

The Development of an Integrated Survey  
of Aboriginal Peoples

Detailed Discussion Document

Special Surveys Group  
Statistics Canada  
March 24, 1994



## TABLE OF CONTENTS

Preface

<b>1.0 Background</b>	<b>1</b>
1.1 Origins of the Project	
1.2 Project Objectives	
<b>2.0 Issues Affecting Survey Design and Feasibility</b>	<b>2</b>
<b>2.1 Target Population</b>	<b>3</b>
2.1.1 "Aboriginal Peoples"	
2.1.2 "On- and Off-reserve"	
2.1.3 Definition of the Target Population	
<b>2.2 Cross-sectional versus Longitudinal Data</b>	<b>5</b>
<b>2.3 Data Requirements</b>	<b>8</b>
2.3.1 Geographic Level	
2.3.2 Population Subgroups	
<b>2.4 Content Development</b>	<b>9</b>
2.4.1 Content Considerations for an Integrated Survey of Aboriginal Peoples	
2.4.2 The Requirement for Integration of Content	
2.4.3 Working Towards a Consensus	
2.4.4 Other Content-related Issues	
<b>2.5 Data Collection Methods</b>	<b>12</b>
<b>2.6 Collection Tools</b>	<b>13</b>
<b>2.7 Collection Periods</b>	<b>14</b>
<b>3.0 Survey Design Options and Issues</b>	<b>15</b>
<b>3.1 Impacts of Data Requirements</b>	<b>16</b>
3.1.1 Cross-sectional versus Longitudinal Data	
3.1.2 Coverage of On- and Off-reserve Populations	
3.1.3 Definition of the Population of Interest	
3.1.4 Levels for which Data are Required	



3.2 Survey Design Options	18
3.2.1 1996 Post-censal Survey	
3.2.2 Follow-up to Another Post-censal Survey	
3.2.3 Sample Survey Conducted Independently of the Census	
3.2.4 Use of other List Frames	
4.0 Survey Outputs	26
5.0 Consultation Process	27
6.0 Costs and Schedules	28
7.0 Other Considerations	29
7.1 Project Team Approach	
7.2 Steering Committee	
7.3 Access to Data	
7.4 Release Policy	
Appendix A -- Overview of Other Surveys	31
National Longitudinal Survey of Children	
Survey of Labour and Income Dynamics	
National Population Health Survey	
Table -- Summary of Major Content Areas of the National Surveys	
Aboriginal Peoples Survey	
Appendix B -- Counts of Aboriginal Persons in Canada	43
Appendix C -- Methodology of Sample Size Calculations	46



## Preface

This report has been prepared by the Special Surveys Group of Statistics Canada, in response to requests from a number of federal government departments with varying requirements for data on Aboriginal peoples in Canada. Its primary objective is to explore options for the integration of the National Longitudinal Survey of Children, the Survey of Labour and Income Dynamics, and the National Population Health Survey for the collection of data on Aboriginal peoples.

The contact persons for this study within the other federal government departments have been:

Marilyn Willis  
Strategic Policy Group  
Human Resources Development Canada

Kathleen McGovern  
Medical Services Branch  
Health Canada

Angus Dalley  
Management Information and Analysis Branch  
Indian and Northern Affairs Canada

The following persons within Statistics Canada have provided valuable assistance in the preparation of this document:

Gary Catlin and Karen Roberts - National Population Health Survey  
Gilles Montigny and Pat Grainger - National Longitudinal Survey of Children  
Maryanne Webber and Alison Hale - Survey of Labour and Income Dynamics  
Adele Furrie - Post-censal Surveys Program  
Ruta Logan - Aboriginal Peoples Survey  
Sylvie Michaud, Christel Le Petit, Marie France Germain, Sheila Krawchuck - Social Survey Methods Division  
Alain Cregheur - Housing, Family, and Social Statistics Division  
Rick Baxter and Brian Williams - SOD  
Scott Murray, Jeanine Bustros, Pat Foster, Tim Foran - Special Surveys Group

The contact person within the Special Surveys Group for this report is:

Joan Coulter  
Special Surveys Group  
Household Surveys Division  
Statistics Canada  
5 (D7) Jean Talon Building  
Tunney's Pasture  
Ottawa, Ontario  
K1A 0T6  
Telephone: (613) 951-6802  
Facsimile: (613) 951-0562





## 1.0 Background

### 1.1 Origins of the Project

Within Statistics Canada, three new longitudinal surveys are currently being developed to explore a variety of demographic, social, economic, and health-related characteristics of Canadians. These surveys are the National Longitudinal Survey of Children (NLSC), the Survey of Labour and Income Dynamics (SLID), and the National Population Health Survey (NPHS).

For each survey, interest has been expressed by a number of interested parties in the possibility of collecting data on Aboriginal persons. However, cost and design considerations have precluded the collection of data for persons living on-reserve within the main operations of the surveys. As well, current sample sizes will not allow the preparation of estimates for Aboriginal persons living off-reserve. As a result, separate collection strategies need to be considered for Aboriginal peoples.

It is clear that carrying out independent data collection operations on Aboriginal peoples for each of the three surveys is neither feasible nor desirable from the standpoints of cost and respondent burden. Instead, possibilities for integrating the three surveys need to be considered. In addition to facilitating data collection, integration would also provide opportunities for the development of a comprehensive data base of information on Aboriginal persons.

Options for an integrated survey of Aboriginal peoples also need to be considered in relation to a potential 1996 post-censal Aboriginal Peoples Survey (APS). Work has begun within Statistics Canada on assessing the interest in conducting a survey similar to that which followed the 1991 Census. The possible relationships between a 1996 APS and the proposed integrated survey will be discussed throughout this document. The two project teams will continue to work closely on consultation and development activities to ensure coordination and effective use of resources.

It is in response to the issues outlined above that the present integration project was established within the Special Surveys Group of Statistics Canada. Its primary objective is to explore options for the integration of the National Longitudinal Survey of Children, the Survey of Labour and Income Dynamics, and the National Population Health Survey for the collection of data on Aboriginal peoples.

Groups within a number of federal government departments have indicated interest in the development of an integrated survey on Aboriginal peoples. These include:

- Health Canada
- Human Resources Development Canada
- Indian and Northern Affairs Canada

In addition, other federal departments, provincial governments, Aboriginal organizations, and others have varying needs for data on Aboriginal peoples. Consultation with interested parties will be essential in the identification of data needs and the development of the integrated survey. The purpose of the present document is to outline for further discussion the issues and options related to an integrated survey of Aboriginal peoples. Background on the existing surveys is provided in Appendix A to assist in the discussion process.



## 1.2 Project Objectives

The objective of the integration project is to investigate options for integrating the NLSC, SLID, and NPHS for the collection of data on Aboriginal persons. The purposes of integration of the surveys would include the efficient use of common resources, the reduction of sample requirements through integration of the samples, the use of consistent content as appropriate, the reduction of respondent burden and the development of an integrated data base of information on Aboriginal peoples. This integrated data base could cover demographic, social, health, economic, and labour market characteristics. Specifically, the data base would be intended for use in:

- determining the prevalence of various social, economic, and health characteristics and risk factors affecting children and young people;
- determining the prevalence of various social and health characteristics and risk factors affecting adults;
- measuring characteristics related to education, labour market activity, assets and debts, and income for adults; and
- studying the effects of the above factors on child development, adult health, and the economic situation of the family.

The aim of the present document is to outline for further discussion the options and issues related to the development of an integrated survey on Aboriginal peoples. (The integrated survey project will ensure coordination with any other Statistics Canada surveys potentially covering the same population, such as the integrated survey on health and children currently being developed for the Yukon and Northwest Territories.)

The scope of the integrated survey and hence of the data base will be determined during the consultation process. In the interim, it will be assumed to cover the widest scope possible, including the total Aboriginal population, living both on- and off-reserve, in all provinces and territories. However, the implications of this assumption, and alternative options, will be discussed throughout the document as appropriate.

## 2.0 Issues Affecting Survey Design and Feasibility

One of the most basic but often most difficult questions in developing any survey is the determination of the data requirements to be met. Who are the intended users of the data, what are their expected applications, and what types of data are therefore needed? Some users may require data to increase their knowledge of certain situations or events and of the factors influencing them. These users are looking to the survey to help them explore and thereby better understand some aspect of the world. Others wish to use the data to support specific program responsibilities, from developing proposals to implementing programs to evaluating effectiveness. In both cases, data may be used to identify issues which require further study.

Through the consultation process, the requirements of the various potential users of an integrated survey of Aboriginal peoples would be explored and defined. Below is an outline of the key questions which would need to be addressed. It should be noted that none of these questions can be considered in isolation, as each has an impact on the options available for the others. It is only when all of the users' requirements are taken into consideration that the best survey design options can be developed.



## 2.1 Target Population

### 2.1.1 "Aboriginal Peoples"

A basic question to be determined in developing a survey of Aboriginal peoples is defining who is included in "Aboriginal peoples."

The Census of Population of Canada, a self-enumeration survey, is conducted every five years. In the 1991 Census, respondents who received the long questionnaire (approximately 20% of the households in Canada, and all households on Indian reserves) were asked to indicate their cultural origin by responding to the following question:

**"To which cultural group(s) did this person's ancestors belong, for example, North American Indian, Métis, Inuit, English, French?"**

The 1991 Census provided an estimate of 1,002,675 persons in Canada who reported at least one Aboriginal origin (i.e., North American Indian, Métis, or Inuit ancestry) (see Table 1, Appendix B).

The 1991 Census marked the first time that a separate question was asked on whether or not the respondent was a registered Indian as defined by the *Indian Act* of Canada. The inclusion of this question enabled data users to isolate Canada's registered Indian population. The total population reporting Aboriginal origins and/or registered Indian status was 1,016,335.

In 1991, the Aboriginal Peoples Survey (APS) was conducted as a post-censal survey to the 1991 Census. The sample population was selected from the Census respondents who reported Aboriginal origins and/or reported being registered under the *Indian Act*. Survey respondents were then asked with which Aboriginal group they identified (i.e. North American Indian, Métis, or Inuit). In addition, they were again asked if they were registered according to the *Indian Act*. Thus, the main difference between the Census and the APS definitions is that the Census measures Aboriginal origins, while the APS measures those with Aboriginal origins who also identify with their Aboriginal origins. Both the Census and APS included questions on registration under the *Indian Act*. It was the "identity" definition which members of Aboriginal organizations indicated during APS consultation as their population of interest. The APS estimated 625,710 persons in Canada identified with their Aboriginal origins and/or were registered under the *Indian Act* (See Table 3, Appendix B).

#### Issues;

- What definition of "Aboriginal peoples" will be employed in the integrated survey? What questions will be used to identify this population? (Note: The feasibility of some questions may depend on the collection methodology used, particularly self-response vs. interview.)

### 2.1.2 "On- and Off-reserve"

Another definition to be discussed is the distinction between on- and off-reserve populations. It must be determined what is considered to be a "reserve" for the purposes of collection methodology and data analysis. In the 1991 Census and the APS, the term "on-reserve" was used for all persons living on Indian reserves or Indian settlements. Indian reserve refers to land, the legal title to which is vested in Her Majesty, that has been set apart for the use and benefit of an Indian band and that is subject to the terms of the *Indian Act*. Indian settlement refers to places, identified by the federal department of Indian and Northern Affairs Canada (INAC) for statistical purposes only, where a self-



contained group of at least 10 Indian people reside more or less permanently. Indian settlements are usually located on Crown lands under federal or provincial jurisdiction. They have not been set apart for the use and benefit of an Indian band as is the case with Indian reserves. Some Sechelt lands in British Columbia are now referred to as "Indian Government District" (IGD) and some reserves in northern Quebec are now referred to as "terres réservées" (TR). These are also considered, for statistical purposes only, as Indian reserves and settlements.

