# DOCUMENT RESUME

| | |
|---|---|
| ED 137 142 | SE 022 444 |
| TITLE | Guidelines for Preparing Subnational Population Projections. Asian Population Studies Series No. 32. |
| INSTITUTION | United Nations Economic and Social Commission for Asia and the Pacific, Bangkok (Thailand). |
| PUB DATE | 75 |
| NOTE | 74p.; Best copy available |
| EDRS PRICE | MF-$0.83 HC-$3.50 Plus Postage. |
| DESCRIPTORS | *Demography; *Models; *Planning; *Population Trends; Prediction; *Research Methodology; *Statistical Analysis |
| IDENTIFIERS | *UNESCO |

ABSTRACT

There has long been a need for a systematic set of instructions for the projection of populations for such subnational areas as states, provinces, districts, or statistical planning areas; and for specific groups of population which may overlap different areas within a country, such as the labor force, various occupation groups, the school age population, literates, or ethnic groups. These guidelines attempt to meet that need. The guidelines are designed to aid development planners by discussing the types of population projections and the ingredients of each to help them decide what the data will permit, and what they should do to produce the numbers required by their work. Sections include the following: (1) General Considerations in Population Projections; (2) Subnational Areas and Population Projection; (3) Measurement and Estimation of Internal Migration; (4) Measurement and Estimation of Natural Population Change; (5) Mechanical Methods for Projecting Subnational Populations; (6) Analytic Methods of Subnational Population Projection; (7) Projecting Urban and Rural Population and Socio-Economic Groups; and (8) Some Methods of Estimation from Incomplete Data. (Author/RH)

ASIAN POPULATION STUDIES SERIES
No. 32

# GUIDELINES FOR PREPARING SUBNATIONAL POPULATION PROJECTIONS

ECONOMIC AND SOCIAL COMMISSION FOR ASIA AND THE PACIFIC

Bangkok, Thailand, 1975

UNITED NATIONS

## THE ESCAP *ASIAN POPULATION STUDIES SERIES*, AND OTHER PUBLICATIONS

### A. Asian Population Studies Series*

No.1** *Administrative Aspects of Family Planning Programmes — Report of a Working Group* (E/CN.11/742; United Nations publication, Sales No.66.II.F.10): 1966

No.2** *Family Planning, Internal Migration and Urbanization in ECAFE Countries — A Bibliography of Available Materials* (E/CN.11/807; United Nations publication, Sales No. E.68.II.F.13): 1968

No.3** *Communications in Family Planning — Report of a Working Group* (E/CN.11/830; United Nations publication, Sales No. E.68.II.F.17): 1968

No.4** *Assessment of Acceptance and Effectiveness of Family Planning Methods — Report of an Expert Group* (E/CN.11/882; United Nations publication, Sales No. E.69.II.F.15): 1969

No.5** *Evaluation of Family Planning Programmes — Report of a Regional Seminar* (E/CN.11/936; United Nations publication, Sales No. E.70.II.F.20): 1970

No.6** *Fertility Studies in the ECAFE Region — A Bibliography of Books, Papers, and Reference Materials* (E/CN.11/992; United Nations publication, Sales No. E.72.II.F.3): 1971

No.7 *Interrelation between Population and Manpower Problems — Report and Selected Papers of a Regional Seminar* (E/CN.11/1015): 1972

No.8 *Research, Teaching and Training in Demography — A Directory of Institutions in the ECAFE Region* (E/CN.11/1007): 1972

Supplement No.1 (E/CN.11/1007/Add.1): 1974
Supplement No.2 (E/CN.11/1007/Add.2): 1974
Supplement No.3 (E/CN.11/1007/Add.3): 1976

No.9** *Report of the Working Group on Training of Personnel in Family Planning Programmes* (POP/TPFP/1; mimeographed): 1972

No.10 *Report and Selected Papers of the Regional Seminar on Ecological Implications of Rural and Urban Population Growth* (E/CN.11/1043): 1973

No.11 *Population Aspects of Social Development — Report of a Regional Seminar and Selected Papers* (E/CN.11/1049): 1972

No.12 *Socio-economic Returns of Family Planning Programmes — Report of an Expert Group Meeting* (E/CN.11/1070): 1973

No.13 *The Role of Voluntary Organizations in National Family Planning Programmes — Report of an Expert Group Meeting* (E/CN.11/1087): 1973

No.14 *Comparative Study of Mortality Trends in ECAFE Countries* (E/CN.11/1108): 1973

No.15 *A Comparative Study of Family Planning Service Statistics Systems in the ESCAP Region* (E/CN.11/1198): 1975

No.16 *Husband-Wife Communication and Practice of Family Planning* (E/CN.11/1212): 1975

No.17 *Population Periodicals — A Directory of Serial Population Publications in the ESCAP Region* (E/CN.11/1176): 1974

No.18 *Some Techniques for Measuring the Impact of Contraception* (E/CN.11/1119): 1974

No.19-A *Asian Resources for a Population Library Information Network — Report of a Working Meeting* (E/CN.11/1114): 1974

No.19-B *Establishment of a Population Library and Documentation Centre Information Network among the Countries of the ECAFE Region — Report of a Task Force Meeting* (mimeographed): 1974

No.20-A *Report of the Workshop of ECAFE Population Correspondents from Indonesia* (mimeographed): 1974

No.20-B *Report of the Workshop of ECAFE Population Correspondents from Thailand* (mimeographed): 1974

No.20-C *Report of the Workshop of ECAFE Population Correspondents from the Republic of Korea* (mimeographed): 1974

No.20-D *Report of the Workshop of ESCAP Population Correspondents from Pakistan* (mimeographed): 1975

No.21 *Comparative Study of Interrelationship between Levels of Literacy and Skill and Fertility Trends* (mimeographed)

No.22 *Report of the Expert Group on Developing Indices for Measuring the Impact of Training on Job Performance of Field-workers in Family Planning Programmes* (IX.ISSN.0066-8451; mimeographed): 1973

No.23 *Comparative Study of Population Growth and Agricultural Change* (E/CN.11/1224; in five volumes): 1975

No.24 *Role of Surveys and Studies for Family Planning*

(Cont. inside back cover)

* No charge, but priority given to Governments and institutions.

ESCAP was previously named the Economic Commission for Asia and the Far East (ECAFE); hence, those titles published before 1975 make reference to ECAFE.

** Available in microfiche form.

IX ISSN 0066-8451

ASIAN POPULATION STUDIES SERIES
No. 32

# GUIDELINES FOR PREPARING SUBNATIONAL POPULATION PROJECTIONS

ECONOMIC AND SOCIAL COMMISSION FOR ASIA AND THE PACIFIC

Bangkok, Thailand, 1975

UNITED NATIONS

# PREFACE

The Working Group on Projections of Population of Subnational Areas met at Bangkok, from 14 to 23 May 1969, and recommended that the Economic and Social Commission for Asia and the Pacific (ESCAP), then known as Economic Commission for Asia and the Far East "explore the feasibility of preparing a manual on methods for subnational population projections, and organizing meetings of expert groups in this connexion, in close collaboration with other members of the United Nations family of organizations."

In pursuance of this recommendation, the guidelines presented in this volume have been prepared as a part of the work programme of the Population Division of ESCAP. The Population Division of the United Nations, in New York, has already published a series of manuals on statistical methods in population projection, which are detailed in the text below. The present guidelines refer to these manuals, and provide ready reference to methods which are appropriate for the preparation of subnational population projections, particularly for the less developed countries.

Subnational population projections are an essential tool in general planning for the development of regions within a country. This subject was prominent in discussions of the Working Group on Population Projections, which met at Bangkok from 30 September to 6 October 1975, and it will receive further consideration in the Expert Group Meeting on Population Growth and Economic Development in Subnational Areas, scheduled for May 1977.

This document was prepared for ESCAP by Walter P. Hollmann, Chief, Population Unit, Department of Finance, State of California, United States of America. ESCAP appreciates the financial support provided by the United Nations Fund for Population Activities for this project.

# CONTENTS



6

## List of tables

Page

# List of figures



9

# INTRODUCTION

The third United Nations manual on population projections[1] is a useful guide to the national planner, to the student of population, and to the analyst who seeks to understand and to anticipate the impact of population trends upon national development. There has long been a further need for a systematic set of instructions in the projection of population for such subnational areas as states, provinces, districts, or statistical or planning areas; and for specific groups of population which may overlap different areas within a country, such as the labour force, various occupation groups, the school age population, literates, or ethnic groups. These guidelines attempt to meet that need.

The guidelines below are addressed principally to the development planner faced with problems of areas or specific groups of population within a country. The planner needs figures upon which to base plans, but he is faced with a variety of constraints imposed by inadequate, inaccurate, incomplete or non-existent data. He may need only approximate numbers of people, or he may require future population composition as well. The guidelines are designed to aid him by discussing the types of population projections and the ingredients of each to help him decide (a) what his data will permit him to do, and (b) within the constraints imposed upon him, what he should do to produce the numbers required by his work.

The growing number of United Nations manuals on demographic topics form a useful small library in this field. The present document refers to them in describing techniques to make subnational population projections. It emphasizes what is unique in subnational projection, while presenting some of the elements common to population projections for any area. A demographer concerned with a single area, or set of geographically related areas, or groups of population within a country, would be well advised to devote some efforts to the other demographic manuals and literature.[2]

---

1/ Manual III, **Methods of Population Projections by Age and Sex** (United Nations publication, Sales No. 56. XIII. 3).

2/ Particularly United Nations Manual III, op. cit.; Manual V, **Methods of Projecting the Economically Active Population** (United Nations publication, Sales No. 70. XIII.2); and Manual VII, **Methods of Projecting Households and Families** (Sales No. 73. XIII.2).

10

# I. GENERAL CONSIDERATIONS IN POPULATION PROJECTIONS

## A. POPULATION PROJECTIONS AND DEVELOPMENT PLANNING

The importance of population projections or forecasts in development planning can scarcely be overemphasized. A rational plan for any kind of development cannot be formulated without at least an idea of the number of people or the institutions to be served. Ideally, future populations should be estimated in the greatest detail possible. The magnitudes of future educational or dependency problems are reflections of the relative size of the younger and the older age groups. The need for additional employment opportunities is directly related to the size of the future labour force, mainly a function of the populations in working-age groupings, and investment in the development of human resources. The future food requirement, assuming stable environmental conditions, is a function of the age and sex groupings of the entire population.

The preceding sentences might imply that population is an independent variable, and that all other variables included in the planning process are dependent. Such is not the case; nowhere is that more apparent than at the subnational level. The development of a facility to meet a need assuredly affects the population size, especially where migration is subject to few restraints. Family planning programmes are of course developed in the belief that they affect future population size through reduction in the level of fertility. Public health programmes of all kinds, designed to improve the health and welfare of people, can be expected to reduce the mortality level. In short, all developmental activities, by affecting levels of migration, fertility and mortality directly or indirectly, can bring changes in population size, growth, composition and distribution.

It is apparent that a complex interrelationship exists between future levels of population and the many actions which constitute the implementation of plans. If the demographer treats population as if it were an independent variable, he does so for convenience. The prudent analyst is ever ready to respond to the planning decisions which his data generate. Only by such feedback can his efforts serve as flexible tools in the planning process, rather than as exercises which become obsolescent as soon as the first responsive planning action is taken.

## B. PROJECTIONS AND FORECASTS

Although no responsible demographer would claim the ability actually to predict the future, projections and forecasts are an attempt to describe what would happen "if". In a projection the "if" is followed by explicit assumptions about the future, related to the immediate past. The Multilingual Demographic Dictionary[3/] distinguishes between forecasts and projections as follows:

> "Population projections are calculations which show the future course of fertility, mortality and migration. They are in general purely formal calculations, developing the implications of the assumptions that are made. A population forecast is a projection in which the assumptions are considered to yield a realistic picture of the probable future development of a population. Generally speaking, these are short-term forecasts, as the margin of error to which they are subject increases considerably with the space of the forecast."

The definition seems to relegate projections to a realm of purely theoretical interest, while forecasts enjoy the advantage of being based upon assumptions designed to yield a "realistic" picture of the future. The distinction seems more apparent than real; whatever the method, the demographer or planner must select carefully an assumption or set of assumptions concerning future changes in the variables. If the selection has been careful, the assumptions should describe conditions which might reasonably be expected to occur. Typically, the planner specifies a set of assumptions leading to series of estimates of future population that would vary from a maximum to a minimum for comparative purposes, indicating the range within which the planner is working.

Were the planner to select a single "most likely" set of assumptions, he would, according to the definition, produce a forecast. Then the term "projection" might properly refer to estimates of future population reflecting the effects of a suggested public health programme, or a possible new policy on migration, or of a family planning proposal designed to avoid a certain number of births. Once the proposed policies or activities were adopted and faithfully implemented, the

3/ (United Nations publication, Sales No. 58. XIII.4), para. 720, (English section).

term "forecast" might be deemed appropriate. A document such as the present one, having concrete goals, must emphasize projections and leave the reader to decide when the nature of his work permits application of the other term.

## C. METHODS OF PROJECTION

Some of the main projection methods are the "mathematical" method, the "economic", the "component", and a miscellaneous class.[4] The distinction is for convenience, and does not imply that the methods are mutually exclusive.

"Mathematical" projections generally are concerned only with total populations, and project total population change as a function of time, on the basis of recent or long-term observation. The functional relationship between population change and time may be as simple as a straight line plotted on linear or semi-logarithmic paper, or it may embody the logistic growth curve. The latter describes a course of population change as an exponential function, shaped like the letter "S", beginning with a growth rate which gradually increases and reaches a maximum in the middle of the projection period, from which it declines as a linear function of the population. The curve approaches an asymptotic value as the population reaches a maximum within the constraints of a limited universe. This is an idealized form of projection; in general, the actual growth of population does not take place in such a smooth fashion.

Other mathematical methods project the ratio of a subnational population to the whole or subnational components of change to the components of change of the whole; or they may allocate total future growth among the parts.

If the mathematical approach treats population as the independent variable, to the practitioner of the "economic" approach it is the dependent variable. In constructing a model, the econometrician develops assumptions concerning future needs for food, raw materials, manufactured products, and services in the total market area. Following assumptions regarding changes in worker productivity and other variables, a future labour force for the area of interest is projected, and on the basis of assumed levels of labour force participation, a derived population is calculated as an end product.

Such an approach has the distinct advantage of directly or indirectly taking more account of the inter-related factors than do other methods. It also suffers two short-comings, one of which is especially critical. The total population is derived from labour force figures through the inverse use of participation rates; however, during rapid social and technological change, variations in participation rates pose projection problems of their own. Furthermore, the data requirements for the economic approach are probably even more stringent than those for other methods.

Another approach treats population as a dependent variable; it derives the future population from a projected pattern of land use. This approach would be quite effective for small-area projections if the pattern of land use could effectively be controlled through zoning, or if it could at least be reliably projected. If the planner has sufficient evidence from the development plans to formulate assumptions of the future pattern of land use, he can assign ultimate and appropriate densities to each zone. It is, however, not likely that such an idealized form of planning and projecting would find more than very occasional and limited applicability.

In cases where no basis exists for estimating rates of population growth or births, deaths and migration, but where a population projection is required, the "analogy" method can be applied. Demographic rates are transposed from areas where social and economic conditions are assumed to be comparable with those of the area under study. Model life tables and model stable populations could also be used in estimating demographic parameters, if some local data were available. The growing body of literature[5] on the estimation of demographic measures from incomplete data may suggest means for establishing a basis for projections. But the importance of migration, age/sex selective underenumeration or overenumeration, and misreporting of ages in subnational areas make the application of such measures imprecise, if not impossible.

The "component" ("cohort survival", or "cohort-component") method of population projection is the most widely used where essential data are available. Because it takes account of future changes in its parts or components, and because under ideal conditions these components can be specific for such details of composition as age, sex, or ethnic group, a degree of "control" is maintained over the calcula-

4/ See Manual III, op. cit., pp. 2-3.

5/ See **The Concept of a Stable Population: Applications to the Study of Countries with Incomplete Demographic Statistics** (United Nations publication, Sales No. 65. XIII.3); and Manual IV, **Methods of Estimating Basic Demographic Measures from Incomplete Data** (United Nations publication, Sales No 67. XIII.2).

tions which should improve the validity of projections.

Reduced to its essentials, this method usually starts with a benchmark population distributed by age and sex. Mortality rates are applied to each age and sex cell and age-specific fertility rates appropriate to each age are applied to the female cells. The migration component is included either by the application of rates of migration specific for each age and sex grouping, or by the inclusion of projected numbers of migrants, i.e. negative for out-migrants and positive for in-migrants, appropriate to each age and sex grouping. In this manner, the population after each one-year or five-year phase in the projection period is calculated from the one preceding.

A separate projection of the migrant population may be desirable, where the migrant population is large and significantly different from the original population in ethnic composition or in such social characteristics as marital status, generating different demographic behaviour. In practice, the future assumptions with regard to the fertility and mortality behaviour of the population are based upon prior events, which in turn reflect the inclusion of migrants in the past. Unless a radical change is anticipated, it is unlikely that such a refinement would improve the projection; and the additional data required would cast additional doubt on its advisability.

The planner who chooses the cohort-component technique is typically faced with the responsibility of formulating assumptions concerning the future patterns and levels of fertility, mortality and migration. If he is equipped with data on the current levels, he must state assumptions of stability, or of a measured increase or decrease with perhaps ultimate values at a selected date. It is usually desirable that a high projection, a low projection, and an intermediate or "most likely" one be developed.

In recent projections developed for the State of California, United States[6], for example, annual net migration levels of zero, of 100,000 and of 150,000 per year were assumed; for each of these, different fertility levels were selected. Age-specific birth rates were used, resulting in completed fertility of 2.11, 2.45 and 2.78, respectively, and mortality by age was assumed to remain nearly constant. The effective age composition of the net migrant population was derived from recent experience by a residual method, i.e. subtracting the population change, by age, which was attributable to natural increase by age derived from the difference between censuses. In all, four sets of calculations were made by computer for each single year of age and sex groupings, for each of 58 counties, for each calendar year to the year 2020. The very detailed sets of state and county population projections are used in development planning, especially for water resources. Recently increased concern for the environment and the future "quality of life" has stimulated greater interest in such projections.

## D. DATA REQUIREMENTS AND AVAILABILITY

In any specific instance, the method selected depends mostly upon the data available. The Working Group on Projections of Population of Subnational Areas concluded that, where possible, the cohort survival method should be used. If this is to be the ideal, certain considerations of data are in order. Data for projection purposes fall into three broad classes: (a) benchmark or baseline population data; (b) vital statistics, and other regularly collected administrative data which can provide measures of components of population change; and (c) data on births, deaths, migration and population through sample registration and sample surveys. In case such data are not available for the subject area, analogical data from other areas might be available. These should be in the degree of detail desired in the subnational projections.

Benchmark data should contain at least as much detail as the final projection is to include. Typically, the census population should be known by sex and by five-year or smaller age group. Moreover, other details which are required in the projection, or which might shed light on likely demographic changes, should be available. These may include urban and rural residence, marital status, ethnic composition or economic status.

Vital statistics should include births by age of mother, and deaths by age of decedent. Additional information such as birth by parity, rates of marriage and breakdown of the foregoing for urban and rural residence would increase the utility of the data for projection. School enrolment statistics, although lacking in complete coverage in many countries of the region, are still of potential value. Furthermore, it is highly desirable to have a comparison, through time, of vital statistics and the rates calculated from them, to facilitate the understanding of recent changes. Unfortunately, development of adequate vital statistics for much of the less developed world re-

6/ California State Department of Finance, "Population projections for California counties, 1975-2020", (Sacramento, June 1974).

quires several decades.This short-coming underlies the selection of the methods described in this document.

Sample surveys have proven useful in providing estimates of variables which are not available from other sources. The data collected are subject to the errors of memory which are characteristic of retrospective information, but useful statistics can be assembled in this manner. For example, the annual number of births by age of mother and deaths by age of decedent, for the past five years in the sample households, might be gathered. Information on residence one or five years earlier, tabulated by age and sex, would yield insight into the migration pattern of those who had not departed prior to being surveyed; this consideration recommends the shorter time frame. If the question is asked within an area which experiences little external migration, very good internal migration data for subareas can be assembled by tabulating information on out-migrants as well as in-migrants. Data on age, sex and marital status permit the calculation of fertility rates, while parity data would make possible the estimation of completed fertility. Information on household size would be of use in the preparation of current estimates, as would information on almost any commodity consumed or service required which is available in administrative statistics for the sample universe.

Of particular interest is the measure of migration. Where registration figures exist, they may be useful; where participation is high, school enrolment may prove of utility. Migration at the subnational level, if it is principally internal, is relatively unhampered and thus poses problems in measurement, both in net volume and in disaggregation by age and sex.

In general, it is recommended that the first two classes of data listed above, the benchmark and that of vital events, be improved in completeness and accuracy, using the third class. Not to be neglected are the general measures of completeness of coverage, the mathematical means of reducing age-heaping, and studies of systematic underenumeration. Before beginning the task of projection, the benchmark should be corrected so that it is as accurate as possible.

## E. REVISION OF PROJECTIONS

So complex are the causalities in the broad field of development planning, so numerous are the determinants of population (and of every other) change and their consequences, that it is not possible to impute independence to any variable. As mentioned earlier, it is for convenience in many or most projection schemes that population is considered the independent variable, but the practitioner must be alert to the probable effect of the implementation of development schemes on demographic variables used as bases for population projections. When conditions have changed, or when new programmes have been introduced of magnitude sufficient to require different assumptions concerning future births, deaths and migration or future growth patterns or rates, then the projections are revised, or a new forecast or set of projections is prepared.

Whatever the method finally selected, the projection process should be regarded as a part of development planning. It may comprise the first survey of the magnitude of population-related needs for the development planners. As the plan proceeds, as recommended schemes are implemented, one or both of two observations are made. The first of these is that the assumptions initially formulated are not the best possible, the relationships between population and its dependent variables are not what was assumed. The second is that the plan itself affects population change. In either case, with the passage of time and with the improvement of the data base, new projections would be demanded by a changing situation.

In any case, since projections of the type under consideration have rarely predicted the exact course of the projected population variable, they should always be regarded as provisional. It is necessary to revise such projections at appropriate times. While a more limited revision is sometimes sufficient, revision usually entails the preparation of a new projection on a later base point and with more up-to-date assumptions. During revision, methodological improvements may be introduced either through greater availability of data or greater experience in empirical studies with alternative methods.

The frequency of revision depends on the degree to which new information changes the apparent validity of projections. The decision to revise depends not only on this consideration, but on a judgement whether the apparent error is due to short-term deviation from the trend. Projections which do not account for planning results, or effects of planned change in the area under study, have been called "autonomous". Those which do involve such reflexive responses have been called "feedback" projections.[7] (See footnote 7 next page)

## F. TIME SPAN OF PROJECTION

Just as there is no way to determine in advance the useful life of a projection, there is no standard time span over which projections should be made. Population projections for the world to the year 2100 have appeared, usually with alarming implications. The above-mentioned California projections use 2020 as a final date, because the major water project for which they were originally prepared requires 50 years. Typically, population projections for educational facilities use a ten-year time frame, presumably a sufficient period to plan, fund and build an educational installation.

In cases of less elaborate requirements for capital outlay, a five-year projection span may be adequate, while support budgeting may require a projection of only one or two years. A useful principle in selecting the span of a population projection is that it should equal but not necessarily exceed the maximum length of time required for completion of the facilities planned.

Projections over a longer period may prove useful in the demonstration of more distant problems, and indicate the need for early planning. The decision should be made on the merits of each case, bearing in mind that the longer the time span, the greater the potential deviation of the projected from the actual population.

## G. CONCLUDING REMARKS

In conclusion, the selection of a method and the frequency of revision depend on local circumstances. A decision reflects the initial availability of data and trained personnel, the speed of implementation of development plans, especially those which have a significant impact upon population, and the progressive improvement of the data base. The latter leads not only to a more varied and reliable input to the projection process, but also to selection of more promising methods. It was the consensus of the Working Group on Projections of Population of Subnational Areas that each country should strive to use the cohort survival method if the essential support data could be assembled. If not, alternative techniques should be selected, and as much effort as possible be directed toward improvement of the data base. Chapters V and VI below describe the preparation of projections appropriate to varying data availabilities. Chapter V discusses methods which can be described as "mechanical" since they require minimum detail, while chapter VI is concerned with "analytic" methods, those which consider changes in the age and sex divisions within the population with the passage of time.

---

7/ See Manual VIII, **Methods for Projections of Urban and Rural Population** (United Nations publication, Sales No. 74. XIII.3), p. 3; which recommends submission of tentative autonomous projections to planners for their consideration. On the basis of such consideration, revised assumptions should be used in producing feedback projections. If time and administrative arrangements permit, this would yield an initial projection of maximum reliability.

# II. SUBNATIONAL AREAS AND POPULATION PROJECTION

## A. DEFINING "SUBNATIONAL"

The term "subnational" refers to any area smaller than the whole country of which it is a part; typically a region, state or province, or a district or village. A country may be divisible into various categories of area such as administrative, ecological, cultural, economic, urban or rural. Each of these types might have to be treated in isolation to meet the special needs of social and economic planning. Although the choice of type and level of subnational area would depend upon the needs of administration and planning, the poor availability of data in sufficient detail for a well defined geographic area is a major constraint in the preparation of population projections.

The definition of urban area, for example, is a common problem between countries. This definition is not only based on different criteria between countries, it varies within each country over time. With the growth of population and annexation of surrounding agglomerations of population, boundaries of urban areas expand. Thus, an area may be classified as urban without achieving the qualitative standards of urbanization.

## B. NEED FOR DELINEATING CENSUS TRACTS

Usually the political subdivisions of a country are far from uniform in size or characteristics; they can be and frequently are far too large and heterogeneous for convenient use as planning units. In the case of Thailand, for example, the Bangkok-Thonburi urban complex contained about 3 million inhabitants in 1970, being 8 per cent of the population of the nation. The next largest city, Chiangmai, had barely 100,000 inhabitants, or less than 4 per cent of those in the capital city; there were no other cities within the 100,000-to-999,999 class at all. Although other countries may exhibit more even distributions, the need for more useful units for comparison is nearly universal.

The development of population data for planning would be made substantially easier if each country were divided into tracts of 5,000 to 10,000 inhabitants bounded by natural or otherwise easily recognized and permanent features such as rivers, major highways, canals, railways and the like. Ideally, insofar as possible the boundaries of the tracts should include populations which are more or less internally homogenous. That is, they should be designed to include within their respective boundaries, inhabitants having common characteristics—ethnic, socio-economic, occupational— which distinguish them from inhabitants of other areas.

By drawing boundaries in this manner, interesting and useful interarea comparisons are possible, e.g. the differential effects of specific programmes on two or more types of population. Such information might prove critical in forecasting the probable results of development programmes, if the composition of the subject population is known.

Furthermore, if boundaries of such areas are kept constant through time, longitudinal studies are possible, enabling the researcher to compare the level of development of an area at the time of one census with that at another time. To insure maximum utility, and especially to provide detailed data, on political subdivisions for administrative purposes, boundaries should coincide with those of the cities and provinces in a manner permitting aggregation of the parts to the whole, using the small areas as building blocks.

