This chapter covers the use of specialized household surveys to collect data to study the determinants and consequences of international migration. Previous chapters have described the data collection systems producing information on international migration in general or on special types of international migrants. Although all the systems considered so far often gather some information that is relevant for the characterization of international migrants— in terms of sex, age, citizenship, education, date of arrival, country of previous residence, or occupation— the information gathered is too limited to allow the in-depth analysis of the likely causes of international migration or of its consequences for the persons involved. In addition, most of the data systems discussed earlier do not collect data on international migrants departing, and use restrictive definitions of international migrant, thus missing many migrants. Finally, none of those systems collects information on the situation of the migrants prior to migration, which will be seen below to be vital to understanding both the determinants and consequences of migration for international migrants. In particular, the most commonly available source of data on international migration, the population census, suffers from a narrow definition of international migrants (the foreign born); lacks information on out-migrants; collects very limited information, and nothing on the pre-migration situation of migrants; and usually identifies only the stock of lifetime migrants rather than recent migrants.

Specialized surveys constitute the best data collection system to gather the information needed to carry out a proper examination of the determinants and consequences of international migration. Although several types of surveys are reviewed, the focus is on household surveys. A household is usually defined as a group of persons who share the same living accommodation; who pool some, or all, of their income and wealth; and who consume certain types of goods and services together, such as food and shelter. Household surveys seek information from households as units of consumption, production, income sharing and decision-making. Most household surveys also include special schedules to collect information regarding particular members of the household or persons linked to a household in special ways. Surveys focusing on international
International migration statistics

migration can include special schedules or sections for members of the household who are international migrants or for persons who used to be members of the household but who have left the country in which the household is located to settle or work abroad. The design of surveys to quantify and assess the factors leading to the international migration of individuals and households, and to understand the effects that migration has on the persons involved, their households and the communities to which they belong is the main topic of this chapter.

Before addressing survey design, section A reviews selected existing survey approaches to the collection of data on international migrants, and their limitations. Section B presents the definition of international migrant recommended here for specialized surveys collecting data for the analysis of the causes or consequences of international migration, and on that basis discusses the identification of appropriate comparison groups for analysis. Section C provides guidelines for sample design for generalized surveys of international migration, and section D discusses certain more limited special approaches. Section E identifies and explains the need for information on many factors for the analysis of the determinants and consequences of international migration. Finally, section F introduces the set of model questionnaires for specialized surveys on international migration, which are presented in the annexes to this book.

A. USE AND LIMITATIONS OF EXISTING SAMPLE SURVEYS

1. General purpose surveys

General purpose household surveys are a potentially useful source of information on international migrants when they have large sample sizes and are conducted in countries where international migrants constitute a sizeable proportion of the population. The term "general purpose survey" is used here to indicate that the main focus of the survey is not the study of international migration. Most existing surveys focus on specific topics, such as labour force and employment, fertility and health, or income and expenditure. In the majority of countries, such surveys have sample sizes that are too small to yield statistically reliable data on international migrants. Thus, if we consider international migrants to be persons born outside the country in which they live, their share of the population is less than 6.5 per cent in three quarters of the independent countries of the world (United Nations, 1995a). Since the typical size of most nationally representative household surveys is in the range of 5,000 to 10,000 households, assuming that there are five persons per household and that the proportion of foreign-born persons in the population is 3 per cent, the expected number of international migrants would range from 750 to 1,500, of whom perhaps 300 to 600 would be economically active adults. While numbers
Design of surveys to investigate international migration

of that magnitude may provide some indication of the characteristics of the foreign born, the problems bias and high standard errors inherent in such small numbers are always present. These risks are magnified if, instead of considering all the foreign born, only those who had arrived within a specified period (such as five years) preceding the survey are to be the focus of analysis. The desirability of concentrating on recent migrants when analysing the causes or consequences of international migration implies that general purpose surveys are usually not useful because of the small numbers of recent migrants covered. In addition, general purpose surveys typically include few questions allowing the identification or characterization of international migrants, nor do they contain questions about the pre-migration experience of migrants or their adaptation. Although it is sometimes possible to add pertinent questions to general purpose surveys to discuss these limitations, this is recommended only in surveys with large sample sizes carried out in countries with high proportions of international migrants. Otherwise, the small numbers of migrants likely to be encountered do not justify the expense.

Examples of national surveys which may be large enough to capture sufficient numbers of international migrants for meaningful analysis include the monthly United States Current Population Survey (60,000 households), the annual National Population Survey (PMAU) of Brazil (65,000 households), the National Sample Survey of India, and the Labour Force Surveys of European Union countries (which have sample sizes of 60,000-100,000 for the larger countries and 10,000-50,000 for the smaller ones). The Current Population Survey (CPS), which has been carried out every month since 1947, illustrates how general purpose surveys may be used to analyse some aspects of international migration. The CPS routinely gathers both basic demographic information and data on labour force participation and employment. The place of birth of each household member is recorded, as is the place of residence 12 months before interview in the survey carried out every March. Starting in January, 1994, country of citizenship has begun to be recorded. The CPS is a panel survey in which each household in the sample is interviewed over four consecutive months, then excluded for four months, and interviewed again for an additional four months before being dropped permanently from the sample. This scheme is meant to minimize interviewee fatigue while ensuring continuity and the comparability of results over time. There is a 75 per cent overlap of sampled households from one month to the next and a 50 per cent overlap from one year to the next. Consequently, the data relative to a particular household can be compared over a maximum interval of 11 months. That is, for the international migrants covered by half the CPS sample – 30,000 households – the data gathered have the potential for allowing an analysis of the short-term consequences of migration. Indeed, an analysis of the annual changes experienced by newly arrived cohorts of international migrants could provide insights about the short-term consequences of migration and the integration process. Changes in the status of migrants could be assessed in relation to the economic conditions prevalent in the United States during the relevant period or in
International migration statistics

relation to equivalent changes among non-migrants. This example illustrates the potential for using existing large-scale national surveys to study some limited consequences of international migration, but this potential has rarely been exploited.

The 1979 Population, Labour Force and Migration Survey of Pakistan (PLM) illustrates how the addition of a special set of questions to an on-going survey can allow a better characterization of international migration (Irfan et al., n.d.). Questions were added to two rounds of the national Labour Force, Income and Expenditure Survey. The head of household was asked to indicate whether any member of the household had ever migrated to live elsewhere since December 1971, when war with India erupted. Anyone moving abroad or returning from abroad within the 8 years preceding the interview was identified as an international migrant. The survey covered 10,242 household members, 0.15 per cent of whom were identified as return migrants and 0.48 per cent as international out-migrants. The information recorded on migrants was limited to age, sex, dependency status, year of departure or return, and labour force participation while abroad. This case suggests that, although there are significant cost advantages in latching onto an existing survey, the peculiarities of international migration demand special approaches because international migrants are rare elements in the population and are seldom represented satisfactorily in general purpose samples. In addition, the questions that can be added without disrupting the main function of an existing survey are usually too limited to allow more than a superficial characterization of international migrants.

Even when general purpose surveys include some questions on international migration, their use to study the consequences of migration for the migrants themselves is limited. In this regard, general purpose surveys have limitations similar to those of censuses. Thus, although it is common to use census or survey data to compare the status of international migrants with that of non-migrants in terms of occupation, income, unemployment levels, fertility, etc., and to interpret similarities between the two groups as indicative of the migrants' success, in reality such similarities reflect, at best, only the process of integration. To assess the consequences of migration for the migrants themselves, data are needed on their status at the time of migration so as to compare it with that at the time of interview (migrants could conceivably suffer a deterioration in their status even if they appear similar to non-migrants). An additional problem is that the status of the two groups can be compared only for the limited set of characteristics covered by general purpose surveys or censuses, and that both data sources fail to reflect the experience of migrants who have subsequently left or died. Studies focusing on return migration have shown that migrants who return to the country of origin tend to be those who are least successful in the country of destination.

To conclude, although large-scale general purpose surveys can provide some limited information relevant for the study of international migration, they have serious limitations in producing estimates of the stock or flow of
Design of surveys to investigate international migration

international migrants into a country, since such estimates are generally subject to high variability. Yet in countries lacking other information on international migration, surveys have occasionally been used for that purpose. That is the case, for instance, of the surveys carried out in seven West African countries in 1993 under the coordination of CERPOD (1995).

2. Surveys of international migrants

Rather than households, certain international migration surveys focus on persons who cross or are about to cross international borders. A major problem in using data from border crossing at exit/entry points is the sheer volume of movements that take place, the vast majority for purposes other than to change residence. It is therefore difficult to find migrants among movers. In the case of the United States, for instance, there are about half a billion entries a year but less than a million persons are admitted as immigrants. The design of surveys of movers must confront the additional problem of the lack of an appropriate sampling frame. The surveys described below use innovative procedures to attempt to deal with these problems.

In Pakistan, the need for information on migration to the Middle East led to the implementation of an International Migration Project which included a survey of out-migrants at the major ports of departure (Gilani et al., 1981a). One of the aims of the survey was to identify families left behind; a second survey was conducted on these families to study the impact on them of temporary labour out-migration (Gilani et al., 1981c). Out-migrants were defined as workers who had departed for the Middle East within the two years preceding the survey. Since it was thought that most travelled by air, it was decided to undertake a survey of all passengers departing from the three international airports of Pakistan and to identify those leaving specifically to work in the Middle East. The survey was carried out during September-November 1979, and led to 12,516 male out-migrants being interviewed. Their age, occupation, place of residence in Pakistan, country of destination, and expected length of stay abroad were recorded. This information was then used to create a sampling frame to select households in Pakistan with migrant workers abroad for the second survey. The household survey gathered information on 1,710 households located in 250 villages and 50 cities and towns throughout Pakistan, 1,153 of which had migrant members (Gilani et al., 1981b). The procedures used to select households are unfortunately not described in the project documents, and the mean number of households per community was less than six, implying a high cost of data collection per household. A major problem in implementing the household survey was locating the addresses provided by departing migrants. Nearly 2,400 households had to be visited to complete 1,153 interviews in households with migrants, casting doubt on the representativity of the sample. Nevertheless, the approach followed has some novel features worth highlighting,
especially the creation of a frame to select a sample of households on the basis of information provided by departing migrants. Such an approach may be useful in cases where most out-migration consists of individuals leaving their families behind and where the aim is to study the effects of migration on those families. The study also has the virtue of including a comparison group of non-migrant households.

Pakistan has also undertaken a survey of male migrants returning to Pakistan on incoming flights from the Middle East (Azam, 1994). The survey was carried out during April 1993 and obtained the following information on Pakistani men who had worked abroad: age; occupation; marital status; community of residence in Pakistan; country of employment; length of stay abroad; whether the migrant's work contract had expired or whether the migrant was returning home only for a short visit; whether the migrant had a written contract with an employer abroad; if so, whether it had been signed in Pakistan before departure and whether it had been processed by the Protector of Emigrants Office. Because in 1992 the Government of Pakistan stopped requiring that Pakistani citizens undergo immigration clearance upon re-entry into the country, the logistical problem of how to identify returning Pakistani citizens at the airport had to be addressed. Flight manifests were used to determine the total number of people on a flight as it arrived; women, children and foreigners were identified as they entered the arrival lounge and the remaining persons were screened to identify migrant workers eligible for interview. Interviewers were instructed to interview at least 20 per cent of the migrant workers on each flight, but managed to interview an average of 48.5 per cent. However, no attempt was made to ensure random selection. Out of a total of 36,155 eligible passengers, 17,524 were interviewed, 13,899 of whom were returning from working abroad. Among the latter, 5,170 were returning with the intention of staying in Pakistan. Although the survey has the virtue of being low cost, it is not based on a representative sample and its results cannot therefore be interpreted as indicating the experience of all return migrants. In addition, by covering only a given month, it cannot reflect differences from one time or season to another, which may be substantial.

Lastly, a survey on the northern border of Mexico, Encuesta sobre migración en la frontera norte de México, has been carried out since 1993 (San-tibáñez-Romellón et al., 1994; Bustamante et al., 1994) and covers both inflows and outflows. The sampling universe is constituted by border crossings, that is, the survey aims at measuring actual flows. To obtain appropriate measures, space and time are divided into specific units each of which is assigned a probability of coverage according to its importance in terms of the flows that take place over that space and during that period. In terms of space, the 28 main border crossing points/areas were selected along the Mexico-United States border and the number of vehicles crossing at each was observed to determine relative weights for the observations to be sampled from each point. Sampling sites where migrants arrive or depart were then identified in each crossing area (e.g. highway points with immigration posts, bus terminals, train stations, and
Design of surveys to investigate international migration

airports). The survey was planned so that each day two crossing areas were selected as locations for the survey, with two interviewers assigned to each site. Results are just beginning to emerge. Information on the dynamics of population movements from Mexico to the United States, including migrants crossing the northern border of Mexico is being collected. The results are likely to be useful for characterizing mobility across the Mexico-United States border, especially temporary and return migration. However, by covering only migrants, the survey lacks an adequate comparison group that would allow the study of the determinants of international migration.

3. Specialized surveys on international migration

A number of surveys focusing on international migration have been conducted over the past decade that illustrate both the potential usefulness of specialized surveys and the shortcomings inherent in the designs adopted by most of them. The Economic Commission for Europe, for instance, recently coordinated surveys in three countries of central and eastern Europe – Lithuania, Ukraine and Poland – aimed at the study of out-migration and short-term international travel originating in those countries. Although there is some variation across the three surveys, they are generally based on samples that are not representative of the population of origin (even of the few communities selected) and also suffer in data quality from using proxy respondents to obtain information on migrants absent from the household (Frejka, 1995; Mullan, 1995; Sipaviciene et al., 1995).

In Western Africa, CERPOD has coordinated a major survey programme entitled Réseau d’enquêtes sur les migrations et urbanisation en Afrique de l’ouest through which migration surveys have been carried out in Burkina Faso, Côte d’Ivoire, Guinea, Mali, Mauritania, Niger, Nigeria and Senegal in 1993 (CERPOD, 1995). All surveys are based on nationally representative samples and are aimed at both measuring migration flows as well as the characteristics of migrants. Full migration histories are recorded for the persons interviewed. The eight country surveys covered nearly 100,000 households but, because the sampling designs made no special effort to identify international migrants, the numbers of the latter are small. In addition, project documents do not explain clearly how samples were selected and problems of sample design may affect the data gathered (see, for instance, Senegal, n.d.). The CERPOD surveys do have the advantage, however, of covering both migrants and non-migrants, thus providing a basis for useful comparisons that may shed light on the determinants of migration. Furthermore, because the surveys simultaneously covered countries linked by major migration flows (such as Burkina Faso and Côte d’Ivoire; and Mali and Senegal), they may permit a more comprehensive analysis of the causes and consequences of international migration than any other data collection effort to date in Africa. However, the lack of detailed data
International migration statistics

on international migrants reduces the potential usefulness of the surveys for studying them.

Another project involving surveys in several countries is being carried out by the Netherlands Interdisciplinary Demographic Institute (NIDI) and Eurostat and is expected to be completed in 1997. The project aims at collecting the data needed to study the determinants of international migration to selected countries of the European Union. Surveys will be carried out in both countries of destination (covering only households with international migrants) and in countries of origin (covering households with and without out-migrants and their communities). The countries of origin where surveys will be carried out include Egypt, Ghana, Morocco, Senegal and Turkey; countries of destination are France, Germany, Italy, the Netherlands and Spain. Each country of origin has significant flows to at least two of the destination countries included in the study, and each country of destination has experienced large inflows from at least two of the countries of origin. Detailed information on international migrants is to be collected in single-round surveys in each country. To keep costs down, areas where international migrants are known to originate or settle are targeted a priori, without formal probability sampling. Sample sizes are to be around 2,000 households in each country of origin and 500-1,000 households per migrant group in each country of destination. The design of the project is consistent with the systems approach to the study of international migration (Zlotnik, 1992; Kritz and Zlotnik, 1992; Bilsborrow and Zlotnik, 1994), and has an overall survey design generally consistent with that recommended in this chapter (see section B). Its results will therefore be of particular interest both from a methodological and a substantive perspective.

The use of longitudinal surveys to study the consequences of international migration should also be mentioned. Australia has begun a Longitudinal Survey of Immigrants which involves interviewing international migrants both before and after they move to Australia. The survey began with annual cohorts of 3,000 immigrants. Principal applicants complete a detailed form prior to being accepted as immigrants. After arrival in Australia, they are interviewed four more times, after one month, one year, two years and five years. The data should be useful for analysing the integration of migrants and the consequences of migration for the migrants themselves if sample attrition is low.

In the United States, a similar survey is being proposed and tests of its feasibility are under way. The survey will have as its universe all persons granted permanent resident status during a given year. A sample of 13,000 immigrants will be selected from a sample frame based on the administrative records of the Immigration and Naturalization Service, which are processed to grant permanent residence. Immigrants selected will be contacted by telephone soon after they obtain permanent residence to ascertain whether they are willing to participate in the survey and to obtain information to trace them later. In-depth interviews will be conducted soon after permanent residence is granted and then annually over a two-year period. Detailed information on the pre-migration and post-migration experience of the migrants will be collected. The
Design of surveys to investigate international migration

data gathered are certain to be useful in assessing the short-term integration of legal immigrants in the United States and the consequences of international migration for the migrants themselves. However, as in the case of other longitudinal surveys covering only international migrants (including the Australian example above), the lack of information on non-migrants in the country of origin prevents an adequate assessment of the causes or consequences of international migration.

B. DESIGN OF SURVEYS FOR THE ANALYSIS OF THE DETERMINANTS AND CONSEQUENCES OF INTERNATIONAL MIGRATION

1. Defining international migrants

The proper design of surveys requires a clear definition of the population of interest. Therefore, a crucial step in designing international migration surveys is defining international migrants for purposes of the survey. As documented in other parts of this book, the characterization of international migrants varies considerably from one data source to another, and existing surveys are no exception. Indeed, given the flexibility that surveys offer in terms of the type and depth of information they can gather, there is ample room for considerable variability in the definitions of international migrant used in survey instruments. There is no single definition that will satisfy all needs and, consequently, the best a survey can do is to make explicit which group is the focus of attention, though data may still be collected for other international migrants and non-migrants as well.

As discussed in Chapter 2, there are three key criteria for the identification of international migrants: place of birth, citizenship and place of residence. Thus, international migrants can be considered persons whose country of birth is different from that in which the survey is being conducted; persons who do not have the citizenship of the country in which the survey is being conducted; or persons who have changed their place of residence from another country to that in which the survey is being conducted. In general, surveys use some version of the third formulation to identify international migrants of interest. For the purposes of analysing either the causes or the consequences of international migration, a change of residence from one country to another is clearly the critical event of interest and must be considered the main criterion for the identification of international migrants. However, leaving aside for the moment the problems inherent in the definition of “residence”, it must be recognized that not all persons who change country of residence are equivalent from the analytical perspective. Thus, a person born in country A, who has always lived in country A and then moves to country Z cannot be considered similar to a person who is born in country Z, moves to country A and then returns to country Z. That is, return migrants must be distinguished from persons arriving
International migration statistics

in a country for the first time. Furthermore, distinguishing the inflows of persons who “belong” (insiders) to a country from those of persons who do not “belong” (outsiders) is crucial from the policy perspective. As stated in other parts of this book, the main marker of “belonging” is citizenship, an attribute of international migrants that should not be ignored.

Surveys gather information directly from the migrants themselves who have moved from one country to another or indirectly from (proxy) respondents who provide information on the persons who have moved and to whom they are related. In terms of retrospective information, “establishing residence” is normally taken to mean living in a place for at least a certain length of time. A period of six months or a year is usually adopted as the cut-off point. In the model questionnaires presented in the annexes to this book, six months has been used as the minimum period of stay to establish residence. Note that this definition tends to equate residence with presence and does not involve any legal considerations. Since the questions actually posed to respondents (see below) are phrased in terms of “having lived in another country for at least six months”, there is little danger that the specific meaning of residence recommended here will be misconstrued in practice.

Having established what “change of residence” means for survey purposes, it is necessary to assess which changes of residence matter the most. To be most useful, the results of a survey should focus on current events so that they can provide timely information on the factors that shape them or on the consequences of such events. It is therefore advisable to concentrate on international migrants who have changed residence over a recent period preceding the survey. The choice of a cut-off point for that period is not obvious: The further the cut-off point from the date of the survey, the less likely that events encompassed are relevant for an analysis of the current situation. However, the closer the cut-off point to the date of the survey, the smaller the proportion of persons who would have changed residence during that period and hence the greater the difficulty of encountering a sufficient number of them. In addition, data quality considerations argue against adopting a cut-off point that is set too far in the past. Since respondents will be asked to provide detailed information regarding both the period immediately preceding the most recent change of country of residence and the period following that change, the further that event is in the past, the more likely the data will be affected by recall errors (Som, 1973; Bilsborrow et al., 1984, Ch. 4). Consequently, despite the problems associated with locating an adequate sample of recent international migrants, it is strongly recommended here that attention be focused on persons who have changed their country of residence within 5 or at most 10 years preceding the survey. The model questionnaires presented in the annexes assume a five-year cut-off point for the identification of international migrants subject to intensive interviews. Those migrants also constitute the main subjects for analysis based on the information gathered, whether to study the determinants or the consequences of international migration.4

246
Design of surveys to investigate international migration

To sum up, it is recommended that specialized migration surveys focus attention on international migrants defined as persons who have lived for at least six months in a country other than that in which they are being interviewed and whose move into the country of interview occurred during the five years preceding interview. This definition assumes that international migrants are being interviewed themselves in their country of current residence. When migrants cannot be interviewed because they have emigrated from the country in which the survey is being undertaken and information on them is being obtained from a proxy respondent, the definition has to be modified as follows: an international migrant is a person who used to live in the country in which the interview is being conducted and was a member of the household of the person being interviewed but who left at some point during the five years preceding the interview to live abroad for at least six months. These definitions provide the basis for the formulation of the questions needed to identify international migrants included in the model questionnaires presented in the annexes.

Once international migrants are identified, it is important for analytical purposes to distinguish return migrants from other international migrants. In previous chapters of this book return migrants have been identified as persons moving into their country of citizenship. Following that approach, if a survey is conducted in country A, a citizen of A who has lived in country B during at least six months and enters A during the five years preceding the interview with the intention of staying would be considered a return migrant irrespective of whether he or she had lived in country A before. Because, according to the nationality laws of most countries, few people can obtain the citizenship of a country without having lived in it, citizenship usually indicates previous presence in the country. Exceptions may arise, however, in the cases of countries that grant citizenship to descendants of previous emigrants or to persons with certain backgrounds. In such cases, the identification of “return migrants” may have to be based on both citizenship and place of birth, with a “return migrant” being a citizen born in the country of interview who has lived outside that country for at least six months and returns to live in it again during the five years preceding the interview. The use of citizenship as an organizing criterion is crucial because there are limits on the extent to which governments may regulate the international mobility of their own citizens. Therefore, the policy instruments that can be used to influence the international migration of citizens are generally different from those used in the case of foreigners. Consequently, if the analysis of factors influencing international migration flows is to have policy relevance, it is essential to make a distinction between those flows that are subject to full government regulation and those that are not.