In the 1991 Census, enumeration was not permitted or was interrupted before it could be completed on some Indian reserves and Indian settlements. Moreover, some Indian reserves and Indian settlements were enumerated late or the quality of the collected data was considered inadequate. There were a total of 78 of these incompletely enumerated Indian reserves and Indian settlements in the 1991 Census. Based on population counts from previous censuses, it is estimated that these incompletely enumerated reserves and settlements represent approximately 38,000 persons. Because the APS sample was selected from the 1991 Census, these 78 reserves and settlements were also not included in the APS. An additional 181 Indian reserves and settlements, representing approximately 20,000 individuals, were incompletely enumerated during the APS because enumeration was not permitted or was interrupted before all questionnaires could be completed. Another 14 Aboriginal communities, representing about 2,000 people, were also incompletely enumerated for the APS. The population estimates provided above from each of the 1991 Census and the APS thus exclude persons on these incompletely enumerated Indian reserves and Indian settlements.

In addition to the legally-defined Indian reserves and settlements, there are other communities which are composed mainly of Aboriginal persons, particularly Métis communities and Inuit communities. Data users might wish to be able to differentiate Métis and Inuit persons living in these communities from those in other areas of Canada. The 1991 APS provided data for these "Aboriginal communities" in their data profiles.

One must also be aware that there are a limited number of Indian reserves and settlements with a large non-Aboriginal population. These areas would likely need to be taken into account in the sample design.

#### **Issues:**

- What definition of "on-reserve" will be employed in the integrated survey for sample design, collection, and analysis purposes?
- Will data be required separately for other "Aboriginal communities"? How will these communities be defined?

### **2.1.3 Definition of the Target Population**

Defining the target population for a survey involves determining the population for which data are desired by the potential users. Having explored the options for defining "Aboriginal peoples" and "reserves," one must determine how the two interact in defining the population of interest.

If the survey were to be restricted to legally-defined Indian reserves and settlements, the population would be composed almost entirely of persons of North American Indian origin. Data from the 1991 Census indicate that less than 20% of the total population of persons reporting Aboriginal origins would be included. Of persons who reported only North American Indian origins on the Census (i.e.





single response to the question on ethnic origin) approximately 48% live on-reserve (See Table 2, Appendix B).

Extending the survey to cover Métis and Inuit communities would add these other component groups of the total Aboriginal population, but would still restrict the survey to communities composed mainly of Aboriginal persons. Coverage of the total population of Aboriginal peoples would require that the survey cover both on- and off-reserve areas.

During consultation prior to the 1991 APS, Aboriginal persons in particular stressed the importance of being able to obtain data not only for the total population of Aboriginal persons, but for the subgroups of North American Indians on- and off-reserve, Métis, and Inuit. Interest has also been expressed in obtaining data for even smaller sub-groups, such as bands or First Nations. Whatever definition is used for "Aboriginal persons," the need to identify and derive data separately for these subgroups will have to be considered in the survey design.

Once the desired target population is defined, the actual population for which it is feasible to collect data, the survey population, must then be determined. Differences between target and survey population could result from a variety of operational restrictions or budget limitations. The potential impacts of the definition of the target population on the survey design options and the overall cost of the survey project will be discussed in greater detail in Section 5.

#### Issues:

- What would be the target population for an integrated survey? Would it include:
  - members of all Aboriginal groups,
  - persons living both on- and off-reserve?
- For what subgroups of Aboriginal peoples will data be desired (e.g. North American Indians, Métis, Inuit, bands, First Nations)?

## 2.2 Cross-sectional versus Longitudinal Data

When a survey is conducted using a representative sample of the population as of date A, it enables estimates to be calculated for current characteristics of the population. These could include the number of persons unemployed on date A, or the number of persons living on-reserve. Estimates of average values, such as the average age of persons in the population, can also be calculated.

When a survey is to be carried out more than once, the door is opened to the collection of two main types of data which allow two very different types of analysis of change. These are cross-sectional data and longitudinal data.

Repeated cross-sectional surveys are based on representative samples of the population at each collection period. They enable the calculation of a series of cross-sectional estimates of characteristics of the population at each point in time, as well as estimates of changes at the aggregate level. For example, the percentage of the population who are unemployed on date A and date B can be estimated, as well as the change in the percentage between the two dates. Cross-sectional surveys are effective and efficient vehicles for monitoring these types of changes at an aggregate level. They incorporate both changes in the occurrence of the characteristic itself and changes in the composition of the population. However, they are limited in their ability to allow analysis of the factors affecting these changes. For example, the surveys may indicate that the



percentage of persons unemployed remained stable between date A and date B; however, one does not know if the same people were unemployed for the entire period or if large numbers of people moved into and out of unemployment.

Social and economic researchers frequently want to focus on the change or stability experienced by individuals, in order to better understand the causes and factors affecting these events. Retrospective questions on cross-sectional surveys (i.e. questions asking about events during a specified period in the past) can provide some indications of individual changes over time, but they are subject to errors in recall. Longitudinal surveys, on the other hand, are designed for analysis of individual change.

Longitudinal surveys are based on a sample of the population which is selected at date A and followed over time. By following the same respondents, longitudinal surveys record important events in a person's life. They allow analysis of the links between changes in demographic, social, economic, or other characteristics, and the factors influencing them. For example, a longitudinal survey could study how persons move into and out of the state of unemployment, and the relationships between these events and other characteristics.

The sample for a longitudinal survey is representative of the total population when selected, but loses its representativeness over time. As a result, the ability to produce cross-sectional estimates also decreases over time. If cross-sectional estimates are required in addition to longitudinal data, the sample must be supplemented at each cycle to account for additions and changes in the population.

Related to the type of data required is the question of the frequency of data collection which will be most appropriate. A key factor in the decision is the stability of the characteristics to be measured, or the expected frequency of significant changes, for both Aboriginal peoples and the total population. In the case of a longitudinal survey, one must also consider the ability of the respondent to recall information for the intervening period. Requiring respondents to provide information for too long a period in the past will affect the accuracy of the data obtained. Survey designs which involve rotation of content areas from cycle to cycle will also impact on length of time between repetitions of questions, and hence on data quality.

The three national surveys, NLSC, SLID, and NPHS, have been designed to provide longitudinal data with some cross-sectional data also available. (Samples are to be augmented as required to ensure the representativeness of the ongoing sample for production of cross-sectional estimates.) As well, sample sizes are large enough to account for expected attrition of the longitudinal sample due to difficulties in tracing or non-response. Sample designs employ techniques such as rotating panels to maximize the benefits of the cross-sectional and longitudinal aspects of the surveys while minimizing the development and collection costs.

The three surveys have differing plans and schedules for their cross-sectional and longitudinal components. Below is an outline of the current plans for the survey cycles. (Note that this chart indicates only whether or not a survey will provide data for a particular year; it does not reflect the level of activity occurring during the year. For example, the NPHS carries out collection four times in each collection year, SLID twice per year.)



Figure 1 – Current Plans for Survey Cycles

Year	Cross-sectional	Longitudinal
1994	NPHS SLID NLSC	NPHS SLID NLSC
1995	SLID *	SLID
1996	NPHS SLID * NLSC	NPHS SLID NLSC
1997	SLID	SLID
1998	NPHS SLID * NLSC	NPHS SLID NLSC

- \* SLID plans to produce a cross-sectional data file each year, but the focus of the survey is longitudinal data.

Determining whether a cross-sectional survey, longitudinal survey, or combination of the two would be most appropriate for an integrated survey of Aboriginal peoples will depend on the types of changes over time which the users wish to study. In addition, implications for survey designs will need to be considered, since cross-sectional and longitudinal surveys facilitate differing types of designs which have their own strengths and weaknesses.

Given that the integrated survey project was initiated in conjunction with the new Statistics Canada longitudinal surveys, it is assumed that some interest in longitudinal data exists. However, it is expected that both cross-sectional and longitudinal estimates will be of interest for a variety of data applications. The implications of producing cross-sectional and longitudinal estimates on the cost and complexity of an integrated survey will need to be explored in detail. One possibility would be to produce cross-sectional estimates on a regular basis (e.g. every 5 years, every 10 years) and carry out a longitudinal survey of a sub-sample of respondents in the intervening periods. For example, a detailed cross-sectional survey such as the Aboriginal Peoples Survey (conducted in 1991) could be carried out on a regular basis, with an integrated survey of Aboriginal peoples following a sub-sample of respondents on a longitudinal basis.

**Issues:**

- What are the requirements for cross-sectional and/or longitudinal estimates?
- How frequently are data required or desired? (This will depend on the frequency and significance of expected changes in the characteristics of interest.)
- If longitudinal data are desired, what would be the preferred length of time of the panel, taking into account periods over which characteristics of interest are expected to change, burden on respondents of repeated data collection, and difficulties of tracing respondents?



- Will the data from the integrated survey be comparable to the main surveys?
- What will be the impacts of cross-sectional vs. longitudinal estimates on survey costs, including:
  - complexity of sample design,
  - sample size required,
  - costs for tracing respondents, and
  - other collection costs, such as interviewer training.

## 2.3 Data Requirements

The geographic areas or population subgroups for which data are desired will vary according to the mandate and application of the user. Any requirements to derive estimates or analyze trends below the level of the total population will have implications on the survey design and methodology. As well, in general the smaller the population for which data are to be produced, the larger the sample size required and hence the greater the survey cost.

### 2.3.1 Geographic Level

With respect to the geographic level of data, differences in the circumstances of Aboriginal peoples living on- and off-reserve would likely be of interest, particularly in making comparisons with data for the rest of Canada. Data for areas below the national level might also be desirable for analyzing regional differences. Aboriginal organizations may be particularly interested in obtaining data for small areas, such as reserves or bands. The ability to provide some information for small areas may be important in obtaining the support and participation of Aboriginal peoples. For example, the Aboriginal Peoples Survey provided data profiles for participating reserves and settlements to respond to needs identified by Aboriginal organizations during the consultation process.

The issue of geographic level for which data are required cannot be considered in isolation from the question of cross-sectional versus longitudinal data. If the main focus of the survey were to collect longitudinal data, emphasis would shift from the production of estimates to analysis of relationships among characteristics and changes over time. Without the requirement for point-in-time estimates, data for larger geographic areas or groups of similar communities might be sufficient. On the other hand, data on changes over time might be of great interest for communities or other small areas. In general, the focus in discussing longitudinal data needs should be on trends and relationships rather than geographic detail.

Certain types of longitudinal survey designs could be considered which would provide data for some small areas. For example, the feasibility study for conducting SLID on reserves proposed a sample allocation which would yield data by reserve for selected reserves (Michaud, Lavigne, & Webber, 1992). Since cost factors would preclude sampling in each small area at each survey wave, the survey would need to be designed in order that inferences could be made for populations or areas not included in the sample. In addition, the potential for complaints of favouritism towards communities included in the sample might require rotation or other means to vary the communities selected.

### 2.3.2 Population Subgroups

The design for an integrated survey will also be influenced by the population subgroups which users wish to analyze. Within the total population of Aboriginal persons, users may be interested in





obtaining data separately for North American Indians, Métis, and Inuit or for bands, First Nations, or other groups within these broad categories. In some cases subpopulations may be defined according to geographic boundaries (e.g. groupings of reserves or communities), while in others they may require specific questions to respondents (e.g. band or First Nation membership).

Data may also be desired for subgroups defined by demographic characteristics. For example, are data needed on gender groups (i.e. male/female) or age groups? Given the focus of the NLSC on children, it is assumed that data will be desired at the very least for children versus adults. As with the geographic level, requirements for data for small population groups will increase the size of the sample needed, and hence must be balanced against cost constraints.

