

BIBLIOGRAPHIC DATA SHEET1. CONTROL NUMBER
PN-AAJ-4542. DOCUMENT CLASSIFICATION (800)
PA00-0000-G570**3. TITLE AND SUBTITLE (240)**

National migration surveys; manual VI: sample design manual

4. PERSONAL AUTHORS (100)**5. CORPORATE AUTHORS (101)**

U. N. Economic and Social Commission for Asia and the Pacific

6. DOCUMENT DATE (110)

1980

7. NUMBER OF PAGES (120)

26p.

8. ARC NUMBER (170)

FEA325.5.U58

9. REFERENCE ORGANIZATION (150)

Census

10. SUPPLEMENTARY NOTES (500)

(Manuals 1-7: PN-AAJ-449 - PN-AAJ-455)

11. ABSTRACT (950)**12. DESCRIPTORS (920)**Pacific Islands
Questionnaires
Planning
SurveysInterviewing
Migrations
Manuals
Asia**13. PROJECT NUMBER (150)****14. CONTRACT NO. (140)**

RS/BUCEN-1-78

15. CONTRACT TYPE (140)**16. TYPE OF DOCUMENT (160)**

FEA
325.5
U58
no. 6

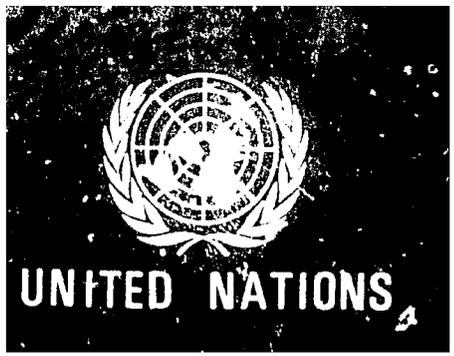
ECONOMIC AND SOCIAL COMMISSION FOR ASIA AND THE PACIFIC



SURVEY MANUALS

**NATIONAL
MIGRATION
SURVEYS**

VI. SAMPLE DESIGN MANUAL



ECONOMIC AND SOCIAL COMMISSION FOR ASIA AND THE PACIFIC
Bangkok, Thailand

**Comparative study on migration,
urbanization and development
in the ESCAP region**

Survey manuals

NATIONAL MIGRATION SURVEYS

VI. SAMPLE DESIGN MANUAL



UNITED NATIONS
New York, 1980

CONTENTS

	<i>Page</i>
Preface	i
I. SURVEY OBJECTIVES	1
II. OBJECTIVES FOR STATISTICAL DESIGN	1
III. SURVEY DESIGN FEATURES: GENERAL METHODOLOGICAL CONSIDERATIONS	2
A. Study population	2
B. Respondent selection rules	3
C. Non-sampling error and validation study	6
IV. SAMPLE DESIGN	8
A. Stratification	8
B. Sample size	8
C. Sample allocation	9
D. Carrying out the sample plan	14
E. Country modifications of sample design	16
Annexes	
I. Discussion of the advantages and disadvantages of using lifetime <i>versus</i> recent migrants for sample design	18
II. Theoretical basis for the optimum sample allocation	19
III. Sources of data for use in sample allocation	21
List of tables	
1. Increase in variance	5
2. Comparison of actual and effective sample sizes (rounded) for varying values of \bar{n} and δ	5
3. Population distribution and migration statistics for participating countries, 1970 round of censuses	11
4. Alternative sample allocation schemes	12
5. Ratio of sampling rates under optimum allocation A	13
6. Possible selection stages	15

PREFACE

The materials presented here have been drafted for field surveys on migration in countries of the ESCAP region. The migration surveys were set up in response to the recommendations made in several population conferences held over the last ten years.¹

The migration surveys are components of a regional project, the "Comparative study on migration, urbanization and development in the ESCAP region". The first phase of this project consists of analyses of census materials on migration and urbanization in the participating countries. Although special tabulations from the 1970 round of censuses have been generated to focus specifically on migration and urbanization, these have limitations in providing information on the relationships between migration, urbanization and development. The main aim of the national migration surveys is to provide the kinds of detailed information that cannot be collected in national population censuses.

The data to be collected in the migration survey will focus on the following four major issues:

- (a) Patterns and types of population mobility;
- (b) Demographic and socio-economic characteristics of migrants and non-migrants;
- (c) Motivations of people to move or not to move;
- (d) Demographic and socio-economic consequences of migration in areas of both origin and destination.

Preliminary versions of the questionnaire were circulated in 1978 to a large number of experts throughout the ESCAP region and in other parts of the world. Their comments were taken into consideration in the production of a questionnaire which was tested in Indonesia, Pakistan and Thailand in 1979. The results of these tests were discussed at the first session of the Advisory Committee on Migration and Urbanization, held at Bangkok from 29 October to 1 November 1979. The discussions and recommendations of this Committee are published in *Asian Population Studies Series*, No. 47, issued by ESCAP in 1980. The Advisory Committee recommended that a small working group should be set up to revise the questionnaire, taking into consideration the results of the pretests and the discussions of the Committee. A revised version of the questionnaire was further circulated among the Committee members for their comments and then tested in a pilot survey conducted in Thailand in May 1980. The results of this survey were used to modify the present document.

The materials presented here are regarded as a "core" and it is hoped that participating countries will follow them as closely as possible, allowing for minor modifications which may be required to meet local circumstances. ESCAP will welcome the opportunity to discuss any modifications which a participating country feels will be necessary. While priority is given to individual country requirements it is recommended that the information collected should be as uniform as possible to allow standardization of materials and procedures. The results should allow a comparison between countries which will maximize the understanding of a common problem affecting the majority of countries in the ESCAP region.

¹ Recommendations were made by the Second Asian Population Conference, held at Tokyo in 1972; the World Population Conference, held at Bucharest in 1974; Habitat: United Nations Conference on Human Settlements, held at Vancouver in 1976; the ESCAP Committee on Population at its first session, held at Bangkok in 1976; the Expert Group Meeting on Migration and Human Settlements, held at Bangkok in 1977; and ESCAP at its thirty-fourth session, held at Bangkok in 1978 and its thirty-fifth session, held at Manila in 1979.

This manual presents a set of guidelines for designing and selecting an appropriate sample for the national migration surveys. It is intended as a framework for discussion and decision making concerning sample design and related aspects of the survey and not as a definitive statement of survey design elements to be adhered to unswervingly. Alternative design modes and assumptions are explored along with their implications, within the context of the stated goals and objectives of the survey. Adaptations of the design modes presented here will need to be made for each country on an individual basis taking into account specific objectives, characteristics and resources. In addition to sample design considerations, some attention is also given in the manual to general survey methodology issues and measurement of non-sampling error for the survey.

I. SURVEY OBJECTIVES

The aim of the national migration surveys is to fill gaps in existing knowledge concerning the determinants and consequences of various types of population movement and their interrelationships with national and regional development patterns. As such, the information to be collected is intended to supplement information available from other sources (population censuses and surveys focusing on other topics), as well as to provide detailed data on aspects of the migration process which are beyond the scope of the other sources of information.

The national migration surveys will focus on (a) the patterns, streams and types of movement within each country, (b) the causes and consequences (broadly defined) of such movement and (c) the characteristics, including attitudes and perceptions, of the survey population as they pertain to migration. There is more interest in studying implied analytic relationships than in making parameter estimates such as the volume of migration, although the latter is certainly a secondary objective. For analytic purposes, the survey is to provide information about non-migrants as well as migrants. The design of the survey will be such that it is best suited to migrant/non-migrant comparisons in the place of destination of migrants (interviews will be taken at current place of residence), although comparisons with non-migrants at the point of origin may also be made by coding and tabulating place(s) of origin for respondents who report moves. Further, the migrant categories of major interest are short-term (less than 12 months at current residence) and long-term, although subcategories within these may be studied using information gathered from migration histories so that, for example, return and circular migrants can be identified.