In conclusion, specialized surveys on international migration should focus on international migrants characterized as persons who have changed their country of residence during the five years preceding the survey. If the survey is meant to cover inflows, it will encompass citizens as well as foreigners moving into the country of interview during the relevant five-year period. The former will be denominated return migrants (though this will also include persons with
International migration statistics

residence rights in country A who migrate to country B then return to country A). When the survey is aimed at obtaining information about migrants who are still abroad, it will focus on persons who left the country within the relevant five-year period, thus including both emigrating citizens and foreigners. Normally, emigrating citizens will be the object of study and emigrating foreigners may be disregarded.

To follow this approach, information on place of birth, citizenship, country of previous residence, and time of the most recent change of residence should be recorded for all persons interviewed. In addition, any change of citizenship and its timing should be recorded for each international migrant interviewed so as to ascertain whether the most recent change of residence took place before or after a change of citizenship.

2. Identification of appropriate comparison groups

A key decision that needs to be made at the outset in designing specialized surveys on international migration is whether the data collected are to be used for the analysis of the determinants or the consequences of international migration. This decision determines the most appropriate locus (country or countries) in which to carry out the survey and the group or groups of non-migrants that are to be used as the comparison or control group. There is a great deal of confusion in the literature regarding the data needed to investigate the determinants or consequences of international migration, with the result that the serious limitations of existing micro-level studies are not recognized. This section aims at elucidating the problems involved and makes explicit how the compromises typically made in the analysis have implications for the validity of inferences that can be drawn from particular types of surveys.

The discussion in this section assumes that international migrants have been precisely defined (for instance, following the guidelines in the previous section) and that a country can be characterized as being either a country of origin or a country of destination for the international migrants of interest. In addition, there is an implicit assumption that the country of origin of the migrant is the country of citizenship, although the framework presented below is relevant also for other conceptualizations of country of origin. In reality, every country has the dual role of being the country of origin for some international migration flows and the country of destination for others. The artificial dichotomy here is common in the analytical literature, and is used precisely to make evident the country perspective.

Perhaps the most straightforward approach for assessing the consequences of international migration for the migrants themselves involves a single survey in the country of destination which interviews international migrants to gather information on their status just before migration as well as at the time of interview. However, the need to minimize recall problems can make this
Design of surveys to investigate international migration

Another approach usable only when migration is a relatively recent event. Another approach to the collection of data to permit a comparison of the status of migrants over time is a longitudinal or panel survey, which reduces problems associated with recall but substantially increases the cost of data collection. Although it is possible to assess the consequences of international migration for the migrants themselves by examining changes in the migrants’ situation over time using either of the two approaches above, such an assessment represents only a partial view of the whole picture and may even be misleading. Thus, suppose an index indicating the migrant’s status in the country of origin just before migration has a value of 100 and that it rises to 150 five years after migration. The migrant is clearly better off. But suppose that for non-migrants remaining in the country of origin the status index changed from 70 to 140 over the same period. A comparison of the migrant’s status five years after migration with that of non-migrants at the same time would still find the migrant better off, but in relative terms the migrant would have gained less from migration than non-migrants had gained in the country of origin. This example illustrates the importance of assessing change not only with respect to the migrant’s own status but also in relation to appropriate reference groups. The remainder of this section discusses the key issue of selecting appropriate comparison groups according to the purpose of the survey, focusing first on the study of the determinants of international migration and then on its consequences.

Survey requirements to study the determinants of international migration

Table 6.1 presents in schematic form the various possible types of analyses relevant for the study of the determinants of international migration, together with what their requirements are in terms of data collection in the country of origin and the country of destination. The term “focal group” is used to denominate the population covered by the survey that makes possible the type of analysis indicated. An assessment of the quality of each type of analysis is provided on the basis of the adequacy of the focal groups involved and the likelihood of obtaining reliable information from those groups. Thus, whenever information on international migrants has to be obtained from proxy respondents because the migrants are not themselves present in the locus of the survey, the option is considered “barely acceptable”. The adequacy of focal groups for a particular type of analysis is determined by whether they represent the population at risk of international migration. The greater the overlap between the combination of focal groups for a particular type of analysis and this population at risk, the more preferable the option.

Nine possible types of analysis of the determinants of international migration are distinguished. The first three refer to the individual level and the next three to the household level. A further two deal with the determinants of return migration and the last one covers the determinants of potential migration. Cases are further differentiated according to whether they deal with a single country of origin, a specific pair of countries linked by migration (one country of origin
### Table 6.1: Focal groups for data collection at origin and destination to study the determinants of international migration

<table>
<thead>
<tr>
<th>Type of analysis</th>
<th>Quality of option</th>
<th>Focal groups for data collection at origin</th>
<th>Focal groups for data collection at destination</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Determinants of out-migration of individuals from country A</td>
<td>Barely acceptable</td>
<td>(a) Households in country A with members who have left the country (proxy respondent necessary)</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Other households in A</td>
<td></td>
</tr>
<tr>
<td>2. Determinants of out-migration of individuals from country A to country Z</td>
<td>Acceptable</td>
<td>Non-migrants in country A</td>
<td>Migrants from country A in country Z</td>
</tr>
<tr>
<td></td>
<td>Barely acceptable</td>
<td>(a) Households in A with members who have migrated to Z (proxy respondent necessary)</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Other households in A</td>
<td></td>
</tr>
<tr>
<td>3. Determinants of out-migration of individuals from country A to main destinations (Z, Y, X, etc.)</td>
<td>Preferred</td>
<td>Non-migrants in country A</td>
<td>Migrants from country A in countries Z, Y, X, etc.</td>
</tr>
<tr>
<td></td>
<td>Barely acceptable</td>
<td>(a) Households in country A with members who have migrated to Z, Y, X, etc. (proxy respondent necessary)</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Other households in A</td>
<td></td>
</tr>
<tr>
<td>4. Determinants of out-migration of households from country A to country Z</td>
<td>Acceptable</td>
<td>Households remaining in country A</td>
<td>Migrant households from country A in country Z</td>
</tr>
<tr>
<td>5. Determinants of out-migration of households from country A</td>
<td>Preferred</td>
<td>Households remaining in country A</td>
<td>Migrant households from country A in countries Z, Y, X, etc.</td>
</tr>
<tr>
<td>6. Determinants of out-migration of individuals (and households) from countries A, B, C, etc. to Z, Y, X, etc.</td>
<td>Preferred</td>
<td>Non-migrant individuals and households in countries A, B, C, etc.</td>
<td>Individual and household migrants from countries A, B, C, etc. in Z, Y, X, etc.</td>
</tr>
<tr>
<td>7. Determinants of return migration of individuals from country Z to country A</td>
<td>Acceptable</td>
<td>Individuals who migrated from A to Z and returned to A</td>
<td>Migrants from country A remaining in country Z</td>
</tr>
</tbody>
</table>
Design of surveys to investigate international migration

Table 6.1. (continued)

<table>
<thead>
<tr>
<th>Type of analysis</th>
<th>Quality of option</th>
<th>Focal groups for data collection at origin</th>
<th>Focal groups for data collection at destination</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Not acceptable</td>
<td>None</td>
<td>(a) Households in country Z with members that have returned to country A (proxy respondent necessary)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(b) Migrant households from country A all of whose members are still in country Z</td>
</tr>
<tr>
<td>8. Determinants of return migration of households from country Z to country A</td>
<td>Acceptable</td>
<td>Households that migrated from A to Z and have returned to A</td>
<td>Migrant households from A remaining in Z</td>
</tr>
<tr>
<td>9. Determinants of potential out-migration of individuals (or households) from country A</td>
<td>Barely acceptable</td>
<td>(a) Households in A containing members intending to migrate (b) Other households in A</td>
<td>None</td>
</tr>
</tbody>
</table>

and one country of destination), or one country of origin and several countries of destination.

The most common type of analysis found in the literature (1) looks at the determinants of out-migration of individuals from country A based upon a survey in the country of origin A only. It allows for comparison of the individual characteristics and the household context of international out-migrants with those of non-migrants. Because the information on out-migrants must be obtained from proxy respondents (usually close relatives of the migrant remaining in the origin country and origin household), it cannot be very detailed or of high quality. Despite the apparent simplicity of this analytical approach, adequate data collection is still not as straightforward as usually assumed, since it should involve obtaining information about the out-migrants pertaining to the time of migration and about non-migrants for approximately that same time. Instead, information gathered on both has almost invariably been only on the time of interview. When the international migrants of interest are defined as those having left country A during the five years preceding interview, the relevant reference date for non-migrants should be 2.5 years preceding interview, that is, the mid-point of the period over which the change of residence of migrants would have taken place.

To study the determinants of out-migration of individuals from a particular country A to a particular country Z, the analyst has two options. In the first, considered acceptable, surveys are carried out in both the country of origin A and the country of destination Z, but the focal groups in each are different. The survey in the country of destination covers only migrants and indeed only
migrants from country A, while that in the country of origin covers only non-migrants (that is, almost the whole population remaining in country A, which was at risk of international migration but did not migrate). Migrants are therefore interviewed directly in Z and not by proxy. The combination of non-migrants in A and migrants from A to Z in Z approximates well the total population at risk of international migration in A at the beginning of the observation period (say, five years before the survey). Those missing include persons who died during the period considered, those who migrated from A to countries other than Z, and those who migrated to Z but then migrated again to another country during the period. Each of these will normally be very small proportions of the migration flow from A to Z. Because the groups missing may differ from those covered in ways germane to the determinants of migration, this option is evaluated here as only acceptable.

The second option to study the determinants of individual out-migration from A to Z is to undertake a survey exclusively in the country of origin obtaining data on out-migrants from proxy respondents. The use of proxy respondents again makes this option barely acceptable. Note that the survey described here is similar to that in panel 1, the only difference being that here (in 2) only out-migrants to Z are of interest. In practice, it is not clear how one should treat households that contain out-migrants to countries other than Z. If any are encountered during the survey, they should probably be excluded from the analysis if the goal is really to focus on the determinants of those migrating to Z. But such a restriction makes little sense. (See discussion of the second option of panel 3.)

The third type of analysis involves the determinants of individual out-migration from a specific country A to several major countries of destination (3). The preferred approach involves undertaking surveys in the country of origin A to interview non-migrants and in each of the major countries of destination to gather information directly from the migrants originating in A. An alternative but barely acceptable approach is to carry out a survey of out-migrants and non-migrants (together with their households) only in the country of origin. In this approach information on out-migrants would have to be obtained from proxy respondents, but all out-migrants are considered since there is no need to focus only on those with a particular country of destination. The data gathered can be pooled to assess simultaneously the determinants of out-migration to various countries in comparison to the option of not migrating by using, for instance, multinomial logit analysis. It is likely that the coefficients of the determinants of the decision to migrate to a specific country Z versus not to migrate will be different from those of the decision to migrate to Z versus Y versus X versus to not migrate. Whether such differences are important or not is an empirical question whose relevance will depend on the degree to which Z is a more important destination for migrants from A than Y or X. This indicates the analytical advantage of this approach compared to panel 1, though the survey requirements are identical.
Design of surveys to investigate international migration

The next two panels are concerned with the international migration of households. The fourth type of analysis (panel 4) is concerned with the determinants of out-migration of households from a specific country A to a specific country Z. Because whole households are being considered as opposed to individuals, there is no possibility of using proxy respondents to report on those that have left a country (neighbours provide notoriously poor information). Consequently, the analysis envisaged requires that surveys be undertaken in both the country of origin A to interview households without out-migrants and country of destination Z to interview migrant households originating in A. It is important to stress that retrospective data on the situation of non-migrant households in country A must be obtained for the mid-point of the period used to define international migration, so that the situation of the migrants before migration can be compared with that of non-migrants at the (mean) time the migration decision was made. Information on households will generally be obtained from the head of household, supplemented by data from all other adult household members interviewed individually. Because this approach demands that only households originating in A be interviewed in Z, the problem of locating those particular households is greater. In the country of origin the households of interest are those that do not have any migrant members, whether the latter are in Z or in any other country. Consequently, if any household with out-migrant members is encountered in the sample, it should be excluded from the analysis. Note that the type of analysis possible in this case can provide an assessment of the effects of differences in individual and household factors on the propensity of households in A to migrate only to Z. Because only one country of destination is considered, nothing can be said about how differences in the factors at destination can influence out-migration from country A.

The fifth type of analysis focuses on the determinants of out-migration of households from a specific country A to several countries of destination. It therefore represents a generalization of (4), and requires that surveys be undertaken in the country of origin A covering households without out-migrants as well as in each of the (presumably major) countries of destination – Z, Y, X, etc. In each of the countries of destination, only households containing migrants from A would be included in the survey. The advantage of this approach over the previous one (4) is that it pools data from several countries of destination and thus permits an analysis of the factors determining the choice of destination by out-migrant households from A. Since the countries of destination may have different policies that affect the magnitude, timing and characteristics of flows of migrants from A, this type of analysis allows an assessment of the effects of policy differences on international migration from A, as well as, simultaneously, of cross-country differences in other factors including economic conditions. It is only when different countries of destination are included that such analysis is possible.

The sixth type of analysis represents the broadest approach to the study of the determinants of out-migration, since it considers both the out-migration of individuals and whole households from several countries of origin to several
International migration statistics

countries of destination. This requires that surveys be conducted in each of the
countries of origin and each country of destination. Given the international
migration linkages existing between countries today, it is often the case that
a few countries of destination attract migrants from a common set of countries
of origin (Zlotnik, 1992), and that the migrants from each country of origin
gravitate to a small number of countries of destination. Therefore, implementa-
tion of this type of analysis is not as daunting as it would appear at first glance,
as it would usually require undertaking surveys in a small number of countries,
numbers that have proved manageable in projects such as the current NIDI and
CERPOD projects (see section A.3). By gathering information on individuals
and households without migrants in the countries of origin and on migrants
from various countries of origin residing in the various countries of destination
considered, this approach makes possible the analysis of the determinants of
migration from several countries of origin to several countries of destination,
thus making it possible to ascertain why out-migration differs from one country
of origin to another as well as why some countries attract international migrants
more than others. The results of such analyses could be very helpful in policy
formulation, since they would encompass both the effects of differences in social
and economic conditions in the countries of origin and destination, as well as
those of policy differences in both countries of origin and destination. The
surveys required should be carried out within a short period of time (preferably
over one or two years) but need not be carried out simultaneously. The current
multi-survey project being executed by NIDI provides a test case for under-
taking these types of linked surveys.

The seventh type of analysis possible involves the determinants of return
migration of individuals from a specific country Z to a specific country A
(panel 7). There are two alternative approaches to gather data for this type of
analysis. The most acceptable one requires a survey of individual migrants from
country A to country Z returning to A and being interviewed in A plus a survey
of migrants from A remaining in Z and interviewed in Z. Disregarding deaths
and the possibility of re-migration to third countries, the two groups combined
constitute the appropriate population at risk of returning to country A after
migration to Z. An analysis of the determinants of return migration based upon
the pooled data will provide useful information about the individual and
household factors leading to return migration. However, surveys of return
migrants should be undertaken only in countries where such migration is
sizable; otherwise it will be very difficult and costly to locate a large enough and
representative sample of return migrants in the country of origin. The barely
acceptable alternative indicated is to gather information only in the country of
destination, which would require using proxy respondents to obtain informa-
tion about migrants who have returned to A. Because it is more likely for return
migrants not to leave any household members in the country of destination
when they return to their country of origin than it is for out-migrants not to
leave family members in the country of origin when they initially leave, the need
to use proxy respondents in the country of destination is likely to result in
Design of surveys to investigate international migration

serious underrepresentation of return migrants. If only for that reason, the second approach to this type of analysis is not recommended.

The eighth type of analysis relates to the determinants of return migration of households from a specific country Z to a specific country of origin A. As in the case of individual migrants, such an analysis requires that surveys be carried out in both countries, comprising a survey in country Z of migrant households from A to Z remaining in Z and a survey in country A of households that had migrated from A to Z but returned to A. In both cases data can be obtained directly from the migrants themselves, though it may be very difficult to locate return migrants in the country of origin, especially if they are rare even with respect to the out-migrant population.

Both types of analysis involving the determinants of return migration (7 and 8) can be extended to include multiple countries of destination or several countries of origin. Thus, to study the causes of return migration from a set of various countries of destination to a single country of origin A, surveys should be executed in each country of destination focusing on migrants from country A while a survey in country A would cover return migrants from all of the relevant countries of destination. The data obtained would allow an analysis of how differences across countries of destination affect return (out-) migration to A. Similarly, a survey in a single country of destination Z could encompass all households with international migrants from various countries of origin and, combined with surveys of return migrants from Z carried out in each of those countries of origin, would provide the pooled data necessary to determine how differences across countries of origin influence the propensity of migrants to return from Z.

Lastly, to study the determinants of potential international migration from A, a survey in A can be carried out that covers all households and asks questions to distinguish those who intend to migrate internationally from those that do not. Although the reliability of the responses may be questionable because they are subjective, attitudinal, and therefore subject to change, studies based upon such data can provide useful results on the determinants of potential migration, and several have been carried out. They also have the advantage of requiring a survey only in a single country. This type of analysis is assessed as "barely acceptable" because of the subjective nature of the data it relies on.

Survey requirements to study the consequences of international migration

Paralleling the treatment of the analysis of the determinants of international migration, table 6.2 presents in schematic form the types of analysis relevant for the study of the consequences of international migration and their implications for data collection in the countries of origin and destination. As indicated in the above discussion on determinants, the preferred approaches are those in which the appropriate population at risk is covered, although the applicability of this concept is less straightforward in relation to the consequences of migration than with respect to the determinants. However, since
### Table 6.2. Focal groups for data collection at origin and destination to study the consequences of international migration

<table>
<thead>
<tr>
<th>Type of analysis</th>
<th>Quality of option</th>
<th>Focal groups for data collection at origin</th>
<th>Focal groups for data collection at destination</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Consequences of in-migration for individual migrants in country Z</td>
<td>Barely acceptable</td>
<td>None</td>
<td>(a) Migrants from countries A, B, C, etc. in country Z</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(b) Non-migrants in country Z</td>
</tr>
<tr>
<td>2. Consequences of in-migration for individuals from country A to country Z</td>
<td>Acceptable</td>
<td>Non-migrants in country A</td>
<td>Migrants from country A in country Z</td>
</tr>
<tr>
<td></td>
<td>Barely acceptable</td>
<td>None</td>
<td>(a) Migrants from country A to country Z</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Non-migrants in Z</td>
<td>None</td>
</tr>
<tr>
<td></td>
<td>(a) Households in country A with members who have migrated to country Z (proxy respondent necessary)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b) Other households in country A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Consequences of in-migration for individual migrants in country Z from main countries of origin (A, B, C, etc.)</td>
<td>Preferred</td>
<td>Non-migrants in countries of origin A, B, C, etc.</td>
<td>Migrants from countries A, B, C, etc. in country Z</td>
</tr>
<tr>
<td>4. Consequences of in-migration for households from country A to country Z</td>
<td>Acceptable</td>
<td>Households in country A</td>
<td>Migrant households from country A in country Z</td>
</tr>
<tr>
<td></td>
<td>Barely acceptable</td>
<td>None</td>
<td>(a) Migrant households from country A in country Z</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(b) Other households in country Z</td>
<td></td>
</tr>
<tr>
<td>5. Consequences of in-migration for households from countries A, B, C, etc. to country Z</td>
<td>Preferred</td>
<td>Non-migrant households in countries A, B, C, etc.</td>
<td>Migrant households from countries A, B, C, etc. in country Z</td>
</tr>
<tr>
<td></td>
<td>Barely acceptable</td>
<td>None</td>
<td>(a) Migrant households from countries A, B, C, etc. in country Z</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(b) Other households in country Z</td>
</tr>
</tbody>
</table>

256
Design of surveys to investigate international migration

Table 6.2. (continued)

<table>
<thead>
<tr>
<th>Type of analysis</th>
<th>Quality of option</th>
<th>Focal groups for data collection at origin</th>
<th>Focal groups for data collection at destination</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Consequences of in-migration of individuals and households from countries A, B, C, etc. to countries Z, Y, X, etc.</td>
<td>Preferred</td>
<td>Households in countries A, B, C, etc.</td>
<td>Migrant individuals and households from countries A, B, C, etc. in countries Z, Y, X, etc.</td>
</tr>
<tr>
<td>7. Consequences of return migration for individuals returning to country A</td>
<td>Barely acceptable</td>
<td>None</td>
<td>(a) Individuals in households that migrated to country Z and who have returned to country A (proxy respondent necessary) (b) Other households that migrated from country A to country Z</td>
</tr>
<tr>
<td>Acceptable</td>
<td>(a) Individuals who migrated from A to Z and returned to A (b) Non-migrants in country A</td>
<td>Preferred</td>
<td>Individuals who migrated from country A to country Z and have returned to country A</td>
</tr>
<tr>
<td>8. Consequences of return migration for households returning to country A</td>
<td>Acceptable</td>
<td>(a) Households that migrated from A to Z and have returned to A (b) Other households in country A</td>
<td>None</td>
</tr>
<tr>
<td>Preferred</td>
<td>Households that migrated from country A to country Z and returned to A</td>
<td>Preferred</td>
<td>Households that migrated from A to Z</td>
</tr>
</tbody>
</table>

International migrants are selected from a pool of potential movers, the best comparison group is always constituted by non-migrants remaining in that pool. Table 6.2 presents eight different types of possible analysis of the consequences of international migration, each of which is discussed below.

The first type of analysis (1), relating to the consequences of international migration for individual migrants in country Z, is also the most commonly undertaken, mainly because it requires only that a single survey be carried out in the country of destination Z and that the current status of migrants from whatever country of origin be compared with that of non-migrants. If migrants
International migration statistics

are doing as well as or better than non-migrants, the consequences of international migration are interpreted to be positive for the migrants concerned; they are considered negative if they are doing less well. Although it is important to control statistically for differences in the basic characteristics of migrants and non-migrants (i.e. age, sex, education, duration of residence in the country of destination, years of labour market experience, etc.), such appropriate controls have unfortunately generally not been used. In addition to such analytical problems, the main drawback of this approach is that it relies on an inappropriate comparison group because, in order to assess the consequences of international migration for the migrants themselves, their situation should be compared with that of non-migrants in the country of origin and not with non-migrants in the country of destination: the former had the potential of sharing the experience of migrants, not the latter. The comparison of migrants and non-migrants in the country of destination can, at best, shed light on the degree of integration of international migrants but not on the consequences of migration. Finally, another important methodological problem in the implementation of this type of analysis is that the survey must take care to ensure that a large enough sample of international migrants is obtained to be representative of that group, which will require specialized sampling techniques (see section C below). Otherwise, the rarity of international migrants in the population may render any type of analysis statistically unsound.

The second type of analysis is on the consequences of international migration for individuals originating in country A and residing in country Z. The most acceptable approach requires surveys in both countries, gathering information in Z on migrants from A and in A on non-migrants. Note that both migrants and non-migrants are interviewed directly, without recourse to proxy respondents. The survey in Z, however, has to contend with the difficulty of selecting a representative sample of migrants from country A, which are likely to be rare elements. The alternative, barely acceptable, option in panel 2 is similar to that described in panel 1 but with migrants from only a single country of origin considered. This will make the sampling problem even more severe.