**Issues:**

- For what geographic level are estimates required or desired?
- What are the links between type of data required (i.e. cross-sectional vs. longitudinal data) and the most appropriate level of geographic detail?
- Would a commitment to provide data for reserves or other small areas increase the support and participation of Aboriginal peoples in the survey?
- If the sample were to be clustered to reduce costs and improve efficiency, how could complaints of favouritism towards particular communities or groups be avoided?
- For what subgroups of the population are data required or desired (e.g. Aboriginal groups/bands/First Nations, gender, age groups)?

## **2.4 Content Development**

### **2.4.1 Content Considerations for an Integrated Survey of Aboriginal Peoples**

Content for each of the NLSC, SLID, and NPHS were developed to meet the needs of each survey's potential data users. Some content areas, such as basic demographic information, are common across the three surveys. Other topics are areas of particular focus of only one survey. It should be noted that even in topics which are common to all three surveys, major differences may exist in the actual content covered. Differences in aspects such as question wording, reference period, and collection method may also have significant impacts on comparability. (An overview of the content of each of the NLSC, SLID, and NPHS is provided in Appendix A).

Some work is in progress to integrate the content of the NPHS with the NLSC to ensure consistency across the two surveys. It will also be necessary to consider the content of the 1991 APS and the potential 1996 APS (see Section 3.2 Survey Design Options).

The content of the integrated survey must be determined in consultation with Aboriginal peoples and the many parties interested in obtaining data on Aboriginal peoples. (The consultation process itself is discussed in greater detail in Section 5 below.) Policy considerations for federal departments may include data requirements for program planning, implementation, and evaluation. The data interests of all potential users will need to be balanced against issues such as cost and respondent burden.



#### 2.4.2 The Requirement for Integration of Content

From a statistical point of view, it would be ideal to have all three surveys (i.e. NLSC, SLID, and NPHS) administered to the same sample of Aboriginal respondents. This would provide direct data on the nature, extent, and direction of relationships between variables collected on the three surveys; that is, between characteristics related to health, child development, and labour force activity and income. However, it is unlikely that an integrated survey of Aboriginal peoples could consider including all questions from the NLSC, NPHS, and SLID in a single data collection. Not only would costs be high, but the burden on respondents would be unacceptable. Currently, the average length of interview is estimated to be two hours for the NLSC, one hour for the NPHS, and fifty minutes in total for the labour and income components of SLID, which would give a total of close to four hours on average per household. Even if duplicate questions were eliminated, the total interview length would still be considered excessive.

Other options could be considered for collecting the full content of the three surveys from all respondents. For example, the interview could be broken into several components to be carried out at different times of the year. This could significantly increase collection costs, particularly if personal interviews were required for each component. An alternate option would be to ask certain base questions with each collection, and rotate the themes of health, children, and labour force and income. Given current plans for SLID to be conducted annually and NLSC and NPHS every two years, and the expected shorter length of the SLID interview, a schedule such as the following could be considered:

- e.g. Collection 1 -- base data + labour + children
- 2 -- base data + labour + health
- 3 -- base data + labour + children
- 4 -- base data + labour + health
- etc.

Annual data collection would be very costly on an ongoing basis, and any less frequently would mean that data on health or children would be collected only once every four years or more. The time difference between collections could significantly reduce the ability to analyze relationships between variables from the components on health and children.

If the total content of the three surveys could not be administered to all survey respondents, another alternative would be to collect data on each of children, health, and labour force and income from different respondents in overlapping geographic areas. This would allow indirect inference of relationships between variables. Sample sizes would be larger since there would be, in effect, separate samples for each of the three components; however, collection costs would not be three times as large due to the geographic clustering of the sample. This approach would require that the variations between respondents within communities be smaller than those between respondents in different communities, an assumption which would need to be confirmed prior to survey design.

A third alternative, which combines the advantages of the above options while avoiding some of the disadvantages, would be to integrate and condense the content of the NLSC, SLID, and NPHS into one questionnaire administered to all respondents. The joint collection would enable the analysis of relationships between variables, while the reduced content would decrease collection costs and respondent burden. This option would require examining the means for combining and reducing the



total content of the survey, while still maintaining the maximum comparability possible with the three national surveys.

It is expected that some content from the three national surveys may need to be combined or eliminated, particularly if additional content relevant to Aboriginal peoples is added. However, the potential alternatives described above indicate how an effective sample design could enable the maximum data to be obtained from the survey while maintaining acceptable collection costs. Exploration of these and other options in conjunction with consideration of other design issues would be required before the best survey design options could be determined.

#### **2.4.3 Working Towards a Consensus**

The integration project is dependent on the willingness of the parties involved in the three surveys to work towards a consensus in defining the content of the integrated survey. This willingness exists already within Statistics Canada, and the managers of the three surveys have agreed to form, at the appropriate time, a working group on the content of the integrated survey. This group would work in parallel with the three surveys, aiming to keep costs for collection of the integrated survey as low as possible while maximizing the comparability of the data on Aboriginal peoples with each of the national surveys. Analysis of data from the first collection cycles of each of the three surveys will assist in identifying correlated variables, and thereby determining the minimum content requirements of the integrated survey.

In the case of the NPHS and NLSC, a working group has already been established to define the minimum content requirements for the collection of data for these two surveys in the Yukon and Northwest Territories. As well, the two surveys are in the process of defining a plan for integration in the case that the respondent to the NPHS is a child of less than 12 years. The output of this work could serve as a point of departure for the development of an integrated survey of Aboriginal peoples.

#### **2.4.4 Other Content-related Issues**

One particular issue to be considered in integrating questions from the three surveys will be differences in reference period and time of collection. For example, the reference period for labour and income data in SLID is the previous calendar year, whereas the NLSC collects information during the autumn on labour force activity during the past two years. Compromises may be necessary in determining questions for the integrated survey which are relevant to the time of collection and the situation of the respondents, while providing the best possible comparability with the national data.

Questions may also need to be revised or developed in order to be relevant to the population of Aboriginal peoples. For example, some questions may not be appropriate for persons living on-reserve or in isolated communities. Others may need to be reworded to be relevant to the situation or experience of the respondents and thereby to facilitate accurate responses. Additional questions may also be desirable to explore in greater detail issues of particular concern for Aboriginal peoples. Different questionnaires for persons living on- and off-reserve may be found to be preferable from data quality and collection methodology standpoints. It must be kept in mind, however, that any development or revision of content specifically for a survey of Aboriginal peoples will need to be balanced against the need for comparability with estimates from the national surveys.



As with any data collection undertaking, it must also be ensured that the content of the questionnaire is appropriate for the survey vehicle and the collection method being employed. Certain types of questions are feasible only in a person-to-person interview; others are better suited to self-enumeration by the respondent. Some topics of interest may require specialized expertise in observation or measurement which an interviewer could not provide.

**Issues:**

- Consistency with questions from the three national surveys would allow estimates from the integrated survey of Aboriginal peoples to be compared with those for the total population of Canada. Survey design options which enable the maximum possible content to be retained from all three surveys could be considered. Compromises will need to be made in determining any questions to be eliminated from the three national surveys.
- Consistency with the national surveys will be affected by any differences in time or method of collection and reference period.
- Questions may need to be revised to be relevant to the target population of Aboriginal peoples; this would need to be balanced against maintaining consistency with the national surveys.
- Additional content related to issues of specific concern to Aboriginal peoples may also be of interest.
- Different content or question wording for persons on- and off-reserve or in remote areas may need to be considered to ensure relevance of the questions to the respondents.
- All content must be appropriate for collection by the survey vehicle.
- New or revised questions will require testing, some in remote or other areas where field costs would be higher than usual.

## **2.5 Data Collection Methods**

As indicated in the overviews of the three longitudinal surveys in Appendix A, each will employ a variety of collection methods for the first and subsequent collection cycles. The best method for each survey component depends on factors such as length and complexity of the questions, need to access specific documents or records, feasibility of proxy reporting by another household member, or use of direct observation by the interviewer. Beyond the direct costs for collection, the methodology selected can also impact on other key issues such as data quality or scheduling.

Where differences in collection method exist among the national surveys, the best method for an integrated survey will need to be determined, taking into account all the factors as outlined above. In some cases, alternate methods may be required to meet specific needs for the population of Aboriginal peoples. The impact on data comparability of any differences from the method used in the national surveys will need to be assessed.

Given the length and complexity of the content of the three national surveys, it is expected that the questionnaire for an integrated survey would also be lengthy and complex. It is hence unlikely that it could be collected by telephone interview, or that the questionnaire could be completed independently by the respondent. It might also be difficult to determine in advance the language in which the respondent would find the survey questions easiest to understand. As a result, it is recommended that collection of at least the initial cycle be carried out by personal interview in the home of the respondent. In subsequent cycles with repeat members of the panel, it is possible that





telephone interviewing would be feasible. In this case, the availability of a telephone on a private line could be an issue for some respondents, particularly those living in isolated areas.

Other collection-related issues may also be different for respondents living on- and off-reserve, and may require variations in collection method. Experienced interviewers may not be available on many reserves, which will increase the costs for staff training. Travel costs will also be higher for collection in isolated areas. The experience of the 1991 Aboriginal Peoples Survey and the Census will be valuable in determining the expected costs and any potential difficulties in collection on reserves and in remote areas.

**Issues:**

- The best collection method for each component of an integrated survey would need to be determined, taking into account issues such as cost, timing, and impact on data quality.
- Differences from the collection methods of the national surveys would be expected, due to the integration into a single survey as well as specific needs of the Aboriginal population. The impact of the differences on data comparability will need to be assessed.
- It is expected that personal interviews will be required for the first cycle of the survey, and possibly for all subsequent cycles for at least some proportion of respondents. This will have a significant impact on the collection costs of the survey.
- Variations in collection methods may be required for respondents living on-reserve or in remote areas. Higher collection costs may also be expected for this portion of the sample.

## **2.6 Collection Tools**

The design of the three national surveys has been based on the use of computer-assisted interviewing for data collection. Interviews will be conducted by telephone or in person or with a combination of the two methods. For an integrated survey of Aboriginal persons, the question of computer-assisted interviewing versus paper-based questionnaires raises numerous issues. Given that computer-assisted interviewing is at the present time just being phased in for the Labour Force Survey (LFS), sufficient experience does not yet exist to fully assess some of these issues. As a result, the following discussion will simply identify some of the aspects which would need to be considered in selecting the most appropriate collection tool.

Computer-assisted interviewing requires access to the equipment. Currently, the computers are used mainly for the LFS and supplements to the LFS, and as a result are in high demand. The total time for collection of the integrated survey would include transport of the machines to the interviewers (some in isolated areas), interviewer training, and completion of interviews. Given the length of time involved, and the complete unavailability of the computers during this period for any other survey, it appears impossible to use the LFS computers. The purchase of additional machines specifically for the integrated survey would be a significant investment for the project. Options such as staggered or rolling collection periods could be considered to reduce these costs, but their impact on other aspects such as data comparability would need to be assessed. Other logistical issues would also need to be considered. For example, the computers operate on batteries which must be recharged after several hours of use. Access to electricity or supplies of batteries would need to be ensured in all collection sites.

Another issue to be considered is interviewer training. The national surveys will be collected by interviewers from the LFS who are accustomed to computer-assisted interviewing. They will require



training only in the specific questions for each survey. For a survey of Aboriginal peoples, many Aboriginal interviewers would be hired who have no previous experience with computer-assisted interviewing. Experience from the LFS will provide some indications of whether training new interviewers in computer-assisted interviewing is more or less expensive than training them in paper-based interviewing. One factor would be the complexity of the questionnaire, and the need to use reference materials or to move from one section to another based on screening questions. Computer-assisted interviewing provides automatic switching between screens and on-line access to manuals, but system development costs would be higher for more complex questionnaires.

Use of computers would introduce costs for systems development and programming, but eliminate costs for printing training manuals and other tools. It would also eliminate the need for an additional data capture stage. Computer-assisted interviewing allows on-line checks of reasonableness and consistency of responses which could increase data quality and decrease the costs for later editing of data files. However, complexity of on-line editing systems must be balanced against systems development costs.