II. OBJECTIVES FOR STATISTICAL DESIGN

A major survey objective, from a measurement standpoint, will be to strive for uniformity of survey and sample design so that the goal of country-by-country comparability can be achieved to the fullest extent possible. This means that concepts and definitions, questionnaires, interviewer manuals and materials, sampling plans, data tabulation and analysis plans will be most effectively implemented if they are standardized, in so far as national requirements and characteristics permit. It is recognized, however, that strict, international comparability among participating nations will not be possible. One reason for this is that each country has its own criteria for defining urban and rural places; hence there will be the problem of variable sizes and shapes of migration-defining areas from country to country. In some countries, the urban areas may be subdivided into size categories to improve this comparability, although the basic problem remains.

The imposed constraints and requirements for the design are as follows:

(a) The survey is to be designed as an independent survey or as the central component of a survey which might otherwise have non-competing objectives;

(b) Only a single round is currently anticipated by ESCAP, but Governments might consider periodic rounds in the future, particularly in connection with the National Household Survey Capability Programme being executed by the United Nations Statistical Office;

(c) Availability of census and other survey information about migration, local definitions and current sampling techniques applied in the different countries must be taken into account.

Moreover, the following design features are needed:

(a) A probability sample;

(b) National estimates, plus estimates by urban (by size groupings) and rural areas, and possibly (depending upon the ultimate sample size) major regions with an urban-rural split within;

(c) Simplicity (the questionnaire is fairly complex and time-consuming to administer). To the extent possible, further complexities of design should be confined to office operations and procedures in order to make the field-work and the interviewers' jobs as easy as possible.

III. SURVEY DESIGN FEATURES: GENERAL METHODOLOGICAL CONSIDERATIONS

There are a number of survey design aspects (distinguished from sample design features) which require deliberation prior to the specification of a sampling plan. This section of the manual addresses several such issues and their implications for over-all survey design as a prelude to the discussion of sample design strategies.

A. STUDY POPULATION

Several issues arise in the course of defining a target population for the survey. One issue concerns whether the target population should be restricted to selected types of respondents within households (i.e., the principal wage earner or the household head). This is not recommended for both substantive and methodological reasons. From a substantive point of view, it is of greater interest to collect migration information for the general population than specifically for principal wage earners or household heads. In addition, the selection of persons filling functional roles within households generates a variety of sampling-related difficulties which, given the substantive interests of the survey, do not seem justifiable. It is recommended instead that the target population should be defined as the adult population nation-wide. Although the decision is somewhat arbitrary, age 15 will be used as the minimum age to define adulthood for the survey.

An upper age limit of 64 years will also be used to determine the eligibility of respondents for several reasons. First, since a major objective of the survey is to study linkages between economic conditions and mobility behaviour, it would be beneficial both conceptually and in terms of the sample size for analysis to exclude from the sampled population elderly persons, who in the aggregate, are only marginally attached to the labour force. Secondly, since recall ability on the part of respondents is known to decay with the passage of time since the events occurred, the exclusion of the elderly from the study population may improve the over-all validity of the survey data. Thirdly, since most population movements occur between the ages of 15 and 45,¹ it is likely that a large proportion of migration events experienced by the elderly segments of the population would have occurred in the distant past under quite different social and economic circumstances, thus detracting from the usefulness of the survey data for policy deliberations. However, the limitations introduced by this upper age limit must also be considered when the analytical objectives are defined in each country. Since the population of 65 years or older may be an important component of the return migrant subgroup, the analysis for this subgroup will be more restricted by excluding this segment of the population. The data on certain historical migration patterns and lifetime migration cycles will also be more limited.

Although the entire population of each of the ESCAP countries satisfying the age requirements specified above are theoretically targeted for the survey, there are a number of population subgroups which might reasonably be excluded from the survey. One such group is nomadic tribes. The United Nations has suggested that the movement of nomads represents a special type of spatial mobility which is beyond the scope of internal migration surveys.² In addition, the inclusion of nomadic tribes

¹ See, for example, R. Paul Shaw, *Migration Theory and Fact* (Philadelphia, Regional Science Research Institute, 1975).

² See *Methods of Measuring Internal Migration*, (United Nations publication, Sales No: E.70.XIII.3), p. 1.

in the study population would substantially complicate the sampling and survey procedures. Given their marginal importance in the survey, it is recommended that such groups be excluded from the domain of study.

A second group to be considered is seasonal or migrant workers. The United Nations has again suggested that these are conceptually beyond the scope of this type of survey.³ However, in so far as they represent an important analytical group for this study with respect to remittances and other variables, it is recommended they be retained in the universe of study and hence be included in survey coverage.

A third group consists of persons in long-term care institutions such as prisons and hospitals. Such persons are considered beyond the scope of the migration survey and will therefore be excluded from coverage. However, people in group living quarters for "non-institutionalized" persons will be included.

Finally, some attention needs to be directed to the issue of whether international migrants (persons who were either born or had resided for a period of time in another country) should be included in the survey universe. While it is customary in studies of internal migration to exclude such persons (or at least those persons born outside the country under study), it is recommended that this group be included in the target population for the survey. The migration history technique permits their internal movements to be studied independently of moves involving the crossing of national boundaries. In addition, there may be interest on the part of some countries in studying the internal migration behaviour of the immigrant population separately from that of the native population. It should be noted, however, that the types of analyses that can be performed using information from immigrants are somewhat restricted because samples of non-migrants in their respective countries of origin are not to be included in the survey design. Nevertheless, it would seem advantageous from a policy perspective to obtain information on the internal mobility behaviour of the immigrant populations of the participating countries.

In short, it is therefore recommended that the target population for the survey should be defined as the non-nomadic, non-institutionalized, resident population between the ages of 15 and 64.

B. RESPONDENT SELECTION RULES

In addition to defining a target population for the survey, criteria must also be established for determining the eligibility of individuals within sample households. Two options are available. One possibility is to use a *de facto* counting rule for defining eligible household members, that is, all persons residing in a selected household at the time of the survey (including those whose "usual" residence may be elsewhere) would be eligible as respondents in the selected households. This option has the important advantage of increasing the likelihood of coverage of types of migrants who are, in general, difficult to capture in censuses and surveys (i.e., short-term and seasonal migrants). Under the alternative approach, a *de jure* counting rule, eligible respondents would be limited to those members of a selected household whose "usual" residence was at that address. While this could be advantageous in some respects, there are several sampling and interviewing problems associated with this option. First, an interviewing problem arises in the case of persons who are away from their usual residence for a long period of time (i.e., working elsewhere temporarily). If self-response is to be used (as is planned), it would be necessary to locate such persons should they be selected as the respondent in their "usual" residence, a very difficult and costly procedure. In lieu of self-response, the only alternative would be to collect information from another household member (proxy response). This is not recommended, however, for two reasons. First, many of the questionnaire items deal with

³ *Ibid.*

"soft" information, such as reasons for moving, and other attitudes and perceptions. Secondly, proxy response has a greater potential for bias when eliciting retrospective information such as historical migration. Given the survey objective of studying the many types of mobility along with their social and economic contexts and the operational difficulties associated with the *de jure* approach, a *de facto* counting rule would seem preferable for this survey.

There is some concern, however, that a strict *de facto* approach would present conceptual problems in collecting and analysing the data in the case of persons who are only visiting the village/town/city (i.e., non-residents who are not working, looking for work or studying). Since the questions on migration use "present place" as a point of reference, the data for such persons would be misleading unless the questions were asked from the perspective of the usual residence. It was not felt that the enumerators could modify the interviewing procedures for this group without adversely affecting the quality of the data. Therefore it was concluded that non-residents of a household who were in the present village/town/city only for a visit should be excluded from the sample. This procedure results in a bias by excluding persons who happen to be away for visits. In order to minimize this bias, a large part of the "visiting" population will be given a chance of selection for the individual questionnaire at their usual residence. The "usual residents" of the household who are temporarily away (within the interviewing period) on a visit will be identified in the household questionnaire and will become eligible respondents for the individual questionnaire (if they meet the age criteria for the survey). Such persons who are only away for a few days may be reached for an interview when they return to their usual residence. Although the non-response rate for this group of visitors may be higher than that for the remaining population, this procedure should reduce the over-all bias of the survey estimates. After the survey is carried out, the data from the household questionnaire can be used to estimate the volume and non-response rate for visitors and test whether this group is significantly different from the remaining adult population for certain characteristics.