Metropolitan areas in the United States, for example, are almost totally divided into census tracts which approximately meet the criteria described above. The initial laying out of the tracts is a local responsibility, usually borne by planning agencies in local government, country or municipal. To insure comparability through time, vigorous attempts are made to keep boundaries intact from census to census; tracts are divided as population growth warrants, frequently during the enumeration process, by making "field splits". Where a city boundary crosses a tract, population data for the portion of the tract within and outside the municipality are tabulated separately to permit separate aggregation of city data.

So far, attempts to make population projections for the census tracts are very infrequent, although there are increasing efforts to assemble non-census administrative data by tract, especially in some larger cities. Of particular interest is the development of the "DIME" (for Dual Independent Man

Encoding)[8/] technique which permits the geographic matching of data files by use of map co-ordinates or "nodes", a system which will enable the planner to compare, by computer, data files by address without reliance upon prior planning areas; and to aggregate, by tract or other area, information from any file of cases for which street address or other location data are available.

Beginning with the census of 1970, Japan has published details of age and sex characteristics for census tracts averaging 10,000 inhabitants, of cities exceeding 200,000 people and of prefectural capitals. Current plans do not entail changing the boundaries of the tracts for censuses, although where significant growth occurs, tracts will be split. In this manner, as in the United States, comparability should be maintained through time, while split tracts insure the continued availability of geographically fine-grained detail.

## C. PROJECTION OF CENSUS TRACTS

Although the census tract is a useful demographic artifact, it is far from essential to the preparation of subnational population projections. In fact, its use in population projections is seldom possible within the capacity of current data collection and tabulation systems. For example, Australia did not use the concept in its census of 1961, yet age and sex data were published for municipalities outside the metropolitan areas of Sydney and Melbourne, the majority of which were less than 10,000 in population. Subnational projections for several of the six constituent States of Australia have been prepared.[9/]

## D. PROJECTION OF POLITICAL SUBDIVISIONS

Use of existing political subdivisions as subnational areas or building blocks for larger subnational areas is quite a satisfactory application, providing only that the continuity of data for geographically equivalent areas can be ensured. The comparison of data through time is very difficult if annexations, new incorporations, or other boundary changes result in place names with different meanings from census to census. Similarly, a projection for a subdivision can only obtain for the area as it was described at the base year. Future changes in population incident to annexations of adjacent territory are ordinarily not provided for in projection work. Whether he uses census tracts or political/administrative subdivisions, the planner should have age and sex data for the selected subnational area; the logical first step is the decision as to what "subnational" denotes in the context of the problem area. What area does he wish to or is he able to use in planning operations?

## E. PROBLEMS OF BENCHMARK DATA

The ideal data set would include a census benchmark, recent measures of birth and death behaviour from vital records, and a basis for formulating assumptions about future migrations, however, these data are seldom at hand. Few countries in the ESCAP* region possess vital registration systems furnishing data of sufficient accuracy to use in projecting without considerable adjustment. The analyst must decide whether his data can be corrected for underregistration or misreporting of age, whether he should estimate vital rates from census or other data, or whether he should use a method which does not require vital rates. Techniques are available for all three courses of action. These guidelines emphasize the third alternative.

Because of its low magnitude, international migration is generally of little significance in projecting national populations. But, if the influx of immigrants or the exodus of emigrants is concentrated in particular subnational areas, international migration may be important in projecting populations of subnational areas, even though it may be small on a national scale.

Since most countries require travellers crossing national boundaries to declare themselves, data on international migration are collected at the point of entry or departure. Thus, up-to-date migration figures should be available, of a quality commensurate with the effort at the borders to ensure registration of entries and exits. But in several countries of the ESCAP region, there is unrestricted back-and-forth movement of large, nomadic populations across national boundaries, rendering the records highly incomplete. In some countries there are unauthorized entrants from neighbouring countries. In recent history, high levels of international migration have coincided with periods of unusual stress, incident to wars or to mass repatriation. Mass movements of this nature generate the very conditions that make for poor registration. Also, most countries in the region do not publish data on international migration, or data are published with insufficient detail.

8/ United States Bureau of the Census. *Census Use Study: The DIME Geocoding System*, Report No. 4, Washington, D.C.

9/ W.D. Borrie, *Population and Australia: A demographic analysis and projection.* (Canberra, Australian Government Publishing Services, 1975), pp. 255-314.

With the above exceptions, international migration is not an important factor in the population change of the less developed countries in the ESCAP region. Far more relevant to subnational projections is the level of internal migration.

Since in general there are no restrictions on the internal movement of population, it is difficult to maintain registers on internal migration. There is also a problem of definition of migrants, and of boundaries of subnational areas within a country. Since internal migration has an important bearing on subnational population projections, chapter III is entirely devoted to these problems.

The preceding paragraphs are an attempt to pose the problem of benchmark data for subnational projections, and to describe very briefly the data needed. The development planner or the analyst who prepares subnational population projections must recast the problem in terms of the type of area or areas selected. The quality and availability of data among ESCAP countries are extremely diverse, as are the magnitudes of the populations involved. The planner is constrained by the data available to him; the less substantial the data, the greater is the challenge to be faced, and inevitably the less rich in detail is the finished product.

Table 1 below collates data from the countries of the ESCAP region regarding the most recent censuses, vital statistics, and population. Some recent statistics on age and sex composition are available for most countries of the region. Such breakdowns are usually available in the less developed countries only for large subnational areas, such as provinces or states. In some cases, data in broad age-groups are available at the district level. As the amount of error is inversely associated with size of area, finer breakdowns for small areas are generally avoided.

The table includes the less populous countries where future residential population, by small area, is of limited importance in development planning. Table 1 also shows that there are seven countries for which data as recent as 1966 do not exist, or which never had a population census; for several, no data suitable for subnational projection benchmarks exist. Also included in the table are the estimated mid-1975 population, and the size-rank of each country within the region, as an indicator of the relative importance of deficiencies in data. In addition, an indication of the completeness of vital registration statistics is shown: "C" stands for "complete", where an estimated 90 per cent of events are recorded; "I" and "-" indicate that they are seriously deficient, or that no vital data are available.

For 93 per cent of the population of the 36 ESCAP countries there are incomplete or no vital data, including the heavily populated People's Republic of China and India, while only 7 per cent inhabits the countries which have usable vital data. More description of the source of birth and death statistics and their measurement problems are given in chapter IV.

In summary, subnational projections can be prepared for most countries from recent benchmark data, although vital statistics may not be of satisfactory quality and the level of disaggregation of the whole may not be quite adequate for development planning. How to accomplish this is the concern of the rest of this document.

Table 1. Population censuses and vital registration statistics of countries in the ESCAP region.

| Country | Most recent census | Est. population mid-1975 (in thousands) | Vital registration a/ | Population size rank |
|---|---|---|---|---|
| Afghanistan | - | 19,280 | I | 14 |
| Australia | 1971 | 13,809 | C | 17 |
| Bangladesh | 1974 | 73,746 | I | 5 |
| Bhutan | - | 1,173 | - | 27 |
| British Solomon Is. | 1970 | 187 | I | 29 |
| Brunei | 1971 | 147 | C | 31 |
| Burma | 1973 | 31,240 | I | 11 |
| China, People's Rep. of | 1953 | 838,803 | - | 1 |
| Cook Islands | 1971 | 25 | C | 35 |
| Fiji | 1966 | 577 | C | 28 |
| Gilbert and Ellice Is. | 1968 | 66 | - | 34 |
| Hong Kong | 1971 | 4,225 | C | 21 |
| India | 1971 | 613,217 | I | 2 |
| Indonesia | 1971 | 136,044 | I | 3 |
| Iran | 1966 | 32,923 | I | 10 |
| Japan | 1975 | 111,120 | C | 4 |
| Khmer Republic b/ | 1962 | 8,110 | I | 20 |
| Korea, People's Dem. Rep. of | 1944 | 15,852 | - | 15 |
| Korea, Rep. of | 1975 | 33,949 | I | 9 |
| Laos | 1973 c/ | 3,303 | - | 22 |
| Malaysia | 1970 | 12,093 | I and C d/ | 19 |
| Mongolia | 1969 | 1,446 | I | 26 |
| Nauru | 1971 | 8 | C | 36 |
| Nepal | 1971 | 12,572 | I | 18 |
| New Zealand | 1971 | 3,031 | C | 23 |
| Pakistan | 1972 | 70,560 | I | 6 |
| Papua New Guinea | 1971 | 2,716 | I | 24 |
| Philippines | 1970 | 44,437 | I | 7 |
| Singapore | 1970 | 2,248 | C | 25 |
| Sri Lanka | 1971 | 13,986 | C | 16 |
| Thailand | 1970 | 42,093 | I | 8 |
| Tonga | 1966 | 128 | I | 32 |
| Trust Territory of the Pacific Is. | 1973 | 117 | - | 33 |
| Viet-Nam, Dem. Rep. of | 1961 | 23,798 | - | 12 |
| Viet-Nam, Rep. of e/ | 1960 | 19,653 | - | 13 |
| Western Samoa | 1971 | 164 | I | 30 |

Population of ESCAP region = 2,186,846
Population of countries with usable vital data = 149,176 (6.8 per cent, excluding Malaysia)
Population of countries with incomplete or no vital data = 2,037,670 (93.9 per cent, excluding Malaysia)
No. of countries with complete data = 10
No. of countries with incomplete data = 25
No. of country with some complete and incomplete data = 1

Sources: United Nations Statistical Papers, Series A, Volume XXIII, No. 4, pp. 16-21, 24-27. United Nations, Asian Population Study Series No. 28, table 1.

Notes : a/ "C" means complete or nearly complete; "I" means incomplete; and "-" means information not collected or not available.
b/ Now Cambodia.
c/ Partial census of the Plains of Vientiane.
d/ West Malaysia has complete registration.
e/ Now Republic of South Viet-Nam.

# III. MEASUREMENT AND ESTIMATION OF INTERNAL MIGRATION

Of the components of population change, i.e. births, deaths and migration, the most difficult to estimate and to project is usually the migration component. The problem becomes especially acute in the case of subnational projections, as migration is frequently the major cause of subnational population change, especially in areas most rapidly gaining or losing population. This problem is seldom experienced in working on population on the national level, since migration in and out of the country is usually minor and may be documented. For a subnational area closed to migration, projections can be made with reasonable accuracy, but may yet be limited by inaccuracies of the benchmark data, weaknesses of the fertility and mortality assumptions, and events such as natural catastrophes or sudden economic changes that would affect demographic trends. The techniques of subnational projection are conditioned by unusual demographic changes. Owing to unique features of such areas, different data input may be required for each case.

## A. DEFINITIONS AND SOURCES OF DATA

Anyone who changes residence is termed a "mover"; a mover is referred to as a "migrant" when he changes his residence from one "migration-defining area" to another. A migration-defining area is a political or other subdivision of a country for which it is possible to measure or to estimate migration. Population projections for areas too small or too specialized to be migration-defining cannot practically be made by a component method. An "in-migrant" is defined in the United Nations Manual VI [10] as "a person who enters a migration-defining area by crossing its boundary from some point outside the area but within the same country. He is to be distinguished from an 'immigrant' who is an international migrant entering the area from a place outside the country". An "out-migrant" is defined as "a person who departs from a migration-defining area by crossing its boundary to a point outside it, but within the same country. He is to be distinguished from an 'emigrant' who is an international migrant, departing to another country by crossing an international boundary". "Gross migration" refers to total in-migrants and total out-migrants; the difference between them is termed "net migration".

10/ *Methods of Measuring Internal Migration* (United Nations publication, Sales No. 70 XIII. 3), p. 3.

Since it is the aim of this document to show the means of projecting the population of subnational areas, and since it is a resident population figure which underlies most but by no means all of development planning, a comment on residence is in order. Whether ultimately tabulated **de facto** or **de jure**, most censuses regard an individual's residence as the place where he sleeps most of the time. In the United States, for example, where the census is mainly **de jure**, a "T" (for "transient") nighttime survey is held during the enumeration period in order to count those in hotels and other transient quarters, so that they may be allocated to their regular area of residence. A single residence must be established for those who may occupy two or more dwellings, on the basis of where they sleep most of the time. Those with no fixed abode are tabulated on a **de facto** basis.

Thus, if a person almost anywhere in the world spends five or six nights each week in the city but returns to his home village for one or two, he is counted as a resident of the city, although he may vote, pay some taxes or till the soil in the village. Similarly, an individual may work in one locality for a majority of the months of the year but regularly return to a seasonal residence elsewhere; he is usually regarded as a resident of the place in which he has lived most of the months, provided he intends to return at the time he is counted. Nomadic people or those who migrate seasonally without fixed abode are typically counted where they are found on the census day.

The foregoing residence considerations are common but by no means universal. Since it is the purpose of population projections to anticipate what a census would find at a future date, the definitions used in the national census should be retained to the maximum extent possible in estimates and projections for that country. Anyone attempting projections for a subnational area should be informed of the residential definitions in force, and whether they are consistent from census to census.

Data on internal migration come from a number of sources, the most important of which is census information. Most censuses include items which facilitate classification of the population into migrants and non-migrants, although the definition of these terms is, within the census context, highly variable. The distinction may be made on the basis of whether an individual has moved into the area at some time in his lifetime, or within a specified period such as one or five years. For example, publications of the 1970 census of Thailand include, for each of the 71 Changwats (provinces), the Changwat of residence of its inhabitants five years earlier; where the numbers are small they are aggregated into an "other" category which precludes construction of a complete matrix of inter-provincial migrants. A recent report of the United States Bureau of the Census 11/ contains a matrix of the migrants from each of 510 economic areas to every other area, as defined by a census question on residence of the respondent five years prior to enumeration.

In a few countries, specialized population registers show promise as sources of data for internal migration. Administrative statistics may reflect changes of residence for registrants or license-holders of particular kinds, shedding light on the migration of special-category populations.

Sample surveys have been used increasingly in the development of postcensal and intercensal demographic information. Periodically, a typical survey should elicit information for measuring migration. Often, the size of the sample precludes its use in estimating migration into or out of the smaller subnational areas, but it may yield clues to changes in migration levels estimated at other times from other data.

A few national sample surveys have collected data on place of birth, specific for small areas such as state or province. The National Sample Survey of India, in its ninth round, inquired into the "native place" of members of the labour force, being defined as that area in which the forebears of the respondent lived and with which he had some connexion. The United Nations Demographic Centre in Santiago 12/ and the International Institute of Population Studies, Bombay 13/ have conducted intensive local migration surveys. A sample survey of in-migrants to Seoul has been conducted by the Institute of Population Problems of the Republic of Korea. 14/

## B. MEASUREMENT AND ESTIMATION

The estimation and projection of internal migration is usually most critical in applying the component method to prepare current or intercensal estimates, and forecasts or projections. Simple net migration without detail is an adequately disaggregated component, where no breakdown of population composition is to be estimated or projected; however, it is not so where age, sex or other categories are required.

### 1. Direct measurement

Typically, the population of a nation at time t is expressed through a balancing equation,

$$P_t = P_0 + B - D + M_I - M_E$$

where $P_0$ is the population at the beginning of the period; B and D are births and deaths occurring to residents of the area for which estimates are being made between the beginning and the end of the period; $M_I$ and $M_E$ represent the numbers of migrants moving into or out of the country, respectively, during the same period. This formula involves data on births, deaths, immigration and emigration. The same equation can be applied to subnational areas, if data on births, deaths, in-migration and out-migration are available. It can be expressed as

$$P_t = P_0 + B - D + M_i - M_o$$

where $M_i$ and $M_o$ represent in-migration and out-migration, respectively, to and from the subnational area.

Internal migration is seldom measured directly through this balancing equation because of the non-availability of data.

### 2. Indirect measurement

There are examples in censuses and surveys concerning prior residence at some point in the past, such items as, "where did the respondent live five years ago?"

Adequate responses to such questions permit the construction of a matrix of prior and present residence, by age, of all persons. Disregarding age and sex, such a matrix would have N x (N + 1) cells, where N is the number of subareas in the country and (N + 1) represents the inclusion of those who lived outside the country at the beginning

11/ "Migration between economic areas", *Census of Population. 1970: Final report*, PC(2), pp. 85-322.

12/ Juan C. Elizaga, "A study of migration to Greater Santiago (Chile)", *Demography*, 3(2) 1966, pp. 352-377.

13/ K.C. Zachariah, *Migration to Greater Bombay*.

14/ Institute of Population Problems, "The survey report on fertility and migration of Seoul City", *Journal of Population Studies*, 3:3-85, (Seoul, 1966), pp. 145-152, (English Summary).

of the reference period. Since censuses are national and not world-wide, the corresponding column (or row) representing those originating in the subareas whose destination was outside the country, cannot be constructed. The data become more problematical where an age/sex distribution, of those leaving each subarea and those arriving, can be constructed from responses to the question. Obviously the record must be restricted to those born before the reference date, the "population five years of age and over", where the time span of the question is five years, for example. If this type of question was asked, a series of gross migration matrices can be constructed, one for each age and sex group. By subtraction of those departing from those entering each age/sex cell for each subarea, a corresponding net migration table can be constructed for any or all subareas within the system. If net migration estimates are available from other indicators, they could be utilized in an estimate of net migration by age and sex, if adequate provision is made for that portion of the migrant population not born at the beginning of the migration period.

Frequently the population of a migration-defining area may experience net in-migration of some age and sex groupings while being affected by net departures of others. Large cities, for example, usually attract young, childless adults, while families with children may move to suburbs. Areas dominated by extractive industries such as lumbering may attract adult males, and supply other areas with younger adults, mainly females. The population pyramids which follow illustrate the effects of prolonged migration upon the age and sex composition of two areas. The pyramids graphically represent the distribution of the population of two revenue districts of Sri Lanka in 1953. The Colombo district comprises the capital city which has continuously served as a destination for migrants, mainly male. Galle, the other district was clearly a source of migrants. For comparison, the composition of the entire country is shown in the superimposed dotted pyramids.

An estimated age composition of the net migrant population is of considerable utility in making population projections. In the typical component projection, where each cohort is handled separately, and where rates of fertility and mortality appropriate to each age and sex group are applied, a level of future net migration must be explicitly assumed. An age distribution of net migrants from recent experience within the area in question is probably the most valid basis for allocating the assumed future net migration among the cohorts.

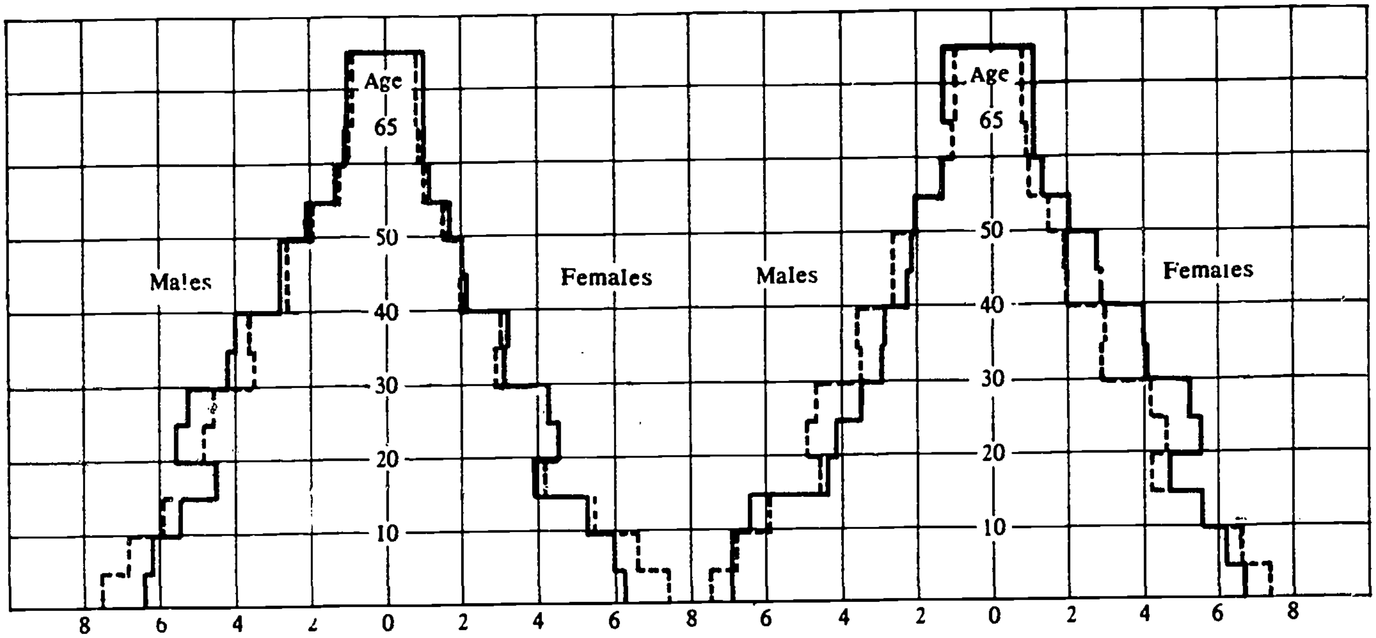
In practice, compromise with the ideal is often unavoidable, but the choice of an age/sex distribution for an assumed level of future net migration should be based upon the most comparable or relevant statistics available. If data for the whole nation is used, in the absence of local data, the subnational projections must take into account the unique factors which might characterize the composition of migrants into or out of the subject area as against the national population. Where the level of migration is low, and where future changes in the population are expected to depend almost entirely upon natural increase, it may be feasible to ignore the effects of the net migrants upon the age and sex composition.

One indirect measure of migration frequently used by demographers is based on a residual method, whereby the population change between two censuses not attributable to the difference between births and deaths is assumed to be migration. To apply this method, two accurate and complete censuses are required. The individual age/sex groupings or cohorts are aged from the earlier census by application of survival factors; the "expected" survivors, including the survivors of those born after the earlier census, are compared with those enumerated in the later census. The difference between the expected numbers and the enumerated numbers is assumed to be the net migrants. If annual population estimates are available, annual numbers of migrants and rates of migration may be calculated, after allocating the migrants to the individual years of the intercensal period either evenly, or (preferably) proportionally to independent estimates of net migration of all ages, prepared independently from administrative data which reflect migration.

Migration of children enrolled in elementary school can be measured by annual comparison of children enrolled in a span of grades, e.g. third through eighth, with those in an equivalent span of grades one grade lower, e.g. second through seventh, one year earlier. Where enrolment in elementary school is substantially less than total for the appropriate ages, the measure may be subject to distortion, as enrolment variations reflect changes in law, in compliance or in economic conditions. It is worth noting that the migration of school-age children is typically accompanied by the migration of adults, especially females, of ages appropriate to parents of those children but whose moving is not usually reflected in other administrative records.

The migration component appears at its simplest in the balancing equation referred to earlier. Since data on in and out-migrants are generally not

**Figure 1. Age and sex pyramids for Colombo and Galle Revenue Districts, Sri Lanka, 1953: showing percentage distribution of each age/sex grouping.**



available, the same equation may be modified to estimate net migration indirectly, expressed as

$$M = P_t - P_o - B + D$$

where M refers to the net in or out-migration between the two censuses; $P_t$ refers to the population at the time of a second census; $P_o$ represents population of an earlier census; and B and D refer respectively to births and deaths during the intercensal period.

Between the census taken in Singapore in June 1957 and that in June 1970, about 753,900 live births and 135,900 deaths were recorded. The population at the time of the later census was 2,074,507 while at the earlier it was 1,445,929. From the balancing equation, net migration for the 13 years can be calculated at about 10,575. It is not possible, on the basis of the data presented, to estimate how the 10,575 are distributed over the 13 years, although administrative statistics such as for school enrolment might be helpful. There is no reason to believe that positive net immigration at a uniform rate took place throughout the 13 years; an examination of the annual numbers of apparent school-age migrants from a comparison of enrolment figures would be a basis for a proportional allocation of the 13-year total. Of course, Singapore is not a subnational area but a small, independent city state. International migration there is significant enough to provide an illustration of the present subject.

It has been possible in areas with several parallel statistical series to construct regression equations from two or more independent variables, with migration as the dependent, but the results of these efforts must always be reconciled to an intercensal estimate of net migration. In Singapore, an unknown amount of the natural increase, B-D, of nearly 618,000 is attributable to net migration; but the exact amount depends upon the distribution, over time, of the net migrants as well as their age and sex composition. For purposes of estimating and projecting population, net migration is a convenient abstraction; it is usually not necessary to distinguish between natural increase during the intercensal or the projection period which is due to migration, and that which presumably would have occurred anyway.

The techniques described in the following sections 3 and 4 are aimed at indirect measurement of migration, among other concerns.

3. Calculation of intercensal migration estimates from place-of-birth data

The rate of in-migration for a subnational area may be expressed as

$$m_l^I = \frac{\Sigma M_{lj}}{N_l} \times 100$$

where $M_{lj}$ is the number of lifetime migrants to area l born in area(s) j, and $N_l$ is the native (to the country) population of area l enumerated in the census.

Migration may also be computed as

$$\Sigma M_{lj} = \Sigma N_{lj} - N_{ll} = N_l - N_{ll}$$

which is the summation of lifetime migrants to area l born in all areas, j, being equal to the summation of those enumerated in area l born in all areas j, minus those enumerated in area l who were born in area l; or more simply, the migrants in area l are equal to all those enumerated minus those born in the area.

In a similar manner, the rate of out-migration for a subnational area may be expressed as

$$m_l^O = \frac{\Sigma M_{il}}{N_l'} \times 100$$

where $M_{il}$ represents the migrants from area l to all areas i, and $N_l'$ is the total population born in area l and enumerated anywhere in the country. Similarly,

$$\Sigma M_{il} = \Sigma N_{il} - N_{ll} = N_l' - N_{ll}$$

which is the summation of lifetime migrants to all areas i from area l being equal to the summation of those enumerated in all areas i whose birthplace was in area l, minus those enumerated in area l who were born there; or more simply, the total population born in area l minus those still resident in the area.

Although an intercensal rate of net migration can be calculated for subnational areas from two censuses having lifetime migration data, it is not often an easy task, primarily because of the effort necessary to assemble the information about those no longer in the area of reference. Place-of-birth data are commonly focused on the area of reference, largely because of the constraints imposed by table design and size. The characteristics of non-migrants can readily be contrasted with those of people who have migrated and are currently residents of the area, and characteristics of migrants from various areas can also be compared; however, it is far more difficult to find the current location and characteristics of those who have migrated from the area in question, prior to the censuses. Where the proper questions have been asked and the answers preserved on tape, a special tabulation can solve the problem, but this requires examination of the national records by computer.

Once the data are assembled, intercensal levels of net migration can be calculated in accordance with the discussion which follows. For simplicity of explanation and notation the in-migrants and the out-migrants are of all ages and sexes; with additional effort, the procedure can be used for age and sex cohorts, handled separately. If two successive censuses are available for the subnational area and of the entire country, it is preferable to use the method of cohort survival ratios, a description of which is given in the next section.