Any requirements to translate the questionnaire into Aboriginal languages could raise additional issues, such as computer memory needed to store multiple versions of the questionnaire, and difficulties displaying the characters used in written forms of some Aboriginal languages.

The above discussion highlights the fact that the choice of computer versus paper-based questionnaires can affect many aspects of a survey from cost and timing to data quality. Experience from current implementation of computer-assisted interviewing in the LFS, as well as past collection of the APS and Census on- and off-reserve, will be valuable in studying potential options and impacts. Investigation of each of the issues raised above would be necessary before the best option for an integrated survey of Aboriginal people could be selected.

#### Issues:

- Should an integrated survey of Aboriginal peoples be collected using computer-assisted interviewing or paper questionnaires, taking into consideration potential impacts on:
  - availability of collection tools (e.g. computers, batteries),
  - ability to provide materials in multiple Aboriginal languages,
  - training of interviewers,
  - data quality,
  - data capture and editing,
  - overall cost and timing of survey?

## 2.7 Collection Periods

The collection periods of each of the three national surveys have been determined taking into consideration issues such as availability of the respondent, reference periods of the questions, and access to sources of data (e.g. annual pay statements). At the present time, collection is planned as follows:

NLSC -- December 1994/January 1995

SLID -- January and May

NPHS -- May, August, November, and February



The NLSC and NPHS will be carried out every two years, while SLID will be conducted on an annual basis. Given that an integrated survey will incorporate content from each of the three surveys, it will need to be determined what collection period will maximize the overall quality of data which could be obtained for all content areas.

An additional complicating factor is the fact that some Aboriginal peoples may be difficult to contact at certain times of the year. Some reserves are difficult to access during the winter. In other areas many people leave the community for hunting, fishing, or jobs in other locations during particular seasons. Young people may be studying at schools or post-secondary institutions at a distance from their homes. Thus, it may be difficult to select one collection period which would maximize the probability of successfully contacting all respondents. However, given the possible need to consider staggered collection periods due to limited inventories of collection tools, it may be possible to schedule collection periods which address availability of both respondents and tools. Collection costs may also be affected by collection period due to travel costs during certain seasons.

The survey design itself may also impose certain restrictions on the collection period of the integrated survey. For example, if the sample selection is linked to the 1996 Census, the date that the required input from the Census will be available must be considered.

**Issues:**

- A collection period must be selected which takes into consideration the survey content, availability of respondents, collection costs and availability of collection tools, and links with other surveys. Differing periods may be required for certain segments of the survey population.
- Any links between the integrated survey and the 1996 Census (e.g. for sample selection) will impact on the collection period.
- Impacts of the choice of collection period on collection costs must be balanced against impacts on response rates and data quality.
- Differing collection periods will increase the complexity of field operations, and may increase collection costs

### **3.0 Survey Design Options and Issues**

In order to develop the optimum sample design for a survey, one must take into consideration all the other aspects of the survey as described above. Objectives and restrictions with respect to data requirements, content, collection methods, and all other survey components must be analyzed, and the best means for integrating them determined. Given the interactions among these components, this normally requires a process of balancing advantages and disadvantages of each option, and identifying means for maximizing benefits and minimizing compromises.

At the present time there are numerous outstanding issues regarding all aspects of an integrated survey of Aboriginal peoples. Until these issues, as raised in preceding sections of this paper, have been addressed, the input to the survey design process is incomplete. As a result, it is difficult to present specifics at this time for survey design options. Instead, the options will be explored in general terms, and the main issues surrounding each option will be discussed.

An important point to be considered in evaluating potential sample designs for an integrated survey are the differences in sample design among the NLSC, SLID, and NPHS as described above.



Variations in aspects such as panel structure and overlap or time in the sample could have major impacts on the comparability of the data. Thus, reconciling these differences, and finding the design which best meets the needs of the integrated survey plus the requirements for comparability with the national surveys, will be a significant challenge.

In order to provide an indication of potential sample sizes, examples are given for certain survey options under particular sets of assumptions of data requirements.

### **3.1 Impacts of Data Requirements**

Decisions that are made regarding the level of detail and the type of information that are required will have a direct impact on the nature and design of the survey. The issues to be considered and their impact on survey design options are outlined below. In order to enable the calculation of sample sizes for particular survey options, a number of assumptions are made regarding coverage and definition of the population of interest, and the geographic level for which data will be required.

#### **3.1.1 Cross-sectional versus Longitudinal Data**

The decision as to whether the integrated survey is to provide cross-sectional or longitudinal data or a combination of the two will have a major impact on the sample design required and its complexity. Looking at the question from the opposite perspective, these types of surveys each facilitate different types of survey designs with their own advantages and disadvantages. The question of data type will need to be considered as part of the survey design process. For the survey design discussion below, it will be assumed that the survey is to be longitudinal, with the ability to produce cross-sectional estimates in its initial cycle.

#### **3.1.2 Coverage of On- and Off-reserve Populations**

As was described in the Target Population section above, the question of coverage of the on- and off-reserve populations of Aboriginal peoples could be summarized into three main options:

- 1) coverage of persons living on Indian reserves and settlements only
- 2) coverage of persons living on Indian reserves and settlements as well as Métis and Inuit communities
- 3) inclusion of all Aboriginal persons living both on- and off-reserve

Given the links between registered Indian status and residence on-reserve, the first option would provide a survey population composed almost entirely of North American Indians. (The small number of Indian reserves and settlements with a large non-Aboriginal population would need to be taken into account in the survey design.) Under the second option the survey would include persons of North American Indian, Métis, and Inuit background, but only those living in communities composed mainly of Aboriginal persons. The final option would include all persons of Aboriginal origin in Canada.

The implications for the survey of collecting data for persons living on- and off-reserve have been discussed in each of the sections above. However, it is in the design stage that the impact of the decision will be most evident, as all aspects of the survey are brought together in the development of the optimum survey design. Some design options are appropriate or advantageous for only one population or another; others could be applied to any population, but with differing costs and





benefits. A combination of options might be considered as the means to best meet varying requirements of the survey.

Given its impact on all aspects of the survey design, the choice of target population will have a major effect on the overall cost of the survey. As a result, it may not be a question of selecting the preferred target population, but rather of determining the population for which it is feasible to carry out the survey. In other words, the cost and feasibility of carrying out the survey on- and off-reserve under varying design options may determine the target population which can be considered for the survey.

Given these links between target population and survey design, it is important to consider each design option as it applies to each potential population. In the discussion below, sample frames and designs will be discussed separately for each of the on- and off-reserve populations in any cases in which they raise differing survey options.

### **3.1.3 Definition of the Population of Interest**

In addition to the question of coverage of on- and off-reserve populations, the definition of "Aboriginal peoples" must also be considered. The major existing options are the Census definition of persons with Aboriginal origins and/or Indian Registration and the APS definition of persons who identify with those origins and/or are Registered Indians. Alternative definitions could also be developed, but would require testing for their applicability under the survey methodology. For the purposes of the discussion below, the APS definition will be used; since this is the more restrictive definition, it could lead to greater cost in selecting a sample with this characteristic.

### **3.1.4 Levels for which Data are Required**

In order to calculate sample sizes for survey design options, one must determine the population subgroups for which estimates will be required.

It is expected that some geographic areas will be of interest, whether provinces and/or regions, or smaller areas. Estimates for persons living on- and off-reserve are also expected to be required, whatever definition of "reserve" is selected. Other groupings based on geography could also be considered, such as residents of north vs. south regions, or isolated vs. non-isolated communities. For all geographic areas, clear definitions will need to be developed.

Separate estimates would also be expected to be required for subgroups of the population defined by demographic characteristics. Given the relationships between age and topic areas such as child development, health, and economic status, it is expected that estimates will be required for specified age groups. Consistency with age groups used in the national surveys will need to be considered if data are to be comparable. Distinction by gender may also be important for certain content areas.

Identification of North American Indians, Métis, and Inuit within the total population of Aboriginal persons would be expected to be required. In addition, other subpopulations within these large groups may be of interest. For example, data users might be interested in separating members of particular First Nations or language groups.



As described above, any requirements to derive estimates or analyze trends below the level of the total population will have implications on the survey design and methodology. As well, in general the smaller the population for which data are to be produced, the larger the sample size required and hence the greater the survey cost. Thus, the data requirements will need to be balanced against the survey complexity and cost.

For the survey design discussion below, the following subgroups have been considered in calculating potential sample sizes for the survey:

- provinces, or regions for areas in which the target population is very small (Regions have been defined as:
  - Atlantic (Newfoundland, Prince Edward Island, Nova Scotia, New Brunswick),
  - Quebec,
  - Ontario,
  - Prairies (Manitoba, Saskatchewan, Alberta),
  - British Columbia, and
  - Territories (Yukon, Northwest Territories)
- age groups (0-1, 2-5, 6-11, 12-64, 65+), (0-11, 12-64, 65+) and (0-11, 12-19, 20+)
- Aboriginal groups -- North American Indians, Métis, Inuit
- on- and off-reserve populations

It should be noted that these subgroups are by no means recommendations of data requirements for the survey. Data needs and the resulting sample requirements will be determined through consultation with the potential users.

## **3.2 Survey Design Options**

### **3.2.1 1996 Post-censal Survey**

#### **3.2.1.1 Sampling Frame**

An integrated survey could be conducted as a post-censal survey following the 1996 Census of Population. The survey would be carried out in the autumn of 1996 using the census data to identify the target population. Data from the census would also be linked with the survey database to supplement the information collected in the survey. The experience of the Aboriginal Peoples Survey (APS), conducted as a post-censal survey in 1991, would be invaluable in developing the integrated survey under this scenario. It would be assumed that the survey would be used to provide benchmark data (i.e. cross-sectional estimates) as well as some estimates of change. In this section the design to produce cross-sectional estimates is discussed. The Follow-up to a Post-censal survey (section 3.2.2) will provide information on the design of a longitudinal component, which is used to estimate changes.

#### **3.2.1.2 Survey Design and Allocation**

Data from the census would be used to identify Aboriginal peoples and to control for age groups and other population subgroups of interest (i.e. to ensure that the sample includes sufficient people in each of these subgroups of interest to enable estimates to be produced). Allocation of the sample on- and off-reserves will be discussed separately.



#### On-reserve:

Two options for the sample allocation are available. For the purposes of discussion, it will be assumed that a province is a stratum (i.e. level) for which it is desirable to be able to produce some cross-sectional estimates. There are a number of reserves in that province.

#### Option 1:

The first option assumes that a simple random sample would be selected from across that stratum (i.e. the province). This implies that, on average, most reserves would have at least some people interviewed. However, since the sample would be spread across all the reserves, most reserves would not have a sample big enough to produce cross-sectional estimates at the reserve level. Every reserve would have a small number of individuals that would be followed over time, so each reserve would be able to see how their sample is evolving. It is possible that one could use other data sources (such as census data) to make some inferences at the reserve level. In general, it would be costly to go to all reserves and in very small reserves, only a very small number of people would be interviewed. However, overall estimates at the stratum level would be better than with option 2. Since Option 1 is close to a random sample, a design effect of just 1.8 is used for the calculation of sample sizes shown below.

#### Option 2:

An alternative would be to use a two stage sample design. The underlying assumption being that one could group similar reserves together, and then only select some reserves in each group. In the selected reserves a sample would be selected, which would be large enough to produce estimates at the reserve level. For the non-sampled reserves, one would have to assume that the behaviour observed in the selected reserves could be applied to the non-sampled ones. The advantage of this option is that it might be easier to get the collaboration of the selected reserves since they could obtain cross-sectional estimates from the first interview. As well, collection cost per interview could be lower since interviews would be clustered on fewer reserves. Disadvantages are that it may be difficult to group the reserves in homogeneous groups, and that for the non-sampled reserves, no information at all would be available. Also, since this allocation is further from a random sample than option 1, a larger sample size would be necessary to get reliable estimates at the stratum level. This constraint is expressed by the larger design effect. For the example of sample sizes given below, a design effect of 3 is used compared to a value of 1.8 for option 1.