In summary, the eligibility for individual respondents will be based on a *de facto* approach for all the target population except for visitors, whose eligibility will be determined on a *de jure* basis. The following list identifies each type of eligible respondent among the persons listed in the household questionnaire:

(a) All "usual residents" (section 1) aged 15 to 64 who slept in the household the night before the interview (question 009, code 1); also, those who were away the night before the interview (question 009, code 2), if the purpose of being away was for a visit or "other" reason (question 010, codes 1 or 5) and they plan to return within the interviewing period;

(b) All "non-usual residents" (section 2) aged 15 to 64 whose purpose for being in the place of enumeration was to work, to look for work, or to study (question 026, codes 2, 3 or 4).

In addition to the issue of which household members should be eligible for inclusion in the survey, there is also the question how many respondents should be selected in each household. This issue is, perhaps, one of the most crucial features of the survey design. While it would be cost-efficient to interview more than one respondent per household (especially in rural areas where travel costs are higher), the issue of intraclass correlation within households must also be considered. It is likely that this intraclass correlation is highly positive for a statistic such as migration status. (The correlation may be either positive, because families tend to migrate together, or negative, in the case where the household head migrates alone for seasonal employment.) If this is the case, it may not be efficient to sample more than one respondent per household.

As an illustration, it is assumed that the only intraclass correlation to contend with is that within households. Table 1 shows the increase in variance over simple random sampling of persons when more than one person is chosen per household, for various possible values of δ (the within-household correlation). When a single respondent is chosen at random from each household the sam-

Table 1. Increase in variance

\bar{n} = number of interviews per household	δ = within-household correlation on migration status						
	.05	.1	.2	.25	.3	.4	.5
1	1.0	1.0	1.0	1.0	1.0	1.0	1.0
2	1.05	1.1	1.2	1.25	1.3	1.4	1.5
3	1.1	1.2	1.4	1.5	1.6	1.8	2.0
4	1.15	1.3	1.6	1.75	1.9	2.2	2.5
5	1.2	1.4	1.8	2.0	2.2	2.6	3.0

Computed as $1 + \delta(\bar{n}-1)$

pling variance is unaffected. For two respondents, however, the variance goes up by 5 to 50 per cent, depending on how large the intraclass correlation is.

Perhaps a more graphic depiction of the impact of different numbers of household respondents can be seen by considering the effective sample sizes. Table 2 illustrates clearly that, unless δ is very small, it would be wasteful to interview more than two persons per household. It shows, for example, that if four persons were interviewed in each household (if, in fact, eligible adults average four per unit) and the intraclass correlation were as high as .3, the effective sample size in terms of reliability of the estimates is only about 6,300 cases, even though 12,000 interviews were carried out. The critical decision point, therefore, is between one or two persons per household. There are not any firm measures of δ (it is not known whether it is closer to .05 or .5).

Table 2. Comparison of actual and effective sample sizes (rounded) for varying values of \bar{n} and δ

\bar{n} = number of interviewed persons per household	Actual sample size (number of interviews)	Effective sample size for					
		$\delta = .05$.1	.2	.3	.4	.5
1	3 000	3 000	3 000	3 000	3 000	3 000	3 000
2	6 000	5 700	5 500	5 000	4 600	4 300	4 000
3	9 000	8 200	7 500	6 400	5 600	5 000	4 500
4	12 000	10 400	9 200	7 500	6 300	5 500	4 800
5	15 000	12 500	10 700	8 300	6 800	5 800	5 000

possibility that non-sampling error could also result from conducting more than one given household argues in favour of selecting only one respondent per household, given nature of the questionnaire. With more than one respondent per household, the in-unit

enumeration time is substantially increased. There is also the very real possibility of a conditioning effect which would bias the migration statistics in the direction of undercoverage. This could come about in the following way. When each respondent is present for the entire interview, the second one learns quickly the pattern that the questioning and responses take. He or she sees, for example, that every mention of a move leads to a string of questions to elicit details. If the second respondent were eager to terminate this lengthy interview, short of doing it directly, he could do so by purposely failing to report any or all moves. One suggestion would be to have the interviewers strive to conduct the interview privately with each household member, but this will not be possible in most situations.

Given these non-sampling error considerations, the safe and recommended course would be to select one person per household for the individual questionnaire. Household information, however, should be gathered from a knowledgeable household respondent. Hence, the interview could require gathering information from two respondents in a given unit.

C. NON-SAMPLING ERROR AND VALIDATION STUDY

There are a number of survey design aspects (distinguished from sample design) that bring into consideration the very real problem of non-sampling error. Such errors may be potentially larger and more damaging to the survey data than the sampling errors. One source of non-sampling error is respondent conditioning of the type discussed in the preceding section. Others are discussed below, with their implication for over-all design.

1. *Retrospective interview for one-time survey*

The survey will rely heavily upon the migration history technique in which respondents are asked to recall in considerable detail each move they have made since the age of 15, including dates, places of origin and destination, employment and family history. Given the large volume of information to be solicited, the possibility of recall errors looms very large, especially for older respondents who will be reporting upon events which occurred many years in the past. Memory failure alone may seriously distort the findings. In order to provide a firm temporal anchor to which the respondent can relate a particular move, questions regarding marital status, duration of stay, activity and number of children have been added to the migration history to serve as a mnemonic device.

Because of the problems of recall bias, it should be remembered that estimates of short-term migration can be improved in future years if periodic household surveys are taken. With short-term migration, the problem of "telescoping" could be serious in a one-time survey such as this one. Telescoping is the respondent's tendency to misreport recent moves outside the reference period in which they actually occurred. One possibility to reduce this effect would be to collect data for, say, the past three years but analyse only the past two years.

2. *Varying survey dates*

The surveys are to be carried out in the countries over a three- or four-year period, starting in 1981. Estimates of short-term migration, therefore, will not be comparably referenced on the time dimension. "Within the last 12 months" may refer to July 1979 to June 1980 for one country to January 1983 to December 1983 for another. This distinction must be kept in mind at the analysis stage when the results across countries are being interpreted.

3. *Reporting accuracy on place of origin*

Currently the plans are to ask the respondent the name of the village, town or city he migrated and code this information in the office, using available census geographic information. A concern here is making sure that the interviewer records the precise name of the area.

correct spelling. The other concern is the validity of the respondent's answer, beyond the problem of memory discussed above. Respondents may mention a larger place that is near the actual place, because of prestige value. If a respondent names a large identifiable city (e.g., Jakarta, Bangkok, etc.), the interviewer is instructed to probe with "Was it actually (place), or just close by?" (See *Manual for Interviewers*, instructions for question 005.) This procedure should get at the exact place rather than a nearby primary city or provincial capital. An additional complexity arises because some countries have several villages with identical names.

4. Length of interview

The questionnaire will take about an hour or so per respondent, especially for those who report moves. This places a great burden on respondent and interviewer alike. However, the recent pre-test in Thailand showed that, with adequate training, the interviewers can handle a long interview with relatively little trouble.

Given the importance of the survey objectives, it is imperative that steps be taken to evaluate the validity of the survey data. It is recommended that a reinterview programme (post-enumeration survey) should be carried out to measure the degree of consistency between responses in the original and post-enumeration surveys. This would seem to be the most viable means of assessing validity in the present survey since independent information on many (if not all) of the survey topics is not available from which to evaluate the survey data. The reinterview survey programme would be limited to about 1,000 of the original respondents, representing all geographic areas (an urban-rural break would be desirable). A shortened version of the original questionnaire would be used to obtain second measures on important items such as the sequence and timing of migration, places of origin and other details. This method will not provide a measure of the "truth" of the original responses, but rather will reveal the degree of consistency between responses given to the same question at two points in time and, for that reason, will give an indication of the volatility of responses due to recall error and other sources of reporting inaccuracy.