Let $I_t$ and $I_{t+n}$ represent the number of lifetime in-migrants enumerated in area 1 at the time of two censuses, one at time t, the other n years later, and let $O_t$ and $O_{t+n}$ represent lifetime out-migrants from area 1 enumerated at the two censuses, while $S_I$ and $S_O$ are the n-year survival ratios of the in-migrants and the out-migrants, respectively.

Net migration for the period is

$$M = (I_{t+n} - O_{t+n}) - (S_I I_t - S_O O_t)$$

being the difference between the in-migrants and the out-migrants at the end of the intercensal period, minus the difference in the survivors of the in- and the out-migrants recorded in the census taken n years earlier. The equation may be rewritten as

$$M = (I_{t+n} - S_I I_t) + (S_O O_t - O_{t+n}) = M_I + M_O$$

thus analysing the intercensal net migration into that attributable to in-migration, $M_I$, and that attributable to out-migration, $M_O$.

The calculation of unique values of $S_I$ and $S_O$ is frequently impossible, owing to a lack of essential data, but a number of satisfactory approximations are possible. The over-all census survival ratio is calculated as:

$$\frac{P_{n,t+n}^{+}}{P_t}$$

or the ratio of the population n years of age and over at time t+n to the total population at time t. If an appropriate life table is available, the over-all life table survival ratio, $T_n - T_O$ may be used with the following proviso. Use of life table survival values may introduce a serious error, in that the survival ratios of all ages of migrants to subnational areas may depend far more upon past patterns of fertility and selective migration, and upon their consequences for the age composition, than upon the age-specific mortalities underlying the life table. Such considerations are especially germane to migrant populations, with their frequently atypical age compositions.

Neglecting mortality and survival ratios entirely, net intercensal migration can be written:

$$M' = (I_{t+n} - I_t) + (O_t - O_{t+n}) = M'_I + M'_O$$

Net intercensal migration among both the in-born $M_I$ and the out-born $M_O$ is underestimated by ignoring the effects of mortality. The error is of greater relative importance in the two portions, $M_I$ and $M_O$, than in their sum because they tend to cancel each other by having opposite signs in the equation. The error is inconsequential only when the in-migrants and the out-migrants are nearly equal, or when the level of net migration relative to natural increase is very small.

If lifetime migration data by age (i.e. place of birth cross-classified by place of residence, by age) are available for the latter of two censuses, a refinement of the above is possible. United Nations Manual VI [15] demonstrates one such method, using data from the nine geographic divisions used by the United States Bureau of the Census. Ten-year survival ratios were calculated for white males: for each division of birth, the number of white males ten years of age and older in 1960 was divided by the number of white males of all ages in 1950. These survival ratios were subsequently applied to white males born in the appropriate division but actually resident in the division of New England, in 1950, to calculate an "expected" number of survivors of prior migrations present in that division in 1960. When these estimates were compared with the enumeration of white males ten years of age and older in 1960 in New England by division of birth, it was possible to calculate the ten-year net migration into and out of that division from and to the other eight. The method assumes that the survival ratio of migrants to New England from, for example, the Pacific Division is the same as that of all white males born in the Pacific Division, regardless of

15/ *Op. cit.*

migration status. The Pacific Division exhibits the highest survival ratio precisely because it has the most favourable age composition for survival, owing to recent in-migration. This suggests that the survival ratio of migrants to New England might better have been estimated from the experience of a population having many migrants (such as the Pacific Division) than from the survival experience of the respective regions of origin of the migrants, if more age-specific data were not available.

The method can also be applied to age, or sex, or ethnic segments of a population. If the characteristics are accurately reported, they do not change over time; some can be determined over time, as in the case of age. Although ethnicity is generally regarded as an unchanging characteristic, the "passing" from one group to another can occur; where such events are frequent, distortions in the estimates of net migration can take place. Furthermore, methods dependent on a comparison of census counts with expected survivors can make no distinction between internal and international migration, unless the native (to the country) population is handled separately.

Place-of-birth statistics may be instructive for the light they shed on lifetime population relocation; however, without two comparable sets of data from two censuses, the analyst lacks a time base from which to calculate future population behaviour. In some methods of projecting population for subnational areas, it is important to have some estimate of net migration by period. If the estimate of net migration can be carried out in detail, by age, sex or by ethnic grouping, the corresponding projection for which it is prepared can have the same detail.

### 4. Method of cohort-survival ratios

If two censuses are taken n years apart, the population of age x at the earlier census which survives to the second census will be x + n years old. If it can be determined how many of the cohort died during the intercensal period, t to t + n, the balancing equation takes the form

$$M(x) = P_{x+n,\ t+n} - P_{x,\ t} + D(x)$$

where M(x) and D(x) represent, respectively, the net migration and the deaths experienced by the cohort age x at the earlier census during the period t to t+n; and $P_{x,\ t}$ represent, respectively, the population aged x+n, t+n and the population aged x at t.

The equation requires data on deaths by single year of age for the subnational area for each year of the intercensal period. Since such data are not generally available, methods using an approximation of intercensal mortality must be employed. Moreover, those under n years of age at the time of the second census must have been born during the intercensal period. Table 2 demonstrates the calculation of male and female migrants, by five-year age-grouping, during the period 1960 to 1970 for the central region of Thailand, which includes the Greater Bangkok area. Several critical assumptions underlie the calculations:

(a) Migration to and from the Kingdom of Thailand was negligible during the intercensal period 16/;
(b) The census survival ratios (CSRs) for the entire kingdom, specific for sex and five-year age group, are adequate measures of survival for the central region;
(c) The four fertility ratios calculated from the population enumerated in the central region in 1970 apply also to the "expected" population of that region in 1970, without its estimated net migration. The four ratios relate male children 0-4 and 5-9 years, respectively, to women 15-44 and 20-49, and female children of the respective age groups to the same age groups of adult females.

In table 2, the ten-year survival ratios are calculated by dividing the entries in column 3 by the size of the respective cohorts ten years earlier in column 2; the cohort survival ratios (CSRs) are recorded in column 4. The ratios are applied to the entries of column 5 to calculate the "expected" survivors of the central region, entered in column 6. The expected population is that which theoretically would be enumerated were there no net migration. For example, of the 1,562,832 males aged 10 to 14 years in the whole kingdom in 1960, some 1,321,641 survived until 1970, by which time they were 20 to 24 years old; a survival rate for the decade of 0.8457 was calculated. On the assumption that the same survival rate applied to the 477,389 males in the central region in 1960, an expected population of 403,728 males 20 to 24 years was calculated.

The number of migrants in the population in 1970 or, more precisely, the number of people in each cohort due to migration during the decade, was cal-

16/ If native population, tabulated by age and sex, were available for the two censuses, there is no problem of assuming an essentially closed population.

Table 2. Calculation of ten-year net migration in central Thailand, using the cohort survival method

| Sex and age | Kingdom of Thailand: Census 1960 | Kingdom of Thailand: Census 1970 | CSR | Central region: Census 1960 | Central region: Expected 1970 | Central region: Census 1970 | Number of migrants 1960-1970 | Migration rate based on expected |
|---|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| **Males** | | | | | | | | |
| 0 - 4 | 2,137,379 | 2,862,938 | | 653,534 | 765,607 | 787,775 | 22,168 | 0.0290 |
| 5 - 9 | 2,012,330 | 2,679,168 | | 628,214 | 775,078 | 795,055 | 19,977 | 0.0258 |
| 10 - 14 | 1,562,832 | 2,309,549 | 1.0806 | 477,389 | 706,209 | 714,270 | 8,061 | 0.0114 |
| 15 - 19 | 1,262,683 | 1,832,177 | 0.9105 | 395,017 | 571,989 | 591,035 | 19,046 | 0.0333 |
| 20 - 24 | 1,211,986 | 1,321,641 | 0.8457 | 404,130 | 403,728 | 423,752 | 20,024 | 0.0496 |
| 25 - 29 | 1,024,650 | 1,098,083 | 0.8696 | 326,208 | 343,507 | 344,578 | 1,071 | 0.0031 |
| 30 - 34 | 883,952 | 1,047,323 | 0.8641 | 274,404 | 349,209 | 342,468 | -6,741 | -0.0193 |
| 35 - 39 | 692,094 | 952,959 | 0.9300 | 210,679 | 303,373 | 293,667 | -9,706 | -0.0320 |
| 40 - 44 | 568,637 | 744,328 | 0.8760 | 181,596 | 240,378 | 238,855 | -1,523 | -0.0063 |
| 45 - 49 | 493,725 | 599,118 | 0.8657 | 159,266 | 182,385 | 182,102 | - 283 | -0.0016 |
| 50 - 54 | 401,639 | 472,185 | 0.8304 | 131,608 | 150,797 | 152,363 | 1,566 | 0.0104 |
| 55 - 59 | 321,629 | 388,328 | 0.7865 | 104,630 | 125,263 | 127,461 | 2,198 | 0.0175 |
| 60 - 64 | 228,571 | 300,801 | 0.7489 | 77,606 | 98,561 | 98,575 | 14 | 0.0001 |
| 65 - 69 | 149,000 | 212,957 | 0.6621 | 53,689 | 69,276 | 71,338 | 2,062 | 0.0298 |
| 70 and over | 177,338 | 250,656 | 0.4517 a/ | 65,586 | 88,931 b/ | 89,005 | 74 | 0.0008 |
| Unknown | 25,676 | 21,651 | | 9,912 | | 6,111 | | |
| Total | 13,154,121 | 17,123,862 | | 4,153,468 | | 5,258,410 | | |
| **Females** | | | | | | | | |
| 0 - 4 | 2,101,936 | 2,796,232 | | 635,502 | 744,400 | 765,954 | 21,554 | 0.0290 |
| 5 - 9 | 1,979,811 | 2,605,723 | | 617,246 | 750,503 | 769,846 | 19,343 | 0.0258 |
| 10 - 14 | 1,525,370 | 2,252,650 | 1.0717 | 467,121 | 681,067 | 697,993 | 16,926 | 0.0249 |
| 15 - 19 | 1,236,294 | 1,885,371 | 0.9523 | 384,010 | 587,803 | 607,976 | 20,173 | 0.0343 |
| 20 - 24 | 1,204,153 | 1,361,717 | 0.8927 | 387,209 | 416,999 | 441,074 | 24,075 | 0.0577 |
| 25 - 29 | 1,046,464 | 1,143,377 | 0.9248 | 324,598 | 352,358 | 366,495 | 14,137 | 0.0401 |
| 30 - 34 | 869,876 | 1,077,088 | 0.8945 | 272,164 | 346,358 | 351,725 | 5,367 | 0.0155 |
| 35 - 39 | 679,940 | 957,607 | 0.9151 | 210,671 | 297,040 | 296,262 | - 414 | -0.0014 |
| 40 - 44 | 563,812 | 766,332 | 0.8810 | 181,992 | 239,776 | 241,300 | 1,524 | 0.0064 |
| 45 - 49 | 482,966 | 597,454 | 0.8787 | 158,643 | 158,117 | 187,789 | 2,672 | 0.0144 |
| 50 - 54 | 410,354 | 489,794 | 0.8687 | 132,801 | 158,096 | 161,900 | 3,804 | 0.0241 |
| 55 - 59 | 329,041 | 401,731 | 0.8318 | 107,601 | 131,959 | 136,573 | 4,614 | 0.0350 |
| 60 - 64 | 244,989 | 324,223 | 0.7901 | 83,549 | 104,926 | 108,778 | 3,852 | 0.0367 |
| 65 - 69 | 163,600 | 238,901 | 0.7261 | 60,468 | 78,129 | 83,015 | 4,886 | 0.0625 |
| 70 and over | 244,717 | 353,486 | 0.5411 c/ | 89.849 | 126,545 d/ | 130,449 | 3,904 | 0.0309 |
| Unknown | 20,416 | 21,826 | — | 7,354 | — | 5,974 | — | |
| Total | 13,103,739 | 17,273,512 | | 4,117,778 | | 5,353,467 | | |

Sources: Central Statistical Office, National Economic Development Board, Thailand Population Census, 1960: Whole Kingdom, (Bangkok, 5 July 1962); and National Statistical Office, Office of the Prime Minister, 1970 Population and Housing Census: Whole Kingdom, E-Cr.-Pop.-No.58-73, (Bangkok, 1973).

Notes: a/ Population 70 and over of 1970 ÷ population 60 and over of 1960
b/ The 1960 population 60 and over x 0.4517 (CSR for 70 and over in column 4)
c/ Population 70 and over of 1970 ÷ population 60 and over of 1970
d/ The 1960 population 60 and over x 0.5411 (CSR for 70 and over in column 4)

culated by subtracting the entries of column 6 from those of column 7. In the cohort considered above, there were 423,752 reported in the census of 1970, or 20,024 more than the 403,728 expected. Using the expected population as a base, a ten-year net migration rate of 4.96 per cent was calculated.

To calculate the expected number of males 0 to 4 years old in 1970, the child/woman ratio, (males, 0-4) ÷ (females, 15-44), was employed. In the enumerated population in 1970 were 787,775 males 0-4 years and 2,305,196 females 15-44, or a ratio of 0.34174 male per female within the selected age spans. When the ratio is multiplied by the 2,240,334 women aged 15-44 in the expected population, a product of 765,611 results, which is the number of males estimated in the expected population who are 0 to 4 years of age. The three remaining cells are filled in the same manner, estimating males 5-9 from females 20-49, and females 0-4 and 5-9 from females 15-44 and 20-49, respectively.

The population 10 years of age and older in 1970 consists of those born prior to 1960; the calculation of a decennial rate of migration for them is appropriate, since they were exposed to the risk of migrating during the entire ten years. Members of the male and female cohorts aged 0-4 in 1970 had, by 1970, been exposed to that risk only an average of 2.5 years each, under the assumption that the births of the members were uniformly distributed over the five years preceding the 1970 census; however, those 5-9 years old had been exposed to the probability of migration for 7.5 years, because under the same assumption their births had been uniformly distributed over the five years following the earlier census of 1960. If the purpose in calculating migration rates is the provision of a figure essential to a projection ten years in the future, the 2.5 and 7.5-year migration rates for the youngest groups can be used along with the ten-year rates for the other ages. In development planning, however, five-year projections are often required, while single-year projections are occasionally useful; for this reason it may be necessary to reduce the decennial rates to five-year or one-year equivalents.

The following description is based upon the assumption that the "compounding" of the geometric rate of change, rather than being continuous as in the computationally more convenient model, takes place once each year for the annual rates, and once every five years for the five-year rate. To use the familiar formula, $P_n = P_o (1 + r)^n$, the rates must be converted to ratios of change, where they are not already in that form. For example, where r represents a rate of change, 1 + r is the ratio of the final value to the value with which it has been compared in the calculation of the rate. Cohort survival ratios are already in the form required for reduction to five or one-year multipliers, but the ten-year migration rates are not. The five-year ratio is equal to the square root of the ten-year ratio, while the one-year ratio is the tenth root of the ten-year figure.

There were 591,035 males 15-19 years old in 1970 in the central region of Thailand. They were the survivors of 628,214 males ten years younger in 1960, who by 1970 numbered 571,989 plus 19,046 migrants (net). More precisely, the number in the cohort in 1970 can be divided in two parts: one part is representative of the survivors of 1960, the other a number representing the net effects of in and out-migration during the ten-year period. It is convenient to use the first and less complex interpretation, manipulating the data as if survivors and net migrants were members of separate populations. The ten-year survival ratio is 0.9105; the cohort was subject to mortality equal to (1-0.9105) per cent of the initial population. The migration rate of the cohort is 3.33 per cent, i.e. the total population of the cohort in 1970 was 1.0333 times the expected or actually surviving population of the cohort. The one-year rate of survival is $\sqrt[10]{0.9105}$ or 0.9907; the five-year rate of survival, which is probably of greater utility in subnational projections based only upon five-year data, is $\sqrt{0.9105}$ or 0.9542. The rate of migration, converted to a ratio of estimated total to estimated expected population, is 1.0333. The corresponding five-year rate is 1.0165, and single year rate is 1.0033 17/ .

The migration ratio of the 5-to-9-year-olds, who were exposed to the risk of migrating for 7.5 years, is $(1 + r)^{0.1333}$ where the exponent is the reciprocal of 7.5, for one year and $(1 + r)^{0.6667}$ where the exponent is five times as great, for five years. Similarly, the two ratios for the 0-to-4 year-olds are, respectively, $(1 + r)^{0.4000}$ and $(1 + r)^2$ for a one-year and a five-year period. Survival ratios cannot be calculated directly from the data presented and if they are needed, it is recommended that they

17/ It is apparent from the example that, with a ten-year rate as low as three per cent, very little precision is gained by calculating the one-year and the five-year ratios geometrically, instead of arithmetically. At four decimal places, the result of dividing the ten-year ratio by 10 is identical to extracting the tenth root of the ratio.

be taken from a suitable life table. [18]

The census survival ratio method is a convenient technique for estimating migration where two censuses are available, and where vital statistics are not of satisfactory quality. The method contains a slight negative bias in its estimates of net migration, owing to the way mortality would affect the migrants. Also, the reliability of results of its application is contingent on the extent of accuracy of age/sex data and coverage in the two censuses.

If better estimates of births and deaths are available, even without age-specific detail, the migration estimates might be improved by prorating, by adjusting to a total derived from the balancing equation. The bias is minimal, however, and affects the value of projections dependent upon the migration estimates to a far smaller extent than other estimating and projecting errors over which the analyst has little or no control. Where levels of net migration are modest, the bias will not affect projected population to a marked degree; where the various age groups exhibit both positive and negative net migration, i.e. net in and net out-migration, the errors tend to cancel.[19]

In the illustration of the calculation of the levels and rates of net migration for the age cohorts of males and females in the central region of Thailand, a "forward" survival technique was used; emphasis was upon those in 1970 who had survived from 1960, and those who had joined them through migration. Symbolically, this was expressed as

$$M(x) = P_{x+n,\ t+n} - P_{x,\ t} + D(x)$$

which may be expressed

$$M(x) = P_{x+n,\ t+n} - S(x)P_{x,\ t}$$

where M(x) is equal to the migrants to the cohort which is aged x at the first census; $P_{x,t}$ is the population of that cohort at the first census; and $P_{x+n,\ t+n}$ is its population at the second census n years later when it was n years older and S(x) is the census survival ratio. The second equation might have been written

$$M'(x) = \frac{1}{S(x)} P_{x+n,\ t+n} - P_{x,\ t}$$

which is the application of a "reverse survival" technique: in the first term of the right hand side of the equation, the number was calculated of those who would have been present at the first census without the effect of migration, compared with those actually counted, resulting in an estimate of the migration. The number of persons x years of age at the earlier census is equal to those x+n years of age at the later census, plus deaths within the cohort; the resulting estimate of net migration includes deaths to the migrants. It is equivalent to assuming that all migration occurred at the beginning of the intercensal period; the "forward survival" method assumes that migration took place at the end. The calculation of the mean of the two migration estimates for each cohort would yield an estimate free of this bias. The two estimates are related by equation

$$M'(x) = \frac{1}{S(x)} M(x)$$

from which it is clear that, as the survival ratio approaches unity, the difference between the two estimates of migration vanishes. The labour of calculating the migration rates by the average method is more than double that of the forward method; therefore the planner should determine whether the benefit of the improvement is worth the additional effort.

---

18/ Methods of fitting model life tables are described by Henry Shryock and Jacob S. Siegel, *The Methods and Materials of Demography*, two vols., (Washington, D.C., U.S. Government Printing Office, 1971), pp. 813-815; and in United Nations Manual IV.

19/ For a detailed discussion of this bias and comparisons with other methods, see United Nations Manual VI, *op. cit.*, chapter II.

# IV. MEASUREMENT AND ESTIMATION OF NATURAL POPULATION CHANGE

At global level, population change consists of the difference between births and deaths, or what is known as "natural change." At country level, this change may be further affected by migration, i.e. movements of people across international boundaries. In general, the volume of such movement may be too small to affect the rate of natural change significantly. But in view of the large movement from rural to rural area, rural to urban area, and small to large towns, the effect of migration on the growth of population in subnational areas cannot be ignored.

Although migration is an important factor in the growth of subnational areas, natural increase accounts for most of the growth in rural areas, and at least half of the growth of urban areas, in the ESCAP region. The excess of births over deaths may either be increased by adding excess of in-migration over out-migration, or be decreased by subtracting the excess of out-migration over in-migration. In projecting subnational populations, consideration should be given to all three factors of births, deaths, and migration.

## A. SOURCES OF DATA AND PROBLEMS OF MEASUREMENT

As shown in chapter II and table 1, birth and death data generally come from the vital registration system, or are obtained or estimated through censuses and sample surveys. Detailed population projections should be used upon rates of birth and death. A rate consists of a numerator, the number of events; and a denominator, the "population at risk" of dying or of producing children.

Ideally, the event of birth or death is recorded shortly after it takes place. The recording document, the birth (or death) certificate, contains the age of the mother (or the decedent), the place of occurrence of the event, and usually the residence of the individual concerned, his sex, and some ethnic and public health data. Unfortunately, in most developing countries the registration of births and deaths is far from complete (see table 1); extensive efforts are directed not only toward improving the completeness of registration, but towards the development of methods for estimating births, deaths, and their rates from incomplete data. Moreover, in some countries births are tabulated by year of recording, not year of occurrence; where registration is virtually complete, as in Australia and New Zealand, the discrepancies introduced by this system are trivial and tend to cancel each other, year by year. Where registration eventually takes place only when the registrant is starting school, or when it is advantageous for some other reason, it is unlikely that those who die young are recorded at all.

If the numerator of the rate reflects the sum of the inaccuracies of incomplete recording, the denominator combines the errors incident to census enumeration and tabulation. All censuses are subject to underenumeration, and the overlooked population is generally not distributed proportionately among the age, sex, and ethnic groupings, nor geographically over the area for which the census was taken. Two methods have been employed to estimate net census undercount: a) controlled re-enumeration, or post-enumeration survey (PES); and b) a "demographic" method.

In the PES, specially trained enumerators are used to count very carefully selected sample areas, properly stratified in accordance with the urban/rural, ethnic and geographic makeup of the entire area.

The demographic method is an application of survivorship techniques. The totals are calculated for those in each age, sex, and ethnic grouping who would be expected to survive from census to census, on the basis of death rates. The "expected" populations are compared with those actually enumerated in the later census.

Neither method yields net undercount information by small area, but both are useful for the insight they give into the systematic underenumeration of the more mobile segments of the population. However, estimates obtained through these methods for the country could be distributed by rationally applying the apportionment or ratio method.

All censuses show "age-heaping" and other age biases in their tabulations. The former is a reflection of a widespread tendency, especially when an estimate is involved, to report an age divisible by ten or five. Other biases in the age data are perhaps more wishful. The extremely old and the very young have been known to exaggerate their ages. Doubtless considerable understatement of age exists out of vanity, or for economic reasons where age is an impediment to employment. Such systematic biases are very evident in a single-year age pyramid, and if not corrected they can create serious errors in the vital rates estimated through the use of age/sex data. Biases can be removed by a number of methods;[20] typically, the numbers reported for "popular" ages are averaged with those reported for other ages.

Census data are usually tabulated on a **de jure** basis, reflecting the place where a person is regarded as living, rather than the place he happens to be at the time of enumeration. But where residence is not established at the time of enumeration, the census data can only be tabulated on a **de facto** basis. When births and deaths are tabulated by residence, it is possible to calculate meaningful rates, specific for areas.

## B. METHODS OF ESTIMATION UNDER VARIOUS CONDITIONS OF DATA

In Manual IV[21] various methods of estimating mortality and fertility rates are described for various conditions of data. Although census survival can be used to calculate mortality rates, the latter are often erratic. Under these circumstances, model life tables can be selected from among those available, or a unique one can be constructed by interpolation from two standard tables. These methods of estimating vital rates are discussed in Manual IV and in Shryock and Siegel.[22] Where two or more censuses have been taken, and especially where data on vital events are less than satisfactory and where the projection horizon is not more than ten or fifteen years in the future, it is often feasible to use survivorship rates specific for age and sex groups and calculated from the censuses, ignoring births, deaths and migration entirely. Under these circumstances, it is recommended to use the child/woman ratio for estimating the population in age groups 0-4 and 5-9.

Where results of two censuses are not available, it is sometimes possible to estimate mortality and fertility under a stable population assumption. Stable population theory [23] can be used for populations which a) have undergone only minor variations in fertility over the past five or six decades, b) have experienced only moderate and gradual changes in mortality, and c) are closed to migration. Relatively stable fertility has until recently been the common experience of most less developed countries of the region. The mortality assumption was also justifiable in those countries. The use of this method to compute subnational populations is explained in chapter VIII.

## C. CRUDE RATES OF BIRTH AND DEATH

A "crude rate" is one which relates events (births, deaths and marriages) taking place within one year to the midyear population, without any refinement or standardization. Assuming linear growth or decline, the population at midyear is also the average population for the year, and crude birth and death rates measure the frequency of the events to the "person-year" lived during the year by the populaton at risk. The expressions for the rates are, respectively for births and for deaths,

$$CBR = \frac{B}{P} \times 1{,}000 \text{ and}$$

$$CDR = \frac{D}{P} \times 1{,}000$$

It is customary to express both rates in events per thousand population. Thus a crude birth rate of 40.3 for Burma estimated for 1965-1970 [24] embodies the ratio of one-fifth of the estimated live births for the five-year period to the estimated population of Burma at the mid-point of the period. It implies that for every thousand people in Burma on that date, an annual average of 40.3 live births took place within the calendar year of which that date was the mid-point. The crude rate is widely understood, easy to calculate if the number of events is known, and quite useful as an indicator. Because populations vary substantially in their age and sex compositions, it can only be used for comparative purposes with some caution. In projections, especially for subnational areas experiencing the effects of migration, its use should be confined to those directed at rather close horizons.

The algebraic addition of the two rates, counting births positively and deaths negatively,

---

20/ Shryock and Siegel, op. cit., pp. 204-211.

21/ Op. cit.

22/ Op. cit., p. 813 ff.

23/ See Manual IV, pp. 14 ff.

24/ United Nations, **Demographic Yearbook 1970.**

yields a crude rate of natural increase expressed either as

$$CRNI = \frac{B}{P} \times 1{,}000 - \frac{D}{P} \times 1{,}000$$

or as $$\frac{B - D}{P} \times 1{,}000$$

depending whether the rates of birth and death are being manipulated, or the crude rate of natural increase from recorded or estimated events is calculated. A negative rate of natural increase is a rate of natural decrease.

## D. AGE-SPECIFIC BIRTH AND DEATH RATES

Since crude rates apply to the entire population, they mask the diversity of each population; it is this diversity which in many cases makes proper comparisons impossible. Although there are many ways in which rates may be refined, the most common and most useful is calculation specific for age and sex. Obviously, the likelihood of an individual dying varies with age; the likelihood of a woman giving birth also varies with age. When rates are calculated in a manner distinguishing characteristics other than sex and age, e.g. ethnicity, socio-economic status or occupation, they are usually referred to as "differential" rates.