#### Off-reserve:

For the off-reserve component, either of the above two options could be used. A simple random sample with an allocation like option 1, where the sample is spread across the whole stratum, could be used. If it were necessary to get estimates for certain specific areas with a high density of Aboriginal peoples, the sample could be concentrated in some selected areas, as in option 2. The choice of allocation would have an impact on the sample size. The more the sample is spread evenly across the stratum, the smaller the design effect and, therefore, the smaller the sample has to be for reliable estimates. However, collection cost per interview would increase the more the sample is spread, largely due to costs associated with travel and training.



### 3.2.1.3 Adjustment for Non-response

An adjustment for non-response has been made to the sample sizes given below, assuming a response rate of 80%. This is based on the response rate to the Aboriginal Peoples Survey, which was a supplement to the 1991 Census.

### 3.2.1.4 Sample Size Examples:

Below are four tables providing sample sizes for on- and off-reserve samples, assuming estimates are required by specified geographic areas and age groups, for each of the two options for allocation described above. Details regarding the mathematical formula used to calculate sample sizes are provided in Appendix C, Section 1.

Sample size estimates with allocation option 1 - Simple random sample across all reserves

#### On-reserve Population

Table 1a

On-reserves -- Target CV: 16.6%			
	Geography level for estimates		
Age groups for estimates	Province	Region <sup>(*)</sup>	Total Canada only
0-1, 2-5, 6-11, 12-64, 65+	22 055	14 820	3 530
0-11, 12-64, 65+	14 775	9 650	2 135
0-11, 12-19, 20+	16 300	10 485	2 175
Total Population only	6 540	3 915	586

(\*) as defined in Section 3.1.4

#### Off-reserve Population

Table 1b

Off-reserves -- Target CV: 16.6%			
	Geography level for estimates		
Age groups for estimates	Province	Region	Total Canada only
0-1, 2-5, 6-11, 12-64, 65+	29 410	18 845	3 610
0-11, 12-64, 65+	18 980	11 595	2 165
0-11, 12-19, 20+	21 110	12 575	2 195
Total population only	7 925	4 345	587





Sample size estimates with allocation option 2 - Two stage sample with selection of reserves from groupings of similar reserves

**On-reserve Population**

Table 2a

On-reserves -- Target CV: 16.6%			
	Geography level for estimates		
Age groups for estimates	Province	Region	Total Canada only
0-1, 2-5, 6-11, 12-64, 65+	31 810	21 960	9 835
0-11, 12-64, 65+	21 905	14 715	3 490
0-11, 12-19, 20+	24 660	16 320	3 600
Total Population only	10 125	6 265	974

**Off-reserve Population**

Table 2b

Off-reserves -- Target CV: 16.6%			
	Geography level for estimates		
Age groups for estimates	Province	Region	Total Canada only
0-1, 2-5, 6-11, 12-64, 65+	44 090	29 175	5 940
0-11, 12-64, 65+	28 960	18 285	3 570
0-11, 12-19, 20+	32 740	20 320	3 645
Total population only	12 675	7 175	978

**3.2.1.5 Issues Regarding a Post-censal Survey**

- Data collected from the census could be used to supplement the survey database, and improve robustness.
- Selection from the census will eliminate the need for pre-screening to identify Aboriginal peoples since they will have been previously identified from census data, therefore reducing cost.
- Using census data could also allow more control of how the sample is allocated, for age groups or other population of interest, again reducing sample size and cost.
- If another Aboriginal Peoples Survey similar to the one carried out in 1991 is to be conducted as a post-censal survey in 1996, the two projects would need to be coordinated to ensure maximum cost effectiveness of joint efforts and minimum respondent burden.



### 3.2.2 Follow-up to a Post-censal Survey

#### 3.2.2.1 Sampling frame

For this option, it is assumed that cross-sectional estimates are provided by a post-censal survey, and that the integrated survey is designed to provide estimates of change. The post-censal survey can either be the first wave of the integrated survey, a post-censal survey similar to the one which followed the 1991 census, or a combination of the two. A sub-sample of the post-censal survey on Aboriginal peoples could be selected, and followed over time.

#### 3.2.2.2 Survey Design and Allocation

Data from the post-censal survey would be used for the selection of the sub-sample. Since detailed information would be available for the selection, it would be possible to focus the sample on specific groups of interest identified during the consultation process (eg. age groups, Aboriginal groups), for a minimal cost. The sample could be allocated based on the groups of interest identified. Groups could also be defined using the on- and off-reserve criteria. Since the same methodology could be applied to both the on- and off-reserve components of the sample, the allocation is not discussed separately by on- and off-reserve for this option.

With this option, the goal of the survey would be longitudinal, and the data could not be used to produce cross-sectional estimates. It would be assumed that the cross-sectional estimates would be obtained from the post-censal survey. Instead, one would be able to carry out longitudinal analysis looking at individual changes (i.e. are the same people employed at each wave, or are there many new people finding jobs while others are leaving the labour market?) Only a longitudinal survey can answer these questions.

A longitudinal survey is also designed to measure overall total population changes. For example, a longitudinal survey can effectively measure whether the employment rate increased between two survey periods. Since it is the same respondents in the two waves, the two employment rate measures are correlated. As a result, a smaller sample size can identify a change in the employment rate with a longitudinal survey than with two independent cross-sectional surveys.

The sample size necessary to be able to identify changes is a function of four factors: the magnitude of the change one wants to identify (do you want to identify a 5% or 10% difference in a proportion?), the precision required (do you allow a 5% or 10% error rate in the measure?), the stability of the measure and the level for which inferences have to be made. If one only wants to evaluate some changes for a reserve or a specific target group, then one doesn't need such a large sample. That small sample can then be used to study the impact of different factors, and, if it is appropriate, conclusions can be extended to other similar reserves, or target groups. On the other hand, if one wants to be able to make inferences for many different target groups, then the sample has to be much larger.

#### 3.2.2.3 Adjustment for Expected Attrition of the Panel

With a longitudinal survey, a certain amount of attrition, or erosion, of the sample is anticipated due to non-response and difficulty tracing respondents over time. The following attrition rates to the first three waves of the longitudinal panel are assumed: 80%, 85% and 90%, which leaves 61% of the



initial sample after 3 waves. These rates are based on the experience of the APS; however, given the lack of experience with longitudinal surveys these rates are estimates. Sample sizes given below are adjusted for the expected attrition.

### 3.2.2.4 Sample Size Examples

Given below in table 3 are some sample size examples, assuming one wants to identify changes with 95% confidence (i.e. 5% error rate), for two different levels of change (5% and 10%) and two types of measures, one stable (correlation of 0.8, similar to the correlation of the employment rate measure) and one less stable (correlation of 0.5). The allocation assumes a design effect of 3. Details regarding the mathematical formula used to calculate sample sizes are provided in Appendix C, Section 2.

Sample size of each sub-group of interest  
Table 3

		Level of change you want to identify	
		Change of 5%	Change of 10%
Stability of the measure over time	Corr. = 0.8	385	95
	Corr. = 0.5	964	241

The sample sizes in the table above are the sizes for each sub-group for which one wants to be able to evaluate a difference. A subgroup could be a combination of an Aboriginal origin by an age group by on- off-reserve category by an isolated/non-isolated category (i.e. North American Indians between 0 and 15 years of age, who live on reserves in isolated areas). The total sample size is the multiplication of the number of subgroups for which one wants to measure change, times the sample size necessary for each sub-group. For example, if there are three aboriginal groups, 6 age groups, 2 categories for on-/off-reserve, 2 categories for isolated/non-isolated, that gives  $3*6*2*2 = 72$  subgroups. If one wants to identify a change of at least 5% for a stable characteristic (correlation = 0.8), one needs a total sample of  $72*385 = 27\,720$ . Other sample sizes can be derived following the same rationale.

### 3.2.2.5 Issues Regarding a Follow-up to a Post-censal Survey

- A Post-censal survey on Aboriginal peoples would provide cross-sectional estimates, as well as detailed information which could be used to ensure the selection of groups of interest, for which one wants to measure changes.
- Data from the Post-censal survey could be linked to the longitudinal survey.
- Given that one is interested only in longitudinal measures, sample size can be smaller.
- The longitudinal sub-sample would not be representative of the population for the subsequent waves, since the population is not updated, and cross-sectional estimates would hence, not be available from it.



### **3.2.3 Sample Survey Conducted Independently of the Census**

#### **3.2.3.1 Sampling Frame**

An area frame would be used which identifies areas in which Aboriginal peoples live. This frame could be based on data from the census or another source.

#### **3.2.3.2 Survey Design**

The survey would be based on a multi-stage design.

##### On-reserve and High Density Areas:

Census data or another source could be used to identify reserves and other areas with a high density of Aboriginal peoples and hence focus the sample selection. For the on-reserve and high density off-reserve areas, some areas would be selected. All persons in these areas would have to be listed and then some households would be selected. The initial contact would be used to identify those households containing Aboriginal peoples.

##### Off-reserve:

For persons living in urban areas or outside of areas with a large proportion of Aboriginal peoples, it would be more difficult to locate Aboriginal people. Significant screening would be required, using a technique such as random digit dialling. This would have a major impact on the cost.

Once the first sample is selected, that sample would be used to produce cross-sectional estimates. Then a sub-sample could be selected for the longitudinal component. The methodology for the selection of that sub-sample would be similar to the methodology described in section 3.2.2 (follow-up to a post-censal survey).

#### **3.2.3.3 Sample Size Examples**

The initial sample would have to be very large to account for the proportion of households which do not include any Aboriginal member. Aboriginal peoples comprise approximately 3% of the total population of Canada. The size of the sample of Aboriginal peoples would have to be larger than the sizes given for allocation option 1 (section 3.2.1.4), because there would not be accurate data available to identify subgroups of interest for selecting the sample (e.g. no age data available to select the required number of persons in specific age groups).

#### **3.2.3.4 Adjusting for Non-response**

An adjustment for non-response has been made to the sample sizes given below, assuming a response rate of 80%. This is based on the response rate to the Aboriginal Peoples Survey, which was a supplement to the 1991 Census.

#### **3.2.3.5 Issues Regarding a Sample Survey Conducted Independently of the Census**

- The timing of the survey would not depend on the Census schedule.
- The cost will be larger because no frame will be available from which to select the sample.





- With no frame available, a multi-stage design will have to be used, therefore giving a larger design effect, which implies a larger sample.
- Selected areas will have to be listed and screened which will also have an impact on costs.

### 3.2.4 Use of Other List Frames

Sections 3.2.1 and 3.2.2 above discuss survey designs based on the use of the Census or a post-censal survey as frames from which the sample for the integrated survey could be selected. Other lists which provide names of persons in the target population (i.e. Aboriginal peoples) could also be considered as survey frames.

The department of Indian and Northern Affairs Canada maintains the Indian Register of all individuals who are registered as Indians under the *Indian Act*. The register thus includes only North American Indians, and would not be an appropriate tool for development of a sample frame if the integrated survey were to include all Aboriginal peoples. Even if it were decided to restrict the survey to the population of North American Indians, the Indian Register has other characteristics which limit its appropriateness for this application. For example:

- The register does not contain the names of all persons who have the right to be registered under the *Indian Act*, but only those who have registered.
- The register may underestimate the number of births and deaths in the population of Registered Indians due to delays in reporting.
- It is possible that other changes, such as moves to a new place of residence, are also not reported promptly.