Although the design of the proposed reinterview programme is at the very early stages of development, a number of general guidelines may be specified here:

(a) It is imperative that the post-enumeration survey take place shortly after the main survey (within three months) in order to minimize the effects of sample attrition (respondents moving from the place of enumeration in the main survey).

(b) Since the reinterview programme will seek to evaluate the consistency of responses to identical questions at two points in time, it will be necessary to reorient the post-enumeration survey questionnaire items to refer to the same time reference period that was used in the main survey questionnaire. Thus, respondents will be asked to recall situations at some point in the past, making it all the more important to conduct the validation interviews as close in time to the original interviews as is feasible.

(c) The reinterview questionnaire should consist of both "hard" items (requiring the recall of events) and "soft" items (soliciting perceptions and reasons underlying events) so that response consistency may be assessed for different types of questions separately.

(d) The respondents selected for the reinterview programme should constitute a probability sample of the original respondents so that response consistency may be evaluated for the entire survey universe.

IV. SAMPLE DESIGN

In this section, sample design strategies for a national migration survey are considered. Census data from each of the participating countries are used in the presentation in order to provide concrete models vis-à-vis specific design modes and assumptions and for the purpose of illustrating the implications of the various options for survey implementation under varying conditions. It is to be emphasized that the sample design strategies discussed here do not exhaust the full range of possibilities, but rather represent what seem to be the most plausible strategies for the present survey.

A. STRATIFICATION

In order to maximize sampling efficiency, it is necessary to take account of the general characteristics of the migration phenomena as well as the specific goals and objectives of the survey. One such important characteristic is that migrants typically constitute a minor segment of national populations. This has two implications for the design of the survey: first, relatively large samples will be needed (the issue of sample size is discussed in detail in the next section), and secondly, it will be necessary to use all available information on the geographic distribution of the target population and migrants in each country to allocate the sample among the strata.

In defining strata for the survey, it is important to make a distinction at the outset between strata for analytic purposes and strata for sampling purposes. Analytic strata, or domains, represent the categories for which survey estimates are to be obtained, while sampling strata refer to the categories used in the sample selection procedures to increase the efficiency of the design for obtaining the desired estimates. These need not be the same. Sampling strata may (and probably will) vary from country to country depending upon the criteria delineating geographic/administrative units in each country and other considerations to be discussed later in this manual in connection with carrying out the sample plan.

In order to ensure comparability, it would be useful to define a minimum set of analytic strata for which survey estimates are to be produced for each participating country. From a substantive point of view, there are three strata (representing geographic categories) which are of primary importance in the present survey: the major metropolitan areas (defined as major city proper and surrounding areas of urban character), all other urban areas and rural areas. It is recommended that, at a minimum, the sample design adopted in each country should provide reliable estimates for these three categories. Some participating countries may also subdivide the other urban stratum into size categories, which should improve the cross-national comparisons. Additional analytic strata or categories may, however, be built into the sample design strategies, depending upon national requirements. The extent to which this is feasible is contingent upon factors such as the ultimate sample size, the degree of clustering to be employed in the sampling procedures and the level of reliability desired for the resulting estimates. These issues are discussed later in the manual.

B. SAMPLE SIZE

A key determinant to over-all sample size is the number of subnational groupings for which fairly reliable data are required. The size of the population in each country would not have much effect on the over-all sample size since the precision of the survey estimates depends mostly on the number of observations in each analytical cell. The population size only affects the precision through the finite population correction factor, which has a negligible effect when the sampling fractions are small. The proportion of the population sampled for the migration survey in each participating country will be quite small.

There does not appear to be a single isolated statistic which would be regarded as the most important. If, for example, the primary interest were to measure the migration rate (proportion) by

major area with a relative sampling error of 10 per cent, a very small sample (200 or so interviews, depending upon clustering factors) would be allocated to the major metropolitan stratum and about the same to the other urban stratum, but several thousand to the rural stratum, in a country with migration patterns such as the Republic of Korea. This allocation of the sample seems unwise since the volume of in-migration to the rural areas in most countries is quite small. It is not analytically justifiable to have a disproportionately large rural sample.

Of course, the survey must provide estimates not only of volume but also a variety of characteristics of migration, such as causes and consequences. In effect, it is a type of multipurpose migration survey. Therefore, the sample within each major geographic category should be as large as possible. At the same time, the sample size should not be so large that effective control of the field operations would be lost. Since non-sampling error could have a serious impact on the data, another major factor in determining the sample size will be the resources available for quality control.

As mentioned above, the suggested major geographic sampling and analytic strata for the survey are major metropolitan, other urban and rural. However, in some countries, the other urban stratum may be subdivided by size categories in order to enhance the comparability of the data for cross-national analysis. To cope with the opposing objectives of detailed analysis (large sample size needed) and management of operations to control non-sampling error (small sample size needed), a practical minimum sample size would seem to be 3,000 interviews per major geographic category. This would indicate a basic minimum over-all sample size of 9,000 with correspondingly larger sample sizes for additional geographic analysis groupings.

Another consideration is the size of the sample for the migrant subgroup within each geographic stratum, since a detailed analysis is planned for this subgroup. The proportion of migrants varies considerably among the strata and is quite low in the rural stratum in some countries. This issue is discussed later in the section on country modifications of sample design.

Taking these factors into account, as well as country resources and comments from the Advisory Committee meetings, 14,000 is chosen as the target sample size for the participating countries. This figure is used in illustrating the alternative sample allocations in the next section. The precise over-all sample size will vary by country depending on individual characteristics, resources and objectives. In addition, it is suggested that 1,000 interviews be taken in connection with the validation-reinterview study (see chapter III, section C above). Altogether, therefore, 15,000 sample interviews would be sought. (It should be noted that the figure 14,000 would refer to designated units rather than to completed interviews. The number of completed interviews would be somewhat less than 14,000 because of non-interviews for various reasons.)

C. SAMPLE ALLOCATION

One objective of the survey will be to compare migrants and non-migrants in different geographic subdomains. However, the main emphasis of the survey will be to analyse and compare different types of migrants and examine their impact on development. In order to increase the precision for estimates of migrant characteristics, it is advantageous to oversample in geographic areas with a higher concentration of migrants. At the same time, an optimum allocation procedure must take into account the differences in cost and population variance among the different strata. An optimum allocation formula involving these parameters is given in annex II, along with its derivation. One advantage of this formula is that the parameters can be estimated in each country to adjust the allocation optimally to the specific conditions in the country.

The definition of migrant to be used for sampling will be determined by the major objectives of the survey and the kinds of migration data available. It is understood that this definition chosen for sampling has no necessary implications for definitions employed in the analysis phase of the

survey and is useful only in so far as it increases sampling efficiency. There appear to be two viable options. The first option would be to use migrants who had lived in a place different from their place of origin at least once during their lifetime (i.e., lifetime migrants). The second option would be to use those who had migrated to the area less than five years prior to the census (recent migrants). Annex I contains a discussion of the advantages and disadvantages of each option.

Data depicting population distribution and the migration phenomenon in the participating countries in 1970 are used in the present manual to illustrate different sample allocations and their implications for survey implementation under various assumptions. Special tabulations from the 1970 censuses were produced by the ESCAP secretariat for four countries, Indonesia, Malaysia, the Republic of Korea and Sri Lanka, while census data in published form and various survey estimates are used for the Philippines and Thailand (see annex III). Shown in table 3 are the distribution of the target population by stratum and the proportion of the population of each stratum classified as lifetime migrants for each country. (The lifetime reference period for migration is used in this manual because of data availability; recent migration data would be used in the same manner.) It is to be borne in mind that these figures reflect the cumulative patterns of population distribution and lifetime migration up to 1970 and will ideally be updated using the 1980 census results for sample selection purposes. In so far as they reflect cumulative patterns, however, they should provide a reasonable approximation of the distributional characteristics of the participating countries for illustrative purposes. It should also be noted that the migration figures for the Philippines and Thailand are not strictly comparable with those for the other countries owing to the unavailability of special tabulations and should thus be viewed as very rough approximations. The sources and limitations of these figures are given in annex III.