Age-specific birth and death rates, respectively, are expressed as

$$ASFR = \frac{b_i}{p_i} \times 1{,}000 \quad \text{and}$$

$$ASDR = \frac{d_j}{p_j} \times 1{,}000$$

In the former, the numerator $b_i$ represents the number of births taking place among women of age i; the denominator represents the population of women of age i, the population at risk. Similarly, the age-specific death rate is expressed as the ratio of deaths to members of an age (and sex) grouping j, to all members of that grouping. Since the sum of the deaths $d_j$ to members of the group is equal to D, the total deaths during the year in the population, and since the sum of the population $p_j$ is equal to P, the total population, then the crude death rate may be regarded as the weighted average of the age-specific death rates in a population. The ratio expressions of both age-specific rates are typically multiplied by 1,000, and the rates are customarily expressed in terms of events per thousand in the relevant age groupings. Age-specific birth and death rates must be prepared as sets or schedules for all age groupings. A single age-specific birth rate, or a birth rate specific for a five-year age group of women, is of limited value for comparisons, because of the possibility of minor variations in age patterns of child-bearing. Age-specific rates may be calculated and used for a single year of age, but their principal application is in five-year groups. The mortality of those under one year of age poses special problems which will be discussed in a later section.

## E. SUMMARY MEASURES OF FERTILITY

If the entire schedule of age-specific birth rates for single years of age were added, a summary measure would result, reflecting the behaviour of women of all ages during the reference period. Since it is a sum of rates, this summary measure (the "total fertility rate," or TFR) does not reflect population differences among age groups. If the constant used is 1,000, the rate might be regarded as expressing the number of children born to 1,000 women passing through the child-bearing period, who each year exhibit the fertility behaviour of the age-specific schedule, _if_ no mortality were experienced during the period. If the numerator of the fraction, i.e. births, be replaced by female births, the measure becomes the "gross reproduction rate," or GRR, an indication of the extent to which women reproduce themselves during a generation, again assuming no mortality.

Table 3 compares five-year age-specific fertility rates for Japan for the two years 1963 and 1967. To calculate the TFR, the sum of each column of age-specific rates was multiplied by five to approximate the sum of one-year rates. The percentage female of all live births registered in Japan during each of the two years was applied to the TFR, yielding the GRR. A comparison of the GRR suggests that, while about 1963 daughters were born to 1,000 Japanese women in 1963, this rate had by 1967 exceeded replacement at slightly over 1,056. Had 1,000 women passed through the child-bearing period without mortality among either the adult women or their daughters, and had the surviving daughters entered the child-bearing period, the assertion above would have been true. The GRR is merely a summary measure describing the reproductive behaviour in one year without regard for mortality. It overlooks not only mortality but changing fertility, as an actual cohort passes through the fertile years; it is a period rate, not a cohort rate, and although it is a useful measure of things as they are, the distinction should not be forgotten in projecting population.

A more laborious calculation, and one which requires an understanding of the life table discussed later in this chapter, is the net reproduction rate (NRR). Prevailing age-specific mortalities, as well as age-specific fertilities, are applied to an hypothetical cohort of women. The NRR is best understood within the context of the life table. The following formula borrows the notation of the life table:

$$NRR = \sum_{15}^{49} \frac{B_x}{P_x} \cdot \frac{L_x}{l_0}$$

where $\frac{B_x}{P_x}$ is the age-specific female fertility rate and $\frac{L_x}{l_0}$ is the proportion of newborn girls surviving to the mid-point of age interval x. The summation for all intervals in the child-bearing period represents the aggregate fertility within the age groups of the survivors $L_x$ from birth, when they numbered $l_0$. A form of the expression describing the calculation for five-year age groups is

$$NRR = 5 \sum_{15}^{45} \frac{{}_5B_x}{{}_5P_x} \cdot \frac{{}_5L_x}{{}_5e_0}$$

Although the NRR is a measure of replacement, since it embodies the expected mortality of the mothers and the daughters, it is still a period rate. It measures the number of daughters who would be born to the survivors of a cohort of newborn females if current (or assumed) levels of fertility and mortality were to continue. The cohort is, of course, hypothetical and the calculation makes no allowance for change with the passage of time. For this reason it can prove deceptive for comparative purposes. An abnormally low NRR can be calculated from age-specific birth rates in a unique year, one in which economic conditions are not favourable or one, in some cultures, regarded as inauspicious. It is not likely that the analyst making subnational projections would have the detailed data necessary for the calculation of the NRR, nor would the additional precision which it offers be critical to the relatively short-term projections most useful at the subnational level.

Table 3. Age-specific birth rates, TFR and GRR, of Japan between 1963 and 1967

| Age | 1963 | 1967 |
|---|---|---|
| 15 - 19 | 3.7 a/ | 4.7 a/ |
| 20 - 24 | 98.1 | 102.6 |
| 25 - 29 | 191.0 | 212.7 |
| 30 - 34 | 80.8 | 91.4 |
| 35 - 39 | 18.7 | 19.6 |
| 40 - 44 | 3.5 | 2.7 |
| 45 - 49 | 0.2 | 0.1 |
| Total | 396.0 | 433.8 |
| 5 x total TFR b/ | 1,980.0 | 2,169.0 |
| % female | 48.6 | 48.7 |
| GRR c/ | 962.9 | 1,056.5 |

Source : United Nations, Demographic Yearbook 1969.

Notes : a/ Five-year rate calculated from source by proration

b/ TFR - total fertility rates

c/ GRR - gross reproduction rate

A much more useful and far simpler summary measure of fertility, and one which is extremely useful in subnational population projecting, is the child/woman ratio. In most cases, the selection of analytical techniques would be confined to those depending upon census data. The child/woman ratio is the ratio of young children to women of age groups which include the mothers. A commonly used ratio is that of children under five to women 15 to 49 years of age, inclusive. In the projection of age groups by a survival method, as shown in chapter VI below, the child/woman ratio is used to calculate the numbers of people in age groups born during the projection period, for age groups without starting populations against which to apply survival rates. 25/

In the example in chapter VI below, the female population 0-4 years at the end of the projection period is calculated on the assumption that the ratio of girls 0-4 to women 15-44 is fixed, and that the ratio of girls 5-9 to women 20-49 is also fixed. Although the technique is convenient, it assumes no change in fertility, mortality or the rate of underenumeration of the young. If a rate of decline in fertility is assumed, this can be applied to the child/woman ratio, as is demonstrated later in the chapter.

## F. APPLYING MORTALITY RATES

The crude and age-specific death rates have been

25/ Survival ratios are calculated by dividing the number in each age and sex group by the number in the group n years earlier, when they were n years younger. The ratio cannot be calculated for those less than n year of age at the time of the later census, because they had not been born at the time of the earlier census.

described above with the birth rates that correspond to them. Like the birth rates, they are "central" rates, that is they relate the number of events to a population at risk whose magnitude is measured at the mid-point of the period in which the events are counted. A central death rate does not describe the probability of death within a period to members of a cohort entering that period. Such a probability statement has enormous importance in the manipulation of the death component in population estimation and projection, as well as in other applications. An age-specific probability of dying answers the question "what is the probability that an individual of exact age x will die before he has attained the exact age x + 1?". To distinguish them from the death rates or "central" rates, the latter variety are called "mortality rates", specific for age and sex.

The "conventional infant mortality rate", the term used by Shryock and Siegel, [26] approximates such a probability of dying. It is the ratio of deaths to persons under one year of age to live births, multiplied by a constant (commonly 1,000), and is usually sex-specific. It only approximates a mortality rate because the population at risk (i.e. those born), and the events (the deaths) do not have the same time base; in a given year, infants who die may have been born in the preceding year, but this is compensated by the likelihood that as many of those born in the reference year will die as infants under one year of age in the following year.

For most ages it is sufficient to move the base population from midyear to the beginning, in order to calculate a mortality rate by the conversion of a death rate; the effect of the shift is minimal. However, during the first year of life, such is not the case because of the rapid decline in mortality: in countries where registration is complete, one-third or more of total infant deaths are reported as taking place on the first day of life. Shortly after the first year, the decline tends to become essentially linear. But special steps must be taken in establishing mortality rates for the early years, since inaccuracies of the basic data and non-linearity of the death rates are of most significance then. Essentially, the steps involve a comparison of mortality estimates based upon recorded births and deaths, with estimates based upon deaths and population. [27] Once linearity has been established, the formula

$$q_x = \frac{2m_x}{2 + m_x}$$

is used, where $q_x$ is the age-specific mortality rate or probability of dying; and $m_x$ is the central age-specific death rate.

After deriving a schedule of the values of $q_x$ for each year of life, based upon births, deaths and the composition of a particular population, it is possible to construct the widely used model "life table", sometimes called the mortality table. The techniques of life table construction, especially the details of conversion of observed death rates to mortality rates referred to above, and the methods of constructing abridged life tables are outside the scope of this manual, a number of standard texts contain explanations of life table construction. The discussion which follows is confined to the basic calculations of the numbers found in the columns of the life table, to facilitate use in particular areas and functions. A further discussion of life tables in chapter VIII is devoted to the selection of model life tables for areas having none.

Table 4 is a life table reflecting the average mortality among Indian females from 1951 to 1960. It is an "abridged" table because it presents data for five-year age groups, excepting the two youngest where change is most rapid. It is a "period" not a "cohort" life table, because the mortality experience is specific for a certain time, in this case an average over one decade, as opposed to a description of the experience and changed conditions that a cohort meets as it ages. Although the cohort table is an interesting development, there are few bodies of experience of sufficient historical duration to permit the construction of cohort life tables. For example, the mortality experienced by a cohort born during 1900 can only be projected, since large numbers of those born then are still alive.

The left column of table 4 lists the periods of life between exact age x and the exact age at the upper end of the interval, x + n. For example, the first line begins with the birth of the cohort at age 0, and ends an instant before the cohort members are one year old. The second line presents the values in the various columns for the period 1-5 years, being the period of the second, third, fourth and fifth years of life and ending just before the fifth anniversary of the birth of the cohort. The mortality rate $q_x$, discussed earlier in the chapter, becomes ${}_nq_x$ of column 2 of the table, which refers to the probability of dying during the interval beginning with x and ending with x + n. Column 3 lists the

---

[26] Op. cit., p. 410.

[27] Ibid., pp. 435-437.

Table 4. Abridged life table for the female population of India, 1951-1960

| Age interval | Proportion of persons alive at beginning of age interval who die during interval ${}_nq_x$ | Of 100,000 born alive: number living at beginning of age interval $l_x$ | Of 100,000 born alive: number dying during age interval ${}_nd_x$ | Stationary population: In the age interval ${}_nL_x$ | Stationary population: Survival ratio $\frac{{}_5L_{x+5}}{{}_5L_x}$ | Stationary population: In this and all subsequent age intervals $T_x$ | Average years to live: Average number of years of life remaining at beginning of age interval $\overset{o}{e}_x$ |
|---|---|---|---|---|---|---|---|
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 0-1 | 0.1383 | 100,000 | 13,826 | 87,273 | 0.8200 a/ | 4,055,000 | 40.55 |
| 1-5 | 0.1002 | 86,174 | 8,637 | 322,713 | 0.9245 b/ | 3,967,727 | 46.02 |
| 5-10 | 0.0379 | 77,537 | 2,937 | 379,026 | 0.9777 | 3,645,014 | 47.01 |
| 10-15 | 0.0197 | 74,600 | 1,469 | 369,269 | 0.9710 | 3,265,988 | 43.78 |
| 15-20 | 0.0269 | 73,131 | 1,964 | 361,039 | 0.9691 | 2,896,719 | 39.61 |
| 20-25 | 0.0284 | 71,167 | 2,019 | 350,603 | 0.9390 | 2,535,680 | 35.63 |
| 25-30 | 0.0421 | 69,148 | 2,913 | 339,770 | 0.9149 | 2,185,077 | 31.60 |
| 30-35 | 0.0742 | 66,235 | 4,915 | 319,052 | 0.8908 | 1,845,307 | 27.86 |
| 35-40 | 0.1002 | 61,320 | 6,142 | 291,923 | 0.8802 | 1,526,255 | 24.89 |
| 40-45 | 0.1132 | 55,178 | 6,244 | 260,056 | 0.8632 | 1,234,332 | 22.37 |
| 45-50 | 0.1276 | 48,934 | 6,244 | 228,909 | 0.8319 | 974,276 | 19.91 |
| 50-55 | 0.1514 | 42,690 | 6,463 | 197,615 | 0.7955 | 745,367 | 17.46 |
| 55-60 | 0.1848 | 36,227 | 6,694 | 164,414 | 0.7530 | 547,752 | 15.12 |
| 60-65 | 0.2262 | 29,533 | 6,679 | 130,801 | 0.7000 | 383,338 | 12.98 |
| 65-70 | 0.2737 | 22,854 | 6,255 | 98,498 | 0.6429 | 252,537 | 11.05 |
| 70-75 | 0.3290 | 16,599 | 5,461 | 68,945 | 0.5742 | 154,039 | 9.28 |
| 75-80 | 0.3920 | 11,138 | 4,366 | 44,327 | 0.6016 | 85,094 | 7.64 |
| 80-85 | 0.4680 | 6,772 | 3,169 | 25,454 | -- | 40,767 | 6.02 |
| 85 and over | 1.0000 | 3,603 | 3,603 | 15,313 | -- | 15,313 | 4.25 |

Source: Based on data from United Nations, **Demographic Yearbook**, tables 21, 22 and 23.

Notes : a/ Proportion surviving from birth through 4 year of age (0-4).

b/ $\frac{{}_5L_5}{{}_5L_0}$

successive values of $l_x$, the number living at the beginning of the age interval; starting with a conventional population of 100,000 newborn, called the "radix", the cohort declines in number through deaths throughout its span. Column 4 shows the number dying between exact age x and exact age x + n, and is the product of the number alive at the beginning of the interval and the probability of dying during the interval.

Columns 5 through 7 are headed "Stationary population". The life table may be regarded from two different viewpoints: (a) as the history of 100,000 newborn passing through life, subject to a fixed schedule of mortality, or (b) as a stationary, i.e. unchanging, population which experiences 100,000 births and 100,000 deaths each year. It is the latter viewpoint that is most helpful in understanding columns 5 and 7; column 5 represents the average number of persons alive during the interval from the cohort point of view; or, from the viewpoint of the stationary population, the population of the specified age. Column 7 represents a cumulative addition from bottom to top of the ${}_nL_x$ values. Since the values represent the number of years to be lived by the survivors of the original cohort, from the beginning of each interval to the end of the life of the cohort, an average expectation of life for each year can readily be calculated by dividing the $T_x$ value by the corresponding value for $l_x$. This quotient is recorded in column 6 and designated $e_x^o$. The table is brought to an arbitrary close by assuming total mortality at the age group 85-90, a simplifying assumption which has a negligible effect on the parameters calculated from the table. Column 6, the survivorship ratios, is not an essential part of the life table. It has been added for the light it sheds on changing survivorship with aging, and for its value as an estimating tool.

Life tables differ from one another in a number of important respects, because mortality patterns differ between the two sexes and at various times and places. Populations tend to exhibit a characteristic pattern of mortality, which can be summarized in a life table, permitting the development of model life tables as an analytical tool where the lack of data would otherwise prevent effective estimation of population parameters. The application of model tables is discussed in chapter VIII.

In table 4, the "crude" female death rate of the population is calculated by dividing the hypothetical annual number of deaths by the stationary population or 100,000 x 1,000 + 4,055,000 = 24.7 per thousand. This death rate is the reciprocal of $e_x^o$.

Of greater potential use in estimating and projecting populations are the survival rates. With the life table for a single benchmark population, for example, the survivors of each interval may be calculated using the survival ratios of the table:

$$\frac{{}_nL_{x+10}}{{}_nL_x}$$

If the table is reasonably current, no change in age-specific mortalities need be assumed during the projection period. In table 4, of the 339,770 women 25-29 years of age in the life table population, 291,923 were present ten years later at the 35-39 interval. The survival ratio 0.859 can, in the absence of other information, be applied directly to Indian women of 25-29 years in a district or a state to calculate the expected number who would be ten years older in another ten years, without considering migration. The phrase "in the absence of other information" must qualify the statement, because the model being used might be less than ideal, although it might well be the best one available.

## G. PROJECTION OF NATURAL CHANGE

No discussion of natural change is complete without some attention being devoted to changing values during the projection period. Although death rates at all ages in most places have declined substantially in the past, it is unreasonable to assume the decline will continue as rapidly. If declines in a subnational area lag or lead those for the nation as a whole, it is reasonable to assume that the two patterns are converging, that areas which lag will catch up while those which lead cannot be expected to maintain their advantage indefinitely.

Projections entail a future level of the vital rates and a time when that level is reached, and they also entail selection of a pattern, i.e. the extent to which declines differentially affect the various age and sex groups in the population. If infant mortality in the specific area exceeds the national average, and if maternal and child health schemes are being developed, it might be reasonable to assume that infant mortality would converge with the national trend and attain its level in ten years, with the difference reduced by a tenth each year and the improvement to be applied arithmetically. As long as the assumed changes are modest, the errors which the planner may introduce by a faulty assumption would likewise be modest. If the projection horizon is not distant, subnational populations would not be strongly influenced by errors from this source.

## H. METHODS OF APPLYING CHANGES

(1) Declining fertility affects future values of the child/woman ratio, yet use of the ratio is most advantageous when the analyst is unwilling or unable to work with birth rates or fertility rates. By assuming a degree of stability or nation-wide uniformity in the ratio, it is not necessary to know the underlying fertility rates. In an illustrative projection appearing in chapter VI, the child/woman ratios are assumed to remain constant throughout the projection period. Table 4 below, however, illustrates a method of calculating a probable response of the child/woman ratio to an assumed decline in fertility. In the illustration a 0.5 per cent annual decline is assumed, and its progressive effects upon succeeding cohorts is demonstrated. The entries in the second and third columns show the cumulative effect of the assumed rate of fertility decline. Those aged 0, for example, can be reduced by 5 per cent, since the assumed decline has been in effect for ten years; those 9 years old, on the other hand, were born at a time when the decline had only been in force one year. Those 9 years old at the second projection year, 20 years from the benchmark, were born when the decline had been in effect for 11 years. Dividing each of the five-year sums by five gives the percentage reduction in the child/woman ratio of the four cohorts. At the ten-year projection, the child/woman ratios should be multiplied by 0.960 for the 0-4 year old, and 0.985 for the 5-9 group, while for the 20-year projection the ratios should be multiplied by 0.910 and 0.935, respectively. The above calculation assumes linearity not only of the decline in the rate, but in the array of sizes of the single-year cohorts. Under these conditions, the decline appropriate to the mid-age in the array can be used for the five single-year age groups.

Table 5. Adjusted projections of child/woman ratio, assuming 0.005 annual decline in fertility

| Projected age | 0.5 per cent decline in fertility: Ten-year projection | Twenty-year projection |
|---|---|---|
| 0 | - 5.0 | -10.0 |
| 1 | - 4.5 | - 9.5 |
| 2 | - 4.0 | - 9.0 |
| 3 | - 3.5 | - 8.5 |
| 4 | - 3.0 | - 8.0 |
| | sum -20.0 | -45.0 |
| 5 | - 2.5 | - 7.5 |
| 6 | - 2.0 | - 7.0 |
| 7 | - 1.5 | - 6.5 |
| 8 | - 1.0 | - 6.0 |
| 9 | - 0.5 | - 5.5 |
| | sum - 7.5 | - 32.5 |

(2) Chapter V discusses a number of "mechanical" methods of population projection, including straight-line, exponential and logistic models. The same techniques can be applied to the projection of vital rates. For example, if ${}_5M_x^t$ is equal to mortality of those in the five-year age group x to x+5 at time t, and ${}_5M_x^{t-5}$ is the mortality of those in the same age group[28/] five years earlier at time t-5, a ratio of change in mortality over the five-year period can be expressed as

$$\frac{{}_5M_x^t}{{}_5M_x^{t-5}}$$

If ${}_5M_x^{t+5}$ represents the rate of mortality of those in that age group five years later, a formula can be written by which mortality can be projected five years in the future, by assuming that the relative decline in each age group would be the same during the five years following the benchmark as it was in the five years preceding the benchmark; thus

$${}_5M_x^{t+5} = \frac{{}_5M_x^t}{{}_5M_x^{t-5}} \quad {}_5M_x^t$$

The assumption is useful, but is justifiable only for relatively short projection periods. Both fertility and mortality rates have been declining, therefore the geometric assumption of change is a "conservative" one. Where rates are increasing, the number of events implied by rates projected under a geometric assumption would soon become improbably high, and an arithmetic assumption would be preferable. Using the same notation, the value of mortality for the age interval projected five years, by an arithmetic assumption, is

$${}_5M_x^{t+5} = {}_5M_x^t + ({}_5M_x^t - {}_5M_x^{t-5}) = 2\,{}_5M_x^t - {}_5M_x^{t-5}$$

(3) A more sophisticated method of projecting mortality is to assume a designated improvement in the expectation of life at birth, say five years during each ten-year interval in the projection period. This technique requires the fitting of a model life table to

28/ It should be emphasized that the populations subject to the rates of mortality being discussed are NOT members of a single cohort. Rather, mortality at a specific age level is changing according to the assumed ratio for all cohorts as they attain the age.

the current female population [29] and selecting a corresponding one from the same series having an $\overset{o}{e}_o$ which is five years older. The survival values from the ${}_nL_x$ column of the appropriate table can be used; or if an incremental value not furnished by the model series is required, values can be interpolated

(4) Age-specific fertility rates can be projected according to two different schemes, by the period rates or by the behaviour of cohorts. In the first approach, the period rates at the beginning of the projection period are multiplied by a constant adjustment factor, selected as a reasonable measure of the expected decline for each phase in the whole projection period; or, more realistically, different rates of decline for different ages of mother. Birth order and age of motherhood are closely associated in all populations, and higher order births are the first to decline when summary measures show a decrease. A drop in the level of fertility at all ages is characterized by modest declines in the principal childbearing years, and much more precipitous drops in the fertility of women over 30. A typical projection of age-specific fertility rates might carry the assumption that a reduction of 10 per cent for the 15-19 and 20-24-year group, no change for the 25-29 years old, 15 per cent for the 30-34 year group, 20 per cent for the 35-39, and a 30-per cent drop for those 40 and over. The decline may be projected over five years, over ten or a longer period, depending upon what data is treated under the assumptions.

(5) Changes in fertility rates are also associated with social and cultural changes that take place through time, affecting potential mothers of all ages, although not necessarily in equal measure. A model reflecting this aspect of changing behaviour would manipulate the rates in such a way that succeeding cohorts were affected differently as they attained successive ages, rather than applying reduction factors differentially to the various ages as in the preceding illustration. This scheme is best understood through a diagram showing how each age cohort meets a different set of period rates as it ages. In the illustration, t, t+5, t + 10,

| Time: | t | t+5 | t+10 | t+15 | t+20 |
|---|---|---|---|---|---|
| Age group | | | | | |
| 15 - 19 | $f_1^a$ | $f_1^b$ | $f_1^c$ | $f_1^d$ | $f_1^e$ |
| 20 - 24 | $f_2$ | $f_2^a$ | $f_2^b$ | $f_2^c$ | $f_2^d$ |
| 25 - 29 | $f_3$ | | $f_3^a$ | $f_3^b$ | $f_3^c$ |
| 30 - 34 | $f_4$ | | | $f_4^a$ | $f_4^b$ |
| 35 - 39 | $f_5$ | | | | $f_5^a$ |
| 40 - 44 | $f_6$ | | | | |

etc... represent times in the projection period. The letter f is used to represent the age-specific fertility rates, further identified by a numerical subscript representing the age group to which the rate applies and a letter superscript which identifies the cohort. Thus $f_1^a$ represents the fertility rate which applies to cohort a during the first child-bearing period. Five years later, cohort b experiences fertility at a rate $f_1^b$ which has undergone a change from the rate representing fertility behaviour of the same age category five years earlier. Meanwhile cohort a has aged, reaching the 20-24 year group at t + 5, and is experiencing the rate $f_2^a$. The task is to assign assumed change factors to the various rates as they respond to social changes; to develop employment opportunities for women, extended education, family planning facilities, changing life styles and economic conditions.

29/ A discussion of model life tables appears in chapter VIII. For a more complete treatment, see United Nations Manual IV, *op. cit.*, p. 8 ff. The odd-numbered tables in the model series differ by five years in $\overset{o}{e}_o$ of females.

# V. MECHANICAL METHODS FOR PROJECTING SUBNATIONAL POPULATIONS

The basis for the selection of a method includes data availability, costs and benefits of added precision from added efforts, the purpose of the projection and the length of the projection period. Although there is no unanimity in the literature, it is useful to consider those projections to a horizon less than ten years distant as short-term, those to horizons between 10-25 years as medium term, and those of a longer as long-term.

The less detailed methods of extrapolating population growth trends are often quite adequate, especially for short-term projections, if the composition of the future population is not a requirement and if the rate of change is moderate. For more rapid growth, it is important to maintain as much detail as possible in order not to lose sight of the individual determinants of future population. This suggests use of the cohort survival or component method. In medium and long-term projections, the choice of method becomes especially important.

Inclusion of an introductory proviso is customary to qualify the assumptions underlying the projections and forecasts of population, and to note that unforeseeable conditions in the area under study would be likely to invalidate those assumptions. Such a proviso may be useful in emphasizing the limitations of all efforts to peer into the future.

Clearly, an assumed level and age pattern of fertility, migration and mortality would be invalidated by a war, a famine or an epidemic. While a development planner or demographer can attempt the projection of the probable effect of a family planning programme or improved public health or economic development, he cannot be expected to project the severity of a disaster or a sudden economic depression.

In the following pages, several non-analytical methods are illustrated which depend upon the assumption that the total population or designated part thereof follows a pattern of growth defined by an equation and represented by a straight line or a curve. From the simple assumption that the population changes each year by the amount it has in the immediate past, elaboration may take one of two courses. One of these is in the adoption of relatively complex curves describing change, while the other is in the direction of relatively complex application of past data to the curve.

In the first illustration which follows, linear projections were made of the population of subnational areas from data of two recent censuses. These were compared with projections made from an earlier intercensal period, but no attempt was made to relate linear change from one period to the other in arriving at a preferred future pattern. The illustration of geometric change, however, was made with reference to only one intercensal period. In the demonstration of the ratio method, two intercensal periods were used and the change in ratios was projected. A simpler assumption would have been that the ratios would remain unchanged at any one of the three censuses for which data are presented. The utility of such an assumption can readily be evaluated by a glance at table 6, which demonstrates their rapid change with the passage of time.

The selection and fitting of curves to describe future growth is a special case of the interpolation problem, to the extent that extrapolation is a special case of interpolation. Variables do not typically behave in a uniform, straight-line manner. It is theoretically possible to fit a curve to past behaviour, using an expression one degree less than the number of points to be fitted; e.g., a first-degree curve or straight-line can be fitted between two points, a quadratic curve through three points, and a cubic curve through four. The calculations are extremely laborious; although the effort may be warranted where intermediate data between points in a past series are required, the illusory precision it lends in extrapolation will not improve projections. [30]

If mechanical methods are to be used, a curve should be selected on the basis of what is known about past and probably future demographic behaviour, and the data available. It is certain that efforts devoted to data improvement yield far greater benefits than those of equivalent efforts devoted to seeking the curve which most accurately fits past data.