At the present time, the monthly Labour Force Survey (LFS) does not include any questions by which Aboriginal persons could be identified. The feasibility of adding an identification question or questions on a regular or occasional basis could be considered in discussion with the LFS project team. This would enable a sample of Aboriginal persons living off-reserve to be selected from among LFS respondents. (The LFS excludes persons living on-reserve.) However, given that the LFS covers a sample of approximately 58,000 households, and that persons with Aboriginal origins living off-reserve comprise approximately 3% of the off-reserve population of Canada (according to the 1991 Census), this would be expected to result in a sample of less than 1800 households with Aboriginal persons. Even with selection from numerous rotation groups of the LFS, the sample available for the integrated survey would be too small to support the calculation of the required estimates. (Addition of a question to identify Aboriginal persons on the LFS could, however, be of assistance in providing general indicators of trends or changes between survey collections. Data could be employed in modelling or other systems to augment data provided by the integrated survey. The possible benefits of a question to identify Aboriginal persons on the LFS would need to be discussed, and weighed against the increased response burden, cost, or other risks for the LFS.)

Other potential sources of list frames include Aboriginal organizations, Band councils, and health or social service providers. None of these sources would cover the complete target population of all Aboriginal peoples; hence, an aggregation of lists would be required. Difficulties would be expected in ensuring complete coverage of the target population without duplication. The major difficulty in using such list frames for survey-taking purposes, however, would derive from the fact that they have not been developed for statistical purposes. The administrative or jurisdictional functions for which they have been created and maintained would have determined factors such as content, definitions employed, frequency of updates, level of coverage, and overall list quality. Given the differing scope



and focus of the organizations compiling the lists, these factors would be expected to vary widely. As a result, preparation of an aggregation of the lists with sufficient coverage and data quality for use as a sampling frame would require significant time and effort. It is expected that some sectors of the target population would still not be effectively covered as no list would exist in which they were included. As a result of all of these potential problems, use of other lists as a sampling frame is not considered to be a feasible option for the integrated survey. (If a pilot survey were to be conducted in selected areas or of selected subpopulations for which good quality lists existed, this option could be considered for this limited application.)

#### **4.0 Survey Outputs**

The exact types and media of data products from an integrated survey would be determined as part of the consultation process. The expected applications of the potential user groups would need to be understood in order to ensure that the data were provided in ways which were appropriate and accessible. For example, some users might require overview or highlight products which incorporate analysis; others might prefer detailed microdata files which allow them to carry out their own in-depth analysis.

Given the expected variety of data needs and intended applications of the potential user groups, it will be important to ensure that the data are provided in formats which are appropriate and accessible. Formats which could be considered include:

- microdata files
- summaries of survey highlights
- profiles of smaller areas (if feasible under the selected survey design and sample size)

Data could also be provided on a variety of media, including paper, microcomputer diskette, and CD-ROM. The addition of analysis, data manipulation software, or other value-added components may be of particular interest to certain groups of data users, such as Aboriginal organizations. It is expected that comparisons between Aboriginal and non-Aboriginal populations will be a key area of interest.

Through the consultation process the level for which estimates are to be prepared will also be determined. Depending on the needs of the users, the survey design and sample size could be established to enable estimates to be prepared for specified geographic areas or population subgroups.

#### **Issues:**

- In what formats and on what media will output products be required by the potential user groups? What value-added components will be of interest?
- The greater the analysis and value-added included in data products, the higher the development costs.
- Longitudinal data are normally more complex to manipulate and analyze than cross-sectional. Users are less likely to have experience with longitudinal surveys since fewer exist, and thus are more likely to need value-added data products.



## 5.0 Consultation Process

As with any data collection operation, a key stage in the initial planning process is the consultation with potential data users. The aims and objectives of the survey, and the applications envisaged for the data outputs, must be clearly understood if the survey is to be designed to meet the needs of the users. Potential data users, who should be involved in the consultation process, will include Aboriginal peoples and organizations, federal government departments, and provincial governments.

Extensive involvement of Aboriginal peoples as representatives of the target population will also be required. The response rate, and hence the success of the data collection operation, will be dependent on the involvement of Aboriginal peoples in the planning and development of the survey. Statistics Canada's experience with a variety of survey vehicles has shown that Aboriginal peoples must be partners in development if the survey is to be successful in achieving a response rate which allows the preparation of accurate and useful estimates.

Experience has also shown that support of the official representatives of the Aboriginal peoples does not necessarily ensure the participation of individuals or specific reserves. The consultation process should include contacts at the local or regional level to identify and address specific concerns or issues which may affect participation in the survey.

The expected benefits of the data to Aboriginal peoples would be emphasized, and their involvement would be sought in ensuring that the survey content and question wording reflect the realities and concerns of Aboriginal peoples. It is expected that the major topic areas of health, child development, and economics would be of interest to Aboriginal organizations and individuals, and would be a factor in motivating their involvement in the survey.

Advisory committees or other structures will be required to ensure the incorporation of expertise from the following groups:

- 1) Aboriginal persons who could advise on issues such as content, question wording, collection methods, and communication with local and regional groups, and who would provide support to the survey operation. This committee would be proposed to include representatives of the national organizations, such as the Assembly of First Nations, as well as local groups.
- 2) Experts in the subject areas to be covered by the survey, who could advise on the integration of the content of the three national surveys while also taking into account issues specific to the population of Aboriginal peoples.

It should be noted that an Advisory Committee on Aboriginal Issues has been proposed for Statistics Canada, and could potentially provide expertise to the survey, if convened.

Statistics Canada also requires, for special survey projects of this type, the establishment of a steering committee to oversee the development and implementation of the survey. Members would be proposed to include representatives of Aboriginal peoples, the three national surveys, sponsoring departments, and senior management of Statistics Canada. In addition to formally approving design details and cost estimates, the committee would also be responsible for ensuring Statistics Canada's credibility and image as Canada's national statistical agency. If the integrated survey were linked to any other data collection vehicle (e.g. the Census), the committee would also ensure that the response rates, timing, and data quality of the other survey were not unduly affected.



The integrated survey of Aboriginal peoples would also be represented on the internal Statistics Canada Committee on Aboriginal Data Coordination. This committee includes programs within Statistics Canada which collect data on Aboriginal peoples, and ensures integration and communication across programs.

**Issues:**

- How can the maximum possible involvement and support of Aboriginal persons be achieved?
- What administrative structure will be most appropriate for managing the development and implementation of the integrated survey?
- What structure of advisory or other committees would be best for accessing the expertise of Aboriginal persons and other content experts?
- Consultation costs are expected to be significant. What will be the balance point for time and cost spent on consultation, beyond which benefits are no longer sufficient to warrant further contacts?

## **6.0 Costs and Schedules**

At this initial stage of the survey process, there are far too many outstanding issues and questions for any discussion of costs and schedules to be meaningful. However, it is important for potential sponsors to have a global idea of the possible magnitude of the project being considered.

Factors which will influence the costs and timing of the survey have been mentioned in the sections above. These include:

- length and scope of the consultation process,
- number of players involved in the survey development process,
- type of data desired (i.e. cross-sectional vs. longitudinal estimates),
- level of detail for which estimates are desired (e.g. geographic level, population subgroups),
- number and complexity of questions to be asked
- survey design,
- sample size,
- collection methodology, and
- data output products.

As a general indicator of potential collection costs, the expected cost for collection of the NLSC (which does not include reserves or settlements) is \$150 per household. Collection costs for personal interviews on isolated reserves would be expected to be higher than for interviews off-reserve due to costs such as interviewer training and travel. As well, the total interview time per survey cycle could easily be longer for the integrated survey than for the NLSC given the integration of content from the three surveys.

Additional non-collection costs, which would also themselves be significant, include consultation, survey development and design, data capture, editing, and production of output products. Thus, an integrated survey of Aboriginal peoples, if it were to be carried out in a manner that would produce meaningful and accurate estimates, would be an expensive proposition.

The time required for initial development of the survey will be determined by the extent of the consultation process. As described above, effective consultation, particularly with Aboriginal peoples, will be crucial to the success of the project. As a result, the time and resources dedicated to this





initial stage can be expected to be significant, but will be worthwhile investments to the overall outcome of the survey. The schedule for development and implementation of the integrated survey of Aboriginal peoples will also be greatly influenced by any linkages with other survey vehicles, such as the 1996 Census.

Until fundamental decisions regarding the integrated survey are made the development of a detailed survey schedule is not possible. However, key stages in the integrated survey can be expected to include the following:

- Establishment of steering committee and advisory committees
- Consultation with Aboriginal peoples
- Consultation with other potential data users
- Determination of the survey content
- Determination of the survey design and sample sizes
- Development of the collection instruments
- Testing of the collection instruments
- Survey collection
- Data capture and editing
- Certification of data and completion of data files
- Release of public microdata, highlight summaries and other data products

Depending on decisions regarding future cycles of the survey, and the development of a longitudinal component, certain stages in the schedule would then be repeated on an ongoing basis.

## **7.0 Other Considerations**

There are a number of conditions to which clients of Statistics Canada are subject when sponsoring a survey.

### **7.1 Project Team Approach**

All projects managed by the Special Surveys Group of Statistics Canada are designed and implemented by an inter-disciplinary team. Apart from the sponsoring agency(ies), this team includes representatives from Statistics Canada field operations, survey methodology, survey management, data processing, and analysis divisions. This approach requires a considerable "in-person" input on the part of the sponsoring agency at a sufficiently high level to make decisions concerning questionnaire content, survey methods, and other aspects of the survey.

### **7.2 Steering Committee**

As described in the Consultation section above, the day-to-day work of the project team would be guided by a Steering Committee of line managers from Statistics Canada and senior representatives from the sponsoring agencies and other team members.

### **7.3 Access to Data**

All Statistics Canada surveys are collected under the Authority of the Statistics Act. Under this act, Statistics Canada is obliged to protect the identities of individual respondents. As a result, microdata



files must be screened, and selected variables removed or collapsed to ensure that individual respondents cannot be identified. The most problematic variables to data confidentiality are geographic identifiers, with the amount of characteristics detail made available inversely related to the amount of geographic detail provided. Some options exist for gaining access to unscreened microdata under strictly controlled conditions. These include collection under Section 12 of the Statistics Act and contractual agreement between Statistics Canada and the sponsoring departments. These options, the rigid terms and conditions under which they would be implemented, and any potential impacts on response rates could be discussed with the sponsoring agencies.

#### **7.4 Release Policy**

Statistics Canada policy is to make the certified survey results available to all users at the same time. Because of this principle, it is suggested that the release of summary highlights and of the public microdata be simultaneous.

In order to ensure that users match estimates published by Statistics Canada and generate estimates in a consistent manner, they are required to apply specified guidelines before undertaking any publication or other release of the data derived from the survey. The guidelines cover policies for timing of release, sampling variability, rounding and weighting. Details regarding these policies would be provided to sponsoring departments or other data users.



## **Appendix A**

### **Overview of Statistics Canada Surveys Discussed in the Report**

#### **NATIONAL LONGITUDINAL SURVEY OF CHILDREN**

##### **Background**

In May 1992, the federal government announced "Brighter Futures" - a series of initiatives to improve the health and well-being of Canada's children. One component of the program is the "What Works for Children - Information Development Program". Its purpose is to develop information for policy analysis and program development on the critical factors affecting the development of children in Canada. (Note: This program previously fell under the mandate of Health and Welfare Canada but is now part of Human Resources Development Canada following a reorganization of government departments.)

The initial activity under "What Works" is the National Longitudinal Survey of Children (NLSC), being developed and implemented by the Special Surveys Group of Statistics Canada, under contract to Human Resources Development Canada. The purpose of the NLSC is to measure characteristics of a sample of children in Canada, as well as the risk and protective factors in their lives, and to monitor the impact of these characteristics, factors and experiences on children's development into adulthood. Data will be collected every two years. The findings will be used to develop effective policies and programs for children at risk.