Table 4 shows the distribution of the sample and the estimated number of migrants and non-migrants in the sample for the geographic strata in each country, based on proportional allocation and optimum allocation under different assumptions. The proportional allocation, shown for purposes of comparison, would be used to obtain a self-weighting sample of households (i.e., each sample household would have the same expansion factor). Although this would simplify the processing of household data, the individual respondent questionnaires would still have differential weights. It can be seen in table 4 that the proportional allocation yields a considerably smaller sample of migrants than the optimum allocation and would therefore result in less precision for migrant characteristics. For example, in the case of Indonesia, the sample under optimum allocation scheme A has 1,038, or 86 per cent, more migrants than that under proportional allocation.

Since the optimum allocation formula in annex II depends on the relative costs, population variances and proportion of migrants in the geographic strata, it is necessary to obtain estimates for these parameters in each country before actually allocating the sample. The different assumptions for the optimum allocation in table 4 are used for illustrative purposes and to show the effect of changes in the assumptions. However, an attempt was made to have each set of assumptions intuitively reasonable.

The cost per sample household for transportation, listing and interviewing is usually higher in the rural areas because the housing units are more dispersed, and sometimes fewer maps are available. Experience in various countries has shown that survey costs in rural areas are sometimes two or more times higher than in urban areas. The actual relative costs depend on population density and transportation costs between and within the sample areas. In table 4 the assumption for optimum allocation scheme A is that the cost per sample household in the rural areas is twice that of the major metropolitan and other urban areas, while the assumption for optimum allocation B is that the rural-to-urban cost ratio is 3:1. In comparing these two allocations, it can be seen that the higher cost in rural areas results in further oversampling the major metropolitan/urban areas.

The assumption for the population variance of migrant characteristics is less straightforward. For some characteristics, such as income and education, one can expect there to be more homogeneity

Table 3. Population distribution and migration statistics for participating countries, 1970 round of censuses

	<i>Distribution of population aged 15+ by strata (Number and percentage)</i>			<i>Percentage of strata population aged 15+ classified as migrant</i>			<i>Strata share of total migrants (Percentage)</i>		
	<i>Major metropolitan area</i>	<i>Other urban</i>	<i>Rural</i>	<i>Major metropolitan area</i>	<i>Other urban</i>	<i>Rural</i>	<i>Major metropolitan area</i>	<i>Other urban</i>	<i>Rural</i>
Indonesia	2 576 524 (3.9)	9 365 131 (14.1)	54 112 650 (82.0)	62.1	16.4	4.8	27.9	26.8	45.3
Malaysia	324 600 (7.7)	1 113 200 (25.9)	2 846 100 (66.4)	59.4	37.6	38.9	11.7	24.3	64.5
Philippines	3 966 695 (11.4)	7 711 125 (21.7)	25 006 666 (66.9)	58.6	38.8	25.2	20.8	26.5	52.7
Republic of Korea	3 500 000 (19.5)	4 500 000 (24.5)	10 000 000 (56.0)	64.0	59.0	4.4	42.5	49.5	8.0
Sri Lanka	712 340 (9.9)	902 680 (12.4)	5 618 870 (77.7)	51.0	35.9	33.2	14.2	12.8	73.0
Thailand	1 400 297 (7.7)	1 142 138 (6.0)	15 964 767 (86.3)	55.0	50.0	25.0	14.4	10.7	74.8

Note: See annex III for sources and limitations of data.

Table 4. Alternative sample allocation schemes

	Proportional allocation scheme			Optimum allocation								
				Scheme A For: $\sigma_1 = \sigma_2 = \sigma_3$ and $2C_1 = 2C_2 = C_3$			Scheme B For: $\sigma_1 = \sigma_2 = \sigma_3$ and $3C_1 = 3C_2 = C_3$			Scheme C For: $\sigma_1 = 3\sigma_2/4 = \sigma_3/2$ and $2C_1 = 2C_2 = C_3$		
	No. of interviews	Expected No. of interviews		No. of interviews	Expected No. of interviews		No. of interviews	Expected No. of interviews		No. of interviews	Expected No. of interviews	
	Migrants	Non-migrants		Migrants	Non-migrants		Migrants	Non-migrants		Migrants	Non-migrants	
Indonesia												
Major metropolitan area	546	339	207	2 001	1 243	758	2 244	1 393	850	3 135	1 947	1 188
Urban area	1 974	324	1 650	3 738	613	3 125	4 191	687	3 504	4 392	720	3 672
Rural	11 480	551	10 929	8 261	397	7 865	7 565	363	7 202	6 472	310	6 162
Total	14 000	1 214	12 786	14 000	2 252	11 748	14 000	2 444	11 556	14 000	2 978	11 022
Malaysia												
Major metropolitan area	1 078	640	438	1 601	951	650	1 789	1 063	726	2 520	1 497	1 023
Urban area	3 626	1 363	2 263	4 368	1 642	2 725	4 882	1 835	3 046	5 157	1 939	3 218
Rural	9 296	3 616	5 680	8 032	3 124	4 907	7 329	2 851	4 478	6 322	2 459	3 863
Total	14 000	5 619	8 381	14 000	5 717	8 283	14 000	5 749	8 251	14 000	5 896	8 104
Philippines												
Major metropolitan area	1 596	935	661	2 543	1 490	1 053	2 818	1 651	1 166	3 838	2 249	1 589
Urban area	2 716	1 054	1 662	4 023	1 561	2 462	4 457	1 729	2 728	4 552	1 767	2 786
Rural	9 688	2 441	7 247	7 434	1 873	5 561	6 725	1 695	5 030	5 609	1 414	4 196
Total	14 000	4 430	9 570	14 000	4 924	9 076	14 000	5 075	8 925	14 000	5 429	8 571
Republic of Korea												
Major metropolitan area	2 730	1 747	983	5 065	3 241	1 823	5 249	3 360	1 890	6 391	4 090	2 301
Urban area	3 430	2 024	1 406	6 252	3 689	2 563	6 480	3 823	2 657	5 917	3 491	2 426
Rural	7 840	345	7 495	2 683	118	2 565	2 270	100	2 171	1 693	74	1 618
Total	14 000	4 116	9 884	14 000	7 048	6 952	14 000	7 283	6 717	14 000	7 655	6 345
Sri Lanka												
Major metropolitan area	1 386	707	679	2 133	1 088	1 045	2 440	1 244	1 196	3 459	1 764	1 695
Urban area	1 736	623	1 113	2 268	814	1 454	2 594	931	1 663	2 757	990	1 767
Rural	10 878	3 612	7 266	9 599	3 187	6 412	8 966	2 977	5 989	7 783	2 584	5 199
Total	14 000	4 942	9 058	14 000	5 089	8 911	14 000	5 152	8 848	14 000	5 338	8 662
Thailand												
Major metropolitan area	1 078	593	485	1 941	1 068	873	3 255	1 790	1 465	4 789	2 634	2 155
Urban area	840	420	420	1 509	755	755	1 899	949	949	3 724	1 862	1 862
Rural	12 082	3 021	8 820	10 550	2 637	7 912	8 846	2 212	6 635	5 487	1 372	4 115
Total	14 000	4 034	9 725	14 000	4 460	9 540	14 000	4 951	9 049	14 000	5 866	8 132

Note: The migrant category in this table refers to an approximation of lifetime migrants. The numbers are for illustrative purposes only to provide a comparison of the alternative sampling schemes. See annex II for derivation of the optimum allocation and annex III for sources and limitations of the data.

within rural areas, i.e., a higher population variance in urban areas. However, in the case of proportions, the population variances in the urban and rural areas would vary considerably with the size of the proportion. For this reason, optimum allocation schemes A and B, which have an assumption of equal population variances among the strata, should be used as the main examples of optimum allocation for each country. The population variances used for the actual sample allocation will depend on the main characteristics to be studied in each country. Optimum allocation scheme C shows the effect of differential population variances among the strata, when compared with A. It can be seen that a higher variance in the major metropolitan and other urban strata results in further oversampling in these areas.