---

30/ See Shryock and Siegel, *op. cit.*, volume 2, p. 681 ff., for a discussion of various types of interpolation, with demonstrations. A result of the extrapolation of a curve fitted to seven data points is shown on p. 692.

## A. SELECTED MECHANICAL METHODS

### 1. Arithmetic projection

The very simplest model of population change would assume a straight-line pattern, or assume that increments to or decrements from the population each year are equal throughout the projection period. Since minimal effort is required, the method is justified where data are crude where detail is unnecessary and where the horizon is near. If the population $P_n$ is projected n years in the future from the present population $P_o$, and from the population m years in the past $P_m$, then

$$P_n = P_o + \frac{n(P_o - P_m)}{m}$$

or, to the present population is added, algebraically, the annual level of change between the most recent census and the one immediately preceding, multiplied by the number of years in the projection period.

### 2. Geometric projection

A useful and familiar formula is identical in form to that used in calculating the effects of compound interest upon a principal sum:

$$P_n = P_o (1 + r)^n$$

where $P_n$ represents the projected population after n years, and $P_o$ the population at the most recent census. A table of compound interest offers a ready means for rapid population projection, if the rate r and a horizon n years away can be calculated or assumed. Typically, selection of a rate of change is based upon the most recent performance of the population being projected; $\frac{P_o}{P_m}$ represents the ratio of the population at the benchmark point to that m years earlier at the time of a prior census. Alternatively, when the projected rate is assumed to be equal to that just prior to the benchmark date

$$P_o = P_m(1 + r)^m, \text{ and } (1 + r) = \frac{P_o^{\frac{1}{m}}}{P_m} = \frac{P_n^{\frac{1}{n}}}{P_o}$$

leading to $P_n = P_o\left(\frac{P_o}{P_m}\right)^{\frac{n}{m}}$

a convenient form for continuing a fixed rate of growth.

### 3. Exponential projection

In the application of the compound interest formula and in the population projection form with which it is identified, it is necessary to select a compounding interval as well as a rate. There is no reason other than convenience to compound annually. A limiting case, that of continuous or instantaneous compounding, is expressed by

$$P_n = P_o e^{rn}$$

here e is a mathematical constant equal to 2.71828, the base of the natural or Napierian logarithms. The analyst using the exponential model would find a table of $e^x$ convenient, which would provide a multiplier for calculating projections r.n from the benchmark population i.e. selecting x from the table equal to a chosen r.n. All of the change models just described must be used with caution.

During periods approaching the later phase of the demographic transition, from high vital rates to low, or during the latter part of a period of heavy in-migration, growth is diminishing from year to year. The fixed increment trend line is a straight line, sloping from lower left to upper right on a conventional graph, while the fixed rate is described by a curve, growing steeper from left to right. These patterns are apparent in the growth curves for Seoul and the Republic of Korea illustrated in figure 2.

The following examples illustrate the application of the straight-line and the simple geometric model in subnational projections, with recent data from the Republic of Korea. The census of 1970 reports figures for nine provinces (do) and two cities. The data from the census of 1960 were adjusted to reflect the 1963 annexation to the Special City of Seoul of areas containing an estimated 1960 population of 155,000, while those of 1966 and 1970 are presented as they were reported. [31] Table 6 shows the result of arithmetic or straight-line projections of the Republic of Korea, its provinces and two major cities, while table 7 presents projections of the same areas under the assumption of geometric change. It is interesting to note that a projection of the entire country to the 2000, based upon the simple assump-

31/ Economic Planning Board, Republic of Korea, *1970 Population and Housing Census Report*, volume 1, "Complete enumeration"; and *ibid.*, 1966.

tion of a continuation of the geometric mean rate of growth experienced between the censuses of 1966 and 1970, falls within the range suggested by Cho [32] The result of the application of a straight-line growth model based upon the experience of the same four years is 48.5 million, a level well within the realm of possibility, should fertility decline more rapidly than expected. The subnational projections contained in both tables are merely illustrative, and embody none of the insights a development planner would be expected to use.

The notation of the formulas shown in the earlier paragraphs of this chapter has been employed in the column headings. Occasionally an operation performed upon the items in a column is indicated as if it were performed upon the column, identified by its number in parentheses. The total populations of the country and its subnational areas as of the dates of the two censuses appear in table 6 while table 7 begins with the percentage change in the population during the intercensal period.

A comparison of the projections contained in the two tables is instructive. Those in table 6 contain no surprises; the subnational parts add to the whole, and their shares change over time as the rapidly growing cities and provinces increase their shares, at the expense of the more slowly growing parts and of those whose populations are declining.

The dominance of Seoul deserves comment. With 13 per cent of the Republic's population in 1966 and 17.5 per cent in 1970, the straight-line method projects 38.1 per cent in the year 2000, when the city's population would be 18.5 million. Table 7, illustrating the effects of the geometric rate, shows a population of 92.8 million for Seoul 30 years in the future, which would be two-thirds larger than the country of which it was a part! The numbers demonstrate the hazards of the application of fixed patterns of growth. Judging by experience in other countries, Seoul is unlikely to grow to 18.5 million in 30 years; with the passage of time, the competitive advantages of other places would become apparent, industries and population would locate elsewhere, and the pattern of growth of Seoul would change.

In the more slowly growing provinces, the two approaches differ far less, and a useful subnational projection by single calendar year toward a ten-year horizon might prove very useful for plans which do not demand details of population composition. The methods shown are least effective where population growth (or decline) is most rapid, and where the necessary span of the projection exceeds a decade. It would appear from the evidence of the 1966 and 1970 censuses that, excepting Seoul, Busan and possibly Jeju Do, useful projections of the total populations of the provinces of Korea can be prepared by the relatively simple methods shown. It remains to determine what insights an examination of another intercensal period might yield on the question of past stability of annual population change.

Tables 8 and 9 present comparable population projections for the Republic and for the same subnational areas, using the growth experience of the 1960-to-1966 intercensal period. Since the interval is 70 months long, not an integral number of years, the annual change in the arithmetic projection was calculated by taking 12/70 of the intercensal change. In a geometric projection, the calculated rate depends upon the frequency of compounding. Since these demonstrations of geometric projections employ a one-year compounding interval, the annual change ratio (1 plus the rate of change) was calculated in table 9 by the expression:

$$\text{antilog}\left(\frac{12}{70}\log\left(\frac{P_o}{P_m}\right)\right)$$

where $P_o$ represents the 1966 benchmark population, and $P_m$ represents the population at the beginning of the intercensal period.

A comparison of the projections from the 1970 census data with those from the 1966 data is instructive. Information from the 1970 census shows the extent to which the projections from 1966 succeed in their forecast. For the entire country, the 1970 figure projected under the straight-line assumption was 1.8 per cent too high, while the geometric projection was too high by 3.1 per cent. The cities of Seoul and Busan grew far more rapidly than might have been anticipated by the projections alone. The former exceeded its arithmetic projection by 19.8 per cent, its geometric one by 12.4 per cent. The decline in fertility appeared far more precipitious in the later years of the decade, a time when migration to the major cities must have been rising dramatically. The sociological relationship between these two trends is outside the scope of this manual; the fact that there is a relationship increases the hazards of preparing population projections for subnational areas, and imposes upon the analyst the obligation to interpret his work with the greatest care. Table

32/ Lee Jay Cho, "The demographic situation in the Republic of Korea", paper presented at the Annual Meeting, Population Association of America, at New Orleans, 1973.

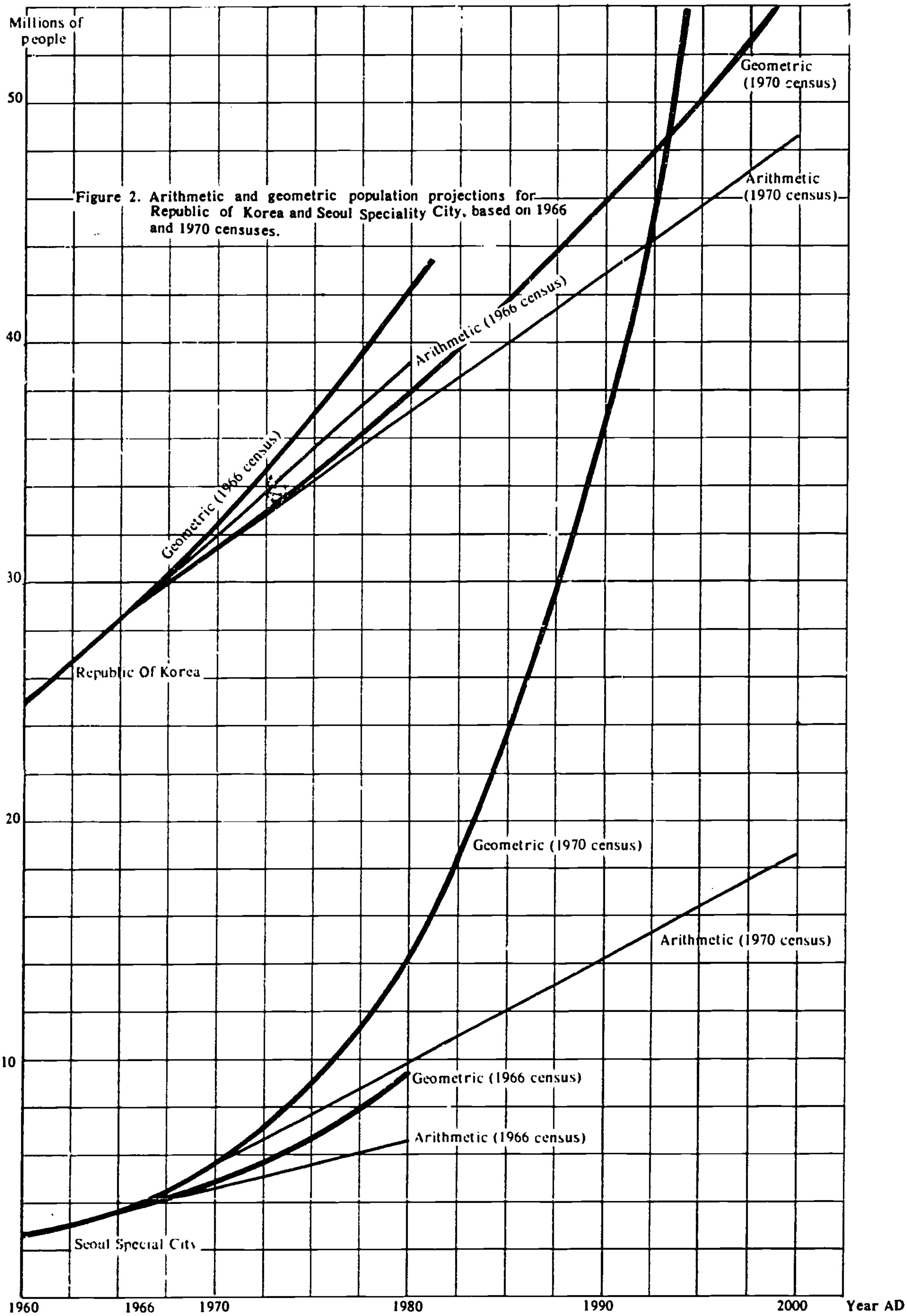



Figure 2. Arithmetic and geometric population projections for Republic of Korea and Seoul Speciality City, based on 1966 and 1970 censuses.

Table 6. Arithmetic projection: Populations of provinces (Do) and two major cities, Republic of Korea, projected to 1980, 1990, and 2000; assuming uniform annual arithmetic rate of change

| City and province | Population: October 1970 $P_o$ | Population: October 1966 $P_m$ | $P_o - P_m$ (1)-(2) | $\frac{(P_o - P_m)}{m}$ (3)÷4 | Projected change: 10 years (4)x10 | Projected change: 20 years (4)x20 | Projected change: 30 years (4)x30 | Projected population: 1980 $(P_n)$ (1)+(5) | Projected population: 1990 $(P_n)$ (1)+(6) | Projected population: 2000 $(P_n)$ (1)+(7) |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Entire country | 31,435,252 | 29,159,640 | 2,275,612 | 568,903 | 5,689 | 11,378 | 17,067 | 37,124 | 42,813 | 48,502 |
| Seoul Special City | 5,525,262 | 3,793,280 | 1,731,982 | 432,996 | 4,330 | 8,660 | 12,990 | 9,855 | 14,185 | 18,515 |
| Busan City | 1,876,391 | 1,426,019 | 450,372 | 112,593 | 1,126 | 2,252 | 3,378 | 3,002 | 4,128 | 5,254 |
| Gyeonggyi Do | 3,353,272 | 3,102,325 | 250,947 | 62,737 | 627 | 1,255 | 1,882 | 3,981 | 4,608 | 5,235 |
| Gangweon Do | 1,865,426 | 1,831,185 | 34,241 | 8,563 | 86 | 171 | 257 | 1,951 | 2,037 | 2,122 |
| Chungcheonbug Do | 1,480,338 | 1,548,821 | -68,483 | -17,121 | -171 | -342 | -514 | 1,309 | 1,138 | 967 |
| Chungcheongnam Do | 2,858,202 | 2,902,941 | -44,739 | -11,185 | -112 | -224 | -336 | 2,746 | 2,635 | 2,523 |
| Jeonlabug Do | 2,431,892 | 2,521,207 | -89,315 | -22,329 | -223 | -447 | -670 | 2,209 | 1,985 | 1,762 |
| Jeonlanam Do | 4,004,832 | 4,048,769 | -43,937 | -10,984 | -110 | -220 | -330 | 3,895 | 3,785 | 3,67 |
| Gyeongsangbug Do | 4,555,866 | 4,472,895 | 82,971 | 20,743 | 207 | 415 | 622 | 4,763 | 4,971 | 5,17[illegible] |
| Gyeongsangnam Do | 3,118,634 | 3,175,146 | -56,512 | -14,128 | -141 | -283 | -424 | 2,977 | 2,836 | 2,695 |
| Jeju Do | 365,137 | 337,052 | 28,085 | 7,021 | 70 | 140 | 211 | 435 | 506 | 576 |

Table 7. Geometric projection: Populations of provinces and two major cities, Republic of Korea, projected to 1980, 1990, and 2000; assuming uniform annual geometric rate of change

| City and province | Intercensal percentage change | Intercensal relative change | $\sqrt{(2)}$ | Annual rate $\sqrt[m]{P_o/P_m}$ | Relative change: 10 years $(4)^{10}$ | Relative change: 20 years $(4)^{20}$ | Relative change: 30 years $(4)^{30}$ | Projected population: 1980 $P_o$x5 | Projected population: 1990 $P_o$x6 | Projected population: 2000 $P_o$x7 |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Entire country | 7.80 | 1.0780 | 1.0382 | 1.0189 | 1.206 | 1.455 | 1.755 | 37,917 | 45,738 | 55,169 |
| Seoul Special City | 45.66 | 1.4566 | 1.2069 | 1.0986 | 2.561 | 6.557 | 16.791 | 14,149 | 36,229 | 92,775 |
| Busan City | 31.58 | 1.3158 | 1.1471 | 1.0710 | 1.986 | 3.944 | 7.834 | 3,727 | 7,400 | 14,700 |
| Gyeonggyi Do | 8.09 | 1.0809 | 1.0397 | 1.0197 | 1.215 | 1.476 | 1.793 | 4,074 | 4,949 | 6,012 |
| Gangweon Do | 1.87 | 1.0187 | 1.0093 | 1.0047 | 1.047 | 1.097 | 1.149 | 1,955 | 2,046 | 2,143 |
| Chungcheonbug Do | -4.42 | 0.9558 | 0.9776 | 0.9887 | 0.893 | 0.797 | 0.712 | 1,322 | 1,180 | 1,054 |
| Chungcheongnam Do | -1.54 | 0.9846 | 0.9922 | 0.9961 | 0.962 | 0.925 | 0.889 | 2,748 | 2,644 | 2,541 |
| Jeonlabug Do | -3.54 | 0.9646 | 0.9821 | 0.9910 | 0.913 | 0.835 | 0.763 | 2,222 | 2,031 | 1,856 |
| Jeonlanam Do | -1.09 | 0.9891 | 0.9945 | 0.9972 | 0.973 | 0.947 | 0.921 | 3,896 | 3,793 | 3,688 |
| Gyeongsangbug Do | 1.85 | 1.0185 | 1.0092 | 1.0046 | 1.047 | 1.096 | 1.147 | 4,770 | 4,993 | 5,226 |
| Gyeongsangnam Do | -1.78 | 0.9822 | 0.9911 | 0.9956 | 0.956 | 0.914 | 0.874 | 2,982 | 2,850 | 2,726 |
| Jeju Do | 8.33 | 1.0833 | 1.0408 | 1.0202 | 1.221 | 1.492 | 1.822 | 446 | 545 | 665 |

Table 8. Populations projected to 1970 and 1980 for provinces and two major cities, Republic of Korea, using 1966 benchmark and assuming arithmetic change

| | Population | | 70 months change | average annual | Projected population | |
|---|---|---|---|---|---|---|
| | October 1966 | December 1960 | 1960-1966 | | 1970 | 1980 |
| | $P_o$ | $P_m$ | $P_o - P_m$ | 12/70 x (3) | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Entire country | 29,159,640 | 24,989,241 | 4,170,399 | 714,923 | 32,020 | 39,169 |
| Seoul Special City | 3,793,280 | 2,600,402 a/ | 1,192,878 | 204,493 | 4,611 | 6,656 |
| Busan City | 1,426,019 | 1,163,671 | 262,348 | 44,974 | 1,606 | 2,056 |
| Gyeonggyi Do | 3,102,325 | 2,593,765 a/ | 508,560 | 87,181 | 3,451 | 4,322 |
| Gangweon Do | 1,831,185 | 1,636,767 | 194,418 | 33,329 | 1,964 | 2,298 |
| Chungcheonbug Do | 1,548,821 | 1,369,780 | 179,041 | 30,693 | 1,672 | 1,979 |
| Chungcheongnam Do | 2,902,941 | 2,528,133 | 374,808 | 64,252 | 3,160 | 3,803 |
| Jeonlabug Do | 2,521,207 | 2,395,224 | 125,983 | 21,597 | 2,607 | 2,823 |
| Jeonlanam Do | 4,048,769 | 3,553,041 | 495,728 | 84,981 | 4,389 | 5,239 |
| Gyeongsangbug Do | 4,472,895 | 3,848,424 | 624,471 | 107,052 | 4,901 | 5,972 |
| Gyeongsangnam Do | 3,175,146 | 3,018,371 b/ | 156,775 | 26,876 | 3,283 | 3,551 |
| Jeju Do | 337,052 | 281,663 | 55,389 | 9,495 | 376 | 470 |

Notes: a/ Includes adjustment for 1963 annexation to Seoul.
b/ With Busan removed.

Table 9. Populations projected to 1970 and 1980 for provinces and two major cities, Republic of Korea, using 1966 benchmark and assuming geometric change

| | Relative change | Log (1) | 12/70x (2) | Relative change | | | Projected population | |
|---|---|---|---|---|---|---|---|---|
| | | | | 1 year | 4 years | 10 years | 1970 | 1980 |
| | 70 months | | | Antilog (3) | $(4)^4$ | $(4)^{10}$ | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Entire country | 1.1669 | 0.067031 | 0.011490 | 1.0268 | 1.1116 | 1.3028 | 32,414 | 42,229 |
| Seoul Special City | 1.4587 | 0.163966 | 0.028108 | 1.0669 | 1.2957 | 1.9102 | 4,915 | 9,389 |
| Busan City | 1.2254 | 0.088295 | 0.015136 | 1.0355 | 1.1497 | 1.4170 | 1,639 | 2,323 |
| Gyeonggyi Do | 1.1961 | 0.077752 | 0.013329 | 1.0312 | 1.1308 | 1.3592 | 3,508 | 4,768 |
| Gangweon Do | 1.1188 | 0.048745 | 0.008356 | 1.0194 | 1.0799 | 1.2122 | 1,977 | 2,397 |
| Chungcheonbug Do | 1.1307 | 0.053350 | 0.009145 | 1.0213 | 1.0880 | 1.2344 | 1,685 | 2,080 |
| Chungcheongnam Do | 1.1483 | 0.059026 | 0.010118 | 1.0236 | 1.0978 | 1.2624 | 3,187 | 4,023 |
| Jeonlabug Do | 1.0526 | 0.022263 | 0.003816 | 1.0088 | 1.0357 | 1.0918 | 2,521 | 2,753 |
| Jeonlanam Do | 1.1395 | 0.056722 | 0.009723 | 1.0226 | 1.0935 | 1.2510 | 4,427 | 5,539 |
| Gyeongsangbug Do | 1.1623 | 0.065050 | 0.011151 | 1.0260 | 1.1081 | 1.2928 | 4,956 | 6,408 |
| Gyeongsangnam Do | 1.0519 | 0.021991 | 0.003769 | 1.0087 | 1.0353 | 1.0907 | 3,287 | 3,585 |
| Jeju Do | 1.1966 | 0.077967 | 0.013365 | 1.0313 | 1.1312 | 1.3603 | 381 | 519 |

10 compares the results of the two projections from 1966 to the reported populations from the census of 1970, and displays the results of the four projections to 1980. It is apparent that projections employing mechanical methods can lead to serious errors if they are based upon data collected during times of unusual redistribution of population.

### 4. Use of a predetermined total population

In the foregoing demonstrations, total population projections for the entire Republic were calculated along with those for the subnational areas. Where acceptable estimates of vital rates are available for an entire country, and where international migration

Table 10. Comparisons of projected populations of two major cities and nine provinces, Republic of Korea, with data from 1970 census and projections to 1980 under four assumptions

(in thousands)

| | 1970 census | Projections to 1970 (from 1966 benchmark) | | | | Projections to 1980 (from two benchmarks) | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Arithmetic | | Geometric | | Arithmetic | | Geometric | |
| | | | Per cent error | | Per cent error | from 1966 | from 1970 | from 1966 | from 1970 |
| Entire country | 31,435 | 32,020 | 1.86 | 32,414 | 3.11 | 39,169 | 37,124 | 42,229 | 37,917 |
| Seoul Special City | 5,525 | 4,611 | -16.54 | 4,915 | -11.04 | 6,656 | 9,855 | 9,389 | 14,149 |
| Busan City | 1,876 | 1,606 | -14.39 | 1,639 | -12.63 | 2,056 | 3,002 | 2,323 | 3,727 |
| Gyeonggyi Do | 3,353 | 3,451 | 2.92 | 3,508 | 4.62 | 4,322 | 3,981 | 4,768 | 4,074 |
| Gangweon Do | 1,865 | 1,964 | 5.31 | 1,977 | 6.01 | 2,298 | 1,951 | 2,397 | 1,955 |
| Chungcheonbug Do | 1,480 | 1,672 | 12.97 | 1,685 | 13.85 | 1,979 | 1,309 | 2,080 | 1,322 |
| Chungcheongnam Do | 2,858 | 3,160 | 10.57 | 3,187 | 11.51 | 3,803 | 2,746 | 4,023 | 2,748 |
| Jeonlabug Do | 2,432 | 2,607 | 7.20 | 2,521 | 3.66 | 2,823 | 2,209 | 2,753 | 2,222 |
| Jeonlanam Do | 4,005 | 4,389 | 9.59 | 4,427 | 10.54 | 5,239 | 3,895 | 5,539 | 3,896 |
| Gyeongsangbug Do | 4,556 | 4,901 | 7.57 | 4,956 | 8.78 | 5,972 | 4,763 | 6,408 | 4,770 |
| Gyeongsangnam Do | 3,119 | 3,283 | 2.69 | 3,287 | 5.39 | 3,551 | 2,977 | 3,585 | 2,982 |
| Jeju Do | 365 | 376 | 3.01 | 381 | 4.38 | 470 | 435 | 519 | 446 |

Table 11. Projection of the populations of nine provinces and two major cities, Republic of Korea, by ratio method

| | Ratio of province to country | | | | | 1970 projected population | Per cent error |
|---|---|---|---|---|---|---|---|
| | 1960 census | 1966 census | 1970 projected | 1970 adjusted | 1970 census | | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Entire country | 100.000 | 100.000 | 99.649 | 100.000 | 100.000 | 32,020 | 1.85 |
| Seoul Special City | 10.406 | 13.009 | 14.849 | 14.901 | 17.577 | 4,771 | -13.64 |
| Busan City | 4.657 | 4.890 | 5.050 | 5.068 | 5.969 | 1,623 | -13.48 |
| Gyeonggyi Do | 10.380 | 10.639 | 10.817 | 10.855 | 10.667 | 3,476 | 3.66 |
| Gangweon Do | 6.550 | 6.280 | 6.095 | 6.116 | 5.934 | 1,958 | 4.98 |
| Chungcheonbug Do | 5.481 | 5.312 | 5.196 | 5.214 | 4.709 | 1,670 | 12.83 |
| Chungcheongnam Do | 10.117 | 9.955 | 9.844 | 9.879 | 9.092 | 3,163 | 10.67 |
| Jeonlabug Do | 9.585 | 8.646 | 8.002 | 8.030 | 7.736 | 2,571 | 5.71 |
| Jeonlanam Do | 14.218 | 13.885 | 13.657 | 13.705 | 12.740 | 4,388 | 9.56 |
| Gyeongsangbug Do | 15.400 | 15.339 | 15.297 | 15.351 | 14.493 | 4,916 | 7.90 |
| Gyeongsangnam Do | 12.079 | 10.889 | 9.666 | 9.700 | 9.921 | 3,106 | -0.41 |
| Jeju Do | 1.127 | 1.156 | 1.176 | 1.180 | 1.162 | 378 | 3.56 |

is either known or assumed to be negligible, a satisfactory projection of the population of the country may have been prepared by more analytical methods than are possible for the subnational areas. Under these circumstances, it is the task of the analyst to distribute the acceptable total among the subnational areas. Having prepared population projections of subnational areas, only the ratio of the total for any future year with that of the independently prepared projection need be calculated, and the subnational figures need be diminished or augmented by the appropriate amount in order to add to the accepted total. This simple technique is demonstrated as a part of the ratio method in table 11, but it can be used wherever there is a pre-existing national total.

## B. RATIO METHODS

The simplest ratio method can be illustrated by uniform and probably moderate growth as reflected in two successive censuses, which report that the proportion of the national population resident in one province was very nearly equal at the two census dates. The assumption might be justified that, in the absence of development programmes specifically for the subnational area under study, the proportion would remain fixed throughout the projection period, and the rate of future growth would be the same as that of the nation. This rather unsophisticated projection might prove quite useful for a short term. Its chief advantage lies in the ease of calculation, especially where a large number of subnational areas are to be projected and an acceptable projection for the country has been prepared.