##### **Objectives**

- to develop information for policy analysis and program development on critical factors affecting the development of children in Canada
- to meet the needs of the "What Works for Children - Information Development Program" of the "Brighter Futures" initiatives of Health Canada

##### **Data Requirements**

- cross-sectional and longitudinal data
- national longitudinal database on children whose initial ages in 1994 are:  
0, 1, 2-3, 4-5, 6-7, 8-9, 10-11
- national cross-sectional estimates on children 0-11 (age range will increase as the children in the initial sample pass through adolescence)

##### **Survey Population**

- children ages 0 - 11 years
- the child is the unit of sampling for purposes of the longitudinal follow-up, not the dwelling

##### **Sample Frame**

- households in the Labour Force Survey
- excludes households not in the LFS frame (i.e. institutions, households on reserves, households in the Yukon and NWT) and full-time members of the Canadian Armed Forces
- sample of children living in the Yukon and NWT will be added to the survey
- children under 12 years of age in the household will be included in the survey (to a



maximum of 4 children)

- data on parents and other family members will also be collected

### **Sample Design**

- selection of households with children under 12 years of age from the current LFS sample (6 rotations) and previous rotation groups

### **Sample Size**

- approximately 25,000 children

### **Content**

- country of birth
- immigration/citizenship
- language
- ethnic origin
- religion
- characteristics of the child (varies depending on age of child)
  - demographics
  - custody
  - child care
  - education
  - physical health
  - newborn indicators
  - behaviour
  - temperament
  - development
  - chronic conditions
  - activities / time use
  - relationships
- characteristics of the parents
  - demographics
  - education
  - labour force activity during past 2 years
  - income
  - physical health
  - activity limitations
  - mental health
- characteristics of other family members
  - demographics
  - education
  - labour force activity
  - physical health





- family characteristics
  - parenting style
  - family functioning
  - social support (of respondent)
- housing characteristics
  - housing conditions
- community characteristics

#### **Collection Methodology**

- face-to-face interviews in the respondent's home for the first two data collections
- main respondent will be the person in the family most knowledgeable about the child (in most cases, the mother)
- some direct assessments will be conducted and questions addressed to the child
- computer-assisted interviewing technology to be used
- work is progressing on the integration of the children component of the NPHS with the NLSC

#### **Output Products**

- public microdata files
  - individuals
  - families
- highlights packages  
(cross-sectional data available after each survey cycle, longitudinal data after second and each subsequent survey cycle)

#### **Schedule**

- beginning December 1994/January 1995
- survey to be conducted every two years

#### **Integrated Survey in the Yukon and Northwest Territories**

A project is currently in progress to develop an integrated version of the NPHS and NLSC for use in the Yukon and Northwest Territories (including Nunavut). Data will be collected using a single questionnaire which combines the key components of the NPHS and NLSC. The statistical bureaus of the two territories are involved in adapting the questionnaire to suit the cultural and lifestyle realities of territorial residents, and will carry out the survey collection as agents of Statistics Canada.



## NATIONAL POPULATION HEALTH SURVEY

### Background

In the fall of 1991, the National Health Information Council recommended that an on-going national survey of population health be conducted. This recommendation was based on consideration of the economic and fiscal pressures on the health care system and the commensurate requirements for information with which to improve the health status of the population in Canada. Existing sources of health data were felt to be unable to provide a complete picture of the health status of the population and the myriad of factors that have an impact on health.

The overall goal of the National Population Health Survey (NPHS) is to provide more comprehensive information on current health status than was available from existing data sources. It is intended to aid in the development of public policies designed to improve health by providing measures of the level, trend and distribution of the health status of the population. The survey will provide panel data that will reflect the dynamic process of health and illness as well as producing periodic cross-sectional estimates. Data content will include the economic, social, demographic, and occupational correlates to health, and the relationship between health status and the use of health services. Also under exploration is a survey design that would allow cost effective supplementation of the content or the sample, and the possibility of linking survey results to administrative data sources. Data for the NPHS are to be collected every two years.

### Objectives

- to provide more comprehensive information on current health status
- specifically:
  - to aid in the development of public policies designed to improve health, by providing measures of the level, trend and distribution of the health status of the population
  - to provide data for analytic studies that will assist in understanding the determinants of health
  - to collect data on the economic, social, demographic, occupational, and environmental correlates to health
  - to increase the understanding of the relationship between health status and the use of health services, not only in the traditional sense but also in areas such as home care, self medication and self care
  - to provide panel data that will reflect the dynamic process of health and illness and produce periodic cross-sectional estimates
  - to provide the provinces and territories and other clients with a health survey capacity which will permit cost effective supplementation of the content or the sample
  - to allow the possibility of linking survey results to administrative data sources for statistical analysis, where viable

### Data Requirements

- will include both cross-sectional and panel survey techniques

### Survey Population

- covers all persons residing in Canada
- long term residents of chronic care hospitals, nursing homes, residences for senior citizens, general hospitals, psychiatric institutions, and centres for persons with physical disabilities are covered through the institutional component of the survey



- residents of non-permanent quarters of military camps, naval vessels, school/training residences, religious institutions, orphanages and children's homes, correctional institutions and young offenders facilities are excluded
- for the first year, survey will exclude full-time members of the Canadian Armed Forces, persons living on Indian reserves, and persons living in institutions

#### **Sample Frame**

- sample will be selected from the redesigned LFS sampling frame in all provinces except Quebec
- the Quebec sample will be selected from a 1986 EA based area frame maintained by the Bureau de la Statistique du Québec
- an additional sample will be selected for the Yukon and Northwest Territories from lists of households maintained by them

#### **Sample Design**

- sample of households will be selected from the redesigned LFS sample frame  
(Note: NPHS will select only households which have not previously been included in the LFS)
- each panel is expected to remain in the sample for 8-10 years or longer
- some data are collected for every person in household (e.g. demographics, utilization of health services, chronic condition)
  - one person in household randomly selected for detailed interview
- only the selected person is traced for subsequent waves of the survey

#### **Sample Size**

- sample of approximately 22,000 households

#### **Content**

- demographics
  - country of birth and year of immigration
  - 2 week disability
- health care utilization
- restriction of activities
- chronic conditions
- language
- ethnicity
- race
- education
- labour force
- household income
- general health
- height / weight
- preventive health practices
- smoking
- alcohol



- physical activities
- injuries
- stress and work stress
- self-esteem and mastery
- sense of coherence
- recent life events
- trauma
- health status (vision, hearing, speech, getting around, hands and fingers, feelings, memory, thinking, pain and discomfort)
- mental health
- drug use
- social support

#### **Collection Methodology**

- four collection periods during the year (May, August, November, and February)
- one quarter of the sample will be interviewed during each period
- no children under 12 will be selected in the first two collection periods (through integration with the NLSC children will be oversampled in the third and fourth quarter to compensate for the first two quarters)
- work is progressing on the integration of the children component with the NLSC
- first visit to each household will involve a face-to-face interview
- use of computer-assisted interviewing (except in the Territories where a paper and pencil questionnaire is administered in person)

#### **Schedule**

- survey to be carried out every two years beginning in May 1994





## **SURVEY OF LABOUR AND INCOME DYNAMICS**

### **Background**

The Survey of Labour and Income Dynamics (SLID) has been developed to support research aimed at advancing the understanding of labour market behaviour and economic well-being. It serves as an enhanced replacement to the Labour Market Activity Survey (LMAS), which collected data for reference years 1986 to 1990. The LMAS provided evidence of the scope and value of information on labour market movement which could be obtained through a longitudinal survey. As well, it demonstrated the effectiveness of methods developed to improve respondent recall, a difficulty with past retrospective labour market surveys. SLID also draws on the experience of other ongoing Statistics Canada surveys, including the Labour Force Survey and the Survey of Consumer Finances.

As a longitudinal survey, SLID will record important events in a person's life, such as family formation and dissolution, migration, and job loss. Individuals and families will be interviewed twice each year, first to collect labour information for the previous year, and second to collect information on income received in the previous year. The sample will be followed for a period of six years. The data are intended to improve understanding of the links between demographic events, labour market events and changes in income. In addition to longitudinal data, the survey will generate annual cross-sectional data.

### **Objectives**

- to follow individuals and families for a period of 6 years, collecting information on labour market experiences, income, and family circumstances, in order to improve understanding of links between demographics events, labour market events, and changes in income

### **Data Requirements**

- longitudinal data
- annual cross-sectional data

### **Survey Population**

- non-institutional population of all ages living in the 10 provinces, excluding residents of institutions, persons living on Reserves, and full-time members of the Canadian Armed Forces living in barracks
- persons 16 years and over are eligible for the questions on labour and income

### **Sample Frame**

- area frame
- first panel, selected in January 1993, consisted of 2 LFS rotation groups

### **Sample Design**

- the first panel was a sample of 15,000 households; future panels are planned to include 20,000 households
  - at time of selection, a preliminary interview is conducted
  - each panel will be representative of the target population at time of selection
- a new panel will be selected every three years; each panel will remain in the sample for 6 years
- sampling units are all persons in the sampled household at the time the panel is introduced
  - all sampled persons remain in the sample for the life of the panel, regardless of whether they



move out of the original household

- anyone who moves in with a person in the longitudinal sample will also be interviewed

#### Sample Size

- approximately 40,000 households, split evenly between 2 panels (including approximately 80,000 individuals 16 years and over)

#### Content

(Note: reference date for both labour and income data will be the previous calendar year)

- demographics
  - date of birth
  - mother tongue
  - whether Registered Indian or member of a visible minority
  - relationship to other household members
  - marital status
  - \* marital history
  - \* number of children born and year of birth of first child
  - \* education of parents
  - \* education of respondent including:
    - number of years and province of schooling
    - post-secondary education
    - diplomas, certificates, degrees received
    - field of study

(\* -- collected in preliminary interview only. Other demographic information updated in each collection cycle.)

- activity restrictions or disability
- labour data (for up to 3 employers in the year)
  - industry
  - class of worker
  - occupation
  - main duties
  - supervisory/managerial responsibilities
  - changes in duties during the year
  - information on firm size
  - usual hours and any changes in hours of work in the last year
  - reason for working part-time
  - work arrangements, including reasons for working on call or an irregular schedule
  - work at home
  - reason for job separation
  - expectation of returning to the job
  - self-employed workers
    - months worked in the reference year
  - paid workers



- wages
- union membership
- pensions
- any absences of a week or longer except paid vacation (excluding on-call workers)
- the reasons for those absences
- how and when the job was obtained
- jobless spells
  - job search
  - desire for employment
  - reasons for not looking for work
  - education received during year
- receipt of UI, social assistance, and workers compensation at any time of the year
- income
  - employment income
  - pension income
  - government income
  - income from investments
  - other monetary income
  - income from tax credits
  - total income
  - total income tax paid
  - alimony/child support received
- wealth (collected with income data, but not every year)
  - tangible assets
  - financial assets
  - equity in a business, farm, or partnership
  - other wealth
  - debt

#### **Collection Methodology**

- computer-assisted interviewing
- interviews carried out mainly by telephone

#### **Output Products**

- linked labour and income microdata files for each survey year
- cumulative longitudinal file will be produced after each year of collection for the first panel
- full longitudinal file will be produced after 6 years of data collection
- once the second panel is introduced a 3 years longitudinal file combining the sample from the 2 panels will be produced

#### **Schedule**

- first collection in 1994
  - labour survey -- January 1994
  - income survey -- April 1994
- survey to be conducted every year



# Summary of Major Content Areas of the Longitudinal Surveys

Content Area	NLSC	NPHS	SLID
Demographics	I	I	I
Language	I	I	
Education	I	I	I
Literacy	I		
Behaviour / Temperament	I		
Child Development / Newborn Indicators	I		
Activities / Time Use	I		
Preventive Health Practices		I	
Health Care Utilization and Drug Use		I	
Chronic Conditions	I	I	
Disability / Restrictions	I	I	I
Smoking and Alcohol Use	I	I	
Physical Activities		I	
Injuries	I	I	
Stress		I	
Health Status	I	I	
Health History	I		
Mental Health	I	I	
Social Support	I	I	
Labour Force Activity	I	I	I
Industry		I	I
Occupation	I	I	I
Family Characteristics	I	I	I
Child Care	I		
Family Custody History and Parenting	I		
Housing Characteristics	I	I	
Income	I	I	I
Neighbourhood	I		

Note: For the NLSC some variables are available for only the child or only the adult.





## 1991 ABORIGINAL PEOPLES SURVEY

### Objectives

- to provide a profile of Aboriginal lifestyles and living conditions, including information on housing conditions, health, employment history, schooling, mobility and the use of Aboriginal languages
- to assist Aboriginal organizations, communities, research groups, and provincial, territorial, and federal governments in understanding the needs of Aboriginal persons in Canada

### Data Requirements

- cross-sectional data only

### Survey Population

- persons who identify with an Aboriginal group and/or are registered under the *Indian Act of Canada*
- residents of institutions and collective dwellings such as hotels or rooming houses were excluded from the survey

### Sample Frame

- chosen from those persons who indicated Aboriginal origins and/or indicated that they were registered under the *Indian Act of Canada* on their 1991 Census of Population long questionnaire
- the first section on the APS was used to screen in only those who identified with an Aboriginal group and/or were registered under the *Indian Act of Canada*

### Sample Design

- to facilitate the coverage of a population that is distributed over a large geographic area, each province was divided into two parts or domains
- domain 1 included communities which had a high concentration of Aboriginal persons; this included Indian reserves, Inuit and Métis communities, as well as some towns and villages that had a large Aboriginal population.
  - a representative sample of persons who indicated Aboriginal origins and/or being a Registered Indian was selected from each community
- domain 2, which covered the remainder of the province, was divided into different parts: depending on the province, major census metropolitan areas; an aggregate of all other census metropolitan areas; an aggregate of all other urban centres not included in domain 1; and an aggregate of all rural areas not included in domain 1
  - a representative sample of persons who indicated Aboriginal origins and/or being a registered under the *Indian Act of Canada* was selected for each of the four parts comprising domain 2
- The first section on the APS was used to identify those persons who identified with an Aboriginal group and/or were registered under the *Indian Act of Canada*; the APS interview continued only with those persons who reported identity with an Aboriginal group and/or being registered under the *Indian Act of Canada*

### Sample Size

- approximately 171,500 persons were sampled for the survey
- approximately 114,800 persons indicated that they identified with an Aboriginal group and/or were registered under the *Indian Act of Canada* and completed the APS questionnaire.

### Content

- content of the adult questionnaire:
  - identity
  - language and tradition



- disability
- health, lifestyle, and social issues
- mobility
- schooling
- work and related activities
- expenditures and sources of income
- housing
- content of the children's questionnaire:
  - identity
  - language and tradition
  - health and disability
  - mobility
  - schooling
  - housing

### **Collection Methodology**

- data for the survey were collected through personal interviews
- for adults, interviews were conducted, for the most part, with the selected respondent
- about 17% of adult interviews were conducted through another household member if the selected respondent was absent during collection
- information on children under the age of 15 years was collected from the parent or guardian; however some children 12 years of age and over could respond for themselves

### **Output Products**

- published reports containing statistical tables, and textual highlights of tables at various levels of geography
- community profiles for domain 1 communities with an Aboriginal identity population of 40 or more
- statistical profiles for each of the 3 Aboriginal groups (North American Indians, Métis, and Inuit)
- microdata files
- custom tabulations
- APS Workshop

### **Schedule**

- data collection took place from October 1991 to January 1992



# APPENDIX B

## COUNTS OF ABORIGINAL PERSONS IN CANADA

TABLE 1.

Population Reporting Aboriginal Origins, On and Off Indian Reserves and Settlements, for  
Canada, Provinces and Territories, 1991 Census

Province	Population with Aboriginal Origins Living On Reserves	Percentage of Population Living On Reserves	Population with Aboriginal Origins Living Off Reserves	Percentage of Population Living Off Reserves	Total Population Reporting Aboriginal Origins by Province
Newfoundland	470	3.5%	12,640	96.5%	13,110
Prince Edward Island	330	17.5%	1,550	82.5%	1,880
Nova Scotia	5,655	25.8%	16,230	74.2%	21,885
New Brunswick	2,750	21.4%	10,065	78.6%	12,815
Quebec	21,275	15.4%	116,345	84.6%	137,620
Ontario	29,040	11.9%	214,510	88.1%	243,550
Manitoba	36,375	31.3%	79,825	68.7%	116,200
Saskatchewan	30,235	45.5%	66,345	54.5%	96,580
Alberta	23,705	15.9%	124,520	84.1%	148,225
British Columbia	37,880	22.4%	131,155	77.6%	169,035
Yukon Territory	340	5.3%	6,050	94.7%	6,390
Northwest Territories	215	0.6%	35,175	99.4%	35,390
Total for Canada	188,270	18.7%	814,405	81.3%	1,002,670

Source: 1991 Census, Statistics Canada

NOTE: The population reporting Aboriginal origins as presented above includes those persons who reported at least one Aboriginal origin in the Census.



TABLE 2.

Population by Selected Aboriginal Origin, On and Off Indian Reserves and Settlements, for Canada, Census 1991

ABORIGINAL GROUP	ON-RESERVE POPULATION	OFF-RESERVE POPULATION	TOTAL POPULATION
North American Indian only	176,620	188,755	365,375
Métis only	1,305	73,840	75,145
Inuit only	65	30,025	30,090
Single Aboriginal origin with non-aboriginal origins	7,235	484,705	491,940
Multiple Aboriginal origins	3,035	37,085	40,120
Total Population with Aboriginal origins	188,270	814,405	1,002,670

Source: 1991 Census, Statistics Canada





TABLE 3.

Population Reporting Aboriginal Identity By Aboriginal Group, for Canada, 1991  
Aboriginal Peoples Survey

ABORIGINAL GROUP	POPULATION REPORTING ABORIGINAL IDENTITY
NORTH AMERICAN INDIAN	
Living On-reserve	166,025
Living Off-reserve	294,655
TOTAL NORTH AMERICAN INDIAN	460,680
MÉTIS	135,265
INUIT	36,215
TOTAL APS POPULATION *	625,710

Source: 1991 Aboriginal Peoples Survey, Statistics Canada

- \* Respondents who identified with more than one Aboriginal group were included in each of the groups with whom they identified. As a result of the double counting the sum of the Aboriginal groups identified is greater than the total population reporting Aboriginal identity. Of those who identified with their Aboriginal ancestry 99% identified with only one Aboriginal group. Identification with more than one Aboriginal group was most common among Métis: 4% identified with more than one Aboriginal group.



## APPENDIX C

### DERIVATION OF THE SAMPLE SIZES

#### Section 1

#### Sample size calculation when the purpose of the survey is to produce cross-sectional estimates

#### Definitions

N = Estimate of the Aboriginal population

p = Minimum proportion of the population which presents the characteristic one wants to estimate.

n = Sample size

$\hat{Y} = N \cdot p$  = Estimated size of the population which presents the characteristic.

r = Estimated response rate

Deff = Design effect

$$Deff = \frac{VAR_p(\hat{Y})}{VAR_{SI}(\hat{Y})}$$

The design effect explains the difference, in term of variance of the estimator, between sampling with the chosen plan "P" and a simple random sample "SI". The larger the difference between the selected sampling plan and simple random sampling, the larger the design effect. The design effect is estimated.

$CV_d$  = Desired coefficient of variation for the estimates. The coefficient of variation measures the precision of the estimates obtained. According to Statistics Canada Policy, estimates with a coefficient of variation of 16.6% should be used with caution.



### Sample Size Calculation

We want to determine the minimum sample size required, such that a given level of precision is obtained.

$$VAR(\hat{Y}) = Deff * N^2 * (1 - \frac{n}{N}) * \frac{S^2}{n}$$

$$S^2 = \frac{N}{N-1} * p * (1-p)$$

$$CV_d^2(\hat{Y}) = \frac{VAR(\hat{Y})}{N^2 p^2}$$

Then the minimum sample size is calculated as:

$$n = \frac{N * (1-p) * Deff}{(N-1) * p * CV_d^2 + (1-p) * Deff}$$

This is the actual required sample size. However, one should adjust for expected non response. Therefore:

$$n_f = \frac{n}{r}$$

### Section 2

#### Sample size calculation when the purpose of the survey is to estimate a net change.

In a longitudinal survey, the prime objective is to observe net change in the population as well as individual change. The focus on change rather than "stocks" (i.e. numbers of persons in particular categories at a point in time) has an impact on sample sizes.



Let

$\hat{P}_1 =$  *estimated proportion of a characteristic at time 1*

$\hat{P}_2 =$  *estimated proportion of a characteristic at time 2*

and

$\hat{P}_1 - \hat{P}_2 =$  *net change between time 1 and time 2*

The variance of the net change can be expressed as:

$$VAR(\hat{P}_1 - \hat{P}_2) = Var(\hat{P}_1) + Var(\hat{P}_2) - 2 * \rho * \sqrt{(Var(\hat{P}_1) * Var(\hat{P}_2))}$$

where  $\rho$  = coefficient of correlation between the characteristics in the two surveys

When two cross-sectional surveys are run and there is no sample overlap, the coefficient of correlation  $\rho$  is equal to 0. When there is full overlap between two collections (i.e. the same sample is used) the variance depends on the correlation between the measured characteristics at the two points in time. If the measured characteristics are highly correlated, then the variance will decrease.





Assuming that the variance of the estimated characteristic is the same at time 1 and 2

$$(i.e. Var(\hat{p}_1) = Var(\hat{p}_2))$$

we get:

$$Var(\hat{p}_1 - \hat{p}_2) = 2(1-\rho) Var(\hat{p}_1)$$

(The assumption would not be true if, for example, there is an attrition problem at the second wave. Then we would have:

$$Var(\hat{p}_2) > Var(\hat{p}_1)$$

Then we should use

$$Var(\hat{p}_1 - \hat{p}_2) = 2(1-\rho) Var(\hat{p}_2)$$

to calculate sample size estimates)

If one wants to be able to identify a difference between the two proportions, then the following hypothesis would have to be tested:

$$H_0: p_1 - p_2 = 0 \text{ versus } H_1: p_1 - p_2 \neq 0$$

$H_0$  will be rejected if:

$$\frac{|\hat{p}_1 - \hat{p}_2|}{\sqrt{Var(\hat{p}_1 - \hat{p}_2)}} > t_{\alpha/2}$$

where  $t_{\alpha/2}$  is the abscissa of the normal curve that cuts off an area of  $\alpha/2$  at each tail. (It is assumed that  $n$  is large enough to use the normal approximation).



Solving for n we obtain:

$$n > \frac{t_{\alpha/2, p_1 q_1, \text{def } 2}^2 (1-p)}{(\hat{p}_1 - \hat{p}_2)^2}$$

This is the actual required sample size. However, one should adjust for expected non response and attrition. Therefore:

$$n_f = \frac{n}{r}$$

The sample sizes given allow one only to identify that a change is significant, but don't guarantee a given CV for either the estimates, or the differences. This implies that, with the sample sizes given, one won't necessarily be able to produce estimates for subgroups. The sample sizes should instead be viewed as those required to follow-up subgroups longitudinally so that one can test whether proportions of a characteristic change.

72 008

STATISTICS CANADA LIBRARY  
BIBLIOTHEQUE STATISTIQUE CANADA



1010213994

•  
•  
7

•  
•  
•  
•

•  
•  
•  
•