Within the major metropolitan and other urban strata, it may be possible to achieve further sampling efficiencies by substratifying the enumeration areas by proportion of migrants and oversampling in areas of heavy migrant concentrations. The key to successful stratification is whether it is possible to identify areas of the cities that are overwhelmingly migrant, say 90 per cent. Probably most gains would be obtained by creating two substrata (within major metropolitan and/or other urban), one with 90 per cent or more migrants and the other with less than 90 per cent migrants. The allocation formula in annex II would still be used, substituting major metropolitan and/or other urban with their substrata.

An issue which must be considered when using migrant information for stratification purposes concerns the manner in which this information is collected in the population censuses of the participating countries. Of the six countries participating in the survey, only three countries (Indonesia, Malaysia and Thailand) collected migration data on a 100 per cent basis, the remaining countries using samples of various sizes. The manner in which the samples were designed has important implications for the present survey. If the samples were drawn using a fixed sampling fraction (for example, one in every 10 households) across enumeration areas, the resulting data could be used directly in the stratification of the sample for the present survey. If, however, the samples were drawn on an area basis (only households in selected areas were asked the migration questions), it will be necessary to combine areas into fairly large strata, thus detracting from benefits to be gained by this stratification.

Table 5. Ratio of sampling rates under optimum allocation A

<i>Country</i>	<i>Major metropolitan to rural</i>	<i>Urban to rural</i>
Indonesia	5.09	2.61
Malaysia	1.75	1.39
Philippines	2.16	1.75
Republic of Korea	5.39	5.18
Sri Lanka	1.75	1.47
Thailand	2.14	2.03

Table 5 shows the ratios of major metropolitan to rural and urban to rural sampling rates under optimum allocation scheme A for each country. This table provides an indication of how the weights would vary among the strata and a comparison between the optimum and proportional allocations. Since the proportional allocation provides for equal sampling rates among the strata, the resulting ratio of sampling rates between all strata would be 1. The highest ratios in table 5 appear in the Republic of Korea, where both the major metropolitan to rural and urban to rural ratios are over 5, owing to the large difference between the proportion of migrants in urban and rural areas. Indonesia also has a relatively large ratio (5.09) between the major metropolitan and rural sampling rates. The remaining ratios in table 5 vary between 1.39 and 2.16, indicating only moderate over-

sampling in urban and metropolitan areas. The sampling strategies for Malaysia and Sri Lanka are very similar. The allocations in the Philippines and Thailand also involve a similar rate of oversampling in urban and metropolitan areas.

The sample allocation was optimized in terms of minimum variance for national estimates of migrant characteristics; therefore, the resulting differential weights among the strata should not adversely affect the precision for these estimates. Since the geographic sampling strata are also analytical categories, a self-weighting sample within each stratum would minimize the variance of estimates for these subgroups.

Optimum allocation scheme A can also be examined to compare the distribution of the sample migrants in the different countries. The total number of migrants in the sample varies from 2,252 (16 per cent of the sample) in Indonesia to 7,048 (about 50 per cent of the sample) in Thailand. The number and distribution of sample migrants is similar for Malaysia, Sri Lanka and Thailand, with more than 3,000 sample migrants in the rural stratum and a good representation of migrants in the sample for the major metropolitan and urban strata too. The estimated 4,924 sample migrants in the Philippines are evenly distributed across the strata. However, the number of rural migrants in the sample for the Republic of Korea and Indonesia is quite low (118 and 397 respectively) because of the small proportion of migrants in the rural areas of these countries. The sample of non-migrants in each stratum appears to be sufficient for the migrant/non-migrant comparison in each country.

After estimating the approximate distribution of the sample in a country based on the optimum allocation, it is important to find out whether any of the subgroups of interest would have an insufficient number of observations for the desired analysis. Although the allocation procedure was designed to maximize the precision for national estimates, it may be necessary to adjust this allocation to ensure a sufficient representation in each important subgroup. Each optimum allocation shown in table 4 provides about 2,000 or more interviews for each stratum in most countries. However, it is quite difficult to control the number of migrants in the sample for a stratum, as illustrated by the small sample of rural migrants in the Republic of Korea and Indonesia. If a country has such a small subgroup which is vital to the analysis, a more complex listing and double-sampling procedure may be required, as explained in section E below on country modifications of sample design.

D. CARRYING OUT THE SAMPLE PLAN

This section can only be sketchily presented at this time because it is heavily dependent on the materials and resources available in the individual countries. Only some possibilities are given.

1. *Frame*

The best sampling frame available for the national migration survey in each country is the population and housing census. In most countries, the 1980 census will be processed before the migration survey is implemented, in which case the 1980 census data and maps will be available for the sampling frame. If the 1980 census materials are not available in a certain country, the 1970 census materials can be used; this should not be a problem for rural areas where no major changes will have occurred. However, in urban areas, auxiliary information will have to be compiled and used to identify newly developed areas with a high proportion of migrants, such as new construction and squatter settlements.

2. *Stages of selection*

To carry out the sample plan, multiple stages of selection will be used, with the primary sampling units (PSUs) selected with probabilities proportional to estimated size (population or households), and the final stage consisting of listing a census enumeration area (EA) or smaller segment. The probabilities of selection will be determined to provide an approximately self-weighting sample

within each major geographic stratum. In the urban areas, the sample households can be randomly (or systematically) selected from this listing. The number of sample households per urban EA may vary somewhat by country, depending upon cost considerations and the intraclass correlation between households, but it will probably be around 10. This ultimate cluster size of around 10 is based on two considerations. First, if the cluster size were much smaller, an inordinate number of PSUs would have to be selected and listed, putting a strain on the available resources. Secondly, experience has shown that the intraclass correlation between households within the ultimate cluster usually makes it inefficient to select many more than 10 households in a sample area. However, in the rural areas it may be desirable to construct compact clusters of about 10 households each from the listing. Every household in a selected compact cluster would then be interviewed. Since the housing units within rural areas are usually dispersed, the use of compact clusters should decrease the travel cost within the sample EAs and lessen the problems of locating sample households.

In most countries, it is not possible to obtain information on volume and proportion of migration for areas smaller than the census EAs. If a country has EAs with 150 or more housing units, the sample EAs should be subdivided into smaller segments so that only a selected segment has to be listed. It is desirable to keep the expensive listing operation to a minimum and to begin interviewing as soon after listing as possible.

The main purpose of a multistage design is to reduce the time and cost of travel between the units at each stage and to minimize the amount of listing that has to be carried out. The stages of selection depend on the conditions and size of each individual country and can only be worked out in consultation with host country statisticians. Table 6 provides an example of the stages of selection in each stratum within a country.

Table 6. Possible selection stages

<i>Area</i>	<i>First-stage units (PSUs)</i>	<i>Second-stage units</i>	<i>Third-stage units</i>	<i>Fourth-stage units</i>	<i>Fifth-stage units</i>
Major metropolitan	Enumeration areas	Blocks or segments ^a	Households	One person per household	
Other urban	Cities/towns	Enumeration areas	Blocks or segments	Households	One person per household
Rural	Villages	Enumeration areas	Segments	Compact cluster of 10 households	One person per household

^a If the enumeration areas in a country have 150 or fewer households each, it would not be necessary to subdivide the EAs into blocks or segments, and this stage would be deleted.

3. Estimation, weighting, imputations and variances

This section is dependent on the final sample design that would be adopted in each country. However, a few points can be made.

To the extent that good independent estimates of the target population (adults aged 15 to 64, perhaps by age groups and by sex) are available from recent census sources or elsewhere, they can be used to ratio-adjust the simple inflation estimates obtained from the survey in order to improve the accuracy of the estimates. It does not seem likely that independent estimates would be available for subnational areas, so only a national adjustment could be achieved. This, of course, needs close scrutiny before a decision could be made.

With the many strata and oversampling to be used, the sample for any one country will not be self-weighting, so that provision will have to be made for variable weights in the processing. Imputations for non-response would best be made through weighting rather than by duplicating interview records. The weight for a given record would thus be the product of three factors: the inverse of the probability of selection, the ratio adjustment factor and an adjustment for non-response. The inverse of the probability of selection would itself be computed by taking account of the product of the probabilities in the various selection stages. The non-response adjustment factor would be computed as the quotient of the count of interviewed persons plus non-interviewed persons divided by the count of interviewed persons, for individual small areas (such as PSUs).

Consideration should also be given to establishing two sets of data records, each with its own weighting procedures. This possibility arises because there will be household data as well as individual data and differing estimates can be made from each set, taking account of the unique weights and non-interview adjustment procedures that would pertain to each.

The variance estimator, to be used in calculating the sampling errors of the survey estimates, cannot be specified until a precise sample plan is settled upon. Supplementary documents should be developed later specifying guidelines for the estimation procedures, such as the weighting and ratio adjustment procedures, as well as alternative variance estimators.

E. COUNTRY MODIFICATIONS OF SAMPLE DESIGN

The basic sample design for the national migration surveys was presented in the preceding sections. This design was based on general survey and sampling objectives common to all the participating countries. It is felt that the basic elements of the sample design put forth in this manual can be used as guidelines by each country in establishing its own specific sampling plan. It is recognized that departures from one or more of the components of the design suggested here will be both necessary and unavoidable in adaptation to national needs or requirements. For this reason, therefore, it would not be productive to lay out a step-by-step approach to implementing the sample scheme. There are too many viable options at each stage of the process for a given nation to choose from, depending on local situations, to make a step-by-step scheme meaningful. For example, the number and type of PSUs and ultimate clusters, even with a fixed sample size of 14,000 households, will probably be highly variable from country to country. It is expected that detailed implementation steps for individual nations will be developed and specified at the stage when each country begins planning for adaptive sampling, in consultation with foreign sampling advisers, when needed.

The purpose of this section is to present sample design modifications which the individual countries may be interested in pursuing because of their specific objectives and conditions.

1. *Sample size requirements for small analytical cells*

It was pointed out in the previous section that after the sample is allocated, certain subgroups which represent a small proportion of the stratum population may have an insufficient number of observations for a detailed analysis, such as the rural migrant category in the Republic of Korea and Indonesia. This problem is even more serious if the subgroup is critical to the analytical objectives of a country, in which case it may be necessary to increase the number of sample cases in this subgroup. It would not be efficient merely to increase the sample size for the geographic stratum, as a large number of observations would include relatively few members of the desired subgroup. Therefore a double-sampling technique may be considered as a means of increasing the number of interviews in the subgroup. This procedure would involve a more complex listing in which the households or individuals within each sample area would be screened for their migration status. The listing would then be post-stratified by migration status. The units belonging to the migrant subgroup would be included in the sample with certainty at the last stage of selection, while the others would be sampled in the usual way.

However, this double-sampling procedure has several major disadvantages. The current survey procedures include the listing of households within sample areas. If the migration status of a household were based on whether a migrant lived in the household, it would be possible that the selected individual within the household could be a non-migrant. However, if the procedures were changed to list individuals within sample areas, the cost and complexity of the survey would be increased considerably. The complexities of these procedures could also introduce larger non-sampling errors into the survey results. Therefore, the double-sampling procedure should only be considered in extreme circumstances, after weighing the costs and benefits.

2. Special migration studies

Some countries may have a special interest in studying a unique aspect of their migration pattern or process which has important implications for national policy, such as transmigration, land resettlement or refugee camps. Although the national survey design may be modified to accommodate the analytical requirements for this special study, striving to maintain the national character of the survey will satisfy the objectives of the regional analysis in the long run.

A separate stratum could be created for the areas with the migration phenomenon of interest. In order to satisfy the sample size requirements for this study, the sample for the special stratum should be supplemented by increasing its sampling rate, after the national sample has been designed. In this way, the national and regional analysis would not be adversely affected.

Although the national migration survey would probably provide most of the data required for the analysis of the special migration phenomenon, additional questions on this specific type of migration may be needed. It is recommended that these questions should be included in a separate module to be administered only in the special stratum. By limiting the additional training and manuals to this stratum, the national survey would not be compromised.

3. Use of existing frames

One of the management objectives is to use, whenever possible, an existing national household sample frame for the migration survey in the participating countries. This would have an enormous advantage in savings on preparation costs, as well as in lead time to get ready for the survey. There are, however, some major technical disadvantages in using in-place samples, among them the following:

(a) The existing sample in a given country may be too thin to produce reliable subnational data, that is, for the primary city, other urban territory and rural areas. If this is so, there must be a contingency plan for drawing the requisite supplemental sample. This would mean that not much lead time would be saved in any case.

(b) There is the likelihood of poor or biased response if the same units have been over-surveyed for other purposes in continuing surveys.

(c) It is quite likely that "differences" in the migration statistics between any two countries would show up merely because of the statistical artifact of the two sample populations having had varying degrees of prior involvement in their nation's surveys.

Nonetheless, the existing frames for national household surveys in each country should be examined and evaluated in terms of their use for the migration survey, taking into account both the cost constraints and objectives of the survey.

Annex I

DISCUSSION OF THE ADVANTAGES AND DISADVANTAGES OF USING LIFETIME *VERSUS* RECENT MIGRANTS FOR SAMPLE DESIGN

The type of migrant used for sample design purposes depends on the analytical priorities and the availability of various kinds of migration data. Given the survey objective of studying in depth many forms and types of migration, perhaps those persons who had lived in a place different from their place of origin at least once during their lifetime (lifetime migrants) should be used for the sample design. Use of lifetime migrants has the advantage of being general in scope and thus of including those persons who would be classified as migrant under most reasonable definitions. Census "place of birth" data were collected by each of the participating countries in 1970 (it is expected that this information will also be collected in 1980) and could be used to classify the population of each country according to lifetime migration status for sample stratification and allocation purposes. Using this information, migrants would consist of those persons whose place of birth was different from their place of census enumeration. It may thus be inferred that such persons had migrated at least once since birth (assuming accurate reporting).

One shortcoming in the use of place of birth data to measure lifetime migration status is that because persons are classified on the basis of a comparison of place of residence at only two points in time (time of birth and time of census enumeration), intervening events are not taken into consideration. Thus, all persons who are enumerated in the same place as their reported place of birth are classified as non-migrants, irrespective of the number of intervening moves which may have been made (i.e., return migrants). For this reason, the place of birth measure of migration status tends to understate the proportion of the population who had "ever moved". In some countries, data on duration of residence were also obtained in the census, which ideally could be used to improve the estimates of the "ever moved" population. One problem with this procedure is that errors in the age or duration data (which are not uncommon in certain countries) could lead to serious classification problems. It is to be recalled, however, that this definition is to be used only for sample allocation purposes in order to increase the efficiency of the sample and not for the analysis. The migration histories completed by sampled respondents will permit return migrants to be identified and reclassified as desired, as well as the use of a number of alternative definitions of various types of migrants. For sampling purposes, data on place of birth and sometimes duration of stay provide reasonable means of identifying a substantial proportion of the population who had "ever moved".

From a policy perspective, a disadvantage in the use of lifetime migrants is that persons migrating recently are indistinguishable (in the census tabulations) from persons who had migrated in the distant past when social and economic conditions might have been quite different. The present survey is also intended to provide guidelines for policy deliberations concerning interrelationships between migration, urbanization and development, and policy makers must concern themselves with the current social and economic realities. Therefore, it may be advantageous to use recent migrants in the allocation of the sample among the strata so as to increase the size of the sample of respondents who had migrated under contemporary conditions. This is an important consideration, as an insufficient number of recent migrants in samples (owing to their relatively small number) has restricted the policy relevance of a number of previous research efforts.

The required information for allocation purposes may be obtained from census questions on place of residence at a fixed prior date (typically five years prior to the census date) or duration of residence in the place of census enumeration. Each of the participating countries collected data on one of these questions in their 1970 census and will probably do so in 1980-1981. With this information, it is possible to classify the population enumerated in each place into two categories: those

who arrived less than five years prior to the census and residents of more than five years' duration. The latter category includes both non-migrants and persons who had migrated to the area more than five years prior to the census. The advantage of this choice as compared with the first option (lifetime migrants) is that the sample may be allocated among sampling strata such that the size of the sample for the recent migrant subcategory will be more suitable for the types of detailed analyses proposed for the survey data.

Annex II

THEORETICAL BASIS FOR THE OPTIMUM SAMPLE ALLOCATION

The following discussion of optimum allocation takes into account the relative costs as well as the population variances in the various strata. A very simple cost function is assumed where the cost per unit in the sample is fixed for each stratum (i.e., does not vary with sample size).

Let C represent the total cost, excluding any fixed overhead costs and let C_1 represent the cost per sample unit for the first stratum, C_2 the corresponding cost for the second stratum, etc. Then the part of the total cost of the survey which will be affected by the sample size is:

$$C = C_1 n_1 + C_2 n_2 + \dots + C_L n_L = \sum_{h=1}^L C_h n_h$$

where

n_h is the number of units sampled from stratum h ,

L is the number of strata.

With this cost function, the optimum allocation (yielding a minimum variance) when the total expenditure is fixed is obtained when the number of units sampled in stratum h is given by:

$$n_h = \frac{N_h \sqrt{t_h} \sigma_h / \sqrt{C_h}}{\sum_h \frac{N_h \sqrt{t_h} \sigma_h}{\sqrt{C_h}}} n$$

where

N_h is the population of stratum h ,

t_h is the proportion of N_h in the subgroup of interest (migrants),

σ_h^2 is the population variance of a key variable for the subgroup in stratum h ,

$n = \sum_h n_h$ is the total sample size.

Proof. The variance of a survey estimate for migrants, from a stratified random sample, is the following:

$$\sum_h \left(\frac{\tilde{N}_h}{N} \right)^2 \frac{\sigma_h^2}{r_h t_h N_h}$$

where

$$\tilde{N}_h = t_h N_h$$

$$\tilde{N} = \sum_h t_h N_h.$$

$$r_h = \frac{n_h}{N_h}, \text{ the sampling rate for stratum } h.$$

To determine the value of n_h which minimizes the variance subject to the fixed total cost C , the following function F with a Lagrangian multiplier is used:

$$F = \sum_h \left(\frac{\tilde{N}_h}{\tilde{N}} \right)^2 \frac{\sigma_h^2}{r_h t_h N_h} + \lambda (\sum_h C_h r_h N_h - C)$$

$$\frac{\partial F}{\partial r_h} = - \left(\frac{\tilde{N}_h}{\tilde{N}} \right)^2 \frac{\sigma_h^2}{r_h^2 t_h N_h} + \lambda C_h N_h = 0$$

Solving for n_h ,

$$n_h = \frac{\tilde{N}_h \sigma_h}{\tilde{N} \sqrt{t_h \lambda C_h}}$$

and

$$n = \sum_h n_h = \sum_h \frac{\tilde{N}_h \sigma_h}{\tilde{N} \sqrt{t_h \lambda C_h}}$$

Substituting for $\sqrt{\lambda}$,

$$n_h = \left\{ \frac{\frac{N_h \sqrt{t_h} \sigma_h}{\sqrt{C_h}}}{\sum_h \frac{N_h \sqrt{t_h} \sigma_h}{\sqrt{C_h}}} \right\} n$$

Annex III

SOURCES OF DATA FOR USE IN SAMPLE ALLOCATION

As was noted in the text, special tabulations from the 1970 round of censuses of four countries (Indonesia, Malaysia, the Republic of Korea and Sri Lanka) were made available for use in this manual. Similar tabulations were not, however, available for the Philippines or Thailand, thus necessitating the use of census data in published form. The available published data are not, however, comparable with the tabulations provided by the ESCAP secretariat and had to be adjusted for use in the present manual. The nature of these adjustments and their implications are discussed in this appendix.

The data available for the Philippines differed from the tabulations provided by ESCAP in two respects. First, the population counts and migration statistics for the Philippines refer to persons aged 10 and above compared with aged 15 and above in the ESCAP tabulations. This does not seem to be a serious problem, but should be kept in mind when the sample allocations for the Philippines are compared with those for the other countries. Secondly, and of greater importance, the migration statistics made available for the Philippines refer to migration in a 10-year interval (between 1960 and 1970) rather than lifetime migration. Thus, the proportion of the population classified as migrant is systematically lower in the Philippines data because of the shortened time reference period (migration interval). In order to make these data comparable with those of the countries for which special tabulations are available, the proportion of the population classified as migrant in each stratum was doubled to approximate a lifetime reference period. It is to be noted that, while this is a somewhat arbitrary procedure, the numbers obtained seem reasonable, for illustrative purposes, within the context of the migration phenomena in Asia. In addition, the use of a constant factor in inflating the estimate of the proportion of the migrant population of each stratum maintains the relative volume (proportionally) of migration across strata inherent in the 10-year migration data. This is an important feature since the sample allocation algorithm is largely dependent upon ratios of migrants to population in the various strata. It is to be kept in mind, nevertheless, that figures on lifetime migration for the Philippines used in the manual are approximations and are used for illustrative purposes only.

The available data on lifetime migration in Thailand also present a problem of comparability. While referring to a lifetime reference period for migration, only moves involving the crossing of provincial boundaries (as opposed to moves involving the crossing of local administrative boundaries) were classified as migration events in the 1970 census of Thailand. Thus, moves that were classified as migration events (i.e., moves within provinces) in the ESCAP tabulations are not so classified in the Thai census. Further, the degree of understatement of the volume of migration is probably not the same in all provinces, but in all likelihood varies by such factors as province size and perhaps proximity to large urban centres (i.e., Bangkok and Chiang Mai). A comparison of estimates of proportion migrant in each stratum from the 1970 census and other sources using the more conventional administrative boundaries reveals that the 1970 figures (based on interprovincial moves) grossly understate the volume of migration based on local boundaries and that the degree of understatement varies considerably by stratum. Thus, it did not seem feasible to use the 1970 census figures for the purposes of this manual.

Estimates of the proportion migrant in each stratum (by the place of birth or lifetime definition) were, however, available from the longitudinal study of social, economic and demographic change carried out in Thailand.⁴ These estimates measure migration using "amphoes" (counties) as the aerial unit of analysis. To the extent that these estimates, however, refer only to selected

⁴ See *The Methodology of the Longitudinal Study of Social Economic, and Demographic Change*, Institute of Population Studies, Chulalongkorn University, Bangkok, Research report No. 6.

subgroups (i.e., ever-married females and male heads of households) of the target population for the present study (the population aged 15-64) and also includes return migration, it was necessary to adjust these estimates. It was felt that the longitudinal study estimates overstated the proportion of the population aged 15 and above who were migrants in comparison with the ESCAP tabulations owing to differences in the study population and the inclusion of return migration; the longitudinal study estimates were accordingly adjusted downwards to yield the numbers used in the present manual. These figures reflect stratum differences in the volume of lifetime migration observed in the several sources of data available and, given the unavailability of appropriate data, seem reasonable for use in illustrating sample design options for Thailand in the present manual.

Comparative study on migration, urbanization and development in the ESCAP region

Survey manuals

- I. Survey organization and monitoring
- II. The core questionnaire
- III. Manual for interviewers
- IV. Manual for supervisors
- V. Manual for interviewer trainers
- VI. Sample design manual
- VII. Manual for office editors and coders

Country reports

- No. 1. *Migration, Urbanization and Development in the Republic of Korea* (Bangkok, ESCAP, 1980).
- No. 2. *Migration, Urbanization and Development in Sri Lanka* (in press)
- No. 3. *Migration, Urbanization and Development in Indonesia* (forthcoming)

In preparation

- Tabulation plan
- Data processing system manual

In preparation

- Philippines
- Pakistan
- Malaysia
- Thailand
- South Pacific

For further information on the comparative study project, contact: The Chief, Population Division, ESCAP, United Nations Building, Rajdamnern Avenue, Bangkok 2, Thailand.

For publications from the project and other Population Division publications, contact: The Clearing-house, Population Division, ESCAP, United Nations Building, Rajdamnern Avenue, Bangkok 2, Thailand.