The third type of analysis relates to studies of the consequences of migration for individual migrants in country Z whose countries of origin are A, B, C, etc. The preferred option requires a survey in country Z of migrants from the various countries of origin considered combined with surveys in each of those countries of origin focusing on non-migrants. The latter represent the appropriate comparison groups and, in principle, involve straightforward surveys that do not need to deal with the problem of rare elements. By comparing migrants in Z with non-migrants in their respective countries of origin, a broad assessment of the consequences of international migration for the migrants to Z can be made, and it can take into account the effects of both differences in the individual and household characteristics as well as of differences in the situations in countries of origin. An alternative, though barely acceptable option, is to carry out a single survey in the country of origin using proxy respondents to report on the current status of international migrants residing in country Z. The
Design of surveys to investigate international migration

data reported would then be compared with that of non-migrants covered by the same survey. Although such an approach has the advantages of involving a survey in only one country and uses the appropriate comparison group, it is in general not recommended because it is unlikely that proxy respondents can reliably provide the type of information needed to assess carefully the status of international migrants who are still living abroad. In addition, the selection of a sample that adequately covers households with out-migrant members living in a particular destination country Z is often likely to be an onerous task that is not justified by the weak results expected from this approach.

The fourth type of analysis involves studies of the consequences of migration for households that migrate internationally from a specific country of origin to a specific country of destination. The acceptable approach in this case requires a survey in Z of migrant households from A and a survey of the appropriate comparison group, non-migrant households remaining in A. This approach is similar to that discussed under type 2 above. A second but barely acceptable approach uses only a survey in the country of destination, and collects data from both migrant and non-migrant households to facilitate comparing the situations of the two at the time of observation. This approach is equivalent to that discussed under type 1 above, and allows the study of migrant integration but not the consequences of international migration.

The fifth type of analysis also centres on studies of the consequences of international migration for households but in this case considers several countries of origin simultaneously. The best approach for this type of studies involves surveys of the relevant non-migrant comparison groups in each of the countries of origin, plus a survey covering migrants from all the different countries of origin in the country of destination. A barely acceptable alternative approach involves a survey in the country of destination Z only, covering migrant households from the various countries of origin as well as non-migrant households. This approach provides a broad perspective on migrant integration but fails to address the consequences of international migration for the households involved.

The sixth type of analysis offers the most general study of the consequences of international migration, involving migrants from various countries of origin to several countries of destination, and comparing them with non-migrants in countries of origin. Just as its counterpart in approaches to the analysis of determinants – also type 6 in table 6.1 – this approach is especially useful to assess the consequences of international migration in countries that are linked by international migration flows (e.g. Kritz and Zlotnik, 1992). It requires that surveys of international migrants (individuals and households) living in various countries of destination be carried out in conjunction with surveys of non-migrants in the main countries of origin. The data gathered allow the examination of why the consequences of migration for international migrants differ according to both country of origin and country of destination, and to the specific origin-destination dyad. Thus, keeping the country of destination fixed but varying the country of origin allows for analysis of the consequences of
International migration statistics

migrant according to country of origin; comparing data on international migrants from a single country of origin across several countries of destination allows an assessment of which country of destination provides the best experience for out-migrants from that particular country of origin; and considering several countries of origin and several of destination simultaneously allows the examination of which specific origin-destination flows have the best or worst consequences for the international migrants involved.

The seventh type of analysis involves the study of the consequences of return migration for individual migrants returning to country A. There are three possible approaches to that type of study which have very different, though rarely recognized, implications for the interpretation of the results obtained. The first one, which is considered barely acceptable here, is based on a survey undertaken in the country of destination Z, covering households that migrated from country A to country Z, some of which have members who have migrated back to A and for whom information is provided by proxy respondents. A second and better approach, which is commonly used, involves a survey undertaken only in the country of origin, where both individuals who migrated to another country and returned and comparable non-migrants are interviewed. The status of the two groups is compared to draw inferences about the benefits or lack thereof associated with international migration. However, the results of such comparisons do not really reflect the effects of return migration since the return migrants are being compared with persons who never migrated: thus any change in status reflects the joint effects of the original migration plus the return move. A better approach to investigate the effects of return migration per se, therefore, involves comparing the status of return migrants with that of other international migrants who have remained abroad, that is, with the rest of the population at risk of return migration. Such an approach requires that surveys be undertaken in both the country of destination Z of the migrants from A as well as in the country of origin A, the latter focusing only on return migrants from Z. This last requirement will require addressing the issue of selecting a representative sample of return migrants which, as a set of rare elements which is itself a subset of rare elements, is likely to be very difficult to locate.

The eighth and last type of analysis focuses on the consequences of return migration for households returning to country A. Since the use of proxy respondents is not possible in the country of origin because whole households have left, the two approaches presented in table 6.2 correspond to the last two of panel 7 and have the same advantages and limitations. Thus the option identified as acceptable cannot provide an assessment of the effects of return migration per se but only of the joint effects of emigration and return. The preferred approach, in contrast, isolates the effects of return migration by comparing migrant households from country A that remain in country Z with migrant households that have returned from Z to A.

As in the cases regarding the assessment of the consequences of international migration for the individual migrant and for the migrant household, the consequences of return migration may be studied by considering several
Design of surveys to investigate international migration

countries of destination and a single country of origin, or several countries of origin and a single country of destination. The purpose of the former would be to assess which country of destination had the most positive impact on the status of return migrants by transmitting skills, allowing the accumulation of savings, etc. As for the latter, its purpose would be to study the differences in the consequences of return migration for migrants returning to different countries of origin. Carrying out surveys in multiple countries of origin and destination would allow both comparisons to be made.

Although the discussion above has assumed that the consequences of interest are those related to return migration per se, in some situations, such as the migration of contract workers, the main interest may be to assess the net overall effects of migration and return. In that case, use of the acceptable options within types 7 and 8 would be appropriate, provided measures are taken to obtain a representative sample of return migrants in the country of origin. An important issue to explore in that case, and in all cases in which migration separates members of the nuclear family (or of the household), is the consequences of migration for the family members remaining in the country of origin. The types of analyses presented in table 6.2 do not deal explicitly with this but, applying the same principles underlying table 6.2, the main comparison group for families or households with out-migrant workers is families or households without them in the country of origin, since taken jointly they constitute the population at risk of experiencing out-migration of family members as migrant workers. One key aspect of the consequences of international migration for the families remaining in the country of origin involves the receipt and use of remittances, a subject treated in detail in Chapter 7.

Implications of the selection of the preferred type of analysis for survey design

As the discussion in the previous sections makes clear, the preferred approach to the study of the determinants and consequences of international migration requires that surveys be carried out in both countries of origin and countries of destination, or at minimum in one such dyad. The question must therefore be addressed of whether the surveys involved can or should be undertaken in both the country of origin and the country of destination simultaneously. If perfect sampling frames allowing the identification of international migrants in the country of destination and of households with out-migrants in the country of origin were available, simultaneity would be desirable. However, while most countries have data on the presence of international migrants within their borders, few have reliable data on out-migration, in part because the migration of citizens is considered an international right. Because of this and because countries of destination are more likely to have the data needed to select a representative sample of international migrants, it is recommended that the survey be carried out first in the country of destination. From this survey detailed information on the place of origin should be obtained from the international migrants interviewed – at the level of the province,
International migration statistics

district and community. These data are then aggregated to get the total number of migrants in the destination country sample coming from each province and district in the country of origin. Then, given population estimates for those provinces and districts (e.g. from a census), ratios of the number of international migrants to the population can be calculated at the province level – and at the district level also if the number of international migrants in the sample in the destination country is large relative to the number of districts in the country of origin so as to provide reasonably reliable estimates of ratios for districts. (In general, this is likely only for those districts which account for most of the out-migration.) Then a national sample of selected provinces or districts may be derived using techniques such as those described in section C.6 below. Although the preferred options for the study of either the determinants or the consequences of international migration require selecting a sample only of non-migrants in the country of origin, to make the comparison with international migrants more meaningful it is important that those non-migrants be selected predominantly from the areas (provinces, districts, and perhaps cities or towns) where the relative propensity to out-migrate internationally to the country of destination is highest. Such a procedure ensures a more relevant comparison group of non-migrants in the country of origin.

An approach of this type has already been followed in a multi-country survey project undertaken in West Africa (Findley et al., 1988). First, areas in the destination country, France, thought to have the most migrants from francophone West Africa were selected by judgement (rather than probability sampling). The West African migrants interviewed in France were then asked to indicate their place of last residence in West Africa (down to the local area in their country of origin). That information was used, albeit again not in a probabilistic manner, to determine the main areas of exodus, which were then selected as the areas in the countries of origin for household surveys to be conducted.

There are also instances in which the reverse approach has been followed, that is, where the first survey is conducted in the country of origin to identify households with international out-migrants. The proxy respondent is then asked the place of residence of out-migrants, which is used to identify sites for the survey in the country of destination. Such an approach has the problem of missing households when the whole household migrates from the country of origin and also tends to overrepresent short-term migration.

Because preferred survey loci and comparison groups for studying the determinants of international migration are so similar to those for the study of the consequences (see tables 6.1 and 6.2), it is recommended that whenever a survey focusing on one of the two is planned it be designed also to collect data for the other topic as well. Substantial effort and expenditure will always be required to design and carry out surveys in two countries, one of origin and one of destination, so the data should be gathered so as to facilitate studying both the determinants and consequences of international migration, even if the intention of the survey design team is only to carry out analyses on one of the
Design of surveys to investigate international migration

two topics. Not only would this strategy be cost effective (the costly steps involved in mapping, sample selection and fieldwork would be carried out only once), it would also make the results more relevant for policy formulation. Studies of the determinants of migration clarify the factors affecting it and can therefore provide guidance regarding the policies to be changed with a view to altering international migration processes. But studies of the determinants provide no information whatsoever on whether this is a desirable thing to do. Studies of the consequences provide the latter information: if the consequences are found to be negative, then policy-makers would seek to decrease them, and vice versa. But studies of the consequences provide absolutely no information about what factors to change to alter the migration flows. Thus studies of both the determinants and consequences of international migration are desirable.

The minimum needed to study both the determinants and consequences of international migration is a survey in the country of origin covering households without out-migrants and a survey in the country of destination covering international migrants (individuals and households). The survey in the country of origin need not make any special effort to cover households with out-migrants since those migrants will be covered and interviewed directly in the country of destination. In the survey in the country of destination, the only population that need be covered is that of international migrants from the country of origin selected. However, in order also to permit an assessment of the integration of those migrants, coverage of a sample of non-migrants in the country of destination is also recommended. In all cases in which non-migrants are interviewed, in addition to recording information on their status at the time of the survey, retrospective information about their status at the mid-point of the period used to identify international migrants should also be recorded to permit the appropriate comparisons to be made. The relevant reference point is 2.5 years before the survey if international migrants are identified as those having changed country of residence within the 5 years preceding the survey.

3. Planning specialized surveys on international migration

Box 6.1 presents a list of the steps involved in designing, implementing and analysing specialized surveys aimed at gathering information for the analysis of the determinants and consequences of international migration, which requires surveys in both countries of origin and destination. The list is provided to ensure that all key steps are taken into account in planning the budget and the schedule of work to undertake surveys in at least two countries. Poor planning and failure to anticipate all the steps involved in carrying out a survey are often the source of serious budgetary and scheduling problems even when only a single survey is carried out. Carrying out linked surveys in more than one country is all the
International migration statistics

more complicated and therefore requires even more careful planning. Because unforeseen problems are likely to arise, the time allocated for the various steps should not be too tightly planned, particularly that assigned to fieldwork and data processing.

To the extent possible, the steps listed in box 6.1 are in chronological order, though several may be undertaken simultaneously. The initial steps in planning the survey, namely, the determination of survey objectives, the establishment of legal, contractual and budgetary parameters, and the development of an appropriate work schedule (steps 1 to 3) are crucial to a successful project.

BOX 6.1. STEPS INVOLVED IN CONDUCTING A SPECIALIZED SURVEY ON INTERNATIONAL MIGRATION

1. Determine survey objectives; establish definition of international migrants and of appropriate comparison groups; decide on countries of implementation (origin and destination).
2. Establish legal, contractual and budgetary parameters for the survey in each country; determine institutional responsibilities for implementation and budgetary accounting.
3. Plan work schedule identifying the role and responsibilities of main institutions and key personnel in the countries where surveys are to be undertaken.
4. Draft questionnaire(s) and manuals for interviewers and supervisors; translate them into the main languages needed to carry out the survey (this step can be carried out simultaneously in countries of origin and destination).
5. Determine who the main users of the survey data will be, solicit their suggestions regarding questionnaire content to ensure that, to the extent possible, their interests are incorporated.
6. Develop sampling frame for survey in the country of destination.
7. Carry out necessary cartographic work to select the sample in the country of destination (using, for instance, the most recent census maps and updating them in the field as necessary).
8. Design and select sample of migrants and non-migrants in the country of destination.
9. Develop documents to organize and monitor fieldwork.
10. Train first group of interviewers and supervisors to pretest questionnaires, and conduct pretests in the country of destination.
11. Evaluate results of pretests and revise questionnaires and manuals for interviewers and supervisors accordingly. Adjust translations of questionnaires and manuals; print final questionnaires and manuals for fieldwork.
12. Develop preliminary plans for the tabulation, analysis and dissemination of survey results.
13. Prepare editing and coding manuals; train editors and coders.
Design of surveys to investigate international migration

14. Prepare instructions and computer programs for data entry and cleaning; test and finalize those programs.
15. Recruit and train additional interviewers and supervisors.
16. Make logistic (housing, transportation etc.) arrangements for fieldwork in country of destination; assign fieldwork to teams comprising several interviewers and one supervisor each.
17. Implement survey in country of destination.
18. Code and edit responses to questionnaires; conduct data entry and cleaning; prepare data files for use in the analysis of results.
19. Prepare tabulations on international migrants interviewed in the country of destination to determine specific locations (province, district, community) of last residence in the country of origin. Use that information to select a sample of households without international migrants in the country of origin.
20. Repeat steps 6 to 18 in the country of origin.
21. Complete tabulation and analysis plans for the survey in the country of destination and for that in the country of origin; carry out the analysis planned.
22. Prepare database with the data gathered by the surveys implemented in the countries of origin and destination for dissemination and further analysis.
23. Prepare report on main findings that includes a clear and concise executive summary focusing on the implications of the results obtained.
24. Publish and disseminate findings through different publications (in the appropriate languages) so as to reach both the research and the policymaking community in participating countries.
25. Organize meetings or workshops to disseminate findings and seek feedback on the quality and relevance of the data gathered. Discuss the methodological issues involved in survey design and the analysis of results to seek ways of improving the approach adopted.

The drafting of questionnaires and manuals (step 4) should be carried out early for both countries of origin and destination. Pretests of the questionnaire to be used in the country of origin (steps 10 and 11) may also be conducted early, although the final printing of questionnaires may be postponed until the results (preliminary tabulations) of the survey in the country of destination are in and an assessment can be made regarding their implications for data collection in the country of origin. Steps 10 to 18 are standard in any survey, but all should be carefully planned and their sequence made clear from the beginning. Note that the preparation of tabulation plans, editing and coding manuals, as well as computer programs to clean and process the data gathered (steps 12 to 14) should be carried out as soon as possible after the questionnaires reach final form instead of waiting until the fieldwork is completed. Also note that data cleaning and processing should occur simultaneously with the fieldwork: as soon as the fieldwork in a given area is finished, the completed questionnaires should be sent to a central location for processing.
International migration statistics

The steps presented in box 6.1 assume that the survey will cover at least one country of origin and one of destination, and that the households without international migrants to be interviewed in the country of origin are to be selected from those areas within the country that are the most likely sources of international migrants to the country of destination considered; hence the relevance of step 19. Also important are steps 22 to 25 on the dissemination of both the data gathered and the findings derived from their analysis. Many data collection projects fail to exploit the full richness of the information collected or to disseminate adequately the analytical results. Further, in general no efforts are made to assess the methodological approach with a view to improving it.

Certain important aspects of planning a survey are not included explicitly in box 6.1. The first relates to the locus and mode of interview (e.g. telephone, mail survey, personal interview). Given the detailed information that will be sought from migrants and non-migrants alike, the personal interview is strongly recommended to obtain data of better quality and a higher response rate (and therefore lower non-response bias). In addition, the confidentiality of responses must be ensured, and interviewers are better able to convince certain types of international migrants (such as those in an irregular situation) that they should respond to the survey.

Concerning the choice of respondent, it is highly desirable to interview the migrant directly, because of the possibility of getting biased answers from anyone else serving as a proxy respondent. As noted above, the use of proxy respondents cannot be avoided when the survey is conducted in a country that the migrant of interest has left, but such a survey approach is not recommended anyway. Proxy respondents should be avoided because they often do not know certain things about the migrant, in particular details about potentially sensitive topics, such as earnings and savings accumulated, personal conflicts with employer, or motivations. In addition, even when they know the facts, they may deliberately distort them to give what they perceive to be more socially acceptable responses (Groves, 1989; Singleton et al., 1988). In the highlands of Ecuador, for instance, male heads of household tended to understate both the frequency of out-migration of their daughters and the extent to which they worked in the place of destination (Bilsborrow, 1993). Consequently, it should not be assumed that the head of household (usually a male) is the best proxy respondent for a migrant household member: the best respondent will often be instead another adult household member of the same sex as the migrant or another adult of the same generation as the migrant (for instance, a sibling).

The length of the questionnaire used and the duration of interview is another matter that needs to be settled early. Surveys undertaken in developing countries rarely encounter problems of resistance or non-response even when they last an hour or so, but there is a widespread presumption that interviews lasting longer will confront those problems and result in biased data. Surveys in developed countries and among upper income groups in large urban
areas of developing countries sometimes encounter resistance even with shorter interviews. This issue must be assessed carefully during the pretesting of the questionnaires.

In designing questionnaires to gather retrospective information, the issue of whether to adopt a life history approach arises. The life history of an individual consists of the full record of dates on which salient events occurred, such as graduating from school, marriage, childbirth, entry into employment, job changes, changes of residence, etc. Events and their dates are sometimes recorded using a monthly calendar covering all years since the age of 12 or 15 years (Freedman et al., 1988). While the use of life history questionnaires has some advantages (especially in aiding the respondent to recall events and put them in the right chronological order), it also has some limitations, including the inability to obtain detailed information about specific events without the use of supplementary schedules and the fact that a wealth of information is gathered that is not used in the analysis. Given the objective of the surveys considered here, full life histories do not seem necessary.

Lastly, it must be noted that the definition of international migrant adopted for survey purposes makes no distinction between the different types of international migrants discussed in Chapter 2. Depending on the country of destination, some types of international migrants—as defined in Chapter 2—will be more common than others among those who changed their country of residence within the five years preceding the survey. Questions to identify specific types of migrants may be included in the questionnaire. For example, it may be important to identify refugees and asylum-seekers; migrants admitted specifically for the exercise of an economic activity; or migrants admitted for family reunification. In countries where international migrants tend to stay for relatively short periods or where they do not live in private households but are rather housed in collective accommodation, special measures may need to be taken to ensure that they are not excluded from the sample selected. Given that the definition of international migrant adopted for survey purposes considers that a change of residence takes place only when a migrant stays in a country for over six months, short-term movers, such as border workers and seasonal workers, will not be covered. If those groups are of interest, different definitions and survey designs will be needed.

C. SAMPLE DESIGN

The quality of any survey depends on the quality of its sample. Sampling aspects are particularly complex in the case of surveys on international migration, mainly because international migrants constitute rare elements among the host population. Given that the preferred approaches to the analysis of either the determinants or the consequences of international migration require that surveys be conducted in both the country of origin and the country of
International migration statistics

destination, it will be assumed henceforth that the international migrants of interest are those originating in a particular country A who are still resident in Z. Extensions to more complex cases involving more than one country of origin or destination can be readily made based on the principles and procedures described below and will not be discussed explicitly.

This section describes in detail how to select the sample for specialized surveys on international migration. Whatever other limitations may be placed on the countries or areas covered by the survey owing to budgetary or logistical problems, it is crucial that the sample be a probability sample in the areas covered by the survey. Only probability samples allow statistically valid inferences to be made from the analyses based on the survey data. A probability sample is one in which every element in the sample (whether a migrant or non-migrant) has a known probability of selection. To ensure that the probabilistic nature of the sample is maintained, personal judgement must be assiduously avoided in the selection of the sample of persons to be interviewed. Thus during the fieldwork stage of the survey, measures should be taken to ensure that interviewers cannot exercise personal judgement in determining who to interview. Too many surveys have been marred by procedures that overtly or in a subtle fashion allow elements of judgement to enter into the selection of the sample, either at the stage of selecting areas to conduct the survey in or in the selection of persons to interview within those areas. Whenever probabilistic procedures are not adhered to, valid statistical inferences cannot be drawn about the populations covered by the survey and the validity of the survey findings cannot be properly assessed. Given the cost and effort involved in carrying out a survey, it is wasteful to compromise its results by disregarding the statistical principles underlying appropriate sampling procedures.

Devising an effective strategy for probability sampling is, however, particularly difficult in surveys focusing on international migrants, since the latter are very rare elements among the populations of most countries. Thus international migrants, as defined for survey purposes in section B.1 above, will usually constitute less than one per cent of the population. Consequently, specialized sampling techniques are required to locate such rare elements efficiently without wastefully dispersing field data collection efforts, which would result from a completely random sampling procedure. Such techniques are discussed in detail below. To make them understandable to the non-sampling expert, we first review basic sampling concepts relevant to the sampling of international migrants.

1. Domains of analysis and sampling frames

The first step in designing a sample is to define the domains of analysis. A domain is a well-defined set of elements about which one wants to draw inferences - such as the population living in a specific geographic region or city.
Design of surveys to investigate international migration

It may be part of a larger population but because the sample will be drawn only to cover a given domain or domains, inferences will only be valid regarding those domains. With respect to surveys on international migration, the domains of interest are the country of origin and the country of destination. However, because of budgetary restrictions, the domains actually used may be limited to small regions within those countries. The survey in the country of destination may, for instance, cover only the cities that attract most international migrants or only those that attract migrants from a particular country. The population covered in a survey thus depends on the domain or domains of analysis selected a priori, and inferences from analyses based on the data collected strictly refer to that population only. Thus, if the domain of analysis is a particular city in the country of destination, the findings regarding the consequences of international migration apply only to migrants in that city and may not be meaningful for migrants in any area beyond it.

The sampling frame provides the basis for drawing a sample of elements belonging to the domains of interest. Elements are the ultimate units being analysed — in the present case, individual migrants and non-migrants or their households. The quality of the sampling frame is a major determinant of the extent to which the sample is representative of the population in the domain of interest. A sampling frame is a listing of elements in the domain of interest. A frame is perfect if every element appears on the list separately, once and only once, if no element is omitted, and if no inappropriate entries are on the list. With respect to international migration, a perfect sampling frame would be provided by a complete list of international migrants living in the country of destination at time \( t \) who had arrived in that country after \( t-5 \) (presuming a 5-year cut-off in the definition of migrant). To be usable, the list should include the current address of each of those migrants. In countries maintaining population registers, such a list is potentially available (see Chapter 3, sections B and C). The list might even include other information about recent international migrants, such as country of citizenship or country of previous residence, that may be relevant for sample selection. (Countries having registers of foreigners may also be able to use this as a sampling frame, although it would exclude all citizens returning after emigration abroad.) Even if the lists obtained from registers are not perfect (some migrants may have left without deregistering and irregular migrants are unlikely to be recorded at all), they provide a very useful basis for constructing a sampling frame, provided issues of privacy and confidentiality do not prevent the use of registration records altogether.\(^7\)

Most countries, however, lack such data sources that can be used directly as a sampling frame. Population censuses, for example, usually do not process and store information on exact addresses. Furthermore, because they take place only once every decade, their information may not be current. If the survey is made soon after the census, however, census information can be used as a sampling frame. Thus, the 1990 census of France was used to identify a sample of international migrants for a specialized survey carried out in connection with the census (Tribalat and Simon, 1993; Tribalat, 1995). In some countries, census
data have been incorporated into a geographical information system (GIS) that, without allowing the identification of the individuals enumerated, nevertheless provides information on the characteristics of the population of small geographic areas. Such information can be used to build a sampling frame when the ideal one does not exist. Indeed, the creation of a population frame requires that data on both population size and the number of international migrants be available for small geographic areas. To select the areas that will be included in the sample, the proportion of the population constituted by international migrants in each geographic area of the domains of interest is calculated and probabilities of selection are assigned as a function of those proportions. Full details of the procedure are provided in sections 4 and 6 below. Note, however, that censuses may not have information on international migrants who have moved into the country during the five-year period of interest. If censuses identify international migrants in terms of place of birth (or current citizenship when place of birth is not recorded), the population frame they yield will be inadequate for the selection of recent migrants to the extent the geographic distribution of the latter differs from that of the foreign-born population, most of whom arrived much earlier. This is but one example of the type of frame problems that may affect the population frame.

The types of problems affecting frames include: the existence of missing elements, implying that some of the international migrants of interest are not covered by the data source being used; misclassification of elements, with some recent international migrants appearing as non-migrants and vice versa; the use of a definition of international migrant in the survey that is not identical to that used in the population frame (e.g. the survey focussing on migrants from a particular country and the frame only providing data on the foreign born, or only on those coming classified by region of origin); errors in recording or processing data on international migration in the population frame (since international migrants are so rare, small errors can cause significant distortions in the proportion of international migrants in small geographic areas); and the less recent the population frame, the more the number and geographical distribution of migrants it shows will differ from the true numbers at the time of the survey. Usually little can be done to alleviate frame problems such as those listed above since there is rarely any data available or other basis for “purifying” the sampling frame. One can only hope that the defects are small relative to the numbers of international migrants correctly identified by the population frame. For a detailed discussion of frame problems, see Kish (1965, pp. 53-59 and 384-433).

2. Fundamental need for probability samples

A probability sample is a sample in which every element (that is, every person or household) has a known, non-zero probability of being selected.
Design of surveys to investigate international migration

Probability samples are designed to permit statistical inferences about means, variances, regression coefficients and other statistical measures relative to the population in the domain of interest. All statistical inferences require that the sample be a probability sample. Inferences from non-probability samples cannot be considered to have statistical validity. Most of the few existing specialized surveys on international migration are not based on probability samples, a fact that vitiates any inferences made from them.

Probability samples do not allow personal or subjective judgement to play a role at any stage in the selection of the sample: not in the selection of the areas in which migrants are to be sought (called primary sampling units or PSUs), nor in the selection of persons or households to be interviewed in those areas. Thus, in a probability sample selection criteria are established a priori on the basis of survey goals and appropriate probabilistic procedures are used to select the areas to be included in the sample. The same is true for procedures to select households within those areas selected. The simplest probabilistic selection procedure is simple random sampling, whereby every element in a set has equal probability of being selected. When the number of elements is large, tables of random numbers can be used to effect the selection provided every element is assigned a unique number. If similar elements can be put into a list, the use of systematic sampling with a randomly selected starting point also results in a simple random sample.

When the elements to be selected are grouped, such as people living in different areas, and simple random sampling is used, the probability of selection of each subset is proportional to the number of elements it contains (this type of sampling is called “probability proportional to estimated size” or PPES). To illustrate PPES with systematic sampling, consider the following example. Suppose the domain of interest is divided into 20 areas with the following population sizes: 27, 42, 50, 100, 200, 48, 35, 24, 33, 21, 34, 37, 55, 40, 59, 80, 120, 410, 74, 32; suppose that four areas are to be selected with probabilities proportional to their population size. By cumulating the reported population sizes, the following list is obtained: 27, 69, 119, 219, 419, 467, 502, 526, 559, 580, 614, 651, 706, 746, 805, 885, 1,005, 1,415, 1,489, 1,521. The total population size is 1,521. To use systematic sampling, a sampling selection interval is calculated by dividing the total number of elements by the number of groups to be selected, namely, 1,521/4 = 380.25. Then, a starting point is selected at random between 1 and 380. Suppose it is 305. The key numbers to effect area selection are therefore, 305, 685 (305 + 380), 1,065 (685 + 380) and 1,445. These cumulated numbers correspond to the four areas which have population sizes of 200 (which accounts for all numbers between 219 and 419), 55, 410 and 74. These four areas are therefore the areas selected for the sample. This method can be used to select with equal probability any number of elements from a list and is most useful if a large number are to be selected since only a single random number needs to be determined as the starting point for the process.

It is important to note, however, that probability samples need not be based on such simple random sampling. Simple random sampling has the
advantage of leading to self-weighting samples, because each element has the same probability of selection as any other. But when the purpose of a survey is to study rare elements, specialized sampling procedures are necessary to ensure that those elements are adequately represented (see section 6 below).

Deviations from probability sampling can occur under a variety of circumstances. A common one arises when budgetary constraints prevent a large number of areas being included in the survey. When only a few areas are to be sampled, it may be more advisable to use expert knowledge to select them than to select them on the basis of random probabilistic principles. The term "judgement sample" is used to describe the non-probabilistic selection of areas or elements to be surveyed. Judgement samples can be useful starting points for the selection of areas to be covered by pilot studies, but even in such cases the use of sloppy non-probability samples in the final selection of elements (households with international migrants) is not acceptable, nor is it necessary. The most common method of selecting non-probability samples of elements (households or individuals) is a quota sample, in which the types of elements to be sampled and the number of each to be selected are determined a priori, regardless of the actual relative prevalence of the different types of elements in an area. Thus, in a survey of international migrants in a city, interviewers could be instructed to each take an area of the city and seek out people (on the street, in dwellings or wherever), interviewing those who are international migrants until some fixed number or quota is achieved. Lack of information about the relative prevalence of international migrants across the areas studied means that there is no way to weight the observations (see section 7 below) to obtain a valid estimate for the city as a whole. In addition, the fact that the interviewer selects the persons to be interviewed inevitably leads to serious biases, even among the most experienced interviewers, since the persons selected tend to be the most accessible, those who wear attractive clothes or live in the nicer houses, who appear less busy or preoccupied, or who are of the same sex, age group, race, or social class as the interviewer. "Serious biases of subjective selection have been demonstrated time and again, whether choosing heads or tails, random samples of integers . . . plants from a field, or people on streets or in homes" (Kish, 1965, p. 29). Thus only in the case of a study limited to a small number of areas is judgement sampling somewhat justifiable, but even then only in the initial selection of areas. Quota samples should never be used in the selection of elements (persons or households).

3. Sample size

To estimate sample means and other statistics, the size of the sample is important. To begin with, the standard error of the sample mean is inversely related to sample size. It does not depend on the fraction of the population in the sample, a common misconception. However, by using stratification (see next
section), a smaller sample can have more information content, implying that sample size is not the key factor in ensuring small variances. In addition, the quality of survey execution in the field is usually more important than sample size per se as a means of reducing total survey error. Non-sampling error, which comprises all sources of error other than sampling error, is usually both larger and more controllable (with careful field procedures) than sampling error.

Determining the desired sample size for a survey is seldom as easy in practice as it appears in textbook examples. As already noted, sampling error (or sampling variance) is inversely related to sample size, so that increasing the size of the sample reduces sampling error. In general, sample size should be determined by considering: the standard error of the key variable or parameter to be estimated from the survey; the size of error in that estimate that is considered within an acceptable range a priori; and the statistical power of the test of hypothesis used for the key variable or parameter. Textbooks focus on examples where the variable to be estimated is a sample mean, a sample proportion or a rate, such as mean income, the proportion of international migrants, or the rate of out-migration from an area. To determine the size of the sample required, a priori knowledge of the population variance of the key variable is needed. Moreover, in stratified samples (see below), such knowledge is required for every stratum. It is extremely rare that a recent survey in a given country would have a similar sample design and would have obtained data about the same key variable so as to provide the information needed.

Another critical factor in determining sample size is the identification of the key variable. In complex surveys, such as those needed for the study of the determinants and consequences of international migration, the measurement of a whole array of variables and relationships is of interest. Consequently, specifying in advance a single key variable or parameter to determine sample size is not only arbitrary but also extremely risky, since the variable selected may not be the most important for the population under study.¹⁰

In practice, budgetary considerations usually prove to be the main determinant of sample size. In view of the problems involved in selecting a sample size on the basis of statistical principles, as discussed above, some examples may indicate orders of magnitude. Bilsborrow et al. (1984, Chapter 4) provide a review of sample sizes and other characteristics of surveys of internal migration in low-income countries, finding sample sizes varying from less than 100 to over 5,000, with only the latter nationally representative. In a series of surveys of internal migration conducted in south-east Asia, ESCAP recommended the use of sample sizes of 6,000 to 8,000 households per country (Turner, 1979, pp. 19-20), though that size was later found to be inadequate and was raised to 15,000 households (United Nations ESCAP, 1980c, p. 9). Surveys of such dimensions are usually too costly for low-income countries and as a result only a few were carried out. To be conservative, surveys of international migration intended to be reasonably representative of a country or of a large region within a country should have minimum sample sizes for each domain of interest of 1,000 to 2,000 households. But there is so little experience in the design of such
medium-sized surveys of international migrants that there are few examples to
cite, and in none has sample size been based on the statistical criteria listed
above. The ongoing multi-country survey project coordinated by NIDI (see
section A.3 above) intends to use samples of 2,000 non-migrant households in
each country of origin and 500 households containing international migrants for
each of the flows covered in each country of destination. The latter will mean
sample sizes of at least 1,000 households with international migrants in each
country of destination. Samples two or three times as large are preferable
whenever budgets allow, especially for countries with large, heterogeneous
migrant populations, so as to permit analyses of differences across groups, such
as for migrants originating in different countries.

Besides cost factors, the extent to which the population of interest is more
homogeneous (requiring a smaller sample) or heterogeneous (requiring a larger
sample) should always be taken into account in determining the desired sample
size in different areas. To the extent possible, in order to compare international
migrants with non-migrants in destination countries, for example, the aim
should be to achieve a final sample in which close to half the households contain
one or more international migrants. Large sample sizes are not necessary
provided specialized sampling techniques are used to locate international
migrants. Different sample sizes and sampling methods may also be desirable for
different areas, such as urban and rural areas.

4. Stratification

Stratification is the division of the population into sub-groups or strata
according to objective criteria or variables already measured for the population
of interest. Once the population is divided into strata, the total variation in it
can be divided into two parts: variation between strata and variation within
strata. Because stratified sampling involves sampling separately from each
stratum, sometimes using different (unequal) probabilities of selection or even
entirely different sampling procedures, this procedure effectively eliminates the
variation between strata from the computation of total variation in the sample
selected, thus reducing total variance to the sampling variance within strata.
Consequently, the gain from stratifying the population can be substantial, and
will be substantial if stratification is carried out so that the elements within each
stratum are as similar as possible to each another (that is, that the stratum has as
little variance as possible), and that the strata differ as much as possible from
each other (that is, that they have means for the stratification variables that
differ as much as possible from each other). To be effective, stratification should
be carried out on the basis of variables that are the main subject of study or that
are closely associated with the variables being studied. Stratification can be
performed on the basis of one or several variables. The latter, called multiple
stratification, requires that the stratification variables used be related not only
Design of surveys to investigate international migration

to the survey objectives but also be as different from each other as possible (uncorrelated). The number of variables used for stratification depends on the information available a priori about the population of interest.

How many strata should be formed? There is no general answer to this question, but certain principles can guide the decision. First, every stratum must have at least two elements to allow the calculation of within stratum variances. Second, (multiple) stratification on the basis of several independent variables that produce \( k \) strata is more efficient than stratification by a single variable also producing \( k \) strata. Thus, it is better to stratify according to place of residence (urban versus rural), socio-economic status (low, medium and high), and proportion of foreign-born (less than 0.1 per cent; 0.1 to 1.0 per cent; higher than 1.0 per cent) therefore producing 18 strata, than to create 18 categories on the basis of any one of the three variables alone.

In choosing the sampling probabilities for strata, an optimum sample design requires selecting a higher proportion of elements in strata where: the variance per element is greater; and the cost of data collection per element is less. Thus, proportions of observations should be selected in strata \( j \) such that

\[
f_j = K \frac{s_j}{\sqrt{C_j}},
\]

where \( K \) is a constant, \( C_j \) is the cost of data collection per element selected in \( j \), and \( s_j \) is the standard deviation per element in \( j \) with respect to the stratification variable being used. Adequate information on either \( s_j \) or \( C_j \) is rarely available a priori, unless a survey with the same sample design has been conducted recently and gathered information on the stratification variables, thus allowing the estimation of \( s_j \). In addition, information on costs (\( C_j \)) from that survey would have had to be collected and analysed, which is virtually never done.

In practice, urban areas tend to have both higher element variance with respect to the variables of interest and lower interviewer costs than rural areas (even allowing for the greater number of callbacks needed), both of which suggest that higher probabilities of selection be used in urban than in rural areas. Desired sample size distributions, derived from survey objectives, may also affect the extent to which the probabilities of selection differ across strata. In all cases, the stratification variables should be correlated with the variables being studied. Thus, if the proportion of households having an international migrant differs markedly between urban and rural areas, or is correlated with population density or socio-economic status, each of those variables should be used in stratification. Different strata may, of course, be used in different domains.

Lastly, stratification allows the use of different sampling frames and different sampling procedures in the different strata. Thus, if adequate maps and sampling frames are available for urban but not for rural areas, different sampling procedures will need to be used.
5. Multi-stage and cluster sampling

The most efficient sampling designs usually involve more than one sampling stage. Up to this point it has been assumed, for simplicity that the sample is selected in two stages: the first involving the selection of first-stage areas, or primary sampling units (PSUs), and the second involving the selection of elements for interview within the PSUs. In practice, it is normally desirable to have more than two stages, thus making use of multi-stage sampling. In population surveys, the first stage in multi-stage sampling is usually the selection of PSUs in proportion to their estimated population size. In a single stage sampling design, areas within a country or region are selected and then all elements (e.g. households) within them are included in the survey. A survey on international migration should never use a single stage sampling design because the number of international migrants likely to be found in PSUs selected at random is so small that the effort would be wasteful.

Multi-stage sampling is particularly useful to ensure that a sample is representative of a whole country or a large region, since it allows a dispersed sample while at the same time keeping down the cost of field operations. In situations where no frame exists to select the final area elements of interest or the ultimate area units (UAUs) at the last stage of multi-stage sampling, a frame has to be created by a mapping operation. Because the cost of undertaking such an operation for a whole country is prohibitive, at the first stage an existing frame (derived from a census or other national data collection system) can be used to select a sample of relatively large area units (PSUs) from the domain of interest. Then mapping operations and the subsequent formation of next-stage units called segments or chunks within each mapped area can be carried out in the field only for those PSUs selected at the first stage.

Multi-stage sampling usually involves the use of cluster sampling, generally at the last stage of sample selection. In cluster sampling, sampling units are clusters of respondents, such as all households in the smallest or ultimate area unit or all clusters of a particular size (e.g. 20) closest to the northeast corner of the UAU, or all members of sample households. Clusters are used to reduce the cost of achieving a given sample size. Thus, the cost of locating 1,000 respondents in a large area is far greater if the 1,000 are widely distributed over the area than if only 50 places have to be visited containing clusters of average size 20.

Choosing the cluster size to use involves two important considerations. First, the larger the cluster, the lower the average cost per element of collecting data in the field, but the larger the average sampling error per element in the survey. There is thus a trade-off between cost and variance. The second consideration involves issues of practicality: the size of clusters used should relate to logistical aspects regarding the organization of the field work and the duration of interviews. Thus, the size of the cluster should take into account the desirable work load for a field team of interviewers and its duration of stay in each UAU. Suppose that the interview is expected to average one hour per household so
Design of surveys to investigate international migration

that, on average, an interviewer is expected to complete three interviews per day (taking into account travel time within UAUs and the need for call-backs when the desired respondent is not at home). Suppose further that a field team consists of four interviewers and one supervisor. The team can therefore complete an average of 12 interviews per day, or 60 per five-day work week. Consequently, 12 might be considered a minimum desirable cluster size and 60 a maximum, assuming the team will have two days of rest a week. In urban areas, where inter-cluster transportation time is minimal, smaller clusters can be used without significantly increasing the duration or cost of fieldwork, but in rural areas, especially those of difficult access or with widely dispersed dwellings, larger clusters might be used to reduce intra-UAU travel time. Since urban areas usually have higher proportions of households with international migrants than rural areas, the use of fewer and larger clusters in rural areas is even more desirable in surveys of international migration than in surveys in general.

Turning now to the other determinants of desirable cluster size, it should be borne in mind that the use of clusters of elements increases the overall survey sampling error or variance. The sampling variance of an element is related positively to the heterogeneity of the population elements (persons) and negatively to the sample size. A basic statistic called the design effect ($deff$) is used to measure the loss in precision (increase in sampling variance) resulting from departures in the sample design actually used in a survey from simple random sampling. More precisely, $deff$ is the ratio of the actual variance of the complex multi-stage cluster design typically used to the variance of a simple random sample with the same sample size (see Kish, 1965: pp. 88, 148-149, 161-163, 217-229):

$$deff = \frac{s_c^2}{s_s^2} = 1 + \rho(b - 1)$$

where $s_c^2$ and $s_s^2$ are the element variances of the complex and simple random sample designs, respectively, $\rho$ (rho) is the intra-cluster correlation coefficient, and $b$ the average number of respondents per UAU (which is the same as average cluster size, or the total sample size divided by the number of clusters). The more the $deff$ ratio departs from (that is, is greater than) 1.0, the greater the "design effect", and the more all the sample statistics based on simple random sampling (including estimates of standard errors of regression coefficients in multivariate analyses) become inappropriate. That is, the standard errors are underestimates, so the statistical significance of regression coefficients is exaggerated. As $\rho$ indicates the average degree of homogeneity within clusters, the more the variable of interest is distributed randomly within clusters -- that is, the less homogeneous the cluster -- the more $\rho$ approaches 0 and the more $deff$ approaches 1, implying that the actual sample design deviates little from simple random sampling. However, $\rho$ is almost always above zero in human populations, reflecting the tendency of neighbouring elements (persons) to be similar. The maximum value of $\rho$, corresponding to complete uniformity or homogeneity within all clusters, is 1.0. The right side of the equation thus indicates that $deff$ increases with cluster size and with homogeneity within clusters.
Values of $p$ are commonly found to range from 0 to 0.2. If clusters of size 21 are used and $p$ is 0.05, $deff = 2$, meaning that for a survey with a complex sample design to have the same precision as a simple random sample, it would have to have a sample size that is twice as large as that of the random sample. Since cluster sampling is less expensive, field costs may well still be lower with the larger sample size and a complex sample design than if simple random sampling is used. Note that if the average cluster size is raised to 31, $deff$ rises only to 2.5, but if $b$ remains at 21 and $p$ rises to 0.1, $deff = 3.0$. Thus changes in cluster size are less important than differences in $p$: a small change (positive or negative) in the value of $p$ results in a large change in $deff$ because it is multiplied by a relatively large $b$, whereas variations in $b$ have less effect on $deff$ since they are multiplied by small values of $p$.

The value of $p$ depends in complex ways on the particular type of sample design. It also varies with the specific variables. For example, family income within urban or rural clusters is likely to have a high value of $p$ because neighbouring dwellings tend to be economically similar, but $p$ is likely to be smaller for variables such as fertility or age of household. In international migration surveys, a larger $p$ is expected because international migrants tend to cluster geographically even more than the general population.

The definition of international migrant used also has implications for cluster size: the more stringent the definition of migrant (e.g. if limited to those arriving within the five years preceding the survey, or to migrants coming from or going to a single country), the rarer international migrants will be and the larger the cluster size should be to ensure that international migrants are encountered in the UAU. Fortunately, $p$ will also tend to be lower, at least with respect to those factors that differ between international migrants and non-migrants. In a survey on internal migration in Ecuador in 1977-78, clusters of 20 households were found to be too small in urban areas to locate a sufficient number of rural-urban migrants arriving during the five years preceding the interview, whereas clusters of 60 households performed adequately in rural areas to capture out-migrants. In the case of international migrants, which constitute far lower proportions of the population, larger clusters will usually be desirable, of say 200 to 1,000 households, in both areas of in-migration in countries of destination and in areas of out-migration in countries of origin.

In conclusion, the interrelations between sample variance and cluster size are complex. The survey objective of achieving a desired number of households with and without international migrants and the survey budget are likely to be the major determinants of both average cluster size and total sample size. But the "cost" of using particular sizes in terms of design effects ($deff$) must be borne in mind in planning the analyses that are to be based on the data gathered.
Design of surveys to investigate international migration

6. Finding "rare elements"

According to estimates of the number of international migrants present in each country of the world, in three-quarters of all countries the proportion of international migrants was at most 6.5 per cent in the early 1990s (United Nations, 1995a) and variations by region and level of development are considerable. Thus, for developed countries the upper quartile of the distribution of the proportion of international migrants was 16 per cent (meaning three-quarters of the developed countries had less than 16 per cent of their population being international migrants), whereas it was about 5 per cent for the developing countries. The median levels were 6 and 3 per cent, respectively, for developed and developing countries. Within the latter there was also wide variation, with the median being about 5 per cent in the Americas and less than 2 per cent in Africa and Asia. Since these estimates are based on the numbers of international migrants defined as persons living in a country other than the one in which they were born (the foreign born), the proportions would be far lower if only those international migrants who had changed country of residence during a five-year period were considered. The numbers above indicate that in almost all countries of the world and indeed in all large countries, international migrants constitute a small proportion of the population. Consequently, they are rare elements, difficult to locate. Similarly, in countries of origin, locating households with international out-migrants is at least as difficult. Indeed, households with international out-migrants are usually more difficult to locate in the country of origin because they are less clustered there than the migrants tend to be in countries of destination.  

The fundamental difficulty of locating international migrants in a survey of international migration is referred to in the sampling literature as the problem of locating “rare elements”. Kish (1965) lists eight procedures that can be used to address this problem: use of multi-purpose samples; cumulation of rare cases from a series of continuing surveys; use of controlled selection; use of stratified sampling with disproportionate probabilities of selection (sampling fractions); use of two-phase sampling; use of large clusters; use of batch testing; and use of special lists prepared as part of multiplicity surveys. The use of multi-purpose samples helps spread the cost of locating migrants, but does not solve the problem of locating a large enough number of international migrants to make their analysis meaningful. The cumulation of international migrants from a series of surveys is impractical for the study of international migration because of the dynamism of international migration itself, which implies that the international migrants captured by one survey may have little in common with those captured by later surveys. The use of batch testing cannot be applied to human populations and is conceptually akin to the use of special lists as part of multiplicity surveys which is discussed in section D below. Consequently, the most viable strategies include the use of controlled selection techniques in conjunction with stratified sampling using disproportionate sampling fractions which can be strengthened by the use of two-phase sampling and large clusters. These are all discussed below.
Developing a sampling design to cover international migrants: The use of disproportionate sampling

Since most countries – particularly those hosting sizeable numbers of international migrants – have census data or data from population registers that can provide some basis for estimating both the number of international migrants and the total population in each area of the country, the following discussion assumes that a population frame exists which can be used to select a sample of international migrants. Although the international migrants identified in the population frame should ideally be the same as those that are to be the focus of the survey, differences in definition will often have to be accepted.

The usual source of data will be that of a population census, which will have information on the geographical distribution of the foreign born (as well as the rest of the population). If a question on time of arrival in the country was also asked in the census, the distribution of those arriving within five years of the census date might be available and should be used in the survey design. Information on country of previous residence, timing of the most recent change of residence and whether that change took place within a certain period prior to the census thus provides a better basis for selecting a sample than using data on all the foreign born. This is especially true if the focus is on recent migrants, as is presumed in this chapter.

Another issue to consider in constructing a population frame is whether the proportions used should refer to international migrants as a proportion of the total population in specific geographical areas or households with international migrants as a proportion of all households in those areas. Since the survey proposed has households as the ultimate sampling elements, the latter would be a better, but it may be difficult to obtain the desirable data. A practical alternative is to seek for each geographical area data on the number and proportion of households whose heads are international migrants.

Once a population frame is available showing the relative proportions of international migrants (or of households whose heads are international migrants) in the different administrative areas of the country (or for the domains of interest), a three-stage sampling design can be implemented. It is assumed that the administrative subdivisions of the hypothetical country under consideration include three levels: provinces, districts and census sectors. (Terminology will differ from one country to another.) The survey design has as ultimate area units (UAUs) the smallest geographical divisions, census sectors. Provided estimates of both the number and the proportion of international migrants are available at each administrative level, sample selection can proceed as follows. At the first stage, provinces constitute the primary sampling units (PSUs), so a sample of provinces is selected with probabilities of selection proportional to the estimated number of international migrants in each PSU (i.e. using PPES as described in section C.2 above). If the number of households whose heads are international migrants is available for each province, PSU selection should be made on the basis of those numbers instead.
Design of surveys to investigate international migration

In the second sampling stage, districts in the PSUs selected are classified into strata according to the proportion of international migrants in the population of each district. That is, the relevant proportions for the districts in the provinces (PSUs) selected are ordered from lowest to highest, appropriate intervals are defined (depending on the number of strata desired, and the extent to which the data fall into natural groupings), and districts whose proportions fall within each interval are classified in the same stratum. Note that districts are included in a stratum irrespective of whether they are urban or rural and irrespective of province.16

Once the strata are created, the key issue is how to select a sample of areas in the second stage from each stratum. In stratified sampling the optimal procedure is to select elements from each stratum in proportion to the estimated standard deviation of the stratum's elements with respect to the variable of interest. Using $P$, the proportion of international migrants, as the key variable of interest, the fraction of the districts to be selected from each stratum should be proportional to the estimated standard error of $P$ for the stratum, namely, $\sigma = \sqrt{P(1-P)}$. Making sampling fractions proportional to $\sigma$ implies that one is using disproportionate sampling, a highly efficient procedure to sample rare elements (see Kish, 1965, pp. 92-98, 142-144, 279-282). To complete the second sampling stage, therefore, districts in each stratum are selected in this manner.

With the data available at all levels as assumed, the third sampling stage is similar to the second. Again, it involves only those census sectors belonging to the districts already selected in the second sampling stage. Census sectors are again grouped into strata according to the proportion of the population constituted by international migrants. Then sectors within each stratum are selected in proportion to the standard deviation of the expected proportion of international migrants in their stratum, as described above.

While the major advantage of using multi-stage sampling is the resulting geographical concentration of field work (mapping, listing households, and interviewing), it can also reduce the work involved in preparing a sampling frame since tabulations on the proportions of population constituted by international migrants need to be prepared for the whole country only at the province level. After that, they are prepared at the district level only for those provinces already selected in the first stage; similarly, they are prepared for census sectors only for the districts selected at the second stage. However, there will be countries where the tabulations necessary for the third sampling stage are not available (either because of problems of confidentiality or because the information about international migrants is considered unreliable at such a small geographical level). In such situations, census sectors can be selected at the last stage simply by using probabilities proportional to the (estimated) population size of those sectors (PPES). In applying PPES in such cases, it is desirable to homogenize the sizes of census sectors by combining those with small sizes and splitting up larger ones so that each sampling unit in the list that constitutes the sampling frame for any given stratum contains UAUs of approximately equal size. From such a list census sectors may be selected randomly (or
International migration statistics

by systematic sampling), preserving the equal probability of selection of households across sectors within each stratum.

Returning to the use of disproportionate sampling, an example of its application may be useful. If four strata are created with mean expected proportions of households whose heads are international migrants of approximately 0.2, 0.05, 0.01 and 0.001, then the standard deviations per stratum are, respectively, 0.40, 0.22, 0.10 and 0.03. Since the optimum sampling fractions for these strata must be directly proportional to the standard deviations, the probability of selecting an element from the first stratum must be approximately twice as high as that of selecting one from the second stratum, four times as high as that of selecting an element from the third stratum, and 13 times as high as that of selecting elements from the fourth stratum.

The actual probabilities (or intervals) to use in each stratum depend also on the overall population sampling fraction for each domain, \( n_d/N_d \), where \( n_d \) is the total desired sample size in domain \( d \) and \( N_d \) is its estimated population size. Notice that although the sampling fractions may be very different across strata, as in the example above, the majority of elements ultimately selected may still be from the stratum with the lowest percentage of international migrants since that stratum will contain by far the largest number of census sectors. Therefore, although the method may assign a very low sampling fraction for that stratum, there should still be enough blocks selected to be representative of the stratum.

How different should the sampling fractions implied by disproportionate sampling be from those implied by the usual simpler, proportionate sampling across strata for it to be worthwhile to use disproportionate sampling? The question is important because proportionate sampling has the advantage of yielding a self-weighted sample that makes all the subsequent statistical analyses simpler since no weights need to be used (see section 7 below). Kish (1965, p. 94) suggests that the sampling fractions should differ by a factor of at least two, a view apparently accepted by Moser and Kalton (1972, p. 94). This means that the proportions of households with heads who are international migrants must differ by a factor of at least four between strata.

One possible variation of the sample design described above is the inclusion of “self-represented” areas, such as a city known to be the major cynosure of international migrants. The city can be included in the sample with certainty, meaning that it is treated as a separate domain where a different sample design may be used, while the rest of the country can be sampled using the three-stage design described above.

An important issue is whether those international migrants who are concentrated in the areas which will be selected with higher probabilities of selection using the sample design described in this section are somehow different in important ways from the international migrants living in areas selected with low probabilities of selection. While we are not aware of any evidence supporting (or contradicting) this widely held view, it does seem to have some a priori plausibility. Suppose that migrants who tend to concentrate in the same neighbourhoods are likely to be less successful than those who live dispersed among
Design of surveys to investigate international migration

the non-migrant population. Even if this is the case, it presents no problem in the present context. In the type of national sample survey recommended here, even areas having low proportions of international migrants would still be included in the sample. Their characteristics and experience as migrants, when properly weighted, will be representative of that population of international migrants in the country. It is only when the domain of a survey is artificially restricted to areas with high concentrations of international migrants that the divergent experience of international migrants settled elsewhere would not be represented, and conclusions would be biased. This is therefore likely to be the case in surveys that select a priori only one city or a few regions where international migrants are thought to be concentrated, rather than taking a probability sample of areas to start with.

The selection of households for interview using two-phase sampling

Following the third stage of the sampling process described above, the census sectors or ultimate area units (UAUs) in which the survey is to be conducted have been selected. The last step is the sampling of the actual elements or households (or individuals) to interview, the goal being to identify reasonable numbers of households with international migrants compared to those without. Even in most of the UAUs belonging to strata with relatively high proportions of international migrants, the majority of the households will not have international migrants and the final sample will probably have a majority of UAUs with few or no households with migrants. It is thus still desirable at this last stage of household selection to use a procedure that ensures that most of the survey effort is not wasted in interviewing households without international migrants. The procedure recommended involves first making a list of all the households in the selected UAUs to determine which have international migrants and which do not, and then to oversample those with international migrants. This procedure is known as two-phase sampling.

Two-phase sampling, also called sequential sampling (an accurate description) or double sampling (a misleading term), involves the selection of a sub-sample of elements (households) from a larger sample (Kish, 1965, pp. 406 and 440-451). There are two main reasons for using two-phase sampling: (a) to collect during the second phase more detailed information on a randomly selected sub-sample of the population covered in the first phase; or (b) to use the first phase as a screening device to identify respondents of special interest (given survey objectives) for the second phase. In (a), more detailed data may be collected from the sub-sample (by using additional questionnaire modules, for instance) and data from phase 1 may be used to provide overall means or other basic characteristics about the larger population, permitting an assessment of whether the phase 2 sample is representative of the whole population surveyed. While purpose (a) has some potential utility in international migration surveys (for example, for identifying migrants of particular interest for intensive
International migration statistics

interviews), the focus here is on (b) since the objective is to screen the entire sample in the UAU selected to seek out households with international migrants.

To implement the procedure, in phase 1 a field operation is used to identify and list all occupied households in the UAUs selected. Using a brief questionnaire, the answers to which are recorded on a single line per household on a sheet (the phase 1 listing questionnaire), the “lister” visits all households in the UAU in some sequential order to record the address, name of head of household, and whether the household contains one or more eligible international migrants (using the definition agreed upon for the survey – in the present case, noting only those who have come from abroad within the previous five years). Then, following a procedure determined in advance, some or all of the households containing international migrants and some of the households without international migrants are selected (sampled) for interview. In the second phase, interviewers administer the detailed questionnaire to the sample selected. The selection procedure used in phase 2 can involve selecting either a fixed number or a large proportion of households with international migrants (depending on the final sample size desired) and similarly a fixed number or a small proportion of households without migrants. Use of two-phase sampling helps ensure that the number of households containing international migrants is not much smaller than the number without international migrants, a group that it is useful to collect data for in order to assess the integration of international migrants (see section B above). Note that if it is decided to exclude households without international migrants altogether, it is not necessary to use two-phase sampling. Instead, interviewers can be instructed instead to go from house to house in the selected UAUs to administer the questionnaire to all households containing international migrants.

To expedite phase 2, the rules for selecting households with and without international migrants in each UAU must be determined beforehand so that the selection of households can take place immediately after the household listing has been prepared at the end of phase 1, while the interviewing team is still in the UAU. Such a strategy minimizes travel time to the UAUs and is therefore particularly important in large countries or where the UAUs are widely dispersed. However, its successful implementation requires that the field supervisors be well trained to follow selection procedures established beforehand for household selection. Thus selection rules must be established a priori, covering all possible situations that may be encountered in the field (in terms of different populations, sizes of clusters, and numbers of migrant and non-migrant households), to prevent interviewers from making their own selection of households in the field, a practice that leads to inevitable biases.

The selection rules for phase 1 can take several forms, depending on survey objectives and on the information available regarding the proportion of households with international migrants in the population. Suppose that the aim is to obtain information from \( M = 1,000 \) households with international migrants and \( N = 500 \) households without, in a population where the estimated
proportion of households with international migrants is \( P = 0.01 \). Assume that the use of disproportionate sampling has raised by a factor of five the mean proportion of households with international migrants in the sample of UAUs selected, making it \( P_u = 0.05 \). If the average size of UAUs is 200 households, each UAU will be expected to contain on average 10 households with international migrants and consequently 100 UAUs would be selected in the country (or domain) to achieve the desired sample size of 1,000 households with international migrants. Suppose four strata have been used with mean proportions of households with international migrants equal to 0, 0.01, 0.05 and 0.2. The mean numbers of households with international migrants in the UAUs of each stratum are therefore 0, 2, 10 and 40. The numbers of UAUs selected in each of the four strata consistent with the above could be 40, 25, 15 and 20. Then the expected number of households with international migrants will be

\[ 40 \times 0 + 25 \times 2 + 15 \times 10 + 20 \times 40 = 1,000. \]

The selection rule in this case would be that every household with international migrants found during phase 1 should be interviewed during phase 2. It remains to be established how many households without international migrants should be selected in each UAU. Since there are 100 UAUs in the sample, 5 households without international migrants should be selected in each UAU. This number should be fixed \textit{a priori}, before the fieldwork in phase 1 begins so that there is no subjectivity in selecting more or fewer households in the field. Systematic sampling can be used to select the sample of households without migrants from the lists made during phase 1, or a cluster could be selected randomly.

Note that, in the example above, if the average size of UAUs had been 500 instead of 200 households, the expected number of households with international migrants per UAU would have been 50 and, consequently, only 20 UAUs would have had to be selected at the third sampling stage. Though a smaller number of UAUs would reduce field costs and facilitate field logistics, it would also lead to larger sampling variance. Hence, to the extent that the size of UAUs can be controlled, opting for average sizes that do not concentrate the sample too much in a few areas is preferable since such concentration increases the design effect of the sample.

It is useful to provide another example to illustrate the value of two-phase sampling compared to other procedures. Suppose that the survey selects a random cluster of households within each UAU instead of relying on two-phase sampling. Suppose the clusters are of size 20 and that the average UAU size is 200 households. If a UAU has 198 households without international migrants and 2 households with migrants located in different clusters, the probability of finding even a single household with international migrants is only \( \frac{2 \times 20}{200} \) or 0.2. That is, only one out of every five clusters selected (and only one of every 5 UAUs) would have a household with international migrants. This means that most of the time interviewers would be visiting households in the sample cluster, wasting their time, not finding any eligible households with international migrants for interview. In contrast, using two-phase sampling and an \textit{a priori} rule stating that all households with international migrants are to be selected,
the two households that qualify would be included in the sample. Furthermore, since the listing process documents the number of households with international migrants and those without in each UAU selected, it permits the calculation of appropriate weights (see section C.7 below).

To conclude, it is worth noting that a two-phase sampling procedure is also useful in designing samples in countries or regions where no data are available on the proportion of households with international migrants (e.g. the census does not contain a question to identify the foreign born). In such a situation, provinces will have to be selected in the first stage with probabilities of selection proportional to total population size (PPES), then districts are selected from those selected provinces the same way (PPES). Then all UAUs in the districts selected at the second stage may be screened through a phase 1 operation to list households and identify those with and without international migrants. Such lists are then processed in a central location and aggregated across UAUs to obtain the overall numbers of the two types of households to determine their respective sampling fractions. The census sectors are then stratified so disproportionate sampling can be used to select UAUs with probabilities of selection proportional to the known UAU sector variance. Once the UAUs are selected, the numbers of households with and without international migrants from each UAU is known. An advantage of this procedure is that the exact numbers of sample households with and without international migrants is known beforehand, which allows precise planning of the time and cost of phase 2 interviewing in the field. A significant disadvantage – and this is why it should be done only when no population frame is available for international migrants – is that it requires an extremely time consuming household listing operation in the districts selected, which are likely to be quite large.

In implementing any two-phase sampling procedure, it is very important to bear in mind that whenever it requires two separate field visits, phase 2 must be carried out as soon as possible after phase 1, particularly when the study of migration is involved. The longer the hiatus between the two phases, the more inaccurate the information gathered during phase 1 may become as individuals and households move into or out of the selected UAUs or die. In a 1975-76 survey of internal migration carried out by the National Statistical Office of Thailand, the migrants to re-interview, identified during phase 1, were sought for re-interviews only 6 to 7 months later; but by this time less than one-third of the households identified as having migrants contained exactly the same migrants as they had initially.

7. Use of weights in the analysis

Probability sampling is necessary in surveys of international migration so that statistically valid tests of hypotheses and inferences can be made regarding the characteristics of international migrants and the determinants and
Design of surveys to investigate international migration

consequences of migration. In a probability sample, every element or observation — every migrant and non-migrant, or every household with an international migrant and every household without — has a known probability of having been selected or included in the sample. In the case of multi-stage sampling, the probability of selection is known at every stage and, consequently, the probability of selecting a particular element (individual or household) is the product of the probabilities of selection at each stage. For example, suppose the PSU or province in which a household is located has a probability of 0.1 of having been selected in stage one, and that the second stage unit (or UAU in this case, which may be a census sector) in which the household is located within the PSU selected has a probability of selection of 0.05. Then suppose further that the UAU contains 100 households of which 6 contain one or more international migrants, and that the a priori selection rule for UAUs in that stratum is that half of the households containing international migrants are to be selected. Then the household's probability of having been selected is (0.1) (0.05) (0.5) or 0.0025. This means that the values of any variable associated with that household need to be weighted (i.e. multiplied) by 1/0.0025 or 400 whenever all the survey observations are aggregated to represent the entire population in the domain. As this example indicates, the general procedure is to weight (multiply) the values pertaining to each element by the inverse of the probability of selection of that element.

Similarly, suppose that the survey also collects data for households without international migrants, and that in phase 1 of the last sampling stage the selection rule states that one out of every 32 households without migrants should be selected. Then, in the UAU described above, (1/32) (100 — 6) = 2.94 or about 3 households without international migrants would be selected, each with an overall probability of selection of (0.1) (0.05) (0.03125) or 0.00015625. The values of all observations for this non-migrant household should therefore be weighted by 1/(0.00015625) or 6,400, a number 16 times greater than that estimated for the household having international migrants and belonging to the same UAU.

In practice, to perform many statistical operations it is desirable to have the weights “normalized”, meaning that each weight for each element is divided by the sum of all the weights for the observations in the final sample. The use of multi-stage stratified samples using clusters in the final stage complicates statistical analyses, but packages such as SUDAAN can handle complex sample designs. Moreover, in surveys of international migration, which inherently involve a serious problem of “rare elements”, it is not advisable to use a simple sample design just so that the sample is self-weighting and statistical computations are easy. Simple sample designs imply very inefficient allocations of field work and must be avoided.

The procedures described in this section — using strict probability sampling methods, developing an appropriate sampling frame, grouping areas into strata according to their expected numbers and proportions of international migrants, using stratified multi-stage sampling with oversampling of areas with
larger proportions of international migrants via disproportionate sampling,
and employing two-phase sampling to identify households with international
migrants in the ultimate area units selected – are appropriate for specialized
surveys of international migration. They apply whether the goal is to interview
only households with international migrants in a country of destination or
households with international out-migrants in countries of origin, or either of
these combined with households without international migrants in the same
country. In most cases, the need to ensure that appropriate comparison groups
are covered will involve undertaking a survey in both the country of destination
and the country of origin. In developing a sample design for the country of
origin, the information gathered through the specialized survey of international
migration in the country of destination should be used to construct a population
frame for the selection of PSUs in the country of origin.

D. OTHER DATA COLLECTION APPROACHES RELEVANT FOR
INTERNATIONAL MIGRATION

1. Multiplicity surveys

Given the problem of identifying and locating international migrants, and
the fact that, once located, an international migrant can often report the location
of other international migrants, especially of those having the same country of
origin as the migrant interviewed or those coming from the same community,
“multiplicity surveys” are sometimes seen as a way to increase the size of the
sample of international migrants at relatively low cost.

Sirken has defined multiplicity surveys as those in which “sample house-
holds report information about their own residents as well as about other
persons who live elsewhere, such as relatives or neighbours, as specified by
a multiplicity rule adopted in the survey” (Sirken, 1972, p. 257). They thus differ
from conventional household surveys, where each individual has only one
chance of being reported on through the dwelling where he or she resides. In
a multiplicity survey, persons have different likelihoods of being reported on and
often more than one chance of being reported upon. Thus, if a survey adopts
a multiplicity rule saying that information about someone may be obtained
from that person’s sibling as well as from that person directly, then each person
has \( n + 1 \) chances of being “observed”, where \( n \) is the number of siblings of the
person concerned that are part of the population sampled. Respondents thus do
not have equal chances of being observed, a deviation from the usual probability
sample. However, in principle, it is possible to adjust for the different probabil-
ities of being reported on by using appropriate weights, provided the survey
collects the information needed to identify the number of possible persons that
could report on a person or an event. In the example just cited, recording the
number of siblings living in the country of interview that everyone has would be
sufficient.
Design of surveys to investigate international migration

An increase in the number of sources that can report an event evidently increases the total number of events reported as well as the opportunities to obtain information about an event or person for any given sample of households. Such a procedure also increases the chances of finding out about rare (and also about sensitive) events in a population. Sampling error can be reduced by the increased sample size obtained from using a multiplicity survey. Certain types of coverage bias may also be reduced by, for instance, obtaining information about whole households that leave an area from the remaining relatives or neighbours. The estimation of emigration based on the place of residence of siblings uses a similar approach (see Chapter 3.A.5).

Variations of multiplicity sampling, known as network or snowball sampling, have existed for many years (Goodman, 1961). All are used to locate or measure relatively rare elements or traits. Multiplicity surveys were originally developed to estimate components of population change, i.e. to measure demographic rates. In terms of the study of international migration, the class of persons allowed to report about an international migrant must be (a) carefully defined, so that there is no ambiguity which requires the interviewer to exercise judgement in the field; and (b) measurable. The need to characterize well the possible respondents severely limits the use of this technique to collect information from “neighbours”, since most definitions of neighbour fail to result in either a fixed number that can be used universally or in a number that can be easily determined in the field for every household. This is unfortunate, since otherwise it would be possible to use at least basic information provided by neighbours concerning the emigration of whole households. The requirement that the number of informers should be measurable can best be met by focusing on the characteristics of persons who have specific relationships with the respondent, such as sons, daughters, mothers, fathers or siblings.

The experience of an exploratory test of the technique to assess the volume and character of movement into, out of and within a given area, may be instructive (Goldstein and Goldstein, 1979, p. 25). The study, carried out in 1978, focused on intra-state movements during 1976–78 reported by a sample of 201 migrant and non-migrant households in a city in Rhode Island, United States. Multiplicity rules were based on accepting information about events from parents, siblings, children and ex-spouses (a very imprecise procedure). Because of recall errors, only information about the migration of relatives in the year preceding the interview could be obtained, and even then only 62 per cent of the moves were reported. Out of these, only 70 per cent of the respondents reported the new address correctly, and only 61 per cent gave the same reason for the move as the migrant gave. The fact that this information was obtained only from immediate relatives in the smallest state in the United States suggests that multiplicity surveys have serious limitations in producing information about absent relatives because of the poor quality of the data obtained. It also provides further evidence that proxy respondents often give distorted answers.

It is still unclear whether multiplicity surveys are useful for sampling international migrants, and some experiments are currently under way in
International migration statistics

Europe. Such procedures need to be used with far more care than has usually been the case, and cannot be considered generally or easily applicable methods.

2. Tracing migrants

Tracing is, in principle, an auspicious procedure for locating rare elements or persons such as international migrants. Even though it is not a sampling technique per se, it is discussed here because it can be considered an alternative procedure — albeit a survey procedure — for locating international migrants. Indeed, it has already been used a number of times for that purpose, primarily in studies of small numbers of international migrants. It has conceptual similarities with two-phase sampling but with the two phases taking place in entirely different geographic areas, in this case different countries. It is also similar to multiplicity surveys because information about how to reach certain persons is obtained from other persons.

There are two types of tracing that may be germane in the present context: (a) first interview households in the country of origin to identify the existence, name and last-known address of former household members who are international out-migrants, and then use the information to seek out and interview those migrants in the country of destination; and (b) first interview the international migrants in the country of destination to obtain the names and addresses of family members remaining in the country of origin, then interview the households of origin.

The purpose of both types of tracing is to obtain information first-hand rather than by proxy from both the international migrant and key household members. Note that tracing is useful only in cases where international migrants have left close relatives in the country of origin whom they are still in contact with, whether or not they are still considered household members. Among the studies that have used tracing, type (a) above is the most common. Not only does the use of tracing improve the quality and completeness of data on international out-migrants and their families in the country of origin, but it also makes possible certain consistency and data quality checks because it is likely that data are collected independently from the different family members in the two countries. Such checks are especially important regarding information on remittances, where the two parties involved (the migrant and family members remaining in the country of origin) may have vested interests in reporting figures that deviate from the true ones in opposite directions.

A survey of internal migration in Sierra Leone involved one of the largest attempts at tracing in a developing country (Byerlee and Tommy, 1976a; Byerlee and Fatoo, 1976b). The survey identified 1,900 rural households with one or more out-migrants to urban areas and then traced 825 migrants to various urban locations in the country. However, only 57 per cent of these were actually
Design of surveys to investigate international migration

traced from households at the place of origin. The other 43 per cent were in-migrants from the same areas of origin reported by those who were successfully traced. The final result is that only 25 per cent of the out-migrants from rural areas were successfully traced, which is almost certain to be an unrepresentative sample.

Despite its attractions, tracing therefore has two major drawbacks. First, the proportion of persons successfully traced is generally too low to avoid serious biases. Taking less than 50 per cent of those eligible usually introduces serious biases in the representativity of the persons traced, and biases may also result when 70 to 80 per cent of those traced are successfully located. The fact that the direction of the bias can often be inferred may or may not help in interpreting the results. Migrants successfully traced will tend to be those who have left most recently, who have known addresses, who maintain close ties with their family members, and who are more economically successful (they send remittances and can afford to visit the country of origin). That is, those traced will usually be positively selected. In Sierra Leone, for instance, the migrants “traced” had a higher mean educational level than that of the larger sample of out-migrants reported by the rural households (five versus years of school completed).

The second major problem with tracing is its cost. A fully fledged tracing survey is likely to cost not only far more than a normal household survey in the country of origin but also more than the combined cost of separate, independent surveys in both the countries of origin and destination. Tracing international migrants from a country of origin to several locations distributed widely across the country of destination is not cheap, especially when the addresses provided may be inaccurate or when maps of the areas of destination may not be readily available or accurate. Similar arguments apply when tracing involves finding the migrants’ households in the country of origin. The geographical dispersion of such households is likely to be greater than that of international migrants in the country of destination and therefore tracing back to the origin is at least equally costly. It is unfortunate that detailed information on the costs of tracing does not seem to be available for any survey. It would be useful to assess the additional cost involved in tracing.

In conclusion, although tracing is occasionally a useful procedure to enrich migration surveys, it cannot be recommended for general use: it is costly and, in most circumstances, it cannot produce a representative sample of the persons being traced. Consequently, it has greater value for exploratory purposes than as a tool to yield confirmatory information on the international migration process.

3. Use of qualitative survey methods to study international migrants

There are several types of procedures involving the collection of qualitative information on a selected group of persons. These include the use of
intensive, in-depth interviews; ethnographic methods; focus groups (see Knodel et al., 1988; Wolff et al., 1991); and so-called "ethno-surveys" (see Massey et al., 1990). Studies based on such approaches are generally based on information obtained from a modest number of persons or a handful of communities that cannot be considered representative of anything larger than the group of surveyed individuals or communities themselves. A further fundamental shortcoming of most of these approaches is that the persons interviewed are not selected according to probabilistic principles and therefore statistically valid inferences cannot be made nor can statistical tests of hypotheses be carried out.

These qualitative methods have another important trait in common: they usually use open-ended questions, allowing the person interviewed to offer his or her version of events, different details and levels of detail about events and their reactions to them. They are thus not constrained by specific questions or precoded responses. Interviewers carrying out in-depth interviews must have the skills needed to keep the interview focused on the topics of interest, using pre-established questions or lines of inquiry to guide the interview so as to make sure that nothing important is completely omitted and that there is some comparability between the approaches used and the topics treated during the interviews of different individuals.

In carrying out specialized surveys on the determinants and consequences of international migration such as those described in previous sections of this chapter, intensive interviews with individual migrants and members of their families can be useful, either before a specialized survey is undertaken or just after the survey. The use of qualitative methods before the survey can provide valuable information about the nature of the international migration process in particular contexts or as viewed by particular types of persons or families, and can therefore be useful in developing questionnaires and in the formulation of hypotheses. After the main survey is completed, once migrants or households with international migrants have been interviewed using standardized questionnaires, some may be visited again for more intensive interviews to learn more about their experiences, their motives and the problems that they have encountered. Such re-interviews can be done for a probabilistically selected subsample of migrants, which would make possible the generalization of the findings of the intensive interviews but would also be expensive in terms of both personnel costs (especially well trained interviewers must be used) and data processing and analysis costs (a great amount of information is obtained in intensive interviews, much of which is not easy to analyse). Therefore, in most practical situations, the number of intensive interviews is usually small (10 or 20) and is not intended to be representative but rather the respondents would be selected by the judgement of the research team. In selecting specific migrants for intensive interview, an effort should be made to include different types of persons: young, middle-aged and older migrants; men and women; migrants from different major countries of origin, or households with out-migrants to different countries of destination; migrants living in different areas (cities, smaller towns, rural areas), etc. The
Design of surveys to investigate international migration

dimensions for selection are evidently numerous. The information collected in the main survey could be used to characterize the main types of interest and to determine which are typical or atypical and therefore worthy of intensive interview.

One qualitative approach that may be used for the purposes above is that based on “focus groups”. It consists of interviewing distinct groups of between 5 and 15 respondents, each group comprising a fairly homogeneous set of persons in terms of sex, age, socio-economic status or other relevant characteristic. The purpose is to learn about commonalities of experience or attitudes within each group by having similar people respond to a set of open-ended questions. The rationale for selecting homogeneous groups is that the more alike people are, the more they are likely to open up in a group interview situation, even stimulating each other to shed light on specific issues. Differences across the different groups in responses to the same general questions can then be revealing about different life experiences or attitudes.

The objective of intensive interviews carried out in conjunction with specialized surveys on international migration is to provide the depth of detail that can help better to understand the migration process, its determinants and consequences than may be possible through data from the usual structured interview lasting an hour or so in itself. In the best of circumstances, skilful interviewers can develop a much better rapport with the respondent during intensive interviews and thus obtain not only more nuanced but also new information, possibly quite revealing, about the person’s migration experience. Of particular interest is the coverage of sensitive issues, such as the problems faced by international migrants (and irregular migrants) to enter or leave a country, bribes paid to officials, illegal activities engaged in (including illegal employment), unfair treatment by employers and family problems. Intensive interviews can thus complement and enrich the somewhat dry quantitative analysis based on large-scale surveys, and thus may provide additional insights that help in the interpretation of the quantitative results.

E. CONTENT OF QUESTIONNAIRES IN SURVEYS TO ANALYSE THE DETERMINANTS AND CONSEQUENCES OF INTERNATIONAL MIGRATION

In previous sections of this chapter, it has been argued that the best approach for the study of both the determinants and the consequences of international migration is to conduct linked surveys in countries of origin and countries of destination focusing on different groups (non-migrants in the country of origin and migrants in the country of destination). Taking that as the basic model for data collection, this section considers the types of information to seek from migrants and non-migrants, their households and their communities of destination and origin to make possible a thorough assessment of the factors
International migration statistics

leading to international migration and its consequences for the migrants involved.

Space restrictions do not allow a careful theoretical discussion or justification of the factors affecting migration at the micro level. Suffice it to say that there are a number of theoretical perspectives from the fields of microeconomics—e.g. human capital theory (cf. Sjaastad, 1962) and the Todaro model (1969)—sociology (e.g. Lee, 1966), geography, political science, and even psychology that can be drawn upon to justify collecting information on a wide range of factors hypothesized to play a role in international migration decisions. Many of the factors involved have been found to be associated with internal migration movements in one country or another. Much less is known about their effects on international migration, but it is clear that in order to understand and quantify specific determinants of international migration, it is important to examine simultaneously (or control for) the effects of the various factors on geographical mobility of individuals when the latter is not completely under State control. Indeed, a major distinguishing trait of international migration is that it is responsive to State policies regarding the admission and residence of foreigners. At the micro-level, individual migrants are normally well aware of the obstacles or barriers that the State erects to prevent the free movement of persons across its borders. Therefore, it is essential in studies of the determinants of international migration at the micro-level that information be collected from migrants and the non-migrant comparison group on contacts and experience abroad, perceptions about State policies, and legal aspects of their status that may be relevant for their international mobility, such as their citizenship or potential right to citizenship of another country.

In contrast to the determinants of international migration, there are no clear theoretical perspectives to guide the formulation of hypotheses pertaining to the consequences of international migration. Most approaches adapt or draw upon theories formulated to explain the determinants of migration. Consequently, there is considerable overlap in the types of factors that are considered relevant for the study of both the determinants and the consequences of international migration. Differences arise mostly in terms of whether the factors of interest are measured with respect to the period preceding the change of residence to another country (for the study of the determinants) or with respect to the current circumstances of the migrant, at the time of interview, after migration (for the study of the consequences). Because the surveys envisaged here generally have the dual purpose of collecting data to assess both the determinants and consequences, it will be necessary to record the status of migrants and their households at both those two points in time in terms of a variety of relevant factors.

Table 6.3 presents schematically a listing of factors thought to be relevant for the analysis of the determinants and consequences of international migration. Three major categories of factors are identified: those measured at the level of the individual migrant or non-migrant; those measured at the level of the person's household; and those relative to the community of residence. Within
Table 6.3. Factors relevant for the analysis of the determinants and consequences of international migration at the micro-level

<table>
<thead>
<tr>
<th>Individual-level factors</th>
<th>Some possible measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Demographic characteristics</td>
<td>Sex, ethnicity, country of birth; age and marital status at the time of migration and currently</td>
</tr>
<tr>
<td>Demographic events associated with migration</td>
<td>Actual or expected birth; actual or expected marriage/divorce; widowhood; orphanhood</td>
</tr>
<tr>
<td>Citizenship</td>
<td>Country of citizenship; right to other citizenship; dual or multiple citizenship; changes of citizenship; citizenship of close relatives; desire for naturalization (if applicable)</td>
</tr>
<tr>
<td>Migrant status in country of destination</td>
<td>Migration status; type of visa, residence permit, work permit; length of validity of permit(s); status upon first admission; changes in migration status; difficulties in obtaining visa or permit</td>
</tr>
<tr>
<td>Education</td>
<td>Level of education in completed years prior to migration; training or school attendance after migration; knowledge of languages before and after migration</td>
</tr>
<tr>
<td>Employment</td>
<td>Years of full time work experience, employment status, occupation and months worked in year prior to migration; current employment status, occupation and months worked in previous year; existence of employment contract and duration after migration</td>
</tr>
<tr>
<td>Earnings, income, benefits</td>
<td>Earnings per month (in money and in kind) in last job before migration or estimated income if self-employed before migration; whether had fringe benefits (health insurance, pension, paid vacation) in last job before migration; and same for current job</td>
</tr>
<tr>
<td>Labour migration</td>
<td>Whether transferred by employer; use of labour recruiter or other intermediary; cost of securing employment abroad; whether underwent employment clearance in country of origin; contacts with potential employers or recruiters in the country of destination before migration</td>
</tr>
<tr>
<td>Contacts with country of destination prior to migration</td>
<td>Close relatives or friends in country of destination prior to migration; previous visits to or stays in country of destination; reasons for those visits or stays</td>
</tr>
<tr>
<td>Aspirations, attitudes</td>
<td>Desire for further education, status, acceptance, material consumption and ownership of goods; marriage aspirations; conflicts with family members or with local norms; reasons for migrating</td>
</tr>
<tr>
<td>Community participation</td>
<td>Membership or participation in local community organizations prior to migration; participation in organizations with links to country of destination prior to migration</td>
</tr>
<tr>
<td>Household-level factors:</td>
<td></td>
</tr>
<tr>
<td>Household size, composition</td>
<td>Number of members by sex and age (children or adult) in household prior to migration and currently</td>
</tr>
</tbody>
</table>
### International migration statistics

Table 6.3. *(Continued)*

<table>
<thead>
<tr>
<th>Household-level factors: (cont.)</th>
<th>Some possible measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Household income</td>
<td>Total household income in money and in kind, in previous month/year and in year prior to migration (distinguishing labour income from other sources of income)</td>
</tr>
<tr>
<td>Household assets</td>
<td>Ownership in country of origin of land or dwelling; of producer durable goods or facilities; of consumer durable goods (type owned, estimated market value); of bank account and intangible assets. Whether goods are location-specific or are marketable or movable. Same for country of destination</td>
</tr>
<tr>
<td>Quality of housing and location in country of destination</td>
<td>Persons per room; whether has electricity, indoor plumbing, flush toilet, potable water, separate private kitchen; access to road, market, school</td>
</tr>
<tr>
<td>Current socio-economic characteristics of household members</td>
<td>Education of head of household, spouse, other adults; school attendance of children and other members; employment status of adult members, earnings and time worked</td>
</tr>
<tr>
<td>Ties to community or country of origin</td>
<td>Duration of residence of family in community or country of origin; presence of friends and relatives in community of origin</td>
</tr>
<tr>
<td>Ties to community of current residence</td>
<td>Presence of friends or relatives of household members in current community of residence prior to migration; assistance received from friends or relatives at the time of migration</td>
</tr>
<tr>
<td>Community-level factors:</td>
<td></td>
</tr>
<tr>
<td>Population</td>
<td>Population size and density of community; composition of population by citizenship, ethnicity, religion, race</td>
</tr>
<tr>
<td>Migration flows at community level</td>
<td>Main countries or communities of origin of international in-migrants; major reasons for international in-migration; proportion of households with international in-migrants</td>
</tr>
<tr>
<td>Employment conditions</td>
<td>Overall average wage level; distribution of labour force by formal and informal sector employment; main occupations and wage levels; unemployment rate by sex, migration status; whether wages or unemployment rate are rising or falling; diversification of employment; whether recent establishment of factories, government development projects; existence of foreign employers; main employers of foreign workers</td>
</tr>
<tr>
<td>Poverty and distribution of wealth</td>
<td>Approximate percentage of income earned or land owned by top 5 per cent and bottom 50 per cent of inhabitants</td>
</tr>
<tr>
<td>Housing conditions</td>
<td>Proportion of dwellings with electricity, indoor plumbing/running water, flush toilet; degree of concentration of international migrant population; existence of ghettos; quality of housing in migrant neighbourhoods</td>
</tr>
</tbody>
</table>
### Table 6.3. (Continued)

<table>
<thead>
<tr>
<th>Community-level factors: (cont.)</th>
<th>Some possible measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communications, transportation</td>
<td>Modes of public transport; use of public transport by and usual commuter routes of international migrants; time required to get to provincial and to national capital; proportion of dwellings with telephones, motorized vehicle; newspaper circulation; existence of newspapers in languages of main migrant groups; number of television and radio stations available; proportion of dwellings with television and radio; access to (proportion of) radio/television programmes in language of international migrant groups; direct international transportation linkages (flights to main countries of origin); media from abroad; international trade and commercial linkages with main countries of origin</td>
</tr>
<tr>
<td>Community facilities and usage</td>
<td>Existence of primary and secondary school; existence of courses in language of international migrants; school enrolment rates by age group and migrant status; existence/number of hospitals; other health facilities per 1,000 population; doctors, other health personnel (per 1,000); use of health facilities by international migrants; existence of banks, post office, government office or police station; existence of cinema, theatre, recreational facility and whether language of international migrants is used; existence of community organizations, migrant associations, cooperatives; existence of welfare services; access by international migrants to welfare services</td>
</tr>
<tr>
<td>Environmental conditions</td>
<td>Existence of or degree of water or air pollution, toxic/nuclear wastes in community; whether community has drinking water, sewerage treatment facilities; garbage collection (per cent of solid waste collected); prevalence of health problems or contagious diseases, mortality levels</td>
</tr>
<tr>
<td>Topography, location, climate, natural disaster</td>
<td>Whether location on seacoast, lake, river; terrain; mean annual temperature, rainfall; whether recent drought, flood, hurricane or other natural disaster and proportion of population affected</td>
</tr>
<tr>
<td>Local governance, community tensions</td>
<td>Responsiveness of local authorities to community needs; degree of democratization; citizen empowerment; local respect for human rights; existence of targeted violence; indicators of social conflict and xenophobia</td>
</tr>
<tr>
<td>Community norms</td>
<td>Degree of openness to new ideas, to outsiders; traditional values fixed or in transition; norms regarding women's education, women's work outside the home, social and sexual equality; degree of homogeneity of society; existence of ethnic, linguistic or religious minorities; differences in norms among minorities</td>
</tr>
</tbody>
</table>
International migration statistics

each category, several sub-categories are identified. In general, table 6.3 is constructed assuming that the information is obtained from the migrants themselves in the country of destination. As explained in previous sections, equivalent information needs to be obtained from non-migrants in the country of origin, making the relevant modifications because of the change of survey site and, more importantly, ensuring that appropriate retrospective information is also obtained.

The complexity of international migration and its diversity from one context to another means that the factors listed in table 6.3 can neither be exhaustive nor necessarily relevant in all circumstances. The list is thus intended to be suggestive of the factors that merit consideration but whose importance will vary from one situation to another. The discussion below, however, endeavours to highlight those factors most commonly considered relevant.

1. Individual-level factors

Most factors listed that refer to the individual level are expected to be relevant in most circumstances and, consequently, information on them should be recorded in any survey which gathers information for the study of the determinants or the consequences of international migration. The demographic characteristics of migrants at the time of migration are indicators of social status that influence the propensity to migrate. Sex, in particular, is a powerful control variable because the experiences of men and women that may relate to international migration often differ markedly according to acceptable sex role perceptions in the society of origin and that of destination (Hugo, 1993b; United Nations, 1995b). Demographic events that themselves can affect the propensity to migrate and the consequences of that migration for the person involved are listed separately so that they are not disregarded. The issue of whether international migration may be the cause of divorce or legal separation should also be borne in mind.

Information on citizenship and migrant status in the country of destination is crucial to assess the consequences of international migration, since migrants with restricted residence or work rights may not reap the same benefits from migration as those who are granted a wider range of rights. Distinguishing between migrants in an irregular situation and those whose presence is fully sanctioned by the State in which they reside is also fundamental. Investigating the problems faced by migrants in obtaining visas and other required permits and documents (for both travel and changes in nationality) sheds light on the effects of State efforts to control international migration. This has not been done adequately at the micro level.

With respect to socio-economic factors, education both affects migration and may be affected by international migration. Thus certain types of international migrants may be selected among the better educated, and some
Design of surveys to investigate international migration

individuals move to improve their educational attainment (to attend a special training course or obtain a higher degree). Knowledge of languages is also likely to have some impact on the decision to migrate and, more importantly, on the selection of country of destination. Improved knowledge of the local language after migration has taken place indicates a more successful integration in the country of destination.

Economic variables are generally considered of key relevance in determining migration decisions, since most migrants seek to improve their standard of living by migration. People are usually less likely to migrate internationally if they are unemployed or poor and more likely to migrate if they have certain skills. Although the effect of differences in wages between the countries of origin and destination is considered a key determinant of international migration, empirical evidence to test its importance has been scant. According to human capital theory, people tend to migrate if predicted earnings are greater in the place (or country) of destination than in the country of origin (provided the difference exceeds transportation, psychological and other costs). Examples confirming the relevance of such variables in influencing international migration are given by Adams (1993) and El-Saadani (1992), among others. Other approaches to the estimation of the determinants of migration use the actual wage prior to migration instead of the predicted wage. In most cases, migrants have some idea of their employment and earnings prospects prior to migration. Finding out more about their a priori expectations, their sources of information, and the extent to which the information was correct is important to assess whether migrants make well informed decisions and what their consequences are. Thus, migration will usually have great effects on the employment, occupation, wages and conditions of work of the migrant. To the extent that the migrant had accurate information before migrating, his or her aspirations are more likely to be realized.

In the specific case of labour migration, a crucial issue is the extent to which migrants are assisted in finding employment abroad by their current or prospective employers, labour recruiters, government agencies, and friends and relatives. Information on the existence of contracts, on the cost and benefits of moving abroad, and on the treatment the migrant received is essential in assessing the consequences of labour migration.

Social networkers are considered to play a crucial role in fostering international migration. They can both stimulate and facilitate changes of residence of persons linked to such networks. Among the many studies that have found support for the role of networks in promoting international migration are those by Hugo (1981), Taylor (1986), Fawcett and Arnold (1987y), and Massey (1990). Investigating the types of contacts that the migrant had with the country of destination prior to migration, both in terms of persons he or she knew there and previous visits to that country or even previous residence, is therefore essential.

An individual's own aspirations, attitudes and motivations are key determinants of migration, but are often ignored because they are difficult to measure.
International migration statistics

objectively. Similarly, the realization of a migrant's aspirations might be used as an indication of success, but is generally not sufficiently explored in assessing the consequences of migration.

Participation in community organizations gives a feeling of belonging and provides more social contacts that may serve either to discourage someone from moving away or prompt migration when the organizations involved have linkages abroad. Furthermore, participation in local organizations in the country of destination can be an indicator of the level of engagement of the migrant with the host community and his or her commitment to stay in it.

2. Household-level factors

Household-level factors also include the demographic, economic and social characteristics of the household which may both affect the decision to migrate and be affected by migration. Demographic factors include household size and composition. Size prior to migration is relevant because large households have a lower propensity to migrate as a unit but at the same time are more likely to experience the out-migration of members. Such a relation may be stronger among lower income households for which migration is part of a survival strategy to diversify sources of income over space. After migration, household composition may reflect restrictions imposed by the receiving State on family reunification, and will evidently have important effects on the psychological consequences of international migration for both migrant household members and those remaining in the country of origin.

Household income can affect the likelihood of international migration in two opposite ways. Higher household income makes it easier to finance the often considerable costs involved in international migration, but to the extent higher income results mainly from higher earnings from employed household members, higher income also indicates a higher opportunity cost of international migration for those members and for the household as a whole. Household assets can also have positive or negative effects on international migration depending on their type (Oberai, 1984). Those that are easily marketable or “liquid” can be sold or realized to finance international migration but assets that cannot be easily liquidated may have a negative effect on out-migration (DaVanzo, 1976). The latter include certain producer assets, such as specialized capital equipment or machinery used in industrial processes, and the value of good will built up over time by businesses and in certain professions, as measured by customer loyalty. In terms of the consequences of international migration, changes in household income and assets associated with a change of residence are important indicators of the economic effects of migration on the household.

The type and quality of a migrant's dwelling prior to migration can be factors slightly influencing the decision to migrate, and after migration they
indicate the status of the migrant’s household. Similarly, the socio-economic characteristics and employment status of members of the migrant’s household may be relevant. In particular, comparing the employment status (or other measure) of members of the household prior to migration and at the time of interview may indicate important effects of migration on household units. Certain household members may be “tied migrants”, whose migration depends on the decisions of others. They may be particularly disadvantaged by household migration. In terms of the determinants of migration, the characteristics of household members prior to migration (including those of members who do not themselves migrate) may contribute to household decisions to migrate or not.

Ties to the community of origin can both determine in which circumstances international migration takes place and the orientation of migrants and their households while abroad (whether they remain attached to the community of origin or not). The strength of such ties is related to the number of years spent in the community of origin as well as to the breadth and depth of personal friendships and relationships with community dwellers. It is also of interest, particularly in relation to an assessment of the consequences of international migration, to explore the ties of migrants and their households to the communities of destination in terms of the presence of friends and relatives and the assistance received from such persons at the time of migration or subsequent to it. The anticipation or expectation of this assistance is widely considered to be a key factor in determining international migration. And the realization of the assistance is often instrumental in the migrant’s initial successful integration.

3. Community-level factors and their utilization

Throughout the social sciences there is increasing recognition of the importance of context in affecting human behaviour. In the case of migration, this has long been recognized, but the specific desirability of collecting data at the community or contextual level to analyse the determinants or consequences of individual or household migration decisions can be traced back only to Wood (1981, 1982), Bilsborrow (1981) and Findley (1982). A discussion of the importance of taking community-level factors into account and of which factors to consider for the investigation of the determinants of internal migration in low-income countries is provided in Bilsborrow et al. (1984, Chapter XI). It is important to take those factors into account in international migration as well, along with State policy factors that may be measured at the community level. In practice, community factors are measured by using data obtained through specialized community-level questionnaires administered to community leaders, selected informants, or groups of residents.

Factors operating at the community level can be conceptualized as intermediate variables between those at the macro or national level and those at the
International migration statistics

individual or household levels. They thereby serve as a filter between national policies and the households of the community, influencing individual and household behaviour and themselves affected by higher level government policies. Table 6.3 presents a list of possible community-level factors relevant for the analysis of the determinants or consequences of international migration. From the point of view of the analysis of the determinants, factors characterizing the community of origin may be particularly important. Those characterizing the community of residence in the country of destination also may have some relevance in determining international migration (mainly the choice of destination) but are more relevant for assessing the consequences of migration.

As in the cases of individual and household level factors, demographic factors may be relevant at the community level as well. Communities with a larger population size and a higher density are likely to offer more opportunities than a smaller place and thus tend to be more attractive for international migrants. The composition of the community in terms of citizenship, ethnicity or religion is also likely to change because of migration, and in a community of origin may put pressure on minorities to leave. The community's migration history is also likely to influence current and future developments, whether the community acts as origin or destination for international migrants, because it will be associated with direct ties and information flows. Investigating the internal migration flows that the community has experienced may also be relevant, since it is argued that international migrants may be substitutes for internal migrants in communities of destination or that they may drive out native residents in some communities. In addition, international migration may be the last step in a multi-stage process, following people's arrival in a large city with strong international ties after a series of internal migration movements.

Employment conditions in the community as a whole as well as the prevalence of poverty are factors that set the stage for the attraction of international migrants (or the retention of the local population in communities of origin). Recent (or planned) expansion of employment opportunities through investment by private sector employers or through government development projects in the community may have a significant impact on attracting international migrants (or in retaining the local population, when that expansion occurs in communities of origin).

Housing conditions in the community of destination are likely to be a minor factor in migration decisions, but may indicate broader aspects of quality of life, poverty and income distribution, which may promote or retard the integration of international migrants. Differences in housing conditions between international migrants (or particular groups thereof) and the general population of the community of destination may be related to the operation of networks, government housing policies vis-à-vis immigrants, or the location of employment opportunities, as well as the relative economic status of international migrants.

Given the importance of information and networks on international migration, both communications and transportation linkages between the
Design of surveys to investigate international migration

community of origin and the national capital or the world at large are most relevant. Thus the distance to the national capital, modes and frequency of public transport available, and time to get there are important, presuming that the capital is the centre for international contacts. Other potentially useful indicators of international influences in the community of origin are the availability of media programmes produced abroad for the local population, and the presence of persons who have lived abroad and brought in ideas about conditions in another country (return migrants, in particular). For the community of destination, the existence of transportation linkages with specific countries and of newsprint and media programmes in the languages of specific migrant groups is an indication of the cultural and social importance and degree of integration of those groups.

In the community of destination, the types of facilities and services in general, as well as their accessibility to international migrants, are commonly thought to play a role in attracting migrants. In the community of origin, the degree of access to such needed services can be a factor shaping the decision to migrate, though not necessarily internationally. Similarly, environmental conditions in communities of origin are increasingly being perceived as causes of migration, though most result in a change of residence within countries. In Mexico, for instance, a 1986 migration survey carried out in 16 cities showed that, for Mexico City, concerns about the deteriorating environment and the "agitated life style" were the two most commonly expressed reasons for intending to move away (CONAPO, 1987). The influence of environmental factors on decisions to move abroad has not been explored. Nor is there much information on how environmental factors in the community of destination shape the consequences of migration for the migrants involved. In a similar vein, the physical location of the community of origin and the likelihood that it may be affected by natural disasters are factors that probably influence decisions to migrate where they occur. As attributes of the community of destination, their role in influencing locational decisions of international migrants may have some relevance.

A more important factor shaping the decision to move from one country to another is the character of local governance, or the existence of tensions or outright conflicts in communities of origin. From the perspective of the community of destination, the consequences of migration will also be influenced by local governance issues and by those related to the existence of intolerance, discrimination or human rights violations affecting international migrants.

Lastly, community norms are part of the context in which people live and provide structure and familiarity to daily life. However, certain persons may consider the norms characterizing the community of origin as stultifying and may decide to migrate in order to avoid them. In some developing countries, norms regarding the roles and status of women may be a factor in prompting the out-migration of women who seek opportunities in a more egalitarian society. In such cases, the norms of the community of destination also play a role by providing the counterpart to the community of origin. Norms in the community
International migration statistics

of destination that foster prejudice and discrimination against minorities also evidently have an influence on the consequences of migration for migrants.

As noted earlier, current empirical approaches to the analysis of the determinants and consequences of migration at the micro level recognize that migration decisions are influenced not only by the characteristics and prospects of the individual migrant and his or her immediate family but also by factors associated with the community of origin and possible communities of destination. This view derives in part from the place-utility approach espoused by geographers and the earlier work of sociologists such as Stouffer (1940, 1960) and Lee (1966), and is concerned with the perceived utility of alternative places of residence. Since information on perceived utility is difficult to obtain, subjective and usually unreliable, it is rarely sought during data collection. Instead, information on the objective characteristics of places is collected. Thus differences in the characteristics of the community of origin and that of destination are important a priori contextual-level factors that influence the international migration decisions of individual migrants and their households. They should be taken into account explicitly in modelling migration functions to reduce the misspecification that characterizes models in which contextual-level factors are not included. Because contextual variables far more than individual variables reflect factors subject to policy intervention, models that include them (so called multi-level models) are potentially much more useful for deriving policy implications than the usual ones restricted to individual- and household-level factors. In multi-level models involving more than one country, policy factors differing across countries can be explicitly included (see also section B.2 above).

Figure 6.1 illustrates how factors at different levels may influence the international migration decisions of individuals and households. The figure shows the relevance of conditions in both the country of origin and that of destination. At the top left is the characterization of persons (migrants and non-migrants) in the country of origin. Their own characteristics influence who moves and who responds to the differences in conditions between origin and destination. Below are the contextual factors or conditions in the immediate community of residence or local reference area in the country of origin. At the bottom left, "national policies" indicates that the government of the country of origin may have policies to influence, directly or indirectly, the socio-economic situation and value systems of the local community, as well as policies which directly facilitate or restrict individual international migration decisions. The context of the community of origin acts as a filter through which factors from beyond the community, such as government policies, must pass in order to influence the decisions of households and individuals of that community. Similarly, the effects of factors in the country of destination are indicated on the right. In practice, the latter should refer to the initial community of destination of the migrant, which is most directly relevant to the study of the determinants of migration. To study the consequences of international migration, factors related to the current community of residence should be contrasted with those of the migrant's community of origin in order to assess change.
Design of surveys to investigate international migration

Figure 6.1. Illustration of multi-level model of determinants of international migration

**Country of origin**

- Individual/household characteristics (O1)
- Local/regional factors (O2)
- National policies (O3)

**Country of destination**

- Migration characteristics (D1)
- Local/regional factors (D1)
- National policies (D2)

O1. e.g. age, sex, education; skill/work experience, occupation; current marital status; motivation; location-specific capital and other assets; household size and composition; income; attitudes/sex roles.

O2. e.g. employment conditions (overall wage levels, unemployment rate); wage levels of particular occupations; employment opportunities for women, children; availability of and quality of schools, health facilities, cultural and recreational facilities; transportation/communications links with national capital, other countries; ethnicity/religion; norms regarding women’s roles; presence of close relatives, friends.

O3. e.g. level of economic development, wage/income level; policies relating to economic conditions in general; income/wealth distribution, poverty; tax/fiscal policy; exchange rate/foreign trade policies; expenditure policy, including share on social and economic services and infrastructure; policies to encourage people to work abroad, activities of foreign labour recruiters; restrictive policies on out-migration, or on allowing citizens to travel abroad as toursts, students.

D1. As in O2, viz. (relative) employment characteristics (overall wage and unemployment levels, levels for particular occupational groups); availability and quality of social and economic facilities; amenities; presence of friends, relatives, same ethnic or racial group; language; norms regarding women’s employment.

D2. As in O3, viz. overall level of economic development, wage/income levels; economic policies; tax/expenditure policies; foreign trade/aid flows; extent of business fixed investment; policies to recruit foreign workers; visa policies, border controls and other physical barriers to immigration and enforcement; quotas and other direct restrictions on immigration; political opposition to immigration, racism; official policies regarding family reunification; internal restrictions on illegal immigrants (employer sanctions, restrictions on their use of social services, etc.).

F. QUESTIONNAIRES FOR SURVEYS IN COUNTRIES OF DESTINATION AND ORIGIN

1. General questionnaire design and presentation issues

This section describes the content of questionnaires for specialized surveys of international migration. As explained in section B.2 above, it is highly desirable to obtain information for both the appropriate comparison group of non-migrants and for international migrants. Therefore, most issues of questionnaire design and content can be treated together both for international migrants and non-migrants and for countries of origin and destination. Hence, in the interests of saving space, a detailed discussion of questionnaire content for countries receiving international migrants (including the skip patterns necessary
to accommodate non-migrants as well) will be presented first. Then, indications of how the questionnaires presented can be adapted for surveys in countries of origin of international migrants and for surveys of return migrants will be provided.

Prototype questionnaires are presented in the annexes. Because they are intended to serve only as models, the questions they include must be pretested in the field and modified as necessary for any actual application. Those conducting a survey will also want to include additional questions or exclude some questions, depending on the particular circumstances involved or on the objectives of the survey. In addition, prior to pretesting in the field, the questionnaires must be translated accurately into whatever languages are likely to be encountered (and translated independently back into the original, to ensure accurate translation).

The questionnaires are complete and self-contained, including skip patterns that make them appropriate both for households with international migrants and those without. Questions and skip instructions are grouped into sections or modules. Three full questionnaires are provided for countries of destination: a household questionnaire, an individual questionnaire, and a community-level questionnaire. The individual questionnaire comprises sections entitled migration and citizenship, pre-migration situation and activity, arrival in country, and current work. The first and fourth sections of the individual questionnaire should always be used to interview non-migrants as well as the household questionnaire. In addition, two optional modules are provided for individuals, one to be used to interview migrant workers and the other to interview women on fertility and family planning.

The definition of international migrant used has some effects on questionnaire design and especially on the time reference of information sought in both the household and individual questionnaires. In general, for reasons indicated in section B above, data will be sought from both migrants and appropriate non-migrants pertaining to both the time of the survey and the time before migration. If a five-year cut-off is used in defining international migrant, the mean time at which migration took place is 2.5 years before the survey. Retrospective data on the situation of non-migrant households (in the country of origin) should be sought for that reference date, as is done by the questionnaires presented in the annexes. If a ten-year cut-off were used, the mean time at migration would be about five years before the survey and the questionnaires would seek information from non-migrants for that reference data. The cut-off point used should not significantly affect questionnaire design, except that less detail should be sought if a cut-off further in the past is used.

The appropriate respondent for the household questionnaire is the person who knows the most about the household’s economic situation and its sources of income, which will usually be the head of household, although the spouse or some other adult will sometimes be as knowledgeable. The appropriate respondents for the individual questionnaires are persons who may be involved in making migration decisions. That group is assumed to comprise every member
of the household who is at least age 15 years at the time of the survey. While it
takes more time to interview each such person separately compared to seeking
information from a designated household respondent, such as the head of
household, information of greater reliability and richer in detail can be collected
when it is solicited directly from each person. 22

Given space limitations, only selected questions are discussed below and
the reasons for their inclusion are mentioned briefly. Some of the theoretical
rationale for including the questions presented is discussed in section E above.
Note that asterisks preceding question numbers in the prototype questionnaires
included in the annexes are used to indicate those that the question is of
secondary importance, of a subjective nature (“soft” or referring to attitudinal
issues), more difficult to implement in the field, or less likely to produce reliable
responses.

2. Questionnaire design for a survey in the country
of destination of international migrants

Household questionnaire

A basic household questionnaire is provided in Annex 1 (pp. 363–402). As
explained in section E above, it is crucial always to collect information on the
current composition of households both with and without international mi-
grants. Thus all households interviewed in the country of destination (or in the
country of origin) must be asked to fill in the household roster and to provide
other information on housing quality, household assets, unearned income, and
location of dwelling. The respondent to the household questionnaire should be
the head of household or the spouse of the head.

The household roster (questions H.1-H.10) provides for a complete listing
of household members, based on the concept of de jure or legal residence, though
the question on whether each person slept in the house during the past week
makes possible a de facto classification and serves to check whether a person is
in fact a de jure member who is temporarily away. The head of household should
be listed first and each member’s relationship to the head must be recorded. In
addition, the following information is collected on each member: age, sex, level
of educational attainment and current school attendance, current marital status,
and work status. Consequently, at least this information is available for all
adults household members, even if temporarily absent and therefore not avail-
able to respond to the individual questionnaire. The roster also includes a ques-
tion on place of birth, to be coded by province or state for those born in the
country and by country for those born abroad. Answers to that question
provide an indicator of the net lifetime international migration of all current
household members. If there are children in the household, data on their places
and dates of birth provide some information on the international migration of
the mother.
International migration statistics

Questions on household tenure, quality, and assets (H.11-H.18) and on household location (H.19-H.20) follow, the latter indicating the quality of location of the dwelling. Questions on household income from sources other than wage income and income from farm and business activities (both covered also in the individual questionnaire) are included in the household questionnaire (H.21) to provide for complete household income.

The household questionnaire should be administered also on the situation of the household prior to migration. Thus, in a survey in the country of destination, the migrant should be asked the composition of his or her household and its situation in the country of origin just prior to migration (see sections B.2 and E above). This is crucial to assess the extent to which the migrants have benefited from migration and also provides information on the situation of the migrant’s household at the time migration took place. See also the discussion below on question 2.11 in the individual questionnaire.

Most questions in the household questionnaire do not usually encounter problems in field implementation. However, since the quality of data from responses to H.21 is sometimes suspect, the simpler questions H.17-H.18, which indicate socio-economic status, are also recommended.

Individual questionnaire

The individual questionnaire comprises four sections and two optional modules. Section 1 on migration and citizenship is fundamental since it identifies international migrants and thereby screens the sample population for non-migrants. The latter do not answer sections 2 and 3, and hence skip to section 4 of the questionnaire. The definition of international migrant used here is a person who has lived in another country for at least six months, has moved to the country of interview during the five years preceding interview and was at least age 15 years at the time of the move. The age cut-off at 15 years is intended to ensure that only those persons who might have actually participated in making migration decisions are interviewed.

Questions 1.1 and 1.2 identify the date and place of birth for all persons, aged 15 and over, who are the respondents to the individual questionnaire (as identified by the household questionnaire). Place of birth is one of the three aspects of international migration (along with citizenship and country of previous residence) that must be determined early in the survey, and is straightforward to collect. Question 1.3 identifies non-migrants, who skip a number of questions. Persons born in the reference country who moved to some other country (determined by questions 1.3 to 1.6) are native born returnees, while all persons who have lived in the reference country and come back (even if they were born in some other country) are return migrants. Evidently, the latter includes native born returnees as a special case. Anyone who has ever lived for over six months in any country besides the reference country and the country of birth is then routed by question 1.6 to a question on the international migration history of the respondent. Only moves from one country to another for more
Design of surveys to investigate international migration

than six months are to be recorded. Note that no information about activities in previous countries of residence is sought, to be consistent with the focus on the most recent move.

Questions 1.7-1.8 identify (recent) international migrants for purposes of the survey and question 1.9 records current citizenship. The sub-questions in 1.9 record whether the respondent has ever been a citizen of another country and the date of change of citizenship. Question 1.10 records the citizenship of immediate relatives and question 1.11 records their country of residence. Both are important in establishing the potential for further international migration (for family reunification purposes, for instance). Further questions on this topic for international migrants only are found in section 3.

Question 1.13 determines language knowledge and questions 1.14 to 1.17 explore the intentions of migrants and non-migrants for future migration. Seeking to determine whether the person has specific plans for departure – regarding both a time and a destination – has been found important for distinguishing those who are likely to move and those who are not, whatever their answers to 1.14.

Section 2 on pre-migration situation and activity seeks information on the situation of the migrant and the migrant’s household prior to international migration. The first question, 2.1, confirms previous country of residence and establishes whether the migrant lived in an urban or rural area, information that is important in planning the survey in the country of origin. Question 2.2 identifies refugees and asylum-seekers who are routinely neglected in surveys of international migration but are increasingly important (see Chapter 5 above). Moreover, virtually all of the subsequent questions are relevant for those types of migrants as well, since refugees often exercise economic activities in the receiving country. Question 2.3 inquires why the respondent left the previous country of residence, listing conditions applying to that previous residence, thus distinguishing this question from that in section 3 on why the migrant chose the current country of destination. Failure to distinguish the two issues has created confusion in other surveys and led to excessive criticism about the value of data on reasons for migration. The two-part question seeks to identify both reasons spontaneously mentioned and the main reason.

Questions 2.4 and 2.5 ask about contact with the country of destination prior to migration, and questions 2.6 and 2.7 ask about sources of information about the country of destination and employment prospects in it: previous information and contacts play crucial roles in eliciting migration. Question 2.8 records marital status and changes therein in relation to migration. Question 2.10 inquires about who actually made the migration decision, a topic that should be explored before question 2.11 to minimize the contamination that might otherwise occur by posing that question in the context of compiling the roster pertaining to the household situation prior to migration. Note that question 2.11 calls for the full administration of the household questionnaire pertaining to the time immediately preceding migration. The respondent to that questionnaire will normally be the migrant himself or herself. In cases involving
the migration of several adult members of a household who are still living together in the selected household, only the head of household (if also a migrant) need be asked to provide the information on the situation of the household prior to migration, thus avoiding unnecessary duplication. The next bloc of questions, 2.13-2.32, inquires about the economic activity of the migrant prior to migration, including job search activity. The information sought is essentially the same as that on current work, which is recorded for all migrant and non-migrant adults in section 4. It is also desirable to collect basic data on the earnings of all (other) members of the migrant's household just prior to the migrant's departure from the country of origin. In lieu of asking the migrant to provide the same level of detail as in questions 2.13-2.32 for each person, a few questions might be asked about the earnings of the head of household and income of any farm or household enterprise. These questions are 4.2, 4.4a, 4.6, 4.7, and 4.10, for employees, and 4.19 and the columns on estimated monthly income and (if appropriate) total land owned from the household questionnaire. Question 2.33 identifies if the person has had contact with a labour recruiter or contractor so as to identify the migrant workers to whom the optional module described at the end of this sub-section will be administered.

The need for documents, a fundamental aspect of international migration, is covered by questions 2.34-2.38 with respect to the country of previous residence or origin. Questions 2.34a-c explore whether the migrant was in possession of the documents needed to leave the country of origin. Questions 2.36-2.38 probe for further information about documents needed from other countries to leave the country of previous residence and difficulties in getting them. Question 2.38 seeks information about unauthorized payments (e.g. bribes to authorities or border personnel), though information about such topics will often not be forthcoming in an interview of this nature (see section D.3 above). Lastly, questions 2.39 and 2.40 are optional questions seeking the respondent's assessment of the relative status of his or her household in the previous country of residence and participation in community organizations. The former tests the "relative income hypothesis" for migration: are those who migrate from relatively well-off or relatively poor households?

Section 3 on arrival in country is the second major module addressed to international migrants. Whereas Section 2 covered the situation of migrants prior to migration, section 3 covers their arrival and adaptation to the country of destination, except for their current economic activity, which is covered in section 4. The first questions relate to the timing and experience of arrival. Question 3.2 asks why the migrant chose the country of interview as country of destination, and question 3.3 attempts to distinguish international migrants arriving with documents from those without. Questions 3.4-3.5 cover the problems confronted in gaining admission to the country of interview, and question 3.6 refers to documents possessed by the migrant, how they were obtained and whether any problems were encountered in acquiring them. Questions 3.7 and 3.8 ask about job transfers or about how the first job (apart from jobs arranged by labour contractors or recruiters, which are covered in the optional module

310
Design of surveys to investigate international migration

discussed below) in the country of destination was obtained. It is important to
distinguish the different ways of acquiring the first job to assess the extent to
which migrants move because they know of such opportunities in advance or
because such opportunities are expected from information obtained through
relatives or friends (covered in question 3.9) or in the absence of any such
contacts. Migrants without jobs on arrival are asked in question 3.11 about job
search activity. Questions could also be asked regarding the migrant's first job in
the country of destination, which would be useful to study the assimilation
process over time, but this issue is not discussed here further.

Question 3.12 is the first of several questions on familial aspects of the
migration, including marital status and whether other family members accom-
panied the migrant at the time of migration or followed later. Whether the
migrant's family is with the migrant, provided he or she was married or wanted
them to be with him or her, determines the consequences of migration for the
migrant. Questions 3.13-3.15 ask about family members who had come to live
with the migrant but who have since left. Those who had come only for short
visits (less than six months) should not be mentioned here and only the most
basic characteristics are obtained on those re-migrating (or dying). The data on
age recorded should refer to the age at death or re-migration.

Question 3.17 covers the important topic of formal education and on-the-
job training after arrival in the country of destination. Both reflect increases in
human capital, which is a key positive consequence of migration in itself and is
related to the migrant's overall economic success. Questions 3.17a and 3.19-3.20
cover the extent to which the migrant and his or her family use health and
education facilities or receive other government benefits, the cost of which is
a major policy issue in most of the major countries of destination of inter-
national migrants.

Question 3.21 inquires about visits back home, which are important for
migrants away from their family (see question 1.11), but can also reflect lack
of integration into the host society. Question 3.22-3.24 ask about changes
in citizenship and intentions to change citizenship. The latter reflects
long-term intentions to remain and integrate or not. Question 3.25 covers
language ability and changes since arrival, a development that is closely
tied to the migrant's success in the country of destination. Lastly, question
3.26 is an attitudinal question seeking to solicit the migrant's overall evaluation
of his or her experience in the country of destination. It also suggests the
existence of pull factors regarding relatives or friends remaining in the country of
origin.

The last section of the individual questionnaire, section 4 on current work,
seeks information from all migrant and non-migrant persons aged 15 years and
over on their current economic activity. Most of the information is fairly
standard and covers: labour force status; earnings, occupation, time worked in
past year, and benefits; secondary work; age when first employed and total years
of employment; whether the person has a written work contract; and income
and assets from a family business or farm. For the latter, it is desirable to
International migration statistics

administer a series of questions if there is any hope of obtaining data of usable quality (see the enterprise questionnaire at end of section 4). For those respondents who are not currently “working”, questions probe whether they are looking for work and why they are not working. Several questions are included on membership in a labour union and the ethnic composition of the labour force in the place of employment, the latter reflecting enclaves. Employment in such enclaves may not benefit migrants as much as other employment. The “soft” attitudinal questions 4.27-4.28 are optional, included to solicit attitudes towards women’s work, from both women and men, to be able to study the relations between such attitudes and both women’s participation in the labour force and their international migration. Question 4.29 determines if the migrant retains assets in another country, possibly the country of origin, which may provide income and facilitate a return move.

Optional module for migrant workers

This optional module covers a subset of international migrants which is particularly important to a number of countries of both origin and destination, namely, persons who migrate with the specific purpose of exercising an economic activity in the country of destination and who do so using some kind of intermediary to find a job abroad (this is a subset of migrant workers as defined in Chapter 2 and also includes persons migrating in an irregular situation to work abroad who find jobs through intermediaries). In certain contexts, such persons may work in a country other than their own for less than six months and, consequently, according to the definition of international migrant adopted for the survey, will not be covered by it. If persons working abroad for shorter periods are of interest, the survey takers may want to modify the instructions regarding coverage so that short-term migrant workers are interviewed, using mainly the module for migrant workers.

The perspective of the module presented here, as for other parts of the questionnaire, is that of the country of destination. The questions presented cover, first, the contacts the migrant had with recruiting agents and whether they influenced his or her migration decision (questions 1-5); whether any special training was received in the country of origin in advance of migrating (question 6); and the type of contract and its provisions as understood by the respondent (questions 7-8). The latter is vital for understanding the conditions under which people migrate and the consequences of that migration for their welfare and for policy. Question 9 deals with payments made to labour recruiters and question 10 covers sources of help prior to the trip, including cash advances, tickets for transport, and exit and entry documents. Questions 12-13 cover the extent to which the contract was fulfilled, and question 5.14 asks about the respondent’s rights to and interest in bringing other family members into the country of employment. Question 15 deals with visits to the country of origin during the period of the contract. Question 16 obtains information on problems experienced in sending remittances back “home” and can serve as an introduction to the topic (covered in Chapter 7 below). Lastly, questions 17-20 ask about
Design of surveys to investigate international migration

the worker's overall impressions regarding his or her experience as a migrant worker.

Optional module on fertility
A second optional module, on fertility, which also covers family planning and child mortality, is relevant for those interested in studying the interrelations between international migration and changes in fertility and the use of family planning. This module has fairly standard questions and is not discussed here further, other than to note that it obtains information on past fertility by using questions on children ever borne and surviving for each woman, as well as on and date of most recent live birth. It also asks about fertility and the use of family planning prior to and after international migration, perceived fecundity, and family size desires. The questions included have been used with considerable success in several hundred national fertility surveys in developing and developed countries over the past two decades. An alternative approach is to administer a complete birth history, recording the dates of all live births and the country of birth of each, thus obtaining data on the timing of births (and use of family planning) relative to international migration.

Community-level questionnaire
The administration of a community-level questionnaire is desirable both in communities of destination and in communities of origin, since the data thus gathered are useful in establishing contextual variables that are key for the analysis of migration decisions and the consequences of international migration (see section E above). The community questionnaire should be administered to community officials or key informants and covers a wide range of community characteristics, including population size and density, community facilities, housing quality, environmental conditions, education, and production and employment. Special questions pertaining to the study of international migration include the prevalence of foreigners in the community; the extent of international migration into and out of the community; transportation linkages to markets, to the national capital, and to other countries; the existence of foreign publications and frequency of foreign television programs; and the extent of commercial trade with other countries and of foreign business investment.

3. Questionnaire design for a survey in the country of origin of international migrants: The comparison group

As discussed in section B above, studies of either the determinants or the consequences of international migration require that a complementary survey be carried out in the country or countries of origin of the international migrants of interest. To study the determinants of international migration, data are needed from non-migrant households in the country of origin pertaining to the
mean time of out-migration of the migrants involved. When a five-year cut-off point is used in the definition of international migrant, the relevant reference date for non-migrants is 2.5 years before the survey, assuming that the survey in the country of origin is carried out at about the same time as the survey in the country of destination. Consequently, all relevant information on non-migrants should be obtained not for the time of interview but for that earlier reference date, and information gathered through both household and individual questionnaires should refer to that date.

To study the consequences of international migration, however, the appropriate reference date for data collection in the country of origin is the time of interview, so as to make possible a comparison of the situation of international migrants in the country of destination with that of non-migrants in the country of origin at the same time (see upper entries in panels 2 and 4 in table 6.2). An even better analysis of the consequences of migration – valid even if the initial status of the two groups of migrants and non-migrants is different and not fully controlled for in the sampling design – can be carried out if data on both groups are collected for both time periods – the time of interview and the time of out-migration. Thus, in the questionnaire used in the country of destination, data on the situation of international migrants are collected for both the time of interview and for the time just prior to migration (using, for instance, section 2 of the model individual questionnaire). Availability of those data permits consideration of the change over time experienced by international migrants. In the country of origin, a parallel approach implies gathering information for the equivalent two reference dates: the time of the survey and the mean time of out-migration (2.5 years before the survey if the migration of interest is restricted to the five years preceding the survey), so as to be able to compare the changes experienced over time by non-migrants. Then, the changes over time of non-migrants can be compared with those experienced by international migrants. There is, therefore, a fundamental similarity of approach in studying the determinants and the consequences of international migration. The discussion below focuses on the appropriate questionnaire design in countries of origin that is consistent with such a comprehensive approach. Questionnaires should record information on households, working individuals within households, and their communities of residence.

In the country of origin, the survey procedure in the field begins with the household questionnaire (see Annex 2, pp. 403–409), first collecting data on the composition and condition of the household at the time of the survey. Thus, the same household questionnaire used for the country of destination should be administered. Then the same information should be recorded relative to the migration reference date, that is, for about 2.5 years prior to the date of interview. Given the short time interval involved, some data pertaining to the earlier time period could be collected quickly, such as those relative to the quality of the dwelling (questions H.11-H.16 and H.19-H.20). The two screening questions added to the questionnaire for the country of origin, H.10a-H.10b, determine when the short cut is possible. However, interviewers must be
Design of surveys to investigate international migration

carefully supervised to prevent them recording “yes” under H.10a and “no” to H.10b in order to save themselves time and effort. Furthermore, because changes are relevant and likely, in no case should questions on the composition of the household (household roster), assets and income be skipped in regard to the situation 2.5 years prior to interview. With regard to the household roster, for instance, even if all the members are the same, the educational level of children of school age will almost surely have changed, and the answers to questions H.6 and H.9 are also likely to have changed for someone in the household. Trying to get at this information by asking if there has been “any change” in the household composition or in each of the characteristics recorded is more cumbersome and certain to lead to incomplete information on changes.

The one change indicated in the household questionnaire for the country of origin is the transfer to it of two optional questions from the individual questionnaire so as to avoid collecting repetitive and secondary information more than once (questions H.22-H.23). Other similar changes could be considered, such as moving to the household questionnaire questions about relatives who are citizens of other countries or about relatives living in other countries (1.4-1.7). But such changes have their limitations because the relatives involved may vary depending on the respondent, even if the latter is always a household member. Lastly, in order to obtain comprehensive data on household income question H.21a on remittances should be added to the household schedule, even for households without international out-migrants and although it will mainly capture flows to and from internal migrants. If the situation of households in the country of origin that contain no international out-migrants is to be compared with that of households that do, which often receive remittances, the issue of remittance flows should be explored.

With respect to the individual questionnaire for the country of origin, the questions to be included are mostly a subset of those used in the individual questionnaire for the country of destination and therefore require no further explanation. The information in Section 1 of the individual questionnaire for the country of origin (part B of Annex 2) is sought only for the time of interview, assuming it has not changed much over the past 2.5 years. It would be awkward to ask the respondent to differentiate the present from the recent past for most of this information, which is unlikely to change rapidly anyway. Regarding specific questions, since it is possible for non-migrants to be citizens of another country, question 1.4 is needed, questions 1.5-1.10 and 1.17-1.18 are important because they make it possible to compare the responses from non-migrants with those of international migrants with a view to studying the determinants of international migration. Questions 1.10-1.16 relate to potential migration, and sections 2 and 4 (there is no section 3 for non-migrants) record work and income information relative to both the time of the survey and the mean time of migration.

With regard to the “mean time of migration”, the discussion above assumes that surveys are being undertaken simultaneously in the countries of origin and destination. Following the discussion of sections C and E above, the desirable procedure is to link the two surveys, first carrying out the survey in the
International migration statistics

country of destination in order to determine from which part of the country of origin the migrants are coming. Then, having data on the number of migrants to country Z from country A and census data for A, areas in country A can be formed into strata based on the proportion of international out-migrants to Z, and areas in the country of origin can be sampled with probabilities proportional to those proportions (or disproportionately), in a manner similar to that used in the sample design for the country of destination. Such a linked survey procedure requires that the survey in the country of origin be carried out some time after that in the country of destination. If carried out one year later, for instance, an interval that is probably reasonable, then another 12 months must be added to the 2.5 years determining the migration reference date, so that the appropriate reference period for both the individual and household questionnaires used in the country of origin (and for the community questionnaire as well) is 3.5 rather than 2.5 years. Given the human tendency to forget as time elapses after the occurrence of an event (Som, 1973), it is desirable for this lag to be minimized. Thus the survey in the country of origin should be carried out as soon as possible after the survey in the country of destination.

4. Questionnaire modifications for a survey of out-migrants based on proxy respondents in the country of origin

Surveys based on proxy responses cannot obtain data of the same depth and quality as surveys in which the responses come directly from the person experiencing the event of interest (international migration in this case), henceforth called the reference person. Nevertheless, it is sometimes necessary to study the situation of international out-migrants and their households on the basis of a survey conducted exclusively in the country of origin, either because such surveys are less expensive than those involving two or more countries or because it is not possible to carry out a survey in the jurisdiction of the country of destination. Consequently, it is important to discuss how the individual questionnaire for countries of destination can be modified to form the basis for a survey carried out exclusively in the country of origin and using proxy respondents.

To identify international migrants who have left household members in the country of origin, the first question should be “Has anyone left this household to live or work abroad for at least 6 months in the past 5 years?” Then the age of the person at the time of departure and the date of departure should be recorded to ensure that each reference person for whom information is being sought qualifies as an international migrant according to the survey definition being used. From this point on, various questions from the individual questionnaire for countries of destination should be adapted, changing the pronouns and verbs as necessary. Thus, question 1.1 becomes “When was X born?”, question 1.2 is “Where was X born?”, and 1.4 asks when X first left the country (asking when X first left the household is also desirable). Question 1.9 records the
Design of surveys to investigate international migration

current citizenship of X and questions 1.10 and 1.11 inquire about the citizen-
ship and place of residence of the immediate relatives of X. Note that in all cases
the reference is to relatives of X, not to the respondent, although the respondent
will almost always be a close relative of X. While the focus in the data collection
is on X, it is useful to modify question 1.12 to ask if the respondent or other
members of the respondent’s household intend to join X abroad or have
initiated any steps to do so. Question 1.13 may ask both if X understands the
language of his or her current country of residence as well as if the respondent
and other members of respondent’s household understand it. Question 1.14
becomes, “Does X intend to remain in that country?” but since it is an attitu-
dinal question regarding the migrant, the response from a proxy is likely to have
low reliability and should only be followed by question 1.14a.

Questions about the pre-migration situation of the migrant should follow,
changing the pronoun in Section 2 from “you” to “X” and all country references
from “here” to “there”, or to X’s current country of residence. The most relevant
questions are 2.2, 2.3 (adding an asterisk), 2.4 and 2.5, 2.6, 2.8 through 2.8c, 2.9,
2.12, 2.12a, 2.12c, and 2.12e. In section 4, further details about the migrant’s
previous work should be limited to questions 4.4a, 4.5a and 4.5 (changing the
verbs from the present tense to the past tense). Although problematic, questions
4.7, 4.10 and 4.11a are important, the respondent may well know the answers so
they should be posed. Questions 4.19-4.20, 4.22-23, 4.26 and 4.29 on business
activities before the migrant’s departure should be retained. Lastly, question
2.33 is important to establish if the optional module on migrant workers should
be administered to the proxy respondent and 2.34 is useful to inquire about
documents.

Questions on the post-migration (current) situation of the migrant should
be more limited because the respondent usually has no first-hand knowledge
about it, whereas it is more likely that he or she had some first-hand information
on the pre-migration situation of the migrant. Questions suggested for inclusion
are 3.3-3.4 (without the details), 3.6, 3.7 through 3.7c, 3.8, 3.9-3.9c and 3.12-3.12a.
The number and identity of accompanying family members are likely to be
known by the respondent. Thus questions 3.12-3.14 are relevant. Question 3.15
inquires about the return of persons who accompanied the migrant. Questions
3.16-3.17 (but pertaining only to the migrant) provide sufficient information on
education. Questions 3.21-3.21c refer to visits by the migrant to the respondent’s
country and the respondent is likely to know the answers. The migrant’s
acquisition of citizenship, question 3.23, is also likely to be known by the
respondent. Questions on the migrant’s current work and earnings abroad
should normally be limited to those in section 4 cited in the paragraph above.

5. Questions for return migrant workers in the country of origin

The optional module on migrant workers may be modified to form the
basis for an individual questionnaire to be applied to migrant workers once they
International migration statistics

have returned to the country of origin. The questionnaire could be used to interview the return migrant in his or her household of residence or at an entry point, though the latter may involve complex procedures of sample selection. Virtually all the questions included in the module may be posed to return migrant workers, with minimal and obvious modifications. First, the relevant migrant workers have to be identified by asking if they (or anyone in the household, if the questionnaire is being used as part of a household survey) have returned during the past five years from working for at least six months in another country. If they have, the module for migrant workers is used. Question 1 of that module becomes “Before you left this country to work abroad, were you ever contacted by a labour recruiter or contractor?” The questions that follow need to be modified to read “Were you interested in working in that country before you had that contact” and so on. At the end, additional questions should be asked on money and goods actually brought back, and changes in the household economy resulting from the whole process (see also Chapter 7 on remittances).

6. Questions of particular interest for policy analysis

Among the questions included in the model questionnaires that provide information potentially significant for policy analysis are those on citizenship and intentions to become a citizen, those on problems encountered in getting the documents to leave the country of current residence (if a non-migrant) or encountered in actually leaving the country of previous residence and entering the country of destination (if a migrant). Such information is useful to understand the effects of regulations and laws on the entry and departure of international migrants (and on the extent of evasion of those regulations). Questions on the existence of immediate family members who are citizens of other countries or living abroad, and questions on migration intentions are both useful for indicating future migration potential. Comparing the sources of information available to migrants and non-migrants, especially non-family sources, is also likely to be useful in shaping particular interventions. Documenting the activities of labour recruiters from the perspective of international migrant workers will certainly provide insights into the mobilization of those workers. Such information is useful both for countries of origin and of destination. The questions in section 3 on the use of services by migrants and their families in the country of destination collect information that is crucial for the assessment of the larger societal effects of international migration.

The effects of all the above and their consequences for the international migrants involved cannot be understood in a vacuum. It is therefore important to adopt a comprehensive approach to data collection that permits a thorough and scientifically sound analysis of the causes and consequences of international migration. To that end, gathering complementary information at the
Design of surveys to investigate international migration

Community level permits a better assessment of the likely effects of policy and thus contributes to providing a solid basis for the formulation of measures that enhance the effects of international migration and prevent its detrimental consequences.

Notes

1. The Demographic and Health Surveys, for instance, typically have a sample size of 5,000 households and represent one of the major international survey undertakings carried out in recent years, covering some 35 developing countries.

2. The small numbers problem is also evident in the European Union Labour Force Surveys, which find foreign employees to constitute only 2-6 per cent of the labour force. Independent sources suggest that these surveys miss at least 4-6 per cent of the foreign workers (Migration Research Unit, University College London, 1993).

3. The pilot survey is being carried out by the Rand Corporation through a project financed by the National Institute of Child Health and Human Development of the United States National Institutes of Health.

4. The NIDI survey of international migration to several European countries uses a 10-year cut-off for its definition of international migrants (see section A.3 above).

5. In fact, there have been no studies assessing the incidence of recall problems in relation to a change of residence from one country to another, an event that is likely to be memorable for the person involved. It may be that recall problems in that case are less serious than is commonly assumed.

6. The use of the concept "population at risk" is well-established in demography, for example, in computation of birth probabilities, parity progression ratios, and mortality rates in life tables.

7. In the Netherlands, the system is decentralized such that each municipality makes its own decision regarding the provision of information from its register to researchers. This complicates, but does not prohibit, the process of gaining access to a national sample of foreigners from the population register.

8. If data from a recent census are not available, the results of large sample surveys may be used in some countries to construct the necessary population frame. However, if international migrants constitute a very small proportion of the whole population, the data on international migrants generated by a nationally representative survey will likely have large variances (especially if disaggregated into several geographic regions) which may render them useless for the present purposes.

9. In unrestricted simple random sampling (srs) the sample variance is:

\[ s^2 = \frac{\sum(y_i - \bar{y})^2}{(n - 1)} \]

where \( y_i \) is the \( i \)th observation for \( y \), \( \bar{y} \) is the sample mean, and \( n \) the sample size. The standard error is \( s \), the square root of the sample variance. In simple random sampling the sample error is inversely proportional to the square root of \( n \) (ignoring the finite population correction factor, which is trivial in sample surveys).

10. Since most surveys use multi-stage sampling with stratification, the standard error of the key variable (which is needed to determine the desired sample size) depends on the extent to which different weights are used across strata, the number of last stage sampling units (households), their average number of respondents, and the size of the design effect. Most of these parameters will not be known a priori.

11. According to the experience of the World Fertility Survey programme, in most cases costs and variances with respect to fertility variables do not vary much between strata (WFS, 1975), but this is usually not the case for socio-economic variables and other factors relevant for the analysis of international migration (see, for instance, United Nations ESCAP, 1980c, pp. 10 and 13).

12. Including the use of stratification, the extent to which the sampling probabilities per strata are equal or not, etc. For a full discussion, see Kish, 1965, Chapters 5 and 10.
International migration statistics

13 This statement is based on roughly similar areas and populations in sending and receiving countries. The larger the receiving country relative to the sending country, the less likely in-migrants are geographically concentrated.

14 Data on households are typically linked to the status of the head of the household, so information on the status of the head may be easier to obtain. The first line in the census or survey schedule usually refers to the head of household, so only the place of birth of the person identified on the first line need be processed to identify heads who are foreign born.

15 Extension to four or more stages can be made, depending on the circumstances. A two-stage sampling design may be used in cases where a more appropriate sampling frame is available.

16 For most countries, stratification by place of residence (urban versus rural) is not useful nor recommended in surveys of international migrants because a high correlation is expected between the proportions of international migrants and the urban nature of districts anyway. Only if the correlation between place of residence and proportion of international migrants is low, and if at the same time the character of international migration to urban and rural areas is different and itself a subject of interest would it be advantageous to stratify also according to the urban or rural nature of districts. For example, if migrants to Z from A mainly went to urban areas and tended to integrate quickly into the host society while migrants to Z from C went mainly to rural areas and did not integrate quickly, it might then be of interest to have separate strata for urban and rural areas to ensure representative selections from each independently.

17 Each of the numbers is the mean of a category, with each category being the range for the stratum comprising all districts with observed proportions of international migrants within that range throughout the selected PSUs. For example, the values in the text could represent strata of 0.1 and over, 0.02 to 0.09, 0.00 to 0.02, and less than 0.005.

18 Two-phase sampling also makes it possible to collect information from part of a surveyed population using a more expensive procedure than that used for the rest. For instance, all respondents may be asked to give subjective reports about their health status, but only those in a sub-sample undergo a medical examination.

19 To investigate either the determinants or the consequences of international migration for the international migrants themselves, only households with international migrants need be interviewed in the country of destination, provided information is obtained on non-migrants in the country of origin from a separate survey in that country (see section B).

20 Predicted earnings are estimated on the basis of age, educational attainment, and years of work experience.

21 Countries with exchange controls, however, can make it difficult for people to convert assets in bank accounts or other liquid assets denominated in local currency into foreign currency to help facilitate migration. The existence of such controls should be explicitly taken into account in studying the causes of international migration.

22 An alternative approach is being used by the multi-country survey of international migration being conducted by the Netherlands Interdisciplinary Demographic Institute (see section A.3 above). The NIDI approach is to identify the “main migration actor” (MMA), and only interview that person. When a group of people, including several persons over the age of 15 years, migrate as a family, the MMA approach can save interviewing time, though the depth and quality of data must be less for certain items. Whether one approach or the other is used depends on country circumstances and survey budgets, and the frequency among migrants of families that comprise more than two persons over age 15 years.