This third mechanical method is demonstrated in table 11. In the method, the projected variable is the ratio of each part to the whole, rather than a rate of change or an annual increment to population. The total projected population required by the method is that of the linear projection to 1970, based upon the 1960 and 1966 data shown in table 8, column 5 of the projection shown is based upon the single assumption that the changes in the ratios of the populations of the provinces and major cities of the Republic of Korea to that of the whole country, which took place between the censuses of 1960 and 1966, would continue until a third census, that of 1970.

If $P_m$ is the population at the earlier of two censuses, and $P_o$ the population at the second of two censuses m time-periods later, and if the ratios of the subnational populations to the total take the form $\frac{{}_mP_i}{{}_mP}$ and $\frac{{}_oP_i}{{}_oP}$ (equal, respectively, to ${}_mr_i$ and ${}_or_i$), then $\Sigma {}_mr_i = {}_or_i = 1.000$. Each of the ratios can be projected according to a straight-line assumption, as follows:

$$ {}_nr_i = {}_or_i + \frac{{}_or_i - {}_mr_i}{m} \quad \text{and} \quad \Sigma {}_nr_i \neq 1.000 = {}_nS $$

where ${}_nr_i$ equals the projected ratio of the individual subnational area, n periods earlier, unadjusted. The adjusted ratio is represented by $\frac{1.000}{{}_nS} \cdot {}_nr_i$

The formula for the population of the subnational area is therefore

$$ \frac{1.000}{{}_nS} \cdot {}_nr_i \cdot {}_nP $$

where ${}_nP$ is the total national population at time n.

In table 11, the calculated ratios of the populations of the subnational areas to their total for the censuses of 1960 and 1966 are shown in columns 1 and 2, respectively. The differences are calculated and projected by simple linear extrapolation, by an amount proportional to the number of months to the census of 1970, and are shown in column 3. The projected ratios of column 3 were "raked" or ratioed, to add to 100.000 in column 4; and the projected populations as of 1 October 1970 were calculated, column 6. Column 5 is included to show the actual ratios for 1970, for comparative purposes.

A comparison of column 7 with the third column of table 10 suggests some improvement in accuracy over the straight-line projections. In most cases, a smaller error is likely to ensue if a relationship, fixed or changing, is assumed among the parts of the whole, more so than if the parts are assumed to change without regard to each other. For this reason, the ratio methods are usually to be preferred over other mechanical methods dependent only upon arithmetic manipulation of the populations of the parts.

The unusual growth of Seoul and Busan during the period between the 1966 and the 1970 censuses poses problems for the planner not amenable of solution by demographic methods of projection. All three of the demonstrated mechanical methods seriously underestimated the 1970 population of the two cities; the most successful of the three was the geometric method precisely, because of its characteristic compounding of growth. Yet, it is clear from the long-term projection of the geometric growth trend that the geometric method cannot be used except under conditions of modest growth for limited projection periods. There is nothing in the illustrative material from the censuses of the Republic of Korea which would, in 1966, have forewarned of the changes which took place in the four subsequent years. Furthermore, there is nothing in the data which would suggest a future limit to this rapid growth.

**Limitations on growth or decline.** Recent environmental literature is replete with examples of the consequences of unrestricted growth, and these are usually based upon an exponential or geometric model. Extreme examples of an ultimate "ball of humanity expanding outward with the speed of light", or a world with "standing room only", or even a world-wide density approaching present maximum national densities, will not take place because of critical needs for air, water or food.

In more local projecting, the development planner should be aware that the competitive advantages of less populous areas may become apparent to potential migrants as some areas reach densities appropriate to their ecological or economic base. Although tentative, some notion of a maximum or optimum population held by the development planner is an important ingredient of a population forecast or projection for a subnational area. A clear violation of a planning figure by a mechanical projection demands appropriate action. Either the projected trends foreshadow a violation of existing plans, calling for a revision; or they indicate the need for alternative projections; or both. This type of consideration need not concern the planner interested only in the probable future of an area as yet insufficiently populous to be subject to constraint. Nor need it be a major consideration of the planner concerned only with short-term projections.

The various subnational areas of a country grow at different rates, or grow by different annual increments, primarily because of the differential effects of internal migration. This rather unstable component of population change responds to varying opportunities for employment or to other circumstances which motivate people to move. Although migration has always characterized human settlement; it is not reasonable to assume that a current migration pattern necessarily continues indefinitely. Subnational population must be projected at a reasonable level of future growth, in the direction of the restoration of equilibrium among competing areas, in the extent to which they attract migrants. Alternatively, planning decisions may affect a part or all of the net migration in any one area, diverting it to others.

In the preparation of subnational population projections by the mechanical methods, the planner must manipulate annual or periodic change in accordance with his view of the future. In an arithmetic projection, for example, it is frequently assumed that the annual growth experienced by the country as a whole continues, but that it would be differently distributed. It may be assumed, for example, that annual increase to the major metropolitan province would cease in 20 years, that this decline takes place uniformly, and that each year an additional 5 per cent of the annual increment would be diverted to other provinces which are expected to develop later. Similarly, the province which has experienced a decline would probably not become totally depopulated. It is reasonable to assume that the annual decrements approach zero by uniform steps. The calculations describing such would progressively reduce the negative annual population change of the province, by diverting it from growing areas in which annual increments are assumed to decline.

The annual increase in the population of Seoul during the period 1966 to 1970 was 432,996 (from table 6, column 4). In the illustrative straight-line projection, the population of the city in 1980 is 9.855 million. If it were assumed that the annual increments would decline linearly to zero in ten years following 1970, the change in the city's population during the decade could be calculated from the formula for the sum of an arithmetic series:

$$S = \frac{N}{2}(A + L)$$

and substituting:

$$S = \frac{10}{2}(0.433 \text{ million} + \text{zero})$$

where N equals the number of terms, and A and L are, respectively, the first and the last terms of the series. The sum of the ten annual increments would thus be 2.165 million, rather than the 4.330 million in the original projection. The decrease would be allocated to other areas if the total country population is to remain unchanged. Should the projections for the country and other subnational areas be continued beyond the year in which the area in question attains zero growth, all of the growth it would have received must be reallocated. There is no simple formula for such a reallocation. The answer lies in national or regional development plans. The pace of development elsewhere must be such that the expected population can be successfully diverted.

## C. THE LOGISTIC CURVE

The problem of finding a model for plotting a limitation of growth is solved more elegantly by the logistic, or S-shaped, curve. Like the continuous compound-interest model and the Gompertz curve which it resembles, the logistic curve is exponential, i.e. the independent variable in the equation or formula is an exponent. The curve describes the growth of a population which increases ever more rapidly over time, describing a curve which is concave upward until a maximum rate of growth is achieved. At this point of inflection or direction change, the rate of growth diminishes, and the curve becomes concave downward as it approaches asymptotically a maximum value described by a horizontal line. The curve is an interesting model for describing past growth. It has been used in the description of population change in a colony of fruit flies with limited space, as well as bacterial cultures, and has in the past received considerable attention from demographers. It is probably more useful in interpolation than in extrapolation, a point that has been demonstrated by Shryock, Siegel et al. [33] who also

33/ Op. cit., pp. 382-385, 690-691

demonstrate the laboriousness of the calculations which it requires. Its utility in population projecting is by no means confined to plotting population magnitudes; it can describe a changing relationship between two parts of a population, and it can be used to plot intermediate values as a rate of change approaches zero. In the logistic curve, the increments of change can be plotted in a bell-shaped symmetrical form; in the Gompertz curve, the plot of the increments is skewed, but the distinction is not relevant in the limited applications appropriate to subnational population projections. An expression for the logistic curve is

$$Y = \frac{k}{1 + e^{a+bX}}$$

where k represents the upper asymptote; a is a constant value which establishes the position of the curve on the horizontal or time axis; b is a negative constant in an ascending growth curve, of which the absolute value reflects the steepness of the curve; and e is the base of the natural or Napierian logarithms.

The use of the logistic curve in plotting population change can be demonstrated by using figures for Seoul. In the illustration, it is assumed that the city had attained a maximum growth rate midway in the 1966-1970 period, on 1 October 1968, when its population was 4,659,000. Furthermore, it is assumed that Seoul would attain a maximum population 12 years later, in October 1980, and that 99.0 per cent of that figure would have been reached two years earlier in 1978. The growth model would be the latter half of a logistic curve defined by

$$\frac{2 \times 4{,}659}{1 + e^{bX}}$$

where the numerator is twice the starting population (because in the logistic model the maximum rate of growth occurs at the mid-point in the cycle); and where the constant term in the exponent of e is omitted, because the starting point 1968 is regarded as 0.

The parameter b must be evaluated, such that 99.0 per cent of the final population is attained within ten years:

$$0.99 \times 9318 = \frac{9318}{1 + e^{10b}}$$

$$1 + e^{10b} = \frac{1}{0.99} = 1.0101$$

$$10b \log e = \log 1.0101$$

$$b = -0.45952 \text{ or } -0.46$$

The equation for plotting the change in Seoul's population, from a maximum in 1968 to a level in 1980, assuming 99 per cent of the final population in 1978, is

$$Y = \frac{9318}{1 + e^{-0.46X}}$$

Projections of annual populations of Seoul, 1968-1980, using a logistic model.

| Year = X (1) | bX (2) | $e^{bX}$ (3) | $1+e^{bX}$ (4) | Population (5) |
|---|---|---|---|---|
| 0 | 0 | 1.0000 | 2.0000 | 4,659 |
| 1 | -0.46 | 0.6313 | 1.6313 | 5,712 |
| 2 | -0.92 | 0.3985 | 1.3985 | 6,663 |
| 3 | -1.38 | 0.2516 | 1.2516 | 7,449 |
| 4 | -1.84 | 0.1588 | 1.1588 | 8,041 |
| 5 | -2.30 | 0.1003 | 1.1003 | 8,469 |
| 6 | -2.76 | 0.0633 | 1.0633 | 8,763 |
| 7 | -3.22 | 0.0400 | 1.0400 | 8,960 |
| 8 | -3.68 | 0.0252 | 1.0252 | 9,089 |
| 9 | -4.14 | 0.0159 | 1.0159 | 9,172 |
| 10 | -4.60 | 0.0101 | 1.0101 | 9,225 |
| 11 | -5.06 | 0.0063 | 1.0063 | 9,260 |
| 12 | -5.52 | 0.0040 | 1.0040 | 9,281 |

Calculations of the projected population, by year from 1968 to the assumed terminal year 1980 are shown in the table below. Because of the shape of the curve, projected values of the population of Seoul are higher at all times between 1968 and 1980; beyond that year, if the logistic curve were continued it would be horizontal, while the arithmetic curve would continue its upward course.

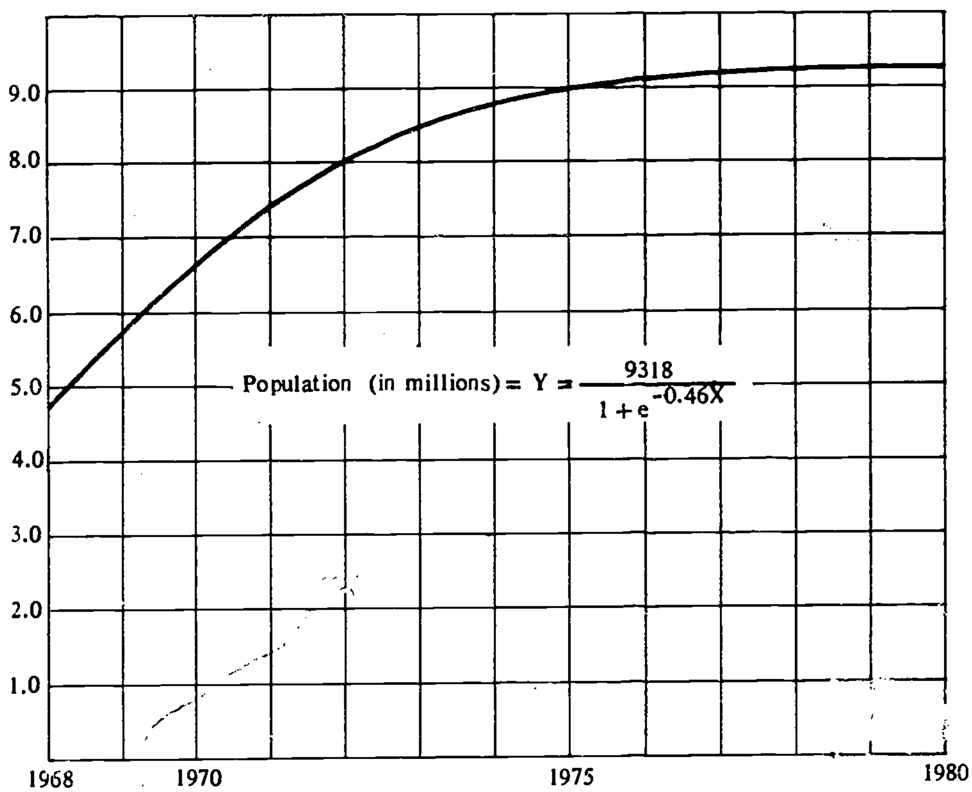
The logistic curve has been used to prepare projections of the urban population of a country, in a method which is instructive in subnational population projecting. But the method requires an existing projection of the population of the country.[34] The percentage of the population which is urban at time t is calculated from

$$\frac{100U_t}{T_t} = \frac{100e^{dt}}{1+e^{dt}}$$

50

34/ United Nations, Manual VIII, op. cit.

Figure 3. Logistic plot: An illustrative projection of the total population for Seoul Special City, 1968-1980.




51

By comparison with the form of the logistic curve presented above, it is apparent that Y, the dependent variable, is here represented by $\frac{100 U_t}{T_t}$; or the percentage of the total population which is urban. The upper asymptote K has the assigned value 100 $e^{dt}$; and dt varies from a large negative value to a large positive value, from -480 to +479 in the example. At the negative extreme, the calculated value of Y is 0.8163 per cent urban; at the 0 value, the population is 50 per cent urban; if dt were calculated for +480 (it was not in the example), it would be 99.1837. Thus, the total time represented by the cycle, from approximately 0 per cent urban to essentially 100 per cent, is 960 "table years". The model could be used for rapid or slow shift in urbanization, depending upon the length of time assigned to the table year. If each were assumed to be a month in length, the complete shift from a virtually rural to a substantially urban population would require 80 years, with attainment of the 50 per cent level in forty years. A logistic curve used in this manner would probably yield a better approximation of a growth pattern for a longer time than either the straight line or the simple geometric. The planner would assumed the appropriate point of the present population on the logistic curve, by an analysis of past population change and an evaluation of the present percentage of the theoretical maximum or "optimum" population being planned for. The logistic transformation would tell him the probable future pattern of growth to the extent that the curve is a valid model.

# VI. ANALYTIC METHODS OF SUBNATIONAL POPULATION PROJECTION

In any forecast, the results are more satisfactory if the task is broken into smaller units for individual study. In population forecasting or projecting, the past behaviour of each age and sex grouping may be examined in order to formulate satisfactory assumptions about future behaviour; or, in a similar manner, the trends in migration, births and deaths, may be examined; or both. Where a population consists of two or more ethnic groups, and data are available for the groups individually, projections may be made separately for each ethnic grouping in order that the assumptions for the future reflect probable differences in ethnic fertility, mortality or migration behaviour.

An analytic method is one which manipulates either the components of change separately or the parts of the population separately. Shryock and Siegel [35] concede that adequate testing has not been carried out to conclude that the analytic methods provide more reliable or "realistic" results than the mechanical methods, but they appear to be superior because they can account for specific information on the components of change, and permit the formulation of assumptions based upon more detailed knowledge than the mechanical methods.

## A. COHORT METHODS

In its deliberations, the Working Group on Projections of Population of Subnational Areas agreed that the cohort survival or component methods were preferable for subnational projections, if they could be used within the limitations of the data bases of the individual countries. They are preferable for the relative richness of planning detail which they yield and being analytic, preferable for the greater degree of control over the variables.

In the demographic literature, a "cohort" consists of a population born in a specific year or span of years, usually of one sex. The "components" of population change are the variables into which population growth or decline can be analysed: births, deaths, in-migration, out-migration, or their rates. A cohort projection is one which projects the individual age/sex grouping separately; a component projection is one which calculates the components separately.

35/ *Op. cit.* p. 796.

A projection is analytical if it distinguishes future change owing to migration from that to natural increase (the excess of births over deaths) or decrease. It is more analytical if the projections of the components of change are specific for the age and sex groups. In the first illustration of an analytic method, the components are disregarded but the projections are made for specific cohorts. The technique is essentially that of Shryock and Siegel, adapted to the female population of one of the four regions of Thailand. The population of each five-year age group is projected from 1970 to 1980, using the experience of the intercensal period 1960 to 1970. Numbers in age groups based on births are calculated from the numbers of women of appropriate ages, by ratio. The term, "cohort survival-ratio" is usually applied to this factor, and the particular method may be called the "method of cohort survival ratios"

## B. METHOD OF COHORT SURVIVAL RATIOS

Using data from the southern region of Thailand, a cohort survival ratio (CSR) between 1960 and 1970 for each five-year cohort of the female population was calculated in table 12. A cohort experiencing no mortality and no net migration, which was accurately counted in the two censuses, would exhibit a CSR of unity. The fact that the CSR is greater than unity in all four Thai regions for both males and females aged 10-to-14 indicates a systematic underenumeration of those 0-to-4 ten years earlier. The method does not correct for this, nor is any attention given to either mortality or migration as they affect the CSRs. Those under ten years of age at the end of the projection period of ten years were born after the later census; consequently their number cannot be calculated by application of a CSR. In columns 2 and 3 of the table, two additional ratios are calculated for each of the two years for which census data are available. These are the ratio of females 0-4 years of age to females 15-44, and that of females 5-9 to females 20-49. The values for 1960 and 1970 suggest that the ratios for 1980 would further increase but perhaps by an amount only half of that experienced between 1960 and 1970, reflecting lower fertility; the effect of reduced fertility is mitigated by lower infant mortality and perhaps lower mortality among woman aged 15-to-49.

The child/woman ratio of 0.430 was used to

Table 12. Female population of the south region of Thailand projected to 1980, using cohort survival ratios

| Age | Enumerated population | | CSR 1960-1970 | Projected population |
|---|---|---|---|---|
| | 1960 a/ | 1970 b/ | | 1980 |
| 0-4 | 246,385 | 349,900 | ... | 482,300 |
| 5-9 | 239,999 | 328,319 | ... | 482,200 |
| 10-14 | 178,644 | 268,638 | 1.0903 | 381,500 |
| 15-19 | 149,637 | 222,439 | 0.9268 | 304,300 |
| 20-24 | 149,719 | 155,507 | 0.8705 | 233,800 |
| 25-29 | 134,838 | 137,665 | 0.9200 | 204,600 |
| 30-34 | 111,872 | 135,715 | 0.9065 | 141,000 |
| 35-39 | 82,350 | 117,184 | 0.8691 | 119,600 |
| 40-44 | 71,545 | 97,511 | 0.8716 | 118,300 |
| 45-49 | 59,141 | 71,124 | 0.8637 | 101,200 |
| 50-54 | 55,511 | 60,991 | 0.8525 | 83,100 |
| 55-59 | 43,910 | 49,216 | 0.8322 | 59,200 |
| 60-64 | 32,825 | 45,367 | 0.8173 | 49,800 |
| 65-69 | 19,665 | 33,064 | 0.7530 | 37,100 |
| 70 and over | 33,163 | 51,598 | 0.6024 c/ | 78,300 d/ |
| Age unknown | 2,870 | 2,878 | – | – |
| 15-44 | 699,961 | 866,021 | | 1,121,600 |
| 20-49 | 609,465 | 714,706 | | 918,500 |
| Total | 1,612,074 | 2,127,116 | | 1,980,000 |
| Ratio 0-4/15-44 | 0.3520 | 0.4040 | | 0.4300 |
| Ratio 5-9/20-49 | 0.3938 | 0.4594 | | 0.5249 |

Notes: a/ Thailand, Population Census 1960, Central Statistical Office.
b/ Thailand, Population and Housing Census 1970, National Statistical Office.
c/ Population 70-and-over in 1970, devided by population 60-and-over in 1960.
d/ Population 60-and-over in 1970, multiplied by CSR calculated according to c/ above.

calculate the 0-4 females in 1980, while 0.525 was used for those 5-9 years of age. All other age groups were computed by multiplying the number in each five-year cohort by the appropriate CSR in force between 1960 and 1970, under the assumption that the combined effects of mortality, migration and census errors would be equivalent in the two decades. To complete the projection, a similar set of calculations should be carried out for the males, using the ratio of males 0-4 to females 15-44, and of males 5-9 to females 20-49, to project the two youngest age groups.

## C. COHORT MIGRATION-SURVIVAL METHOD

A second analytical method permits the examination of intercensal net migration, and permits control over the net migration component, by age, in the projection. In a country such as Thailand, with negligible levels of international in-or out-migration, the CSR measures the effects of mortality and census error. If, in the absence of more detailed information, it is assumed that mortality and census errors specific for age occur at the national rates in all regions, it is possible to calculate ten-year levels and rates of net migration; the appropriate technique appears in chapter III.

The numbers of migrants estimated for each five-year age and sex grouping was the difference, at the end of the intercensal period, between the number actually enumerated and the number that would have been enumerated, had the 1960 population survived to 1970 according to national censal survival ratios. The numbers estimated can serve as numerators for rate calculations which, as in the illustration, incorporate the expected population as a base or "population at risk". The base selected is convenient for projection, if somewhat arbitrary. It might be argued that the population which migrants join is not the population at risk, and that migration is more satisfactorily or logically related to the population from which the migration stream originates. This argument overlooks that the component is net migration, the resultant of transfers in and out, not necessarily to and from the same destinations and origins; the actual migration traffic may be and often is far larger than the net migration which is observed and estimated. For this reason projected rates of net migration, based upon the area of destination of net in-migrants, must be used with caution. If a subnational area is growing unusually rapidly, the projected level of net migration may grow unreasonably, to the extent that more future migrants are calculated than the presumed areas of origin could supply. The problem is less acute where net migration is negative, since the larger, outward component is related to the population at risk, although the inward component is not. A further discussion of projecting net migration appears in a later section.

The essential difference between the migration-survival technique and the census survival ratio method is the separate consideration of migration, rather than as an unevaluated part of the survival ratios. In chapter III the methods of estimating migration numbers and rates were described, as were the mathematics of converting ten-year rates to single-year and five-year rates. Using the example of the central region of Thailand, table 13 demonstrates the projection of five-year age groups of females to 1975 and to 1980, by the migration survival technique. Column 2 presents the reported populations in the census of 1970, while column 3 contains the ten-year survival ratios. The ten-year migration ratios listed in column 5, from which the five-year ratios are derived also appears in rate form in chapter III. The figures of column 4 are the calculations of five-year survival ratios; for almost all of the age groups, it is sufficient to use the square root of the ten-year rate, since it is in force only half as long. Because the conventional position of the rate is opposite the final (survived) population figure, the five-year rate appears in the line above the rate from which it is derived.

In the ten-year projection prepared by the cohort survival ratio method earlier in chapter VI, the population 60-and-over was projected from 1970 to establish the population 70-and-over in 1980. In order to calculate a survival ratio for the 60-64-year-old population to 1975, a special step is necessary, because the national rates from which the other age-specific rates are calculated do not permit a distinction between the two five-year groups between ages 60 and 69. The ten-year survival rate, relating those over 70 in 1970 to those over 60 in 1960, was calculated from national data at 0.5411; the square root 0.7356 describes the probable survival ratio for this terminal grouping for five years. By graphic extrapolation of five-year CSRs of selected five-year age groups, it was established that the survival ratio of the 60-64-year group was about 0.81, as shown in figure 4. After applying this ratio, the 65-69-year-old survivors were subtracted from the 65-and-over survivors calculated by application of the 0.7356 to the 60-and-over in 1970. 36/ A calculation of what the CSR would have been between 1970 and 1975 to yield the implied 70-and-over group established the value of 0.6977 for projection purposes. 37/

The population in 1975 without migration is calculated in the usual fashion, applying the national survival ratios of column 4. The estimated population, including migration, is calculated from the survived population by the application of age-specific five-year migration ratios; i.e. column 6 entries are multiplied by those of column 7.

The entries in columns 7, 8, 9 and 10 for the youngest age groups can be calculated in several ways. The population 0-4 years old in the population can be calculated by applying the migration ratio to the survived population, based upon the child/woman ratio; or it can be found by application of the child/woman ratio after age-specific migration ratios have been applied to the survived population. Since the migration ratio is based only upon 2.5 years of survival, and at an age most subject to errors and therefore inconsistent, it is recommended that the latter approach be used, basing the 0-4 entry in each column upon the women 15-44 in that column, without regard to migration ratios for that age group. In the case of those 5-9, the choice is less clear: the application of survival and migration ratios without regard for the child/woman ratio can lead to unlikely levels of implied fertility changes. On the other hand, application of the child/woman ratio without regard to the migration and survival ratios can lead to unacceptable levels of implied migration and survivorship. In the demonstration, each time the values for a column were developed, both calculations of the 5-9 group were made, and the results were averaged

36/ Calculated of the five-year survival ratio for the 65-69-year population is indicated in table 13.

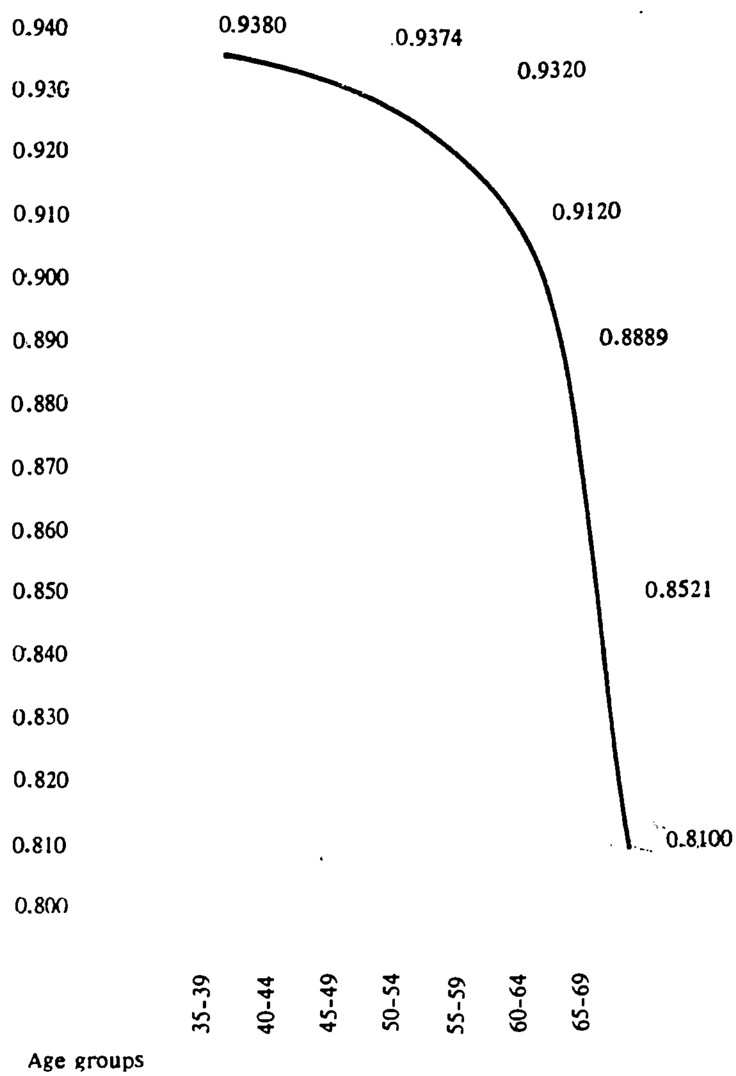
37/ If a life table for females were available for the country, the survival ratio ${}_5L_{65} \div {}_5L_{60}$ could have been used, as could a model life table exhibiting a comparable pattern of mortality. The use of national census survival ratios, as in the present example, is neither necessary nor recommended where adequate life tables are available.

Table 13. Female population of the central region of Thailand projected to 1975 and 1980, by cohort migration-survival method

| Age interval | Enumerated population 1970 | Survival ratios | | Migration ratios | | Projected population 1975 | | Projected population 1980 | |
|---|---|---|---|---|---|---|---|---|---|
| | | 10-year | 5-year | 10-year | 5-year | Survived from 1970 | With migration | Survived from 1975 | With migration |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| 0-4 | 765,954 | | | 1.0290 | 1.0588 | 870,585 | 883,702 | 990,225 | 1,005,424 |
| 5-9 | 769,846 | | | 1.0258 | 1.0171 | 842,747 | 880,970 | 934,514 | 997,375 |
| 10-14 | 697,993 | 1.0717 | 1.0352 | 1.0249 | 1.0124 | 751,293 | 760,609 | 859,738 | 870,399 |
| 15-19 | 607,976 | 0.9523 | 0.9759 | 1.0343 | 1.0170 | 659,464 | 670,675 | 718,623 | 730,840 |
| 20-24 | 441,074 | 0.8927 | 0.9448 | 1.0577 | 1.0284 | 584,691 | 601,296 | 644,988 | 663,306 |
| 25-29 | 366,495 | 0.9248 | 0.9617 | 1.0401 | 1.0199 | 417,168 | 425,470 | 568,706 | 580,023 |
| 30-34 | 351,725 | 0.8945 | 0.9458 | 1.0155 | 1.0077 | 350,589 | 353,289 | 407,005 | 410,139 |
| 35-39 | 296,626 | 0.9151 | 0.9566 | 0.9986 | 0.9993 | 330,129 | 329,898 | 331,597 | 331,365 |
| 40-44 | 241,300 | 0.8810 | 0.9386 | 1.0064 | 1.0032 | 278,057 | 278,947 | 309,246 | 310,236 |
| 45-49 | 187,789 | 0.8787 | 0.9374 | 1.0144 | 1.0072 | 224,892 | 226,511 | 259,979 | 261,851 |
| 50-54 | 161,900 | 0.8687 | 0.9320 | 1.0241 | 1.0120 | 171,264 | 173,319 | 206,578 | 209,057 |
| 55-59 | 136,573 | 0.8318 | 0.9120 | 1.0350 | 1.0173 | 143,912 | 146,402 | 154,063 | 156,728 |
| 60-64 | 108,778 | 0.7910 | 0.8889 | 1.0367 | 1.0182 | 116,374 | 118,492 | 124,749 | 127,019 |
| 65-69 | 83,015 | 0.7261 | 0.8521 | 1.0625 | 1.0308 | 237,041 | 90,823 | 95,978 | 98,934 |
| 70 + | 130,449 | 0.5411 | 0.7356 | 1.0309 | 1.0153 | | 151,209 | 168,866 | 171,450 |
| | 5,347,493[a] | | | | | 5,978,206 | 6,091,612 | 6,774,855 | 6,924,146 |
| Calculation of 65-69 CSR: | | | | | | | | | |
| 60-64 | 108,778 | | | | | | | | |
| 65-69 | 83,015 | | 0.8100 | | | 88,110 | | | |
| 70 + | 130,449 | | 0.6977 | | | 148,931 | | | |
| Women 15-44 | 2,305,196 | | | | | 2,620,098 | 2,659,575 | 2,980,165 | 3,025,909 |
| Ratio females 0-4/ females 15-44 | 0.3323 | | | | | | | | |
| Women 20-49 | 1,885,009 | | | | | 2,185,526 | 2,215,411 | 2,521,521 | 2,556,920 |
| Ratio females 5-9/ females 20-49 | 0.4064 | | | | | | | | |

Note: [a] 5,974 of age unknown have been omitted.

Figure 4. Five-year cohort survival ratios for females 35-and-over, Thailand, 1960-1970




58

and the mean entered in the proper cell. The second projection period carried the mean value of the 5-9 group from 1975 to the 10-14 cell in 1980, where it was subjected to the migration ratio appropriate to its age level.

To illustrate, the projected population 5-9 years old in 1975 without migration is 792,915; based upon the child/woman ratio, it would be 892,577. The mean of the two estimates, i.e. 842,747, was entered in the table. When the migration ratio was applied, the projected population using the adjusted number was 857,158, while the number calculated from the child/woman ratio was 904,782. The mean of 980,970 was entered, maintaining a degree of conformity with the populations in the child-rearing age groups of women. The 10-14-year survivors of the cohort in 1980 numbered 859.738.

In the demonstration of the cohort survival method in table 12, the increase in the child/woman ratios observed by a comparison of the results of the 1960 and 1970 censuses was reduced by half in 1980, under an assumption of a decline in fertility partially offset by a drop in infant and early childhood mortality. In the demonstration of table 13, no change was imposed upon the child/woman ratios but neither of these arbitrary decisions relates in any way to the methods in which they played a part.

## D. OTHER CONSIDERATIONS IN ANALYTIC METHODS

The preceding paragraph raises a fundamental issue in subnational population projection. Internal migration, often the greatest of the components of population change, is extremely volatile in response to social and economic developments. How far, then, should the development planner project? Although, subnational projections are occasional made 50 years into the future, the most useful period is probably shorter, from 10 to 20 years. Significant changes can be accomplished in a decade or two: facilities can be built, specialized manpower can be trained to attack the problems posed by 10-year or 20-year projections which are specific for age and sex. Although most of the people in an area ten or twenty years hence are, in the majority of cases, already there, this is not true of two categories of population which generate particular needs for planning and for facilities: i.e. the children and the migrants.

The effects of an assumed annual decline in fertility of 0.5 per cent is demonstrated in table 5, chapter IV. Using the same rationale, a similar table can be constructed converting any assumed fertility decline into corresponding child/woman ratios. Alternatively, if the annual number of births or the crude or the age-specific birth rates are known, they can be projected according to any scheme deemed reasonable or suited to the possibility to be demonstrated. The decision about what is reasonable would be facilitated if data from two or more census dates were available, in order to establish a trend; but the solution to a problem of this nature often lies in the use of a model, the experience of another province, the nation as a whole or a neighbouring country. Population projections for the whole of France by A. Sauvy[38] prepared in 1928 demonstrate what would happen to the population of the country if it were to follow the fertility trend already exhibited in the department of Seine.

Although the manipulation of future fertility affects only the projections for younger age groups for less than three decades, changes in mortality affect all ages, although not necessarily proportionately. Moderate changes in mortality can conveniently be applied evenly by increasing survivorships in an incremental fashion until they reach a maximum suggested by a reasonable model. If a model life table has been "fit" to the population according to the procedure sketched in chapter VIII below, the planner can select successive model tables in a series, improving the life expectation at birth with the passage of time. If feasible, this procedure is to be preferred, since the relationship among the survivorship ratios within each table has already been established from actual populations.

Although the mathematics of migration projection are no more complex than for fertility and mortality, a greater burden of choice remains on the planner. In the cohort migration survival method, migration was added after the expected or survived population had been developed, as in chapter III. For convenience in computation, the migration rate was calculated as survived population plus migration divided by the survived population. The individual values, representing the rates of migration for the various age/sex groupings, can be increased to a maximum, or decreased to zero or to an assumed minimum, depending on the stated assumptions. Restricted age/sex groupings seldom migrate by themselves: if the planner is attempting to forecast the attraction

---

38/ Cited in Roland Pressat, *Demographic Analysis*, (Chicago, 1972), p. 363, (English edition). Because of its purpose, to show "what if?", Pressat refers to the projection as a conditional forecast.

of a specific labour force in response to planned industrial development, he must allow for family dependents and for the development of secondary employment. The migration attracted by a development scheme depends not only upon the type of work and conditions of employment, but also living conditions including housing, transportation, leisure-time activities, educational facilities and other public services, which are also a part of development planning.

60

# VII. PROJECTING URBAN AND RURAL POPULATION, AND SOCIO-ECONOMIC GROUPS

Sound development planning, as mentioned earlier, requires maximum population detail. Insofar as possible this should include detailed projections of the population by special subpopulations requiring particular services or facilities. This is as true for subnational populations as for those of entire countries.

A population projection which includes age and sex detail is especially important when socio-economic projections are required, because they usually provide the base upon which the socio-economic projections are developed. Virtually every type of facility or service, every school, hospital, food supply, transportation or public health scheme, including family planning, bears different relationships to each of the various age and sex groupings.

Although they do not exhaust the possibilities for socio-economic projections, four major subject areas deserve explanation. These are the urban and rural divisions of total population, enrolments in educational facilities at all levels, labour force and employment, and future numbers of households. 39/ All of these vary with the population, but they seldom change in accordance with a simple proportional relationship to total population. The particular utility of these four types of projections is apparent upon consideration of the primary importance to city and regional planning, of transportation, of schools and teachers, of industrial development and jobs and of housing. A further consideration leads to an appreciation of the interrelationships among these variables in planning and administration, and of their primary importance in countries with youthful populations.

## A. PROJECTING URBAN AND RURAL POPULATIONS

The requirements of breaking down future populations into rural and urban portions pose constraints for data which are more stringent than those limiting national projections. As a result the choice of technique may be limited to one of the more mechanical models, since age and sex details, specific for urban and rural residence, are not likely to be at hand.

The hazards of application of growth rates, either arithmetic or geometric, were demonstrated in chapter V, where calculations of future total populations were shown. In virtually every country in the world, the urban population is growing more rapidly than the rural, largely through the effects of migration. For this reason, the application of growth rates to urban and rural projections introduces hazards which are even more severe than for total populations, either subnational or national. On the other hand, assuming the data were available, the most painstaking application of a cohort survival method to the urban and to the rural populations within a subnational area would lead to erroneous future figures, because future urban residence areas differ from those defined at the time of the base year. If there is no likelihood of area redefinition, the problem of an urban projection is equivalent to that for any other subnational area, and can be approached in a manner suggested by the availability of data in accordance with techniques treated in chapters V and VI.

Usually, the inhabitants of a subnational area may be viewed as having an extra demographic characteristic: each individual is either urban or rural. This attribute is not fixed like sex, nor does it change in a uniform manner, like age. Within a subnational area, the population classified as urban grows at a different rate than the remainder because birth and death rates are typically not equal to those of the total population. The urban group grows at a different rate also because (a) a disproportionate number of migrants from outside the subnational area settle in the urban area; (b) rural-to-urban migration tends to increase the urban areas' share; and (c) perhaps most critically, when certain density or other criteria are met, areas are redefined as urban which were formerly rural.

In chapter IV of United Nations Manual VIII, three mechanical methods are contrasted for their applicability to urban/rural population projections, the method of urban growth rates, of rural growth rates

39/ Four of the manuals prepared by the United Nations are devoted to these subject areas: *Estimating Future School Enrolments in Development Countries: A manual of methodology*, jointly with UNESCO; and Manuals V, VII, and VIII, *op. cit.* This chapter confines itself to a brief introduction to these types of projection and their relationships to subnational projection. The development planner embarking upon one of these tasks is urged to consult the relevant manual.

and of the ratio to the total. All three of these have been demonstrated here within the context of subnational projections in chapter V above; their application to urban and rural portions of the total population is identical to their application to one or another subnational area. The two growth rate methods involve assumptions of constant growth rates and are subject to all of the dangers described in chapter V. But when used with an appreciation for its hazards, a growth-rate method may be used for illustrative purposes. Manual VIII concludes that for urban/rural projection, during each of the stages in the course of urbanization, one of the methods is most effective. During the beginning of the urbanization process, when levels of urbanization are low, the method of urban growth rates is most satisfactory. When levels of urbanization are high, on the other hand, the method of rural growth rate is preferable. Although it is unlikely that this generalization has been subjected to an empirical test, application of this principle forestalls absurd results, since the geometric curves are not permitted to rise to spectacular heights. During the middle phase of the urbanization process, when the urban percentage is neither very high nor very low, use of the ratio method is probably advisable.

An effective method of projecting urban and rural population is described in chapter V of Manual VIII as the "urban-rural growth differential" method (URGD). Application of the method requires a preliminary populaion projection for the national or subnational area. The success of the method depends on the URGD, or the difference between the urban and the rural rate of population growth, being rather stable; this stability is independent of time, place or degree of urbanization. In addition to the projection prerequisite, a measure of urbanization at two or more points in time must have been made either as a part of the original reports of two or more censuses, or ex post facto. The latter may be done by dividing the population into urban and rural segments, according to a set of fixed criteria such as density, or setting a minimum population size for towns or cities to be considered urban.

Table 14 presents data from two censuses of Iran which demonstrate the need for interpretation of rural/urban changes before application of a projection technique. Karaj Shahrestan is an area which is apparently in rapid transition from rural to urban character, although in 1966 it was still nearly 80 per cent rural. The Central Province, of which it is a part, contains Tehran, the capital city; the province as a whole was about 70 per cent urban in 1966 when Iran was approximately 61 per cent rural. The URGDs

**Table 14. Illustrative projections for all Iran, the Central Province, and Karaj Shahrestan to 1976 with urban and rural subdivisions**

| | | | | Population projected to 1976 | | | |
|---|---|---|---|---|---|---|---|
| | | | | Ratio method | | URGD method | |
| | 1956 | 1966 | Percentage change | Number | Percentage change | Number | Percentage change |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| Iran | | | | | | | |
| Total population | 18,954,704 | 25,078,923 | 32.31 | 33,000,000 | 31.58 | 33,000,000 | 31.58 |
| urban | 5,953,563 | 9,794,246 | 64.51 | 15,407,700 | 57.31 | 15,690,200 | 60.19 |
| rural | 13,001,141 | 15,284,677 | 17.56 | 17,592,300 | 15.09 | 17,309,800 | 13.24 |
| Percentage urban | 31.41 | 39.05 | | 46.69 | | 47.55 | |
| rural | 68.59 | 60.95 | | 53.31 | | 52.45 | |
| Central Province | | | | | | | |
| Total population | 2,717,309 | 4,979,081 | 83.24 | 8,250,000 | 65.69 | 8,250,000 | 65.69 |
| urban | 1,814,914 | 3,505,970 | 93.18 | 6,107,475 | 74.20 | 6,119,700 | 74.55 |
| rural | 902,395 | 1,473,111 | 63.24 | 2,142,525 | 45.44 | 2,130,300 | |
| Percentage urban | 66.79 | 70.41 | | 74.03 | | 74.18 | 44.61 |
| rural | 33.21 | 29.59 | | 25.97 | | 25.82 | |
| Karaj Shahrestan | | | | | | | |
| Total population | 161,594 | 232,024 | 43.58 | 400,000 | 72.40 | 400,000 | 72.40 |
| urban | 14,526 | 49,392 | 240.02 | 134,360 | 172.03 | 169,060 | 242.29 |
| rural | 147,068 | 182,632 | 24.18 | 265,640 | 45.45 | 230,940 | 26.45 |
| Percentage urban | 8.99 | 21.29 | | 33.59 | | 42.26 | |
| rural | 91.01 | 78.71 | | 66.41 | | 57.74 | |

for the country, the province and the **shahrestan** are respectively 46.95, 29.94, and 215.84. Columns 1 and 2 of table 14 show the total, urban and rural populations in 1956 and 1966 for the country, the province and the **shahrestan**, along with the percentages of the total populations which were urban and rural for each year. Column 3 compares growth percentages between the two censuses for the various populations. It is assumed that satisfactory projections of the total population of each of the three areas to 1976 have been prepared, and illustrative figures of 33,000,000 for the entire country, 8,250,000 for the Central District and 400,000 for Karaj Shahrestan, appear in columns 4 and 6. A reasonable future course of urbanization is based upon the assumption that the percentage urban (and rural) follows the same straight line between 1966 and 1976 as it did between 1956 and 1966 in the three cases, and is the foundation of the urban/rural breakdown in column 4. This is an illustration of the ratio method of the United Nations manual. In order to use the URGD method, a pair of simultaneous equations were set up from the table, using figures for the whole country:

$$r + 0.4695 = u$$
$$15{,}284{,}677r + 9{,}794{,}246u = 33{,}000{,}000 - 25{,}078{,}923$$

where r and u represent the rural and urban rate of change between 1966 and 1976; 0.4695 is the urban/rural growth differential; the coefficients in the second equation are the reported rural and urban populations of Iran in 1966; and the right side of the second equation is the 1966-1976 population change. The solutions to these equations and corresponding equation for the province and the shahrestan are found in column 7, while the urban and rural populations implied by the change rates appear in column 6.

An examination of the table suggests that the ratio method and the URGD method yield very comparable results in the case of the country and the major province. In the case of the shahrestan, the projected urban and rural growth rates by the URGD method correspond far more closely to the experience of the preceding decade than does the ratio method. For a short-term projection, the URGD is probably preferable, but factors accounting for any radical departure from an established trend should be examined. In the case of the Karaj Shahrestan, for example, Karaj City with a population of 14,526 accounted for the entire urban population in 1956; by 1966 that city's population had increased to 44,243, and Eshtehad, a rural village of 4,542 in 1956, had attained urban status with a population of 5,149. Further research would reveal that in 1966 the density of population in the **shahrestan** was approximately one-eighth that of neighbouring Tehran, and that five places with populations between 2,500 and 5,000 lie within 15 kilometres of Karaj City. The growth behaviour of 1956-1966 can probably be repeated, and the use of the URGD for Karaj Shahrestan is justified.

## B. EDUCATIONAL ENROLMENTS

Students constitute a subnational population, but projection of their numbers demands very special consideration. A crude ratio can be calculated of those attending school within a certain span of grades or forms, to the population from which they are drawn, preferably specific for relevant age groups. Only the number of students and the numbers in the population are needed; although where seasonable attendance is practiced, the definition of enrolment may require conceptual sharpening. The assumption that such a ratio would remain unchanged throughout the projection period would yield an approximate measure of future enrolment, although sudden changes in age composition or attendance patterns can bring about serious errors. Increasing school enrolments usually accompany an expansion in facilities or an improvement in economic conditions.

Figure 5 is a schematic representation of the movement of successive cohorts[40] through a hypothetical school or school system. Each "e" represents a number enrolled in a particular grade in a particular year. The subscript stands for the final digit of the year in which the cohort started school; the superscript refers to the year of the series, a to represent an arbitrarily designated 1970, b for 1971, etc. A "retention" ratio $e_0^b \div e_0^a$ measures the relationship between the "survivors" in 1971,

---

40/ Although the term *"cohort"* in the demographic literature usually refers to a group having a common birth year or span of years, more generally it refers to a group experiencing a certain event in a specified period of time. In this section it is used to designate a group starting the first grade in a given year, plus all who join those who progress one grade each year, either by returning to school or by leaving an earlier cohort to repeat a grade. This application may be termed an "apparent cohort" approach, since it does not distinguish between those who progress and those who remain or return. A "true cohort" approach would require individualized records in which each pupil's retention status is a matter of concern. See also UNESCO Regional Office for Education, *Progress of Education in the Asian Region*, Statistical Supplement, (Bangkok, 1972), table 31, pp. 97 ff., for "apparent cohort" retention ratios in Asian countries.

Figure 5. Movement of a grade cohort

| Grade | 1970 | 1971 | 1972 | 1973 |
|---|---|---|---|---|
| 1 | $e^a_0$ | $e^b_1$ | $e^c_2$ | $e^d_3$ |
| 2 | $e^a_9$ | $e^b_0$ | $e^c_1$ | $e^d_2$ |
| 3 | $e^a_8$ | $e^b_9$ | $e^c_0$ | $e^d_1$ |
| 4 | $e^a_7$ | $e^b_8$ | $e^c_9$ | $e^d_0$ |
| 5 | $e^a_6$ | $e^b_7$ | $e^c_8$ | $e^d_9$ |
| 6 | $e^a_5$ | $e^b_6$ | $e^c_7$ | $e^d_8$ |

including those remaining from the prior cohort, and the number that started a year earlier at one grade below. The ratio $e^c_0 \div e^b_0$ represents the experience of the same cohort, modified by those retained, a year later in passing to a grade higher. A comparison of the first ratio with $e^c_2 \div e^b_1$ is a comparison of succeeding cohorts making the same grade transition; here the educational planner can apply his assumptions concerning future levels of retention. In the UNESCO methodology, manual[41] this basic idea is elaborated. The manual distinguishes between the retention ratio, as it has been used here, the "progression" ratio, which refers only to those who progressed to the next grade, and the "repeater" ratio, the proportion of those who do not proceed to the next grade.

The other basic method, called the "enrolment-ratio" method in the manual, is also known as the "age-participation" method. Unlike the cohort-survival or grade retention method sketched above, this method depends upon census data, not only for a base from which to calculate enrolment but in some applications for the enrolment counts as well. An absolute prerequisite to the method is a benchmark population, distributed by age and sex, and a projected population with corresponding age and sex categories.

Present ratios of enrolment to population are calculated in one of several ways. If attendance data by age are reported in census publications, the ratios of those attending specified grades or levels of schooling to the appropriate numbers in the population can readily be calculated. Alternatively, enrolment data from the educational authority can also be used. The former source should entail no problems of residence in subnational projections, since the population data and the attendance data refer to the same geographical base. Data from school sources, on the other hand, may refer to different areas than those for which census data are reported, because some pupils attend schools in different jurisdictions than those in which they live.

Useful enrolment projections by either method can be accomplished with varying degrees of age/sex and grade-level detail. Usually in the grade-retention method, males and females by individual grade are moved through the survival pattern. Under certain circumstances, and especially where educational policy changes are being considered, it is useful to handle separately those who progress and those who are retained in a grade for another year. Lacking more precise data, crude projections can be made of expected numbers of pupils in the widely recognized two levels of education. The UNESCO Statistical Yearbook[42] shows broad enrolment ratios by age for the countries of the world, using the age groups close to 6-11 as the denominator for the first level, and groups close to 12-17 for the second level, depending upon the particular national practice. If attendance data from a census or enrolment data from the schools are available, specific for single grades and single years of age for each sex, a very precise projection can be made from a future population by sex and single year of age.

Before formulating any assumptions concerning future trends in retention or participation ratios, the future effects of changes in law or public policy upon a number of critical areas should be considered. These include such actions as the introduction, enforcement or prolongation of compulsory education, the broadening of opportunities in the education of girls, changes in the roles of public and private or religious schools, the development of school facilities, the training of teachers, and many others.[43] The training of teachers poses a related problem in projections: future requirements for teachers are linked to future students by the teacher/pupil ratio, in a manner comparable to the linkage between students and population. In this as in some of the other considerations the assumptions are constrained by the limitations of the time in training teachers or building schools.

41/ Jointly with United Nations, *op. cit.*, especially chapters IV- VI, pp. 22 ff.

42/ UNESCO, *Statistical Yearbook* 1970, (Paris. 1971).

43/ In the United Nations/UNESCO manual, *op. cit.*, p. 17 fourteen factors are listed for consideration in formulating assumptions.

Whichever of the two classes of method are used, the paramount problem is the quantitative expression of assumptions about the future in the grade retention or age participation ratios. The first step should be to calculate the projection as if no change in the ratios were to take place, that all changes in future needs for education would be solely attributable to population changes. Then a terminal data should be selected, based on an estimation of the constraints of time, at that terminal date it is assumed that the ratios will have attained their possible values, values which may have obtained in another subnational area, for example, or in a similar country.

## C. THE ECONOMICALLY ACTIVE POPULATION

In addition to future needs for schools and schooling, sound development planning must also encompass the probable relative size and future activity of the economically active, upon whom all rely for support. The projection of the economically active population usually begins with a satisfactory projection of the total population including age and sex detail. If it is necessary to prepare a population projection solely to serve as a base for a projection of the economically active, the irksome problem of future fertility can often be avoided if the horizon is only a decade into the future, because those born after the benchmark date would not become active during the projection period. One of the methods described in chapter VI provides the detail necessary; or, if a national life table is available or if a model life table can be fitted to the area. the survivorship values described by $({}_5L_{x+5}) \div ({}_5L_x)$ in the relevant table can be used. Migration levels or rates must be estimated, and contributions of the migratory stream to the labour force must be considered, when formulating assumptions about activity or participation rates.

The economically active population, frequently called the labour force, varies in definition from place to place. The concept includes those who are working and those who are seeking work. Among those engaged in a subsistence economy, it is difficult to establish a clear distinction between individuals who are economically active and those who are not. Unpaid family workers are usually included as "employed". However, there are serious questions on how to classify underemployed persons such as involuntary part-time workers, either those who work less than a full day or those who work only seasonally.

There are no universal solutions to the problems posed above. In subnational projections of the labour force, acceptance of the definitions adopted for the whole country would facilitate the supplementary use of national data. The chief local problem is the development of projected activity rates or labour force participation rates. This task may be particularly difficult if a specialized form of economic activity is being developed in the area, requiring a work force disproportionately from one sex or from a narrow span of ages. If full employment is the goal of the economic planner, the development of complementary sources of employment, or the employment potential in areas adjacent to the subnational area, must be explored.

Unless retirement patterns change drastically, the principal years of economic activity or participation in the labour force for males are those between 15 and 64, inclusive. Many highly agricultural countries have significant membership in the labour force from the 10-to-14 year age group. In the central age groups in the male population, for example, those between 25 and 54, participation is high throughout the world, generally 95 per cent or above. Participation of males in the older ages is affected by retirement practices, which in turn depend on the capacity of the economy to employ older workers, as well as the existence of alternative support upon their withdrawal from the labour force. At the younger ages, male participation reflects competing activities in prolonged education, alternative life styles, or military service. Although the latter is usually defined as economic activity, the degree of armed forces mobilization can have profound effects upon the nature of labour force participation.

Female participation patterns are less consistent between and within countries. Rates are generally lower than male rates at all ages. After an initial rise, female rates often decline around age 25, reflecting withdrawal from employment in order to bear and rear children. Another increase in participation often occurs in later life, when some of those who dropped out earlier and some who were never active enter the labour force when their children no longer need full-time care. The total projection problem involves more than the calculation or selection of suitable future rates for the groups currently contributing most of the male and female workers. Of critical importance is the formulation of hypotheses about future social and economic developments, and their effects upon the participation of groups with low current rates, such as women, because the supply of labour (i.e. the demand for jobs by members) appears to be most elastic among these groups.

In general, the methods of projecting the economically active population involve the assumption that rates of economic activity or labour force participation are independent of economic conditions. In fact, the number of people seeking work also depends upon the availability of work, and the need for additional workers to support the family. In the developed countries, this fact is widely recognized, but it is possible to ignore it in the assumption of full employment. In the developing countries, on the other hand, where underemployment and unemployment reach far higher chronic levels, the projection of participation rates without some allowance for changing conditions would be an exercise of no relevance to the planning function. Projection at the subnational level may have to be based on national rates. Before applying such projected rates to populations in calculating the future labour force, the national rates must be scrutinized, bringing to bear as much local and specialized knowledge as possible.

Table 15 compares labour force participation rates (activity rates) for males and females in developed and developing countries in the years 1950, 1960, and 1970. A number of observations made from these data can be useful to the analyst projecting participation rates for a subnational area undergoing development:

(a) The participation of males and females aged 0-14 is declining with increasing opportunities for education and with the discouragement of child labour.
(b) Levels of participation of children are lower for both sexes in developed than in developing countries.
(c) Among the 15-19 group, participation has been declining among both males and females in most countries.
(d) Males 15-19 years old participate at higher rates in developing than in developed countries because of lower opportunity for education
(e) Females of that age group participate at lower rates in the developing countries because of lower opportunities for employment outside the home. With social and economic development, young women in the developed countries would approach the male level of participation, and young women in the developing countries would follow their sisters in the developed lands.
(f) Between ages 20 and 54, male participation is high and stable world-wide; female participation is rising in developed countries, declining in developing countries.

It is apparent that those age and sex groupings wherein participation has been changing most rapidly are also those wherein participation is most responsive to local conditions. The comparisons contained in the table offer clues to the direction of change. The analyst of local labour market conditions must evaluate the opportunities for increased participation by women, and the effects of the school system and child labour laws upon the participation of the youngest groups. With his assumptions established on the basis of his observations, the analyst must decide whether each of the age-specific rates is likely to remain stable, or to increase according to a linear or other pattern of extrapolation, or approach an assumed maximum in 10 or 20 years, or to continue a decline to an assumed minimum. The rates are then to be applied to the population in each respective age/sex grouping, in a population projection prepared by a cohort method.

A simple method of projecting labour force participation rates is demonstrated in table 16. Starting with activity rates for Sri Lanka females, published for 1963 and 1971, an annual average linear change is calculated; these figures appear in column 4. Activity rates for 1976 and 1981 are calculated by adding five times and ten times the annual average change to the rate for 1971. Columns 7 and 8 show the 1970 world-wide activity rates for the developing and the developed countries which appear in table 15.

A comparison of the recorded and projected values for Sri Lanka appearing in columns 3, 5 and 6 with the values in columns 7 and 8 raises questions which the analyst must answer. For example, by 1981 the projected activity rates would have reached a level comparable to the 1970 level for developing countries among the 25-44-year-old groups. Although this seems a reasonable prospect, it is less likely that the downward trend for the three groups starting at age 55, suggested by the 1963 and 1971 data, would continue unchecked. The fact that world-wide rates are substantially higher deserves investigation. Of equal concern is the projected rate for the 20-24-year group. Because it exhibited an extraordinary change between 1963 and 1971, it is projected to the world-wide 1970 level by 1976, and well beyond that only five years later. It should be noted that, had a geometric projection of the rates been employed rather than the arithmetic one, the effect would have been even more pronounced. The low activity rates of the youngest groups re-

Table 15. Specific activity rates by sex and age for developed and developing countries: 1950, 1960 and 1970

| Sex and age | Developed countries | | | Developing countries | | |
|---|---|---|---|---|---|---|
| | 1950 | 1960 | 1970 | 1950 | 1960 | 1970 |
| Males | | | | | | |
| All ages | 60.4 | 58.3 | 57.9 | 57.7 | 54.8 | 52.7 |
| 0-14 | 2.4 | 1.6 | 1.3 | 7.4 | 6.2 | 4.7 |
| 15-19 | 69.7 | 61.5 | 55.5 | 74.0 | 70.9 | 68.2 |
| 20-24 | 90.1 | 89.4 | 87.4 | 92.1 | 90.6 | 89.0 |
| 25-44 | 95.9 | 96.1 | 96.2 | 96.2 | 95.7 | 95.8 |
| 45-54 | 94.4 | 94.0 | 94.1 | 93.7 | 93.6 | 93.4 |
| 55-64 | 83.9 | 82.3 | 80.3 | 84.7 | 83.9 | 82.9 |
| 65 and over | 42.3 | 34.2 | 30.4 | 53.4 | 51.8 | 47.6 |
| Females | | | | | | |
| All ages | 31.3 | 32.6 | 32.5 | 32.5 | 29.8 | 28.5 |
| 0-14 | 1.7 | 1.2 | 0.9 | 4.7 | 4.8 | 3.2 |
| 15-19 | 54.7 | 48.9 | 45.1 | 48.7 | 46.3 | 43.4 |
| 20-24 | 59.0 | 62.6 | 59.8 | 53.0 | 50.8 | 50.2 |
| 25-44 | 44.1 | 50.0 | 51.2 | 53.2 | 50.3 | 49.8 |
| 45-54 | 41.4 | 49.0 | 49.3 | 51.3 | 49.8 | 48.6 |
| 55-64 | 32.5 | 35.5 | 36.3 | 42.6 | 39.7 | 38.0 |
| 65 and over | 13.6 | 13.0 | 12.4 | 21.9 | 22.0 | 19.8 |

Source: J.N. Ypsilantis, "World and regional estimates and projections of labour force", Sectoral Aspects of Projections for the World Economy, in First Interregional Seminar on Long-term Economic Projections, volume III: Discussion Papers, Elsinore, 14-27 August 1966 (United Nations publication, Sales No.E.69.II.C.3).

Table 16. Projection to 1976 and 1981 of labour force participation rates for the female population of Sri Lanka, with international comparisons

| Age group | Reported activity rates | | Average annual change $\frac{(3) - (2)}{8}$ | Projected rates | | Estimated rates 1970 | |
|---|---|---|---|---|---|---|---|
| | 1963 | 1971 | | 1976 (3)+5(4) | 1981 (3)+10(4) | Developing countries | Developed countries |
| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| 10-14 a/ | 4.3 | 4.1 | -0.0250 | 3.98 | 3.85 | 3.2 a/ | 0.9 a/ |
| 15-19 | 21.8 | 26.4 | 0.5750 | 29.28 | 32.15 | 43.4 | 45.1 |
| 20-24 | 29.6 | 42.2 | 1.5750 | 50.08 | 57.95 | 50.2 | 59.8 |
| 25-29 | 28.1 | 39.3 | 1.4000 | 46.30 | 53.30 | | |
| 30-34 | 25.3 | 34.7 | 1.1750 | 40.58 | 46.45 | | |
| 35-39 | 26.2 | 32.6 | 0.8000 | 36.60 | 40.60 | 49.8 | 51.2 |
| 40-44 | 26.2 | 30.9 | 0.5875 | 33.84 | 36.78 | | |
| 45-49 | 26.3 | 29.8 | 0.4375 | 31.99 | 34.19 | 48.6 | 49.3 |
| 50-54 | 21.6 | 24.5 | 0.3625 | 26.31 | 28.13 | | |
| 55-59 | 17.7 | 16.7 | -0.1250 | 16.08 | 15.45 | 38.0 | 36.3 |
| 60-64 | 11.6 | 10.5 | -0.1375 | 9.81 | 9.13 | | |
| 65 + | 6.1 | 5.1 | -0.1250 | 4.48 | 3.85 | 19.8 | 12.4 |

Source: ILO Yearbook of Labour Statistics 1972, (Geneva 1973).

Note: a/ Estimated rates for 1970 are for the 0-14 age group; projected rates for 1976 and 1981 are for the 10-14-year population.

flects the educational opportunities available to women in Sri Lanka. Further analysis is required to determine if the 1963-1971 increase can be expected to continue to over 32 per cent by 1981. With better educational opportunities in the developing world, is it reasonable to expect the world-wide activity rate to decline, or is it likely to increase toward the level exhibited by the developed countries? The answer lies in the relative influence of two trends upon future activity rates: (a) increased educational opportunities which have a depressant effect, and (b) the relaxation of traditional sanctions against women's labour force participation.

These and purely local considerations must enter into the projection of activity rates for subnational areas. Additional methods are described in Manual V noted elsewhere. The International Labour Organization has prepared age/sex-specific participation-rate projections for the period 1960 to 1980, for 23 regions of the world and some of the constituent countries. Manual V suggests that these may serve as patterns for countries wishing to project activity rates.

## D. HOUSEHOLDS AND FAMILIES

Another of the socio-economic projections important to development planning is the number of households, or the number of families. A household is usually defined as a group of people sharing living accommodations. Actually, there are in the literature two concepts of household. The household is regarded as consisting of all occupants of a housing unit; or alternatively, the members of a housekeeping unit, more than one of whom may share a housing unit, as in India.[44] The former is the more useful concept in most countries; where it is used, the number of occupied housing units is equal to the number of households. Household members are usually, but by no means always, related to each other by blood or marriage; frequently households include roomers or lodgers not related to the other members. The United States Bureau of the Census, for example, defines five or fewer unrelated persons sharing a housing unit as a household, while six or more are classed as group quarters along with soldiers in barracks or monks in a monastery. There is no limit to the number of related persons living together in the definition of a household. The term "family" stresses the blood or marriage relationship. In many countries where family life is organized around the nuclear family, the number of families and the number of households reported in the population are quite close.

Each household and each family, whatever the type, usually has a designated head. If headship status or relationship to head is cross-classified by age and sex, projection is a relatively simple matter. Although there are some quite sophisticated methods of projecting households and families based upon life table applications and marital status probabilities,[45] at the subnational level neither data in sufficient detail nor a subject population sufficiently closed is likely to exist. The simple methods described below assume the prior existence of population projections.

### 1. Ratio method

For very short-term projections, future numbers of households can be estimated using the simple assumption that the ratio of households to population remains constant at the value calculated from the most recent census. In effect, this method assumes no change in the average household size, and completely ignores the possible distortion of changes in that portion of the population which is not considered as living in a household *per se*. The average household size is a very useful statistic, since changes therein often follow changes in fertility and shifts in age composition. For example, where recent declines in fertility have occurred, household size (and family size) has declined. In areas where sharp increases in fertility occurred in the aftermath of the Second World War, household size has recently dropped precipitously, owing to the tendency of the survivors of high-birth years to form households of their own upon coming of age. For an improved projection, it is advisable to observe the average household size at two censuses and to treat the non-household population as a separate category. Whatever the nature of the population not in households, it will require different types of housing if it increases in the coming years. At the same time, changes in average household size observed over two censuses should be evaluated in the light of changing fertility, changing patterns of residence, and the effects of past fertility and mortality developments, which also affect the age and sex structure of the population. A comparison of average household size in 1960 with that in 1970 might result in underestimating a downward trend in many places, because the decline in these areas took place entirely after 1960 from a peak higher than the 1960 value.

44/ The Indian census of 1961 defined a "census house" as that which might contain several households, each of which had its own kitchen facility.

45/ For a more complete discussion of the subject, see Shryock and Siegel, *op. cit.*, p. 846 ff.

## 2. Method of age-specific headship rates

The most convenient method of projecting population by household is by a variation of an age-specific activity or participation rate. If headship status by age and sex is available on the basis of one or more censuses, either the most recent rate or an arithmetically projected rate to each age and sex grouping can be applied. Total heads of households, and hence number of households are computed by simple addition of the numbers calculated for each age and sex grouping. If data on the population not living in households is separately tabulated in the census reports, and if an age and sex distribution of this special population is available or can be estimated, it is advisable to remove this group by subtraction before the calculation and projection of headship rates. Typically, populations in group quarters such as dormitories, barracks or institutions are extremely atypical of the general population of their age and sex composition.

## 3. Note on projected household size

Although it is a simple matter to project an average household size, the development planner interested in future housing needs may require a distribution of expected households by size. The use of a cross-tabulation of size of household by age and sex of head, would indicate the increasing importance of the newly formed households with younger heads. An average household size calculated from this distribution would follow the expected trend. Shryock and Siegel[46] cite a number of other efforts directed at this problem, most of which require a degree of detail or cross-tabulations not readily available for subnational areas. However, sample census or survey tapes for the whole country would permit cross-tabulations for the country. The relationship of changing household size to a changing distribution by number of persons can also be established by the experience of other areas which have followed the trend earlier, and for which the relationships are available.

## 4. Note on the use of marital rates

If the data are available, current and projected marital rates can serve as indices of family formation. Relating families to households is a problem which differs from place to place because of different or changing patterns of housing. The population not living in household conditions should first be removed from the benchmark population, before relating families to households. The rest of the problem consists in dividing the household population into segments according to the local definition of household (and family).

---

46/ op. cit., p. 850.

63

# VIII. SOME METHODS OF ESTIMATION FROM INCOMPLETE DATA

In the chapters preceding this one, no attention has been devoted to the difficulties associated with incomplete or faulty data. The mechanical and analytical methods projecting populations for subnational areas were selected partly because they could be used as if data problems did not exist. At the subnational level, it is likely that faulty estimates of migration have a more critical effect upon projections of population than faulty recorded data.

In recent years, many developments have taken place in the methodology of projection. Some of these are summarized below. Even if the information below is not directly applicable to the subnational population projections, a knowledge of the use of the methods described herein should provide some clues to the problems involved in population estimation and projection. A comprehensive manual[47/] is available, and a chapter in Shryock and Siegel[48/] is devoted to the subject.

An ideal collection or "bank" of demographic data for a country and its subnational areas consists of three parts:

(a) periodic censuses;
(b) frequent sample surveys between censuses, focussing on new trends in employment, migration, family and household formation, and other variables subject to sudden and rapid change;
(c) a system for reporting details of births and deaths, marriages and divorces, and immigration and emigration (usually collected separately).

Censuses have made significant improvement in the ESCAP region during the past decade or two; the 1970-1971 round of censuses are adequate in quality and detail for population estimation and projection, with few exceptions. Current surveys of population have become more frequent and, notably in India and Japan, are an increasingly important part of administrative data collection.

In the third area, i.e. the registration of vital events (births, deaths, marriages and divorces), data are satisfactory in only a few of the countries of the region. Compliance by all citizens as well as the health and statistical authorities is essential to an effective vital statistics system. Under ideal conditions it would require two decades to build the considerable administrative machinery within a favourable climate of public acceptance. Meanwhile, national and subnational population estimates and projections must be prepared if development planning is to proceed.

Important work has been done with retrospective data on births and deaths, marriages and divorces from censuses and surveys. This is an effort to estimate current rates from the reports of past events, subject to the imperfect memories of the respondents. More current and timely information, and ultimately more accurate, could come from vital records if public health administrators continue to make improvements in that area. Meanwhile, until satisfactory data from the universal collection of vital information are available, useful estimates can be made by other methods. Population projection at the national level can be improved by using such estimates, which in turn, redounds to the advantage of planning efforts for subnational areas. If on the other hand, a subnational area is unique in demographic structure and national figures are not relevant, the analyst should consider devoting more effort to improving his data, perhaps by measuring his rates on a sample basis, than to elaborate on his methods.

In the estimation of migration specific for sex and age groups demonstrated in chapter III, survival rates were calculated for the entire country and applied to the population of one province as enumerated at the earlier of the two censuses. The assumptions underlying the exercise were that the entire country was unaffected by net migration; and that the rates of survival, specific for each age and sex grouping, were appropriate to the province or region. The annual age-specific death rates were not calculated, nor did the projection require that they be calculated. Similarly, no effort was made to calculate male and female birth rates specific for age of mother.

It was observed in chapter III and again in chapter VI that the reported survivors of the population under 5 years of age typically exceed the number originally reported. This is indicative of the underenumeration of the very young, which seems to characterize all censuses in varying degrees. This is only

47/ United Nations, Manual IV, op. cit.
48/ Op. cit., chapter 25, p. 810 ff.

one of the short-comings of the unadjusted survival rates. There is also a systematic misreporting of age, not necessarily consistent from country to country, nor between the sexes, but resulting in survivorship rates which do not necessarily reflect the reality. In preparing a projection of population, unless the interest in narrow age-groupings is paramount, such distortions usually are not critical. Substantial errors in age-specific mortality are masked by the method, while errors in fertility rates are concealed by the use of the child/woman ratio to calculate the male and female populations under 10 years. A useful technique is described, which yields birth and death rates from the population reported in two censuses, where vital statistics are absent.

## A. MODEL LIFE TABLES

In chapter IV the anatomy of a life table was presented, and reference was made to the relationship existing among the parameters of a table. A table developed for one population can often be used successfully in another where lack of data prevents construction of a unique table. Yet, life tables can differ from one another in a number of important respects. The fact that women outlive men in many countries is well known; however, in some countries female mortality among the very young exceeds that of males. Moreover life tables with similar $e_0^o$ may have very different patterns of mortality, in terms of the shape of the ascending curve from the minimum, typically in childhood, to the maximum in old age.

The United Nations Population Division prepared a set[49] of model life tables based upon the recognition of the high correlation between the adjacent values of $_nq_x$ in a population. An improved set of model life tables was later developed by Coale and Demeny,[50] the organization of which reflected regional variations. Starting with over 300 pairs of life tables for males and females, they observed that the patterns of mortality showed systematic differences from world averages in three areas which represented eastern, northern and southern European experience. The sets developed from these were conveniently called "East", "North" and "South". The remaining set, reflecting experience in the remainder of the world, was termed the "West" series, and is recommended for use where no strong argument exists for adoption of one of the other three. In the use of model life tables, the subject population is compared to the sets, and the most appropriate set is selected; mortality rates for the subject area are taken from the model tables by interpolation.

Shryock and Siegel[51] have included an application of the above method to the female population of Fiji. Using survival values $_5L_{x+10} \div {}_5L_x$ from four of the "West" tables, the age groups of the 1956 population were projected to 1966. The four model tables were selected from the total set, such that when the projected populations were "cumulated", based upon their respective survivorships, the results of the four sets of calculations bracketed the values cumulated from the figures reported in the 1966 census.[52] Each of the model life tables has a number to designate its mortality level. In the female "West" series, the four tables chosen were levels 13, 15, 17 and 19, with respective $e_0^o$ values of 45.0, 50.0, 55.0 and 60.0 years. For each of the nine cumulated projected populations from 10-and-over to 50-and-over, inclusive in five-year increments, the level of mortality was selected by interpolation of the levels of the table values falling on either side of the census value. Of these nine values, the median was selected as the value most characteristic of Fiji in the period 1956-1966, and model female deaths $_nm_x$, were calculated for all ages by interpolation between the two tables yielding the median value. These were converted into rates from the life table populations, including the otherwise absent age groups under 10, and applied to the mean female population by age between 1956 and 1966, to yield average annual deaths. When the average annual deaths was divided by the population of Fiji at the mid-point of the decade, a crude female death rate was the result, designated $d_f$ with a value of 0.0135 or 13.5 per thousand.

The crude rate of natural increase is readily calculated by methods discussed earlier. The formula selected by Shryock and Siegel is

$$e^{rt} = \frac{P_{1966}}{P_{1956}},$$

which describes a continuous compounding of growth rather than the periodic pattern discussed in chapter

49/ United Nations. Age and Sex Patterns of Mortality, Series A. Population Studies No. 22; and Methods for Population Projections by Sex and Age, Series A, Population Studies No. 25.

50/ Ansley J. Coale, and Paul Demeny, Regional Model Life Tables and Stable Populations, (Princeton, New Jersey, 1966).

51/ *Op. cit.*, p. 821 ff

52/ Cumulation is the process of adding those over age 10, those over age 15, those over 20, etc. This process is in effect a method of smoothing out a range of misreported ages.

V. In the formula, $r$ represents the rate of increase, t the ten-year period, and the expression at the right is the ratio of the population in 1966 to that in 1956. Thus estimated, the rate of natural increase works out to be 0.0336 or 33.6 per thousand, and it is designated $i_f$. The crude female birth rate, $b_f$ is equal to $i_f + d_f$ or 0.0336 + 0.0135 = 0.0471 or 47.1 per thousand female population.

## B. MODEL STABLE POPULATIONS

A stable population is one with an unchanging relative age structure. It can be shown[53] that any schedule of fixed age-specific mortality rates, fixed fertility rates and the absence of migration, ultimately results in a population which may be growing, declining, or stationary but with a distribution by age and sex which is fixed. What was for many years an interesting demographic principle, has more recently been applied to the estimation of population parameters, where the quality of recorded data would not otherwise have permitted it. This application of the stable population principle was made possible by the realization that, until the relatively recent decline in mortality, many of the populations of developing countries had been experiencing birth and death rates that were comparatively stable over time, in spite of occasional disturbances. Also, it was realized that the selection of an appropriate stable population model on the basis of a few estimated parameters would enable the analyst to estimate many of them. If possible, the analyst should attempt to check the "stability" of the population by a comparison of the age/sex structure at two or more successive censuses, or by a comparison of the rates of growth of the population during two or more intercensal periods.

Shryock and Siegel[54] apply stable population analysis to Fiji, in spite of evidence that stability is only approximate. Working again with the female population, the "ogive" of the distribution is calculated, i.e. the successive percentages of the population C (x), below 5 years of age, below 10 years, below 15 years; at the ninth point on the ogive, the percentage under 45 C(x) equals 88 per cent of the female population. From a set of stable populations, pairs can be found for which respective values of C(x) fall on either side of the calculated values for Fiji. Those above 15 years are shown to be internally consistent, in that they fit the same two model tables; since lower ages seem to conform to entirely different tables, there may be errors in the data, especially at lower ages. Minor misreporting and age effects tend to disappear in the ogive technique. The model standard populations which best fit were levels 13 and 15 which had pre-calculated values for a birth rate, a death rate, $e_0^o$ and $e_5^o$. Selection of a best value yet remains, since the various C(x) values conform to the stable models to varying degrees. The recommended strategy is to select the pattern conforming to the median C(x). The estimate seems "robust" to the extent that the selection of one of the other central values would not have a critical effect upon the magnitude of the estimated parameters, although selection of a model conforming to the lower extreme of C(x) would result in substantial differences. This is the area where errors in the enumeration of children and where age misreporting have the most profound effects, and perhaps more significantly, where recent demographic changes such as declines in infant mortality or in fertility would partially invalidate the stable population assumption.

## C. ESTIMATES OF FERTILITY

Additional methods of estimating fertility appear in the literature.[55] A conceptually simple method involves the technique of reverse survival alluded to earlier. Survival ratios can be calculated with two censuses, and from the survival ratios a life table can be constructed. Using the survival ratios from the youngest intervals, the number of births implied can be calculated i.e. the number which must have occurred in order to account for the surviving populations of young children. When the number implied for the five-or ten-year intercensal interval are converted into an annual average, and after this number is divided by the estimated total population of the same sex at mid-interval, a single-sex crude birth rate results. When the same operation is carried out for the other sex and the numbers combined, the crude birth rate of the population can be calculated.

Another method requires only one census and a set of stable populations. In chapter IV of Manual IV[56] several values of the life expectation at birth are compared, and annual rates of increase of stable populations are compared with those of the actual 1961 population of Indonesia. The underlying idea behind the technique is that the age composition of an actual

53/ See Alfred J. Lotka, "Relation between birth rates and death rates", Science, New Series 26 (653) 21-22, 5 July, 1907. The rate of natural increase of the ultimate stable population, implied by the age-specific rates of a population, is its "intrinsic rate of natural increase".

54/ Op. cit., p. 824 ff.

55/ For example, see Shryock and Siegel, ibid., p. 826 ff.; and Manual IV, *passim*.

56/ Table 2, p. 29.

population is affected to a greater degree by changing fertility than by changing mortality, and that a population with a large percentage of children would be a high fertility population. Furthermore, if the population had a high percentage of children and high mortality, then fertility must be very high; whereas if a comparable population had low mortality, its fertility would be only moderately high. It remains to establish numerical relationships and some degree of confidence in the estimates. The manual accomplishes by using stable population theory to estimate the birth rate for Indonesia. Various indices of the age distribution, e.g. proportions of the female population under age 5, under age 10, etc., were plotted in combination with selected life expectations at birth and with rates of population growth. Patterns of age misstatement revealed by the plots were characteristics of India, Pakistan and many African countries, which suggested the rules of estimation devised for these age distributions. A total of 18 plausible birth rates emerged from an examination of three $e_0^o$ values and three rates of increase each for three age distributions, as measured by proportion under specified ages. It may be tentatively stated that the birth rate was not less than 45 per thousand nor more than 56 per thousand. The analyst of subnational populations who needs an estimate of fertility would be advised to apply whichever of the indirect measures he is able to apply to the total country. However, if warranted in a particular region, a different estimate of fertility should be made and taken into account. Application of indirect measures to subnational areas may result in serious miscalculations, unless the effects of migration can be removed.

The total fertility rate was defined in chapter IV as the sum of age-specific birth rates, a summary measure of fertility measured by the number of live births experienced by a hypothetical cohort of women passing through the child-bearing years without suffering any mortality. A total fertility rate might be constructed from the responses to questions put to those mothers, if (a) fertility behaviour in a population had not changed very much in recent decades; (b) the fertility history of the population was relatively unaffected by migration; (c) differential mortality according to number of children ever born had not had an important effect upon reporting. Assuming accurate reporting, the average number of children ever born per woman in the cohort 45-49 years of age would be the total fertility, but reporting is often not accurate. An analysis of patterns of reporting reveals an interesting tendency whereby older mothers fail to report proportionately more of their births than do the younger mothers. An alternative hypothesis, that fertility rates have been rising markedly, is not reasonable.

William Brass and others[57/] have devised a method of estimating total fertility and the age-specific fertilities underlying it, by fitting a curve to the data on number of children ever born by age of mother. The curve depends upon the relatively higher accuracy of reporting by the younger women, and it also assumes that there exists a pattern in parity, i.e. birth order by age of mother. The older women apparently "forget" some births, fail to count some children who have grown up and left home, and under-report births of children who died in infancy or childhood, to a greater extent than the younger women do. Younger women are required to remember only more recent events occurring in lower numbers, and they are more likely to have all of their children with them.

Unfortunately, differences in age at first marriage, marriage patterns, customs governing remarriage of widows and divorcees, all preclude a standard behaviour within countries even where the effects of contraception have been minimal. However, two sources of data may exist upon which a relationship between fertility rates at different ages can be based for a given population: (a) registered births, and/or (b) the responses to a survey or census question on births during the preceding period, usually one year. Generally, in developing countries registration data are not adequate, so the alternative must be considered. Moreover, the assumption that understatement is not age-selective is probably more valid in the use of the survey question than is reliance on registration where such factors as age, urban-vs.-rural residence, and literacy are associated with failure to register.

The principal source of error in survey data of this nature is that of the reference period, because respondents may not understand the period for which births are to be reported; but this is presumably not age-selective. Using the results of a number of surveys from Africa which provide tabulations of children ever born and children born during the preceding year, both by age of mother, Brass et. al.[58/] has devised a method of fertility estimation.

57/ W. Brass, A.J. Coale, P. Demeny, D. Heisel, F. Lorimer, A. Romaniuk and van de Walle, *The Demography of Tropical Africa*, (Princeton, New Jