SURVEYS OF INTERNATIONAL MIGRATION: ISSUES AND TIPS

Richard Bilsborrow
Carolina Population Center

A. INTRODUCTION: WHY USE SURVEYS

Most countries collect information on international migration using traditional data collection systems, such as population censuses, continuous population registers, and border or admission statistics. These systems provide data on the number of international migrants, and occasionally on certain limited characteristics of the migrants, such as age and sex, and also country of birth or previous residence (if immigrant) or country of destination (if emigrant). However, none of these systems can collect the type of detailed data necessary for characterizing international migration in more depth or for studying either the determinants of international migration or its consequences for migrants and their households. Only a household survey can provide such data. In a household survey it is also possible to recruit and train highly qualified interviewers.

Apart from the lack of detail, existing national statistical systems also cannot collect information on the situation of migrants prior to migration, which is vital for investigating either the determinants or consequences of migration (Bilsborrow et al., 1997). A population census, for example, enumerates the population living in households at the time of the census. Accordingly, censuses usually do not collect information about emigrants, since they are not present at the time of enumeration.

Because of the limitations of censuses and other national data collection systems for collecting data on international migrants, specialized surveys of international migration constitute an invaluable complement to those systems. Surveys can involve the use of questionnaires that are long enough to collect data to identify international migrants on the basis of place of birth, country of citizenship or previous place of residence if different from the place of residence at the time of the survey. They can also collect detailed data on the situation of the migrant and the migrant's household before and after migration, permitting the study of the determinants or consequences of international migration (op. cit.). It is also easier to include specific questions in surveys enquiring about emigrants from the household—those who have left to live in another country—as well as about former emigrants who have returned.

Since traditional national data collection systems gather data on the number of international migrants, and surveys can provide detail on their socio-economic characteristics, timing of move, reasons for migrating, etc., the two should be seen as important complements. Both are needed to properly understand the role of international migration in society.

B. KEY CHALLENGES IN DESIGNING A HOUSEHOLD SURVEY ON INTERNATIONAL MIGRATION: WHERE ARE THEY AND HOW CAN THEY BE FOUND?

Unlike a census, which covers the entire population of a country, surveys need to have some way of selecting persons or households to interview the population. This selection is called the sample. To select a sample, some form of listing or arrangement of data must be drawn upon, called the sampling frame. An ideal sampling frame covers the entire population of interest and should not include persons who are not of interest. Thus for a survey on international migration, the first challenge is the preparation of a complete list of international migrants.

For migration surveys, the most commonly used sampling frame is a population census, although continuous population registers can be excellent sources in those countries which have comprehensive registers, mainly in Europe, as can registers of foreigners in the few countries with complete data. The above are sources of data on stocks of international migrants. The usual sources of data on international migration flows are border or admission statistics which are often incomplete or population censuses when they include questions on the date of arrival.

In a census, the most common way of identifying immigrant stocks is the "place of birth" question asked for everyone living in the household. Respondents pointing out that they were born in a different country from that of enumeration then constitute the foreign-born population. Census enumerations are also the most common source of data used by the United Nations in preparing estimates of the stock of international migrants for all countries in the world.

Unfortunately, data on the foreign-born population generally do not identify when the migrants arrived, and also measure only the survivors of immigrants. Therefore, it is very useful to include an additional question on the date, that is month and year, when the foreign-born person arrived in the country. An alternative is to include a question on the country of residence of all persons five years prior to the census, which identifies all persons migrating to the country in the previous five years.

It is rare for any of the data collection approaches described above to collect data on emigration: the stock and flow data usually available only cover immigrants and not net international migration. However, with the increasing interest in migration, some countries are including questions in their censuses to ask about household members who have emigrated in a recent time period, such as in the five years prior to the census.

Because censuses are so common among the countries of the world and because of their universal coverage of the population, population censuses are the most common source of data for creating a sampling frame for a survey of international migrants. A census allows determining the proportion of migrants in the population for all areas of the country, whether based on the proportion of migrants who arrived from another country in the previous five years or just the proportion of the foreign-born. This makes it possible to form strata based on these proportions, and to over-sample areas with high proportions of international migrants. Field work in the survey can then be concentrated in such areas, making for an efficient allocation of interviewer time, while maintaining a nationally representative sample of areas.

The second challenge posed by migration surveys is finding migrants, since they are usually rare elements in the population of most countries. Thus there are few countries where the foreign-born constitute more than 10 per cent or 20 per cent of the population. Moreover these persons are lifetime migrants, who may have arrived at any time in the past up to their current age, whereas most Governments are interested in recent migrants, especially, say, those arriving or departing in the past five years. For example, if migrants constitute 10 per cent of the population, but recent migrants are three per cent, then a moderate-sized household survey of 10,000 households (of four persons on average) will yield only around 600 recent adult migrants—too small for most types of analyses of interest.

C. WHAT KIND OF SURVEY IS NEEDED?

The least expensive option is to use, when possible, data from an existing household survey. While most surveys in most countries are too small to provide data on an adequate number of recent migrants, there are exceptions, the most common being labour force surveys. They have the advantages of usually being carried out more or less regularly, even in many developing countries; implemented by national statistical offices; based on large sample sizes (50,000 to 100,000 households or more), and using questionnaires that already include useful information for studying migration, including demographic characteristics, marital status, employment and wages, etc. Thus adding questions to the labour force surveys to identify immigrants or emigrants can be done parsimoniously.

While labour force surveys can often provide some basic information about migrants, the migration-specific questions that can be added are limited, restricting what can be learned about the determinants or consequences of international migration. Moreover, in many countries there is no recent, large-scale labour force survey.

Alternatively, a specialized household international migration survey can be designed to efficiently collect data on a sample of migrants and appropriate non-migrants, using specialized sampling methods such as stratification, oversampling of areas with high proportions of migrants of interest, and two-phase sampling of households (Bilsborrow et al., 1997).

D. WHO SHOULD BE INCLUDED IN THE SURVEY? THE ISSUE OF APPROPRIATE COMPARISON GROUPS TO STUDY THE DETERMINANTS OF MIGRATION

Migration can be viewed as a social experiment, in which people, by moving in a particular fashion are subjected to a “treatment”. In order to study the effects of this “treatment” on them and their households, it is necessary to compare immigrants with an appropriate group of persons who are not subjected to the “treatment”.

The appropriate comparison group for studying the determinants of migration is other persons and households that were “at risk” of migration but who did not (e)migrate. This group must have lived in the same area as the emigrants during the same time period whose emigration is being studied. This group is called the non-migrant comparison group population. Together with the emigrants, this group constitutes the population at risk of emigration from a country or area of a country, referred to as origin country, area of origin, or simply origin. For individual emigrants, the comparison population is similar individuals in place of origin; and for households that emigrate, the reference or comparison group is constituted by households that did not emigrate (during the time interval of interest, for example, during the five or ten years prior to the survey).

The ideal data set for studying the determinants of emigration by individuals or households then contains emigrants, or emigrant households, interviewed directly at the place of destination and equivalent individuals in the country of origin who did not emigrate in the time period. This implies conducting a survey in each country–to interview those who emigrated in their country of destination, where they are viewed as immigrants, in a destination country survey, and those who did not emigrate in a survey at the place of origin. Pooling the two groups provides a population for which the statistical factors determining why some persons emigrated and others did not can be readily estimated with multivariate analysis. The same approach is used to study the factors affecting the emigration of households, that is to say pooling data on emigrant households surveyed in the destination country with data on non-emigrant households remaining in the origin in a survey in the country of origin.

The above approach is ideal since it obtains data directly from the persons making the decision, to migrate or not to migrate. But it involves a higher data collection cost as a survey since it must be carried out in two countries. An alternative is to conduct a survey only in the country of origin, in which households with emigrants and households without emigrants are interviewed. In the former, a "proxy respondent" is asked to provide data about the emigrant from the household. This person is the household head, spouse of head, or some other adult knowledgeable about the emigrant. Even so, the information that can be asked about the emigrant is evidently less reliable than that the emigrant could provide directly if interviewed in a destination country survey.

The second problem with collecting data to study emigration from a country only in the place of origin is that when entire households emigrate, there is usually no one left to provide reliable information about them. Thus the more people emigrate from a country as entire households, the more a survey in the place of origin will fail to cover emigrants from that country, and the less useful the survey conducted only at the place of origin becomes.

In order to study the determinants of emigration, data are needed on the situation of the emigrant and his or her household at the time of migration and just before, since this defines the context—the individual's characteristics, the household's characteristics and indeed those of the community or region from which the person emigrates. In an origin country survey, these data are obtained from the proxy respondent. But what should the time reference be for collecting data on non-migrants? The answer is, at the mean time of emigration of emigrants, since this is, on average, the time at which they made the decision not to emigrate. For example, in a survey studying the determinants of emigration in the previous five or eight years, the mean time would be 2.5 or four years prior to the date of the survey.

Note that data referring to the situation of emigrants or non-emigrants at the time of the survey are not relevant for studying the determinants of migration, and instead reflect the situation at the time of interview, which may well have changed for emigrants and non-emigrants and their households as well.

E. WHAT ARE THE APPROPRIATE COMPARISON GROUPS FOR STUDYING THE CONSEQUENCES OF EMIGRATION?

For studying the consequences of emigration, the usual approach in migration research is to conduct a single survey in a destination country, to collect data on immigrants and on non-immigrants ("natives"), then compare them, and conclude that the consequences are favourable or not for the immigrants depending on how their situations are similar to those of natives or inferior to those of natives (Bilsborrow et al., 1997). In fact, such data, albeit convenient to collect since they involve only a survey in the destination country, provide no evidence at all on the consequences of migration for the migrants or their households. Instead, they indicate the degree to which the immigrants are successfully integrating or assimilating in the host country.

The problem with the above approach is that it does not collect data for the appropriate comparison group for the immigrants, which is the other persons "at risk" for migration, specifically those who did not receive the "treatment", namely non-migrants remaining at the place of origin. To study the consequences, it is necessary to conduct a second survey in the country of origin to collect data on the status of non-migrants (a) at the mean time of emigration of the immigrants for whom data are collected in the destination country survey, and (b) at the time of the survey. The difference in their situation at time (a) and time (b) shows the extent to which they improved their situation (or not) by remaining in the origin country.

For the migrants, viewed as immigrants in the destination country or emigrants in the origin country, the change in their situation following migration is, similarly, the difference between their situation (i) at the time of emigration, in the origin country, and (ii) their situation at the time of the survey, when they are interviewed in the destination country. Then the proper assessment of the extent to which migrants benefited from international migration is to compare the difference between (ii) and (i) and that between (b) and (a). If the former, say, positive change is greater than the latter, it can be said that the migrants benefited from migration.

This requires two surveys in two countries, one on immigrants in the destination country, and the other on non-migrants in the origin country. It should be stressed that the collection of data on non-migrants in the destination country is not useful for a proper evaluation of the consequences of migration. However, it is useful for studying integration, as indicated above.

The advantage of the two-country approach above for analyzing the consequences of migration is that data are collected directly from the relevant persons themselves and may hence be presumed to be of adequate quality. Thus, for example, the immigrant provides not only his or her current information but also retrospective information on his or her other situation in the country of origin prior to departing. The discussion here applies equally to the situation of immigrant households observed in destination countries and non-migrant households observed at the place of origin.

A less costly but less satisfactory data collection approach is to carry out a survey only in the origin country, in which a proxy respondent in households with an emigrant is asked about (i) and (ii), that is, about the situation of the emigrant in the origin country just prior to migration and the situation at the time of the survey in the country of destination. This is less satisfactory since the proxy respondent may not have accurate information on the current situation of the emigrant in the destination country. This is especially true the more years that have elapsed since the emigrant has been in the origin household–another reason for focusing on recent migration.

The above discussion is limited to only two countries, that is, specifying what countries should be used for studying the determinants of a particular migration flow from a single country of origin to a single country of destination. Evidently, such an approach is more useful if such a migration flow is a major outflow from the origin country and a major inflow in the country of destination. However, in many if not most situations, those emigrating from a country have several major countries of destination. Similarly, most immigrants to a country of destination usually come from several countries of origin. Therefore, the ideal approach would be to conduct linked surveys of migrants and non-migrants in both countries of origin and destination, in particular, those linked in a migration system (Kritz and Zlotnik, 1992).

F. EXAMPLES

There are very few examples of such origin-destination linked country surveys to investigate either the determinants or consequences of international migration, which collect data directly from the migrants and non-migrants themselves. Some studies exist that involve surveys in one country of origin and one of destination. The number of linked surveys is likely to increase in the near future, given the attention for international migration movements and their potential role in socio-economic development of low-income countries.

One interesting example of a multiple-country survey project linking countries of origin and destination is the push-pull project of the Netherlands Interdisciplinary Demographic Institute (NIDI) and Eurostat in 1997-1998, which examined the processes and determinants of migration flows from selected countries of Africa and the Middle East to the European Union (Schoorl et al., 2000). Surveys were carried out in Egypt, Ghana, Morocco, Senegal and Turkey, as countries of origin, and in Italy and Spain as countries of destination. Common methodologies were used, in terms of sample design and questionnaires, with sample sizes being generally between 1,000 to 2,000 households. Migrants were defined as persons leaving or arriving in the 10 years prior to the survey.

In origin countries, any household reporting an emigrant in the past ten years was considered a migrant household, regardless of destination. But only two immigrant flows were covered in the surveys in the two destination countries, those from Morocco and Senegal in Spain, and those from Egypt and Ghana in Italy. Thus to study the determinants or consequences of migration from Morocco to Spain, data are available from the survey in Spain on immigrants from Morocco, which may be compared with (pooled together, for statistical analysis of determinants of consequences) non-immigrants interviewed in Morocco, the country of origin. The same is true for Senegalese immigrants in Spain, who may be compared with non-migrants in Senegal. And a similar situation exists for Italy, where migrants from Egypt and Ghana may be combined with non-migrants interviewed in those countries, respectively, for analysis.

G. CONCLUSION

The field of migration is still in its infancy, partly due to the lack of adequate data sets for investigating either the determinants or consequences of migration, especially international migration. Data collection in more than one country is strongly recommended but more expensive and complex to coordinate, so almost all research to date is based on less than satisfactory data sets from only one country. In addition, existing studies rarely focus on migrants who move as both individuals and households, and almost never collect data at the correct times, which involves additional challenges in collecting retrospective data.

Collecting adequate quality retrospective data in turn calls for studies to focus on *recent* migration, which can also further complicate the “rare elements” problem inherent in designing samples for migration surveys. However, these problems can be addressed (Bilsborrow et al., 1997), and have been occasionally addressed (Schoorl et al., 2000; Groenewold and Bilsborrow, In press). More experience is needed to improve the methodology and collect appropriate data sets for the analysis of international migration, which is attracting ever more interest.

REFERENCES

Bilsborrow, Richard E., Graeme Hugo, Amarjit S. Oberai, and Hania Zlotnik (1997). *International Migration Statistics: Guidelines for Improving Data Collection Systems*. Geneva: International Labour Office.

Groenewold, George and Richard E. Bilsborrow. In press. Design of Samples for International Migration Surveys: Methodological Considerations and Lessons Learned from a Multi-Country Study in Africa and Europe. In *International Migration in Europe: Current Trends and Issues*, C. Bonifazi and M. Okolski, eds. Rome: Universite di Roma.

Kritz, Mary M. and Hania Zlotnik (1992). Global Interactions: Migration Systems, Processes and Policies. In *International Migration Systems: A Global Approach*, Mary M. Kritz, L. L. Lim, and H. Zlotnik, eds. Oxford and New York: Clarendon Press and Oxford University Press, pp. 1-18.

Schoorl, Jeannette, Liesbeth Heering, Ingrid Esveldt, George Groenewold, Rob van der Erf, Alinda Bosch, Helga de Valk, and Bart de Bruijn (2000). *Push and pull factors of international migration: A comparative report*. Luxembourg: Eurostat and NIDI.