Learning Objectives: Chapter 5

Survey/Census and Administrative Data Collection
Development: Planning and Measurement Issues

After reading this chapter, the reader should be able to:

1. Describe the planning cycle for censuses and surveys
2. Outline issues in designing disability data collection
3. Understand the issues in survey sampling design

“Blind moves” Main Uddin
WHO Photo Contest “Images of Health and Disability 2002/2003”
5. SURVEY/CENSUS AND ADMINISTRATIVE-BASED DATA COLLECTION DEVELOPMENT: PLANNING AND MEASUREMENT ISSUES

5.1 Overview of the survey and census cycles
   5.1.1 Survey cycle and costs
   5.1.2 Census cycle

5.2 Planning for data collection

5.3 Asking the right questions

5.4 Determining data user needs
   5.4.1 Consult with data users
   5.4.2 Statistical advisory groups
   5.4.3 Consult persons with disabilities

5.5 Privacy implications of collecting data

5.6 Other basic management issues in planning

5.7 Quality of survey data and total survey error

5.8 Issues in collection design
   5.8.1 Context of disability questions
   5.8.2 Proxy versus non-proxy survey responses
   5.8.3 Face-to-face versus telephone interviewing
   5.8.4 Special questions for sub-populations
   5.8.5 Collection of disability data about children
   5.8.6 Collection of disability data about institutionalised people

5.9 Planning for administrative-based data collection
   5.9.1 Identifying information needs
   5.9.2 Developing data items
   5.9.3 Deciding on collection methods and materials

5.10 Sample Design
   5.10.1 Sampling frame
   5.10.2 Sample size and unit
5.1 Overview of the survey and census cycles

5.1.1 Survey cycle and costs

The development of most surveys follows the same cycle. Unless the survey has been conducted before, the survey cycle starts with the identification of the need for information by one or more clients or data user. This step is the most important since, without clear identification of user need, the purposes of the survey will be unclear and the development process will be flawed from the beginning.

This starts the ongoing development process for data collection. User needs guide in the planning process that leads into development and design of the survey. When data is collected and processed, estimation of prevalence and other analyses begin. These are disseminated and evaluated by, among others, the data user. So the process comes full circle. Diagram 5.1 below illustrates the survey cycle.

Diagram 5.1: The Survey Cycle

Several of these phases, and in particular, those that raise issues of special concern to disability data collection, are discussed in this and the following chapters.

Considerations of cost are always relevant to the development of surveys. National statistical offices must always be aware of whether the potential
benefits of the survey compensate for the costs of developing a survey, putting it into the field, and collecting and analysing the data.

National statistical agencies regularly record, not only total costs of data collection, but the costs of each phase. Cost information is essential for budgeting each phase of survey development within the organization, and may also be used to compare costs with those of other national statistical organizations.

The most significant costs in any interviewer-administered data collection are the actual field costs to administer the survey and collect the data. The planning and development costs are significant and of course necessary to ensure high quality disability data. Costs of analysis and data dissemination are also major expenses. Usually, total development costs are roughly equal to total output related costs.

5.1.2 Census cycle

The cycle of phases of censuses is similar to that of surveys, as Diagram 5.2 illustrates. In particular, the census cycle begins and ends with evaluation of previously collected data and user consultation. The major difference between the two is that far more time is required for all phases of census collection, especially for additional consultation, design and testing procedures, including topic selection, government endorsement of final design, and quality assurance.

![Diagram 5.2: The Census Cycle](image-url)
5.2 Planning for data collection

Given the complexity and cost of putting a survey or census into the field, and the necessity of having the data serve the needs of the data users, the planning stage is crucial. Some of the issues that must be addressed during the planning stage are:

- What are the objectives of the data collection?
- What are the questions to ask?
- What should be the scope and coverage?
- How should disability be operationalised?
- What should the screening questions look like?
- Should proxy interviews be used?
- Should we use telephone or face-to-face interviewing, or both?
- Should there be different questions or questionnaires for different sub-populations?
- How do we solve issues with the language and terminology we use?
- How do we solve basic methodology issues of sample design, sample frame, sample unit, stratification, weighting and so on?
- How do we preserve privacy and confidentiality or respondents’ response?

Objectives should be stated as clearly as possible because they determine the scope, methodology and content of the survey/census, and ultimately, its statistical outputs.

The first two of these questions can be ascertained through consultation with data users and those persons with disabilities. These will be discussed in detail in the next sections. The last question is discussed in Section 5.5.

5.3 Asking the right questions

Let’s take a look at a real case on how questions were formulated in designing a tool for gathering disability data.

Consider an Australian survey conducted in 2003 called the Survey of Disability, Ageing and Carers (SDAC):

From the outset it was agreed that there were fundamental gaps in information on three overlapping groups of people: persons with disabilities, older people, and those who provide care for other people because of their disabilities.

What was needed was not only basic prevalence information, but also information that would identify unmet needs for persons with disabilities and the elderly -- in particular, health care and technical aids such as wheelchairs and hearing aids. For carers, what was needed was data on the effects of the caring role on their lifestyle.
Data users expressed the need to identify changes over a period of time in the prevalence and characteristics of persons with disabilities. So the questions in the SDAC needed to be closely similar to the earlier survey conducted in 1998.

Finally, in addition to the need for data at national and state/territory levels, for regional planning of disability services, users expressed the need for regional and small area data as well.

The designers of SDAC sought answers to some very basic and common sense questions.

- Who needs the information?
- Why do the users need this data?
- Which populations should be surveyed?
- What questions need to be answered, and what information would answer them?
- What is the population of interest and what do we want to know about it?
- What is the geographical scope of the survey?
- Should our questions be statistically relatable to those in other surveys?
- When is the information needed, and how often does it need to be updated?
- What form should the results take, how much detail and cross-tabulation are required?

During its discussions about a general disability measure for censuses, the Washington Group (WG) at its first meeting in 2002 identified the first two of these questions as particularly important for the task of finding out the purpose served by any disability data gathering instrument. The WG suggested the following possibilities:

**Who needs the information?**

- Researchers
- The general public
- Policy makers
- Associations and NGO’s
- Trade unions
- Government agencies
- International organizations
- Health service organizations and providers
- Industry
- Device and equipment manufacturers
- Employers
- Disability consumer groups
- Insurance agencies
- Education planners
- Media
Why do users need disability data?

- Planning
- Project and planning evaluation
- Marketing
- Policy development
- Advocacy and political action
- Prevention
- Tracking and enhancing participation
- Improving services
- Creating or improving standards

The answers to the other questions (which populations to survey, what to ask, geographical scope) were considered to be dependent on the needs of the data users in each individual country.

This naturally leads to the question: How do we determine what the needs of disability information users actually are? In the next section we outline a series of practical steps that have been used with considerable success.

5.4 Determining user needs

Suppose your agency is planning to develop a survey relating to disability. There is a limit imposed by resources and the patience of respondents on the number of questions asked. However, you want all your questions to address the needs of the users. How should you proceed?

5.4.1 Consult with data users

Data collections produced by statistical agencies that have not consulted, broadly and deeply, with those who would use the data are, in a word, useless.

There are many techniques of consultation:

- Seek written submissions on data content and purpose of data
- Convene user groups to discuss data needs and survey/census issues
- After documenting data user needs, check back with the data users to confirm that they are correct.

As there are limits to the number of questions that can be included in any data collection activity, not every data need can be addressed. The data collection developer should collect as much information as possible about users' needs, and then check to see if some of these needs can be satisfied by alternative sources of data. Usually, it will be the difficult task of the data collector developer to prioritise needs, or to work out a compromise to satisfy as many users’ needs as possible.
5.4.2 Statistical advisory groups

It is a good practice to put into place an on-going and regular process of consultation through regular contact and discussion with the statistical user community. This establishes a rapport between data collectors and users, so that issues can be raised and addressed as they arise, leaving plenty of time for the data collection development process.

The statistics user advisory group is one such mechanism. Advisory groups can be established for issues relating to a specific topic, or to provide advice about a data collection. Topic-based groups can meet two or three times a year to discuss emerging issues relevant to the survey/census topic. Collection-specific groups typically meet during the survey/census development stage.

Membership in advisory groups should cover all stakeholders. In the case of disability survey/census, it should include representatives for all levels of government, academics and researchers, those with disabilities themselves and their advocacy groups.

Box 5.1 describes the Australian terms of reference for its census advisory group for disability.

**Box 5.1: Australian Census Advisory Group on Disability**

The Australian Bureau of Statistics (ABS) convened the Census Advisory Group on Disability (CAGD) to advise it on the requirements for disability data from the 2006 Population and Housing Census in Australia.

The terms of reference of the CAGD were:

1. To identify and prioritise key requirements for disability data for small areas and for a range of population sub-groups, including relatively small subgroups.

2. To guide the research, development, testing and validity analysis of (a) disability question (s) to provide acceptable disability data at a small area level for inclusion in the 2006 Census.

3. To consider the results of testing and validation, and to recommend to the ABS action to be taken with respect to the inclusion of (a) disability question (s) to the 2006 Census.

The advice of this group assisted the ABS in the development of disability questions for the 2006 Census. During consultations, the group advised that the purpose of the disability questions in the census is to identify the disability population in need of assistance with the core activities of self-care, mobility and communication. The group argued that these data were the most important to collect at the fine geographic level available from the Census as they are different, but complementary to, the data derived from the disability survey.
5.4.3 Consult persons with disabilities

In addition to being important data users and those who most benefit from improved data, persons with disabilities and the organizations that represent them often know, better than anyone else, what data needs to be collected. They should be involved from the initial stage of collecting user needs and developing questions, through planning for information campaigns about survey/census goals and results.

In the past, disability organizations have raised the following objections to surveys/censuses proposed by national and local governments:

- Not enough effort was made to use existing data to improve services;
- Survey/census question results seemed designed to make the case for more residential institutions, and other policies that go against the integration of persons with disabilities into society;
- Surveys/censuses violated the privacy of persons with disabilities, particularly those with psychiatric disabilities;
- Surveys/censuses were carried out without consultation with organizations of persons with disabilities.

To help ensure that surveys and other disability data collecting tools are ‘fit for purpose’, it is therefore vital to engage in the process persons with disabilities and their organizations. This can be done in the following ways:

- Ensure that all advisory groups have representation from disability organizations.
- Provide copies of drafts of the survey/census objectives and draft questionnaires (in accessible formats and means) to disability groups.
- Invite disability groups to make submissions on the content of the survey/census or directly request them to provide advice on specific issues that arise in the survey/census development.

All of these methods of consultation are worthwhile and help make a successful data collection. They help to ensure that the results will suit the needs of those disability data users who have direct experience of disability – persons with disabilities themselves.
Box 5.2 provides some examples of countries that have consulted with disability groups when collecting disability data.

**Box 5.2: Consultation with Disability Groups: Examples**

**New Zealand:** During the development of the 1996 Household Disability Survey, the Deaf Foundation were concerned that the survey would not identify deaf people, because deaf people tend not to consider themselves as having a disability and so would not answer 'yes' to the screening questions in the Population Census (used to develop a sample frame for the survey). The Deaf Foundation had threatened to mount a media campaign to discredit the survey results. The government brought the Deaf Foundation into the development process and convinced them that the methodology they proposed would indeed identify deaf people.

**Thailand:** The Thailand National Statistical Office (TNSO) had conducted five Health and Welfare Surveys since the 1970’s. Estimates of the percentage of the population with disability in Thailand had increased over this time from less than 0.5% to approximately 2% in 2001. The TNSO believed the true population of person with disability was still actually much higher than this. To improve their data, the Office consulted both the major users of disability data and persons with disabilities, who helped them to define disability more accurately.

**Australia:** The Australian Bureau of Statistics has consulted with disability groups before each of its disability surveys, and consequently included many of the data requirements suggested by these groups. For example, the Head Injury Advisory Council lobbied the Bureau before the 1993 disability survey to include more information on people with head and brain injuries. They wanted to ensure that such individuals were identified in the general disability population and that the cause of their disability was correctly captured. The 1993 survey was changed to meet this need.

5.5 Privacy implications of collecting data

Prior to data collection, a plan has to be established as to how privacy and confidentiality will be guaranteed. These factors must be considered whenever data about individuals or service provider organizations are collected or disseminated. Privacy and confidentiality may be compromised by what information is collected, how it is collected (e.g., the manner in which the question is asked, and by whom), how the data is stored (e.g., who has access to the data and what protections are in place to stop unauthorised access), and how the data is analysed and reported.

Convention on the Rights of Persons with Disabilities in its article 31 states the process of collecting and maintaining statistical data should comply with legally established safeguards, including legislation on data protection, to ensure confidentiality and respect for the privacy of persons with disabilities.

Legislation or ministerial or departmental regulations may set out privacy requirements for data collections—it is important to be familiar with these rules as the development phase of the data collection begins. Even if no legislation or regulation exists, it is a good practice to follow the following principles:
1. Agencies and other data-collecting organizations should inform respondents of the information they are collecting and the purposes and intended uses of that data.

2. The respondent’s consent – or in the case of someone unable to consent, the consent of a proxy – should always be obtained prior to collecting data, and the individual should be given the opportunity to refuse consent to the use or disclosure of information about them.

3. Agencies and other data-collecting organizations should give people reason to be confident that their privacy will be respected.

Examples of Australian privacy legislation and guidelines relating to data collection can be accessed on the Internet via the following links:


See the following link for the Australian Institute of Health and Welfare’s document on privacy governing its national medical data sets.


5.6 Other basic management issues in planning

During the planning stage, each phase of data collection, process and analysis, and dissemination should be planned in advance, with each phase flowing smoothly into the next. For example, data collection procedures should be developed in the light of the anticipated methods for processing and analyzing the data. And all stages of data collection must be reviewed in the light of the ultimate user and his or her data needs. Statistical information is useful only if it can help answer important research or policy questions. The succeeding chapters provide detailed guides in each planning stage.

Planning also involves basic management issues, such as consideration of the following questions: What is the budget, and is it enough? How will the staff be selected, trained and managed? What project management mechanism will ensure that the project keeps to its timetable?

The overall budget must be divided, and allocated to each phase of the collection. Inadequate monitoring of budgets may mean that data is collected, but no resources are available to produce and disseminate high quality statistical output.

Staff resources need to be planned to ensure that people with the right skills are available at the appropriate times in the development, collection and output phases of the collection. If conducting an interviewer-administered collection rather than self-administered collection, the interviewer workforce is very likely to be the largest single cost in the collection phase.
Project management is essential to ensure a smooth-running collection within acceptable time frame. In the case of disability household surveys, experience suggests that 12-18 months are required for the planning and development stages, with an equal time for data processing, analysis and preparation of survey output. Between these two phases is the data collection period, which, in the case of disability, is relatively short – perhaps only a matter of weeks or months. Disability data is generally unaffected by seasonal influences, and so collection over a longer period is not necessary.

5.7 Quality of survey data and total survey error

Disability surveys and questionnaires are subject to the same general rules about surveys one would find in any standard textbook on epidemiology or survey methodology. The two familiar textbook requirements of good survey data is that it should be,

Valid (measures what it is intended to measure) and

Reliable (gives consistent results over repeated measurements)

Of course, assessing data is not a simple matter. Though reliability is relatively easy to assess, validity can only be definitely determined if there is a ‘gold standard’ against which the data can be measured. Yet in the case of disability data, other than impairment information, it is doubtful whether a suitable gold standard exists.

This is in part why there are various, less demanding standards of validity (construct validity, being the most prominent) for assessing the quality of data.

Surveys, by their nature, attract several varieties of potential error that affect both validity and reliability. There are two sources of error in survey data: sampling error and non-sampling error:

Sampling error arises because survey estimates are based on a sample rather than a complete enumeration of the population, and the sample may not be, for a variety of reasons, representative of the whole population. Sampling error is minimised by increasing the sample size of collections and improving sample design. Some of these issues are discussed below.

Non-sampling error is bias in survey estimates, not traceable to features of the sample that affect the validity of the data collected. Non-sample error is very difficult to measure, and can only be minimised by paying close attention to every step in the process, from survey development, question design, data collection and processing. In a census, since it has no sampling error, all errors are therefore attributed to non-sampling error. Diagram 5.3 sets out some major sources of non-sampling error, grouped by problem area.
We do not have space in this manual to discuss all of these potential sources of error. Specialists in survey methodology are the best people to be on guard against errors associated with frame, non-response and processing. Non-sample errors associated with the specification of the underlying concepts, objectives and data elements are problems that arise in the early development phase, some of which we have already discussed. Finally, measurement errors linked to respondents’ characteristics, interviewers and instruments, are all familiar problems to statisticians and we will mention these only in passing in what follows.

The development and implementation of any survey involves countless decisions, each of which have the potential to enhance or detract from the validity (or precision) of survey estimates. Not all decisions are a matter of choice, however. Better data usually costs more and takes more time and resources.

This is especially true for specialised surveys, such as a disability survey. The more resources that can be invested in training interviewers in the concepts of disability, the better the response rates and data will be. Face-to-face classroom type training is highly effective, but costly. When disability modules are added to existing household surveys, it may be possible to get away with less costly methods such as producing and distributing self-training manuals.
In short, there is a trade-off between cost and high quality data. Inadequate funding is the primary cause of the following major predictors of poor quality survey data:

- Lack of adequately trained staff and interviewers;
- Lack of specialised statistical and methodological expertise;
- Failure to understand and define user needs for data;
- Inability to develop a data collection instrument tailored to meet user needs;
- Inability to put into place procedures to ensure adequate sample design.

As we shall see in Section 8.1, interviewing skills are particularly important for surveys/censuses. Resources spent on training usually pay off in terms of increased quality of data. Additionally, pre-testing of collection instruments, discussed in Chapter 6, is a necessary part of any data collection in order to avoid errors.

5.8 Issues in collection design

Chapter 4 presented a variety of disability question types and made the point that different questions produce different disability statistics. Although determining which type of question is preferable for the purpose of the survey/census is a central design problem, other factors can equally distort estimates of disability prevalence. These are discussed in this section.

5.8.1 Context of disability questions

When disability questions are embedded in non-disability surveys, the background context of the survey (health, income, employment) can affect the disability estimates. It has been shown that when disability data is collected in a health survey, the prevalence rates tend to be much higher than when it is collected in a census, a general social survey or even a survey devoted specifically to the topic of disability. This phenomenon is illustrated in Box 5.3 with the case of New Zealand.

<table>
<thead>
<tr>
<th>Box 5.3: The Context in Which Disability Questions are Asked</th>
</tr>
</thead>
<tbody>
<tr>
<td>The following were disability prevalence rates for New Zealand produced in the same year from three data sources:</td>
</tr>
<tr>
<td>1996 Population Census 14.4 percent</td>
</tr>
<tr>
<td>1996 Household Disability Survey 19.1 percent</td>
</tr>
<tr>
<td>1996/97 Household Health Survey 26.6 percent</td>
</tr>
</tbody>
</table>

The high prevalence derived from the health survey resulted from the fact that respondents, having been asked general questions about their health, health risk behaviours, and their use of health services, were sensitized to health problems and the difficulties they experienced, and were more likely to respond positively to the disability questions. One way to moderate this effect might be to ask the disability questions first, before any mention of health or health problems.
5.8.2 Proxy versus non-proxy survey responses

When information about one person is obtained from another, we call that a **proxy response**. Allowing or disallowing proxy responses is a basic decision about data collection methodology. Censuses usually allow proxy responses, while detailed surveys of disability often do not, requiring the information to come directly from the individual wherever possible. Persons with disabilities are suspicious of proxy responses since, historically, health professionals and others assumed they had the right to speak for persons with disabilities (it is not without reason that the rallying cry of Disabled Persons International is ‘Nothing about us, without us!’)

In the Australian disability survey, proxy responses are allowed for the screening questions, but once a positive response has been made identifying someone as having a disability, further questions are asked of the individual concerned. Assistance in answering the questions is also allowed under the non-proxy reporting scheme.

Statistics Canada, investigating the effect of proxy responses on prevalence, has concluded that proxy reporting tends to lower prevalence rates. Their analysis also showed that, while proxy and non-proxy prevalence rates were similar for moderate and severe disability, proxy-based prevalence rates for mild disabilities were lower than non-proxy rates.

Proxy interviewing may also result in under-recording types of disability for which there is a high level social stigma, in particular disabilities associated with psychiatric illnesses.

It is therefore recommended that direct, non-proxy interviews be used unless the respondent's health condition of intellectual disability makes it difficult to conduct the interview. Care must also be taken in those cases that the proxy is someone who genuinely knows the subject, not merely from a medical perspective, but as a complete individual.

5.8.3 Face-to-face versus telephone interviewing

Face-to-face interviewing results in higher levels of disability self-identification than telephone interviewing. **Table 5.1** shows this difference, by age group, for three Canadian disability surveys.

It is generally safe to assume that higher levels of identification of persons with disabilities, if reliable, are closer to the population true value. Although the difference in prevalence rates may not be of enough concern to warrant the extra cost of face-to-face interviews, still it is recommended, as a result, that this method be used whenever feasible.
Table 5.1: Face-to-face and Telephone Interviewing: Differences in Prevalence

<table>
<thead>
<tr>
<th>Interview Method</th>
<th>Canadian Health and Disability Survey</th>
<th>Testing of Activity of Daily Living Questions</th>
<th>Calibration Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telephone Interview</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-24 years</td>
<td>3.5</td>
<td>5.1</td>
<td>7.1</td>
</tr>
<tr>
<td>25-44 years</td>
<td>5.9</td>
<td>9.5</td>
<td>10.9</td>
</tr>
<tr>
<td>45-64 years</td>
<td>17.1</td>
<td>25.4</td>
<td>28.7</td>
</tr>
<tr>
<td>65 years and over</td>
<td>36.7</td>
<td>56.6</td>
<td>57.7</td>
</tr>
<tr>
<td>All ages</td>
<td>11.7</td>
<td>17.9</td>
<td>20.5</td>
</tr>
<tr>
<td>Face-to-face</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15-24 years</td>
<td>4.4</td>
<td>7.0</td>
<td>6.5</td>
</tr>
<tr>
<td>25-44 years</td>
<td>8.4</td>
<td>10.6</td>
<td>14.9</td>
</tr>
<tr>
<td>45-64 years</td>
<td>23.5</td>
<td>31.3</td>
<td>35.1</td>
</tr>
<tr>
<td>65 years and over</td>
<td>44.3</td>
<td>58.8</td>
<td>55.8</td>
</tr>
<tr>
<td>All ages</td>
<td>16.6</td>
<td>21.6</td>
<td>21.3</td>
</tr>
</tbody>
</table>

5.8.4 Special questions for sub-populations

All countries are composed of sub-populations, defined by culture, geographical area, age group or some other salient feature. Are there reasons to think that members of different sub-population will respond to disability questions differently than other sub-populations?

This question was raised when Statistics Canada was asked to include disability questions on the Aboriginal Peoples Survey of Canada. On the advice of Canada’s aboriginal peoples, a few of the questions on agility were modified to reflect aboriginal conditions in order to validly test the same motor movement within that sub-population.

The Australian Bureau of Statistics had a similar experience when testing the disability module to be used in the Australian National Aboriginal and Torres Strait Islander Survey. Testing showed that not all of the disability items in the standard module were relevant to the indigenous population. The module was improved to make the questions more understandable and relevant to the indigenous sub-population, while keeping the data as relatable to the general disability module as possible.

5.8.5 Collection of disability data about children

Research suggests that, especially when activity or participation questions are asked, disabilities in children may go undetected. Standard activity questions asking about reading a newspaper, carrying a 4.5 kg object for 10 metres or more, or walking 350 metres may not be appropriate for children, especially young children.
Recognition of this problem has lead to the development of age-specific questions, and in some instances, questionnaires. Though the inclusion of age-specific questions or questionnaires raises the cost of a survey, the alternative is not acceptable. Even when specially designed questions are included, there are systematic problems identifying children with disabilities. Fortunately, this is primarily an issue for pre-school age children: once the child reaches school age disabilities are easier to determine and identify.

The source of the problem seems to be the fact that children develop at different rates. One child will be walking at 9 months while others will not take their first steps until well over a year old; some children can communicate quite well by the age of two, while others are three or older before they can talk or understand when spoken to.

Thus, the parent – as proxy for the child – may not be able to tell the difference between a child whose activity limitation is due to normal variation in development, or a long-term disability or developmental delay. Equally, it may be difficult for a parent to determine whether a child has a condition that will correct itself over time, or one that can only be corrected by a technical aid, such as orthopaedic shoes or braces. Parents of children with disabilities may also simply refuse to admit that their child has a disability and dismiss it as slow development.

Mild or moderate disabilities are less likely to be identified. Serious ones tend to be obvious to parents and professionals alike. Other problems may not be noticed until a child begins attending school – for example, behaviour problems associated with attention deficit disorders, or learning or reading difficulties. Unfortunately, once children are diagnosed and ‘labelled’ with some diagnostic term, this is also less likely to be reported during the interview.

If children are included within the scope of a disability survey, it is therefore recommended that additional questions be added. These questions should be

- **developmentally appropriate**, given the child’s age and abilities expected at that age;

- **culturally appropriate**, asking about abilities expected of children in the child’s culture;

- answered by an **appropriate proxy** (usually the parent) who knows the child well;

- worded in **clear and unambiguous language** easily understood by the respondent.

If it is impossible or too costly to add questions of this sort, then the scope of the survey should be limited so that, say, children under the age of 5 are excluded.
The Australian SDAC collects information about children from birth. Parents of children aged 0 to 5 are asked screening questions relating to their children in language modified for the young population group. The questions are part of the main questionnaire with special sequencing for children aged 0 to 5 years. The question on mobility for children aged less than 4 years, for example, reads: “Does (name) ever need more care or help than other children of (his/her) own age to get in or out of a bed or chair?” The question text refers to the same activity asked for the 5 years old and over population, but is modified to ask whether the need for assistance is greater compared to other children of the same age.

Diagram 5.4 presents disability prevalence rate data from the 2003 Australian survey, by age and sex. Disability prevalence rates increased between the 0-4 and 5-14 years age groups and then decreased again for the 15-24 years group. The disability prevalence rate for males aged 5-14 years (12.4%) was almost double that for 5-14 year old females (7.5%)

Diagram 5.4: Disability rates by age and sex, Australia, 2003

Source: Disability, Ageing and Carers: Summary of Findings, 2003 (Cat. no.4430.0).

5.8.6 Collection of disability data about institutionalised populations

The importance of disability information from people living in institutions is discussed in Section 4.2 in the context of the scope and coverage of disability collections. Excluding this population from surveys affects disability prevalence rates, especially for older people who are more likely to be living in institutions for the aged.

Other sorts of institutions, such as prisons, are very rarely included in sample surveys. This is unfortunate since we have evidence that prison populations have high rates of persons with disabilities, especially developmental and intellectual disabilities. Yet, for most countries this population is small and its impact on national disability prevalence rates may not be great.
Since there is usually good administrative data about people living in institutions, as long as this data includes disability information, it may be possible to combine administrative with survey disability data to yield a composite estimate of disability prevalence for the nation.

5.9 Planning for administrative-based data collection

In planning for administrative-base data collection, the same processes involved in any survey or census are also followed. This means that there should also be stakeholders’ consultation. The representatives, however, of advisory groups in this approach should include the agencies or service providers who will be providing the data as well those who will be using the data. It is strongly recommended that representatives of the group who are the target of the data collection – the recipients of the services or programme – be also represented, either individually or by organization.

5.9.1 Identifying information needs

As with any data collection endeavour, the first question to ask is: What information do we need and who is going to use the data?

As mentioned, typically, administrative data is collected for reasons other than getting disability information in general. Among the variety of planning, programme evaluation and monitoring activities for which administrative data is collected are the following:

- to gather information about clients and the services they use;
- to determine needs and unmet needs;
- to satisfy reporting requirements of government departments, boards or other regulatory bodies;
- to gather information about, and to report on, programme expenditures as related to populations served;
- to establish a benchmark for service providers, and to compare this with comparable benchmarks nationally or internationally.

Knowing broadly what kinds of data are required, the next step is to get potential users of the data to be more specific about the information they need.

If, for example, the data user really wants to know whether the number of people with intellectual disabilities accessing the service is increasing, or how many service users have a primary carer, the questions will have to be designed to identify these sub-populations.

If the only people using the data will be policy planners, then more general information about numbers of service users, in broad categories, will suffice.
If the data will be used by researchers who wish to explore service usage issues not usually highlighted, then perhaps more exploratory questions need to be considered.

Once the overall purposes and the types of information needed are agreed on, the next question is whether this information already exists in some other form. Given the costs of data collection, it is well worth investigating this possibility before launching into another data collection.

The next step is to define the scope of the collection – who or what you want to collect information about. Are you interested in tracking service usage? If so, the relevant population will be people with a disability who use the service. Do you want to monitor the hours staff spent caring for clients? If so, your population should be the staff not the clients.

5.9.2 Developing data items

Once the information needed has been clearly articulated, it is time to decide what data items will produce this information.

Data items should focus on the information actually required. Try not to collect information that is not needed, and aim to collect the minimum amount of information required. Where possible, existing data standards should be used as a basis for developing data items.

Think about the analyses you will want to conduct using the data since this should determine the data items you need, and the coding categories you will need to use to capture information. For example, if you want to know the age distribution of clients, you will need to record date of birth. If all you require is general age information, then it may suffice to use broad age groupings (‘0–14’, ‘15–25’, and so on). Using age groupings is less invasive to privacy, but then again may not support the detailed analysis you would demand.

5.9.3 Deciding on collection methods and materials

Consider the simplest methods of collection first. It might be easier to contact each service user directly and ask them one or two questions than to devise an elaborate three-page survey that is posted out for completion. The key considerations on data collection methods are:

- Should data be collected using pencil and paper (e.g., a standard survey form) or electronically?
- Will data collection be part of an on-going update of an information system or a periodic data capture?
- Who will collect the data (e.g., service users, service staff, or hired data collectors)?
If you plan to collect a substantial amount of data, it is probably better to use an electronic tool to capture and store it. This may entail the use of an off-the-shelf spreadsheet software or a purpose-built database. It is always advisable to test the tool to make sure it is fully operational, complies with your data needs, and is easy to use.

Consider using existing systems, it may be possible to use information collected on standard client intake form, which uses an electronic system to collect data, and simply add questions about staffing, finances, or whatever additional data you need. Alternatively, the data required may not be collected by a single administrative agency, but can be pieced together by joining together several administrative collectives (for example, by utilising a single client identifier for all the systems).

When developing your own administrative data collection service intake form, staff hours form, or whatever you may require, it is important to keep in mind that:

- you need to phrase the questions to make them clear and be able to gather what you are asking – try not to ask too much in a single question;
- you need to consider who will be asking the questions, how the questions will be asked and in what surroundings – tailor the data collection to the real-situation in which the data will be collected;
- all questions should be tested — no matter how obvious the meaning may seem to the person who wrote it, the question may still be ambiguous or vague to the respondent. You can only find this out by testing the questions with people who are part of your target population (detailed discussion on pre-testing is presented in the next chapter).

5.10 Sample design

To carry out a survey of persons with disabilities you have to find them. This is not always straightforward, and involves the fundamental issue of sampling. Sampling is a major consideration in the design of a reliable, and affordable, disability survey whose primary object is to determine prevalence. In this manual, it is only possible to review, in a general fashion, some considerations relevant to disability surveys. The reader should refer to the UN Guidelines and Principles for the Development of Disability Statistics, which reviews in far more detail the complex sampling issues raised by disability.

5.10.1 Sampling frame

Since disability prevalence is relatively low, and certainly less than 20 percent, large samples are usually needed to measure it reliably. If the survey is to
yield other information about distribution, types and causes of disability, or estimates for sub-populations, then the sample size has to be even bigger. Given cost considerations, most national statistical agencies are therefore forced to seek ways of increasing the efficiency of sample design. Usually this means making full and efficient use of existing sampling frames – that is, lists, registries or other materials from which samples are derived.

The most common forms of sampling frames are those based on geographical area. A typical area frame is a set of enumeration areas developed out of the most recent population census, which supplies basic information about the number of people and households in the area. If one estimates a prevalence rate for disability, say 10 percent, then the size of the sample, for that area can be calculated easily. Applying conventional survey sampling techniques such as stratification, clustering and systematic random selection, can yield a national sampling strategy.

Unfortunately, relying entirely on area frames is quite inefficient – and so expensive – for relatively rare phenomena such as disability. Many households have to be interviewed, on average, to locate a single person with a disability. To improve efficiency, it is highly recommended, wherever possible, to combine the use of area frames with others, and in particular list frames.

List frames are derived from administrative and other data that specifically target sub-populations (e.g. the elderly, children in orphanages, the institutionalised, or long-term care service users) and are updated regularly. Persons with disabilities, compared to people living in households, are more likely to be included on these lists. Other lists may also be available, for example a registry maintained by a national organization for the blind.

Since area frames tend to be complete but inefficient, and list frames are incomplete but highly targeted and so efficient, a sensible sampling strategy is to combine both frames. This is not without its own costs, however. For example, effort would have to be made to ensure that the list frames are up-to-date, and some mechanisms for eliminating duplication would have to be designed and used.

Another strategy is to conduct a small screening survey in an area to obtain a sample of households with persons with disabilities for a follow-up survey that is far more detailed. This strategy was used by Thailand and India to obtain a list of households with persons with disabilities. Unfortunately, sometimes the initial screening question used to identify households with persons with disabilities does not work well, and identifies only a small subset of the population of person with disability.

Statistics Canada has developed an approach that improves efficiency of sample selection and reduces survey costs. They use the Population Census to identify respondents for a follow-up survey on disability. In order to use the census to filter a population, a suitable disability question was devised and workable mechanisms to carry out the follow-up survey were put into place.
The census form was modified to inform the respondents that their responses would be used to select people for a follow-up disability survey, and they were asked to provide their telephone number in case they will be selected. The approval of the Canadian Privacy Commissioner had to be obtained to use census results in this way. (See Section 3.5 for more details of the Canadian strategy.)

5.10.2 Sample size and unit

Once a successful sampling strategy has been adopted, the next question is: how big should the sample be in order to yield data with an acceptable level of sampling error? This is a highly technical matter that is the job of an experienced survey methodologist. The UN Guidelines and Principles for the Development of Disability Statistics has a discussion of sample size.

In general, the sample size required to produce estimates of disability prevalence with acceptable levels of error is a function of the prevalence levels of the other characteristics or phenomena of interest to the survey designer, as well as the level of dis-aggregation of the data that is required. As a rough rule of thumb, surveys that have sufficient size to produce estimates of unemployment in a specified geographical area are likely to be large enough to produce rough disability estimates for the same area. The sample sizes of some recent disability surveys are provided in Table 5.2.

Another important decision, concerning sample unit, is whether the survey should sample at the level of the person or the household (or institution). Canada and New Zealand used the person as the sample unit and they were able to identify persons with disabilities by their census screening. When this information is not available, then the sample unit is usually the household, which is suitable for area based samples.
<table>
<thead>
<tr>
<th>Country</th>
<th>Collection</th>
<th>Sample size</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>India</td>
<td>Survey of Disabled Persons</td>
<td>70,302 households.</td>
<td>The sample size is the number of households with someone with a disability identified in the listing sample.</td>
</tr>
<tr>
<td>Indonesia</td>
<td>National Socio Economy Survey, 2003</td>
<td>67,000 households</td>
<td></td>
</tr>
<tr>
<td>The Philippines</td>
<td>2001-2003 Philippine Disability Survey</td>
<td>59,443 households</td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>2003 Survey of Disability, Ageing and Carers</td>
<td>14,000 households, 300 non-private dwellings, 500 establishments</td>
<td>Sufficient sample size to produce prevalence by State, as well as cross-tabulations of population by age, sex and disability status.</td>
</tr>
</tbody>
</table>