates the 2005 reference year
LF	Labour Force Section
S	Spouse/partner
Q	An item asked directly on the questionnaire
02	The second question from the Labour Force Section of the Adult Questionnaire

Variable GPRCS03	Refers to:
G	Indicates the 2005 reference year
PR	Parenting Section
C	Child
S	Score - a positive interaction score on the parenting scale for a child
03	The identification number of the item

8.0   Content of the Survey

The Communities Surveys were 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 National Longitudinal Survey of Children and Youth (NLSCY) expert advisory group (EAG), that consists of researchers in the area of child development and the social sciences;

- federal departments;

- representatives from the provinces and territories responsible for child development programs.

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

8.1   Survey Components

The Communities Surveys is divided into several components; these are described in Chapter 6.0, Data Collection. Below is a summary of each component.

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

Adult   Questions asked about the person most knowledgeable (PMK) and spouse. The adult component is completed once even if there are two children in the household.

Child   Questions about the selected child asked to the PMK.

Direct assessments   Several direct assessments are done with the children; these are described in Chapter 6.0.

8.2   Demographic Variables

The demographic variables are collected in the household questionnaire. As part of the household questionnaire some basic demographic information (e.g., age, gender, and marital status) is collected for all members of the child's household. The relationship grid is also completed as part of this questionnaire i.e., the relationship of everyone in the household to the PMK. Using this information it is possible to create an extensive set of variables to describe the child's family situation. Most of these derived variables are critical to the analyses of Communities Survey data and are described in Chapter 4.0.

It is necessary to perform an extensive series of edits on the data that was collected. The following are some examples of the types of editing that are carried out.

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

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

8.3   Adult Questionnaire

Education (Parent)
The Education Section is completed for both the PMK and spouse/partner. The objective is to gather information on educational attainment.

Research has indicated a link between maternal educational attainment, the home environment and child development.

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

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

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

> **Labour Force Derived Variables**
> Several labour force derived variables have been created for the PMK and spouse/partner of the PMK. They include:
>
> GLFPcD5A / GLFSCD5A: NAICS code for PMK's / spouse's current job
> GLFPCD6A / GLFScD6A: SOC 1991 code for PMK's / spouse's main job
> GLFPcD7A / GLFSCD7A: Standard industry code for current job - grouped
> GLFPcD8A / GLFSCD8A: Standard occupation code for current job - grouped

Income
In the Income Section of the survey, the sources of income and the income are collected for each household. There are also a few questions which ask for the perceptions of the PMK or the PMK's spouse regarding how well they think they are doing financially. This information provides an indicator of the family's economic situation, an essential component of the child's environment.

As family income is an important part of many studies on child development, we impute a value for household income if the respondent did not answer these questions. See Chapter 4.0 for a detailed explanation of how income is imputed.

Two derived variables (GINHD04A and GINHD05A) have been created to compare the household income to the low income cut offs (LICO). LICOs are used to distinguish "low income" family units from "other" family units. A family unit is considered "low income" when its income is below the cut off for its family size and its community. A family at or above the low income cut off falls into the "other" category[6]. The variable GINHD03A gives the value of the LICO by geographic area.

---

[6] For more information about Statistics Canada's low-income measures, please see Low income cut offs from 1994 - 2003 and low income measures 1992 - 2001, Catalogue no. 75F0002MIE2004002.

Adult Health
This section asks the PMKs and their spouses about their general health, chronic conditions and restriction of activities.

Chronic Conditions
PMKs and their spouses are asked whether or not they have any long-term conditions (e.g., allergies, asthma, and high blood pressure). A derived variable (GCHPD01 or GCHSD01**)** indicates that the respondent answered "yes", they have at least one of the long-term conditions.

Restriction of Activities
The PMKs and their spouses are asked a series of questions about whether or not their activities are restricted at home, work, school, etc. A derived variable (GRSPD01 or GRSSD01) is also created stating whether or not the PMK or spouse reported an activity restriction.

Depression Scale
A Depression Scale (GDPPS01) was administered to the PMK as part of the Adult Questionnaire.

Family Functioning
The objective of the Family Functioning Section is to provide a global assessment of family functioning and an indication of the quality of family relationships. This section is asked of the PMK or spouse.

Neighbourhood Safety
This section gathers information about the respondent's satisfaction with his/her neighbourhood as a place to raise children, including perception of the extent of danger and problems, and of social cohesion or "neighbourliness". Two scales are created in this section: Neighbourhood Safety Score (GSFHS5), indicating the degree of perceived neighbourhood safety and the Neighbours Score (GSFHS6), indicating the degree of neighbour cohesiveness.

Social Support
The purpose of this section is to collect information on the level of support the PMK's feel they have from friends, family members and members of the community. This section is asked of the PMK or the PMK's spouse.

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

8.4   Child Questionnaire

Direct Measures
The purpose of this section is to establish the groundwork for the Direct Measures that will be asked of children. If the child does not have the ability to do the direct measures (e.g., does not speak English or French, is colour blind), the measures will not be administered.

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

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

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

Communication
The items have been modified from the New Zealand Competent Children Study. They cover a child's ability to understand oral messages and to pass a message on to someone else, as well as to communicate verbally.

A**ctivities**
This section measures the child's participation in various non-school activities and will give some sense of how the child spends his/her time, of personal interests, as well as the degree of interaction with peers.

Community
This section asks about the existence of educational and recreational resources in the community such as museums, community centres and reasons for not using these resources where they do exist, for example, inaccessibility or cost.

Behaviour
The objective of this section is to assess aspects of the behaviour of the child as well as to measure the prevalence of behaviours such as hyperactivity and physical aggression. The scales derived from these questions are described in detail later in this chapter.

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

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

Sleep
The questions in this section asked about the number of hours of sleep the child sleeps on average, including both daytime naps and sleep at night.

Parenting
Parenting style is considered to have an important influence on child behaviour and development. The objective of this section is to measure certain parenting behaviours. Scales are created from the questions in this section.

The PMKs who have a spouse/partner in the house are asked how often the PMK and spouse/partner agree with each other about parenting decisions. This question is similar to questions in the Strayhorn and Weidham scale, from which the other parenting questions have been adapted.

Family History
This section was designed to provide information on the child's family arrangements; whether or not his/her parents are married, separated or divorced, the age of the child when parents separated/divorced and so on.

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

Socio-demographics
These questions gather socio-demographic information on the selected child. Such information as ethnicity and country of origin is collected.

8.5 Survey Scales

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

For example, on the Child's Questionnaire, it was determined that it was important to have an assessment of certain parenting behaviours. The scale is intended to measure four different constructs or factors related to parenting; positive interaction, ineffective parenting, consistent parenting and rational 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, in the Parenting Scales, for children with higher scores for the "positive interaction" factor, the person most knowledgeable (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 construct or factor with adequate precision.

During the development of the Communities Survey, it was decided to use the same scales that are used for five year old children in the NLSCY. For more detailed information about the calculation of scores, please refer to Chapter 9.0 of the NLSCY, Cycle 5 Microdata User Guide. The remainder of this chapter provides an in-depth description of the sources of the Communities Survey scales.

8.5.1 Behaviour Scale

The objective of the Behaviour Scale is to assess the following behaviours of children:

➢ Conduct disorder,
➢ Hyperactivity,
➢ Emotional disorder and
➢ Indirect aggression

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

Conduct Disorder
Includes items GBECQ6G, GBECQ6X, GBECQ6AA, GBECQ6FF, GBECQ6JJ and GBECQ6NN from the Ontario Child Health Study (OCHS).

Hyperactivity
Includes items GBECQ6B, GBECQ6I, GBECQ6P, GBECQ6S and GBECQ6W from the OCHS and GBECQ6QQ and GBECE69B from the Montreal Longitudinal Survey.

Emotional Disorder
Includes items GBECQ6F, GBECQ6K, GBECQ6Q, GBECQ6V, GBECQ6CC, GBECQ6MM and GBECQ6RR from the OCHS. Anxiety includes items taken from OCHS emotional disorder items (GBECQ6F, GBECQ6Q, GBECQ6V and GBECQ6CC).

Indirect Aggression
Includes items GBECQ6J, GBECQ6R, GBECQ6Z, GBECQ6LL and GBECQ6TT from Lagerspetz, Bjorngvist and Peltonen of Finland.

The following table indicates the items that were included on the questionnaire to measure these various constructs of behaviour.

FACTOR	SCORE	ITEMS
Hyperactivity - inattention	GBECdS6A	GBECQ6B, GBECQ6I, GBECQ6P, GBECQ6S, GBECQ6W, GBECQ6QQ, GBECE69B
Emotional disorder - anxiety	GBECdS8A	GBECQ6F, GBECQ6K, GBECQ6Q, GBECQ6V, GBECQ6CC, GBECQ6MM, GBECQ6RR
Physical aggression - conduct disorder	GBECS09	GBECQ6G, GBECQ6X, GBECQ6AA, GBECQ6FF, GBECQ6JJ, GBECQ6NN
Indirect aggression	GBECS10	GBECQ6J, GBECQ6R, GBECQ6Z, GBECQ6LL, GBECQ6TT

8.5.2 Depression Rating Scale

Objectives and Overview
The Depression Rating Scale was administered to the PMK as part of the Parent Questionnaire. Questions for this scale (GDPPQ12A to GDPPQ12L) are a shorter version of the Depression Rating Scale (CES-D), comprising 20 questions, developed by L. S. Radloff of the Epidemiology Study Center of the National Institute of Mental Health in the United States. This rating scale is used to measure the frequency of symptoms in the public at large. The occurrence and severity of symptoms associated with depression during the previous week are measured. The rating scale was reduced to 12 questions by Dr. M. Boyle of the Chedoke-McMaster Hospital, McMaster University.

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

Items Included in the Depression Rating Scale
The Depression Rating Scale includes 12 questions, each of which contains four response categories. In order for the lowest score value to be zero, the value for each question was reduced by one in calculating the score. As well, the answer categories were reversed for questions having a negative loading (GDPPQ12F, GDPPQ12H and

GDPPQ12J). The total score (GDPPS01) may therefore vary between 0 and 48, a high score indicating the presence of depression symptoms.

8.5.3   Family Functioning Scale

Objectives and Overview
Questions related to family functioning, i.e., GFNHQ01A to GFNHQ01L, were developed by researchers at the Chedoke-McMaster Hospital, 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 GFNHQ01M, drawn from the Follow-up to the Ontario Child Health Study, was added to the original scale to determine whether alcohol consumption had an effect on global family dynamics. However, it was not used in the analysis of the scale.

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

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

Administering the Family Functioning Scale
The Family Functioning Scale was administered to either the PMK or the spouse/partner as part of the Parent Questionnaire. The scale includes 12 questions, each of which contains four response categories. In order for the lowest score value to be zero, the value of the categories was reduced by one in calculating the score. The order of the categories was reversed for questions having a negative loading (GFNHQ01A, GFNHQ01C, GFNHQ01E, GFNHQ01G, GFNHQ01I, and GFNHQ01K). The total score (GFNHS01) may therefore vary between 0 and 36, a high score indicating family dysfunction.

8.5.4   Neighbourhood Safety Scale

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

Questions GSFHQ01 to GSFHQ06E cover the length of residency in the neighbourhood, satisfaction with the neighbourhood as a place to bring up children, safety, social cohesion and neighbourhood problems. They represent a revised version of specific sections of the Simcha-Fagan Neighbourhood Questionnaire used by Dr. Jacqueline Barnes in her studies of neighbourhoods in Boston and Chicago. Revisions were made based on the factor analysis of the sections, in consultation with Dr. Barnes.

The items that comprised each factor are described in the following table:

FACTOR	SCORE	ITEMS
Neighbourhood safety	GSFHS5	GSFHdQ5A, GSFHdQ5B, GSFHdQ5C
Neighbours	GSFHS6	GSFHQ06A, GSFHQ06B, GSFHQ06C, GSFHQ06D, GSFHQ06E

It should be noted that in Communities 2000 and 2001, only the Neighbours Scale was created.

8.5.5   Parenting Scales

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

Questions GPRCQ01 to GPRCQ18 and GPRCQ21 to GPRCQ24 on positive interaction, ineffectiveness and on consistency were provided by Dr. M. Boyle of the Chedoke-McMaster Hospital, McMaster University, based on the work of Dr. Ken Dodge (Vanderbilt University) and an adaptation of the Parent Practices Scale of Strayhorn and Weidman.

Calculation of Parenting Scores
Once the factor structures were analysed and the items included in each factor were determined, scores were calculated. To produce the scores, one was subtracted from each item so that the lowest possible score value would be zero (0). For each of the four factors, a score of zero indicates:

> ➢ the absence of positive interaction for the positive interaction factor;
> ➢ the absence of ineffective interaction for the ineffective factor;
> ➢ the absence of consistent parenting for the consistency factor;
> ➢ the existence of punitive interaction or aversion producing practices for the ineffective parenting factor.

The factor analysis derived four factors for this age group: positive interaction (GPRCS03), ineffective parenting (GPRCS04), consistency (GPRCS05) and punitive/aversive (rational) parenting (GPRCS06). The items making up each factor are listed below.

FACTOR	SCORE	ITEMS
Positive interaction	GPRCS03	GPRCQ01, GPRCQ02, GPRCQ03, GPRCQ06, GPRCQ07
Ineffective	GPRCS04	GPRCQ04, GPRCQ08*, GPRCQ09, GPRCQ13, GPRCQ14, GPRCQ15, GPRCQ18
Consistent	GPRCS05	GPRCQ10, GPRCQ11, GPRCQ12*, GPRCQ16*, GPRCQ17*
Rational	GPRCS06	GPRCQ21, GPRCQ22*, GPRCQ23, GPRCQ24*

*   Indicates that the item value was reversed when computing the score. In Communities 2000 and 2001, GPRCQ21 and GPRCQ23 were reversed rather than GPRCQ22 and GPRCQ24.

8.5.6   Social Support Scale

Objectives and Overview
The original scale contains 24 items from Robert Weiss's Social Provisions Model that describes six different social functions or "provisions" that may be acquired from relationships with others. Due to the length of the scale, and on the advice of Dr. M. Boyle at Chedoke-McMaster Hospital, McMaster University the survey uses the shortened version that was derived for the Government of Ontario's, Better Beginnings, Better Futures Project. This measures guidance, reliable alliance and attachment.

Items GSPHQ01A, GSPHQ01E, GSPHQ01F and GSPHQ01I were reversed in the calculation of the score. All values were recoded from 1 to 4 to 0 to 3. The final score ranges from 0 to 24 with a high score indicating a higher degree of social support. The factor structure imposed was the same as that used in Cycle 1 of the NLSCY.

The items that comprised each factor are described in the following table:

FACTOR	SCORE	ITEMS
Social Support	GSPHS01	GSPHQ01A, GSPHQ01B, GSPHQ01C, GSPHQ01D, GSPHQ01E, GSPHQ01F, GSPHQ01H, GSPHQ01I.

9.0   Data Quality

9.1   Response Rates

The following table summarizes the response rates to the Communities Survey 2005.

Community	Sample Size	Sample Size Excluding Out-of-scope	Respondents	Non-respondents	No Contact	Response Rate (%)
South Eastman, Manitoba	582	567	475	77	15	83.8
Hampton, New Brunswick	697	691	574	73	44	83.1
Abbotsford, British Columbia	582	565	462	65	38	81.8
Mississauga, Ontario	668	640	539	51	50	84.2
Niagara Falls, Ontario	721	694	582	66	46	83.9
Saskatoon, Saskatchewan	584	571	467	48	56	81.8
Montreal, Quebec	575	572	466	50	56	81.5
Total	**4,409**	**4,300**	**3,565**	**430**	**305**	**82.9**

9.2   Survey Errors

For four of the seven communities (South Eastman, Abbotsford, Saskatoon and Montreal) 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 in the survey. The difference between the estimates obtained from the sample and those resulting from a complete count taken under similar conditions is called the sampling error of the estimate.

In the case of the remaining three communities (Hampton, Mississauga and Niagara Falls) a complete census of kindergarten children was taken. In these cases there is no sampling error.

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.

9.3   The Frame

The frame, which consisted of a list of all the children registered in senior kindergarten, was provided by the school boards in the seven participating communities. In some cases, the initial lists that were sent contained many cases where information was missing, such as the child's telephone number or address, or cases where there were duplicate records.

For the cases of missing or duplicate information, the school boards who provided the files were contacted. All of the duplicate cases were resolved and, when available, missing information was obtained.

Since the frame files were being provided by different school boards, districts or divisions from seven different communities, the quality of the files provided by one board could be much different than those provided by another. This was observed when the missing information and duplicate cases were identified. Some communities had many more problem cases which needed to be resolved than others.

The frame file could be considered to be up to date at the time when it was provided, in the fall of 2004. However, the frame can not be considered to be up to date during the collection period. It is possible that children who were registered in a kindergarten class in the fall when the frame was obtained would not be registered with that school board during the collection period. These children would then be out-of-scope when the collection was taking place. There would also be children who may move residences, but not schools, during the school year, in which case their contact information on the frame would not be up to date. Since it is not possible to identify these cases until the collection period, this affects the quality of the frame.

It was also noted that several of the children listed on the frame did not appear to live and attend school in the same community. Since all children attending senior kindergarten in the community were included on the frame, regardless of their place of residence, children who did not live in one of the seven communities studied were included on the frame. Other children, who would have registered with the schools after the frame was sent to Statistics Canada, were not eligible for sample selection and are not represented in the estimation. These may include families that moved into the community later in the year. These situations will likely have an impact when comparing results and estimates to other surveys with a purer geography concept than these communities.

9.4   Out-of-scope Respondents

Many steps are used to ensure the highest quality of information when creating the frame of children eligible for the communities portion of the survey. Lists of registered children in kindergarten from schools in participating school boards or districts were provided to Statistics Canada in the fall of 2004. After several steps to remove duplicate and invalid records, the frame for sample selection is finalized. The frame file used to select the sample for the communities, can only be considered to be up to date at the time when it was provided, that is in the fall of 2004. However, the frame may no longer be up to date at the time of collection.

It is possible that children registered in a fall kindergarten class in the selected schools, would no longer be attending that school during the collection period. In some cases the children may have moved out of the communities of interest and these children would then be out-of-scope when the collection was taking place. Other children, who remained in the communities of interest but had changed schools, were also dropped from the sample because the application could not manage these cases. Children who moved into the community and attended school after the fall registration were not eligible for the sample selection and so had no chance of being interviewed. These cases of undercoverage can not be measured through this survey process. Inferences about the excluded population can only be made based on the number of out-of-scope units if the assumption of net migration is near zero.

Out of the 4,409 children in the sample, there were 109 cases of out-of-scope or dropped children. These children were excluded from data collection as they no longer attended the school from which they were sampled. As soon as it had been confirmed that they no longer attended the school from which their name had been provided, and subsequently sampled, they were coded as out-of-scope units. Although an interview was not conducted for these children,

some information was collected at the point of contact, in order to assess their status.   These out-of-scope children are not included in the data file.

10.0 Guidelines for Tabulation, Analysis and Release

This chapter of the documentation outlines the guidelines to be adhered to by users tabulating, analyzing, publishing or otherwise releasing any data derived from the survey microdata file. With the aid of these guidelines, users of microdata should be able to produce the same figures as those produced by Statistics Canada and, at the same time, will be able to develop currently unpublished figures in a manner consistent with these established guidelines.

10.1 Sample Weighting Guidelines for Tabulation

The sample design used for the Communities Survey is self-weighting within each community. Each community has different sampling weights to reflect the different sampling fractions. Some communities were not sampled and in fact a census of that community's kindergarten children was done. When producing simple estimates, including the production of ordinary statistical tables, users must apply the proper population weight (WGT). Even for communities where a census was performed a weight may exist to correct for undercoverage errors caused by families who did not respond to the survey.

If proper weights are not used, the estimates derived from the microdata file cannot be considered to be representative of the targeted population, and will not correspond to those produced by Statistics Canada. In effect, the weight assigned to each child reflects the number of children represented by a particular respondent. For any analysis dealing with correlation analysis or any other statistics where a level of significance is required, it is recommended that a sample weight be used. This weight is obtained by multiplying the population weight (WGT) by the sample size and dividing this total by the total population which we are estimating. This produces a mean weight of 1 and a sum of weights equal to the sample size. The benefit of this adjusted weight is that an over estimation of the significance (which is very sensitive to sample size) is avoided while maintaining the same distributions as those obtained when using the population weight.

10.1.1 Definitions of Types of Estimates: Categorical and Quantitative

It should be pointed out that the Communities file has been set up so that the child is the unit of analysis. The weight that can be found on each record (WGT) is a "child" weight. Estimates of parents or families cannot be made from the Communities data file. A further discussion of units of analyses can be found in Section 4.1 of this document.

Before discussing how the Communities 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 Communities.

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 proportions of children who are in excellent health 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.

      <u>Examples of Categorical Questions:</u>

Q: In general, would you say (the child's) health is:
R: Excellent / Very good / Good / Fair / Poor

Q: Was (the child) born before, after or on the due date?
R: Before / After / On due date

Quantitative Estimates

Quantitative estimates are estimates of totals or of means, medians and other measures of central tendency of quantities based upon some or all of the members of the surveyed population. They also specifically involve estimates of the form $\hat{X}/\hat{Y}$ where $\hat{X}$ is an estimate of the surveyed population quantity total and $\hat{Y}$ is an estimate of the number of persons in the surveyed population contributing to that total quantity.

An example of a quantitative estimate is, for children who were born before the due date, the average number of weeks before the due date that they were born. The numerator is an estimate of the total number of weeks before the due date that the children were born and its denominator is the number of children who were born before the due date.

      <u>Examples of Quantitative Questions:</u>

Q: How many weeks before or after the due date was (the child) born?
R: |_|_| Weeks

Q: What was (the child's) birth weight in kilograms and grams or pounds and ounces?
R: |_|_|_|_|_|_| Kilograms and grams / |_|_| Pounds |_|_| Ounces

10.1.2   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 $\hat{X}/\hat{Y}$ are obtained by:

a) summing the final weights of records having the characteristic of interest for the numerator ($\hat{X}$),
b) summing the final weights of records having the characteristic of interest for the denominator ($\hat{Y}$), then
c) dividing estimate a) by estimate b) ($\hat{X}/\hat{Y}$).

10.1.3   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 <u>total</u> number of weeks before the due date that children were born, multiply the value reported in question GHLCd4DB (number of weeks the child was born before the due date) by the final weight for the record, then sum this value over all records for with GHLCd4DA = 1 (child was born before the due date).

To obtain a weighted average of the form $\hat{X}/\hat{Y}$, the numerator ($\hat{X}$) is calculated as for a quantitative estimate and the denominator ($\hat{Y}$) is calculated as for a categorical estimate. For example, to estimate the <u>average</u> number of weeks before the due date that children were born,

   a) estimate the total number of weeks ($\hat{X}$) as described above,
   b) estimate the number of children ($\hat{Y}$) in this category by summing the final weights of all records with GHLCd4DA = 1, then
   c) divide estimate a) by estimate b) ($\hat{X}/\hat{Y}$).

10.2 Coefficient of Variation Release Guidelines

Before releasing and/or publishing any estimate from the Communities Survey, 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. However for communities with a sampling plan, 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.

First, the number of children who contribute to the calculation of the estimate should be determined. Usually, if this number is less than 30, the weighted estimate should be considered to be of unacceptable quality. For communities, especially when the sampling fraction is large or if a census was conducted, the minimal number can be much less. However, released or tabular information will have to satisfy confidentiality rules when it comes to the minimum sample counts that can be released.

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 rounded weighted 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, and must not breach the rules of confidentiality.

Special attention should be given to estimates from communities where a census of all children was done. While sampling error is not relevant to the estimates produced, non-sampling error can take on a more significant importance. The weight provided on the file for these communities are actually for undercoverage correction due to the number of non-responding cases. Non-sampling errors reflect other reasons for having an imperfect estimator. While both sample and census surveys can exhibit these types of error, a census estimator is often more vulnerable as it is expected to reflect accurate information.

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 of 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 of 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: "Please be warned that these estimates [flagged with the letter U] do not meet Statistics Canada's quality standards. Conclusions based on these data will be unreliable, and most likely invalid."

11.0 Sampling Variability

The Communities Survey 2005 has been sampled using a simple random sampling (SRS) method. The coefficients of variation (CV) applicable to a wide variety of categorical estimates produced from this microdata file could be readily calculated by the user, using standard functions (formulas) found in most statistical or spreadsheet software packages.

The coefficients of variation should be derived using the variance formula for simple random sampling using the finite population correction (fpc) factor. Before using any statistical or spreadsheet software, you should ensure that the appropriate formula is used. The following table shows various algebraic expressions for variance, standard error, coefficient of variation and confidence interval for a simple random sample.

Estimate	Variance	Standard Error	Coefficient of Variation	Confidence Interval (1-∝) %
Mean $(\overline{X})$	$(1-f)\,S^2\,/\,n$	$((1-f)/n)^{\frac{1}{2}}\,S$	$((1-f)/n)^{\frac{1}{2}}\,S/\,\overline{X}$	$\overline{X} \pm t_{\propto/2}\,((1-f)/n)^{\frac{1}{2}}\,S$
Total $(\hat{N})$	$N^2\,(1-f)\,S^2\,/\,n$	$N((1-f)/n)^{\frac{1}{2}}\,S$	$((1-f)/n)^{\frac{1}{2}}\,S\,/\,\hat{N}$	$\hat{N} \pm t_{\propto/2}\,N((1-f)/n)^{\frac{1}{2}}\,S$
Proportion (p)	$\dfrac{(1-f)\,PQ\,N}{(N-1)\,(n-1)}$	$\dfrac{((1-f)\,PQ\,N)^{\frac{1}{2}}}{((N-1)\,(n-1))^{\frac{1}{2}}}$	$\dfrac{((1-f)\,Q\,N)^{\frac{1}{2}}}{(P(N-1)\,(n-1))^{\frac{1}{2}}}$	$\dfrac{p \pm t_{\propto/2}\,((1-f)\,PQ\,N)^{\frac{1}{2}}}{((N-1)\,(n-1))^{\frac{1}{2}}}$

Notation

N : number of units in the population

n : number of units in the sample

$f = n\,/\,N$: sampling fraction of the population

$fpc = 1 - f$: finite population correction factor

y_i : value of the characteristic y for the ith unit

$P = N_y\,/\,N$: proportion of units in the population which have the attribute y

$Q = \dfrac{1-P}{N}$: proportion of units in the population which do not have the attribute y

$S^2 = \sum_{i=1}(y_i - \breve{y})^2\,/\,(N-1)$: variance of the characteristic y in the population

As stated in the previous chapter, for the community sample the acceptable cell frequency count is 30 responding sample units (or an estimated 164 population count). The following table shows the decrease in precision for lower cell counts.

Sample Size	Estimated Population Size	Coefficient of Variation
29	159	16.47%
28	153	16.78%
8	44	31.80%
7	38	34.02%

12.0 Weighting

The principle behind estimation in a probability sample is that each person in the sample "represents", besides himself or herself, several other persons not in the sample. The weighting calculates the number of individuals in the population represented by a record.

For each community, every respondent was given equal weight since sampling was done randomly with equal probability of selection. The weight for each respondent simply corresponds to the ratio of the population size to the sample size. This value was adjusted to take into account the non-respondents and the out-of-scope units within the sample.

By summing the weights of each respondent in a community, an estimate of the actual population size is measured for that community. Thus this final estimated number represents the number of children who were enrolled in kindergarten in this community in September 2004 and were still enrolled during the collection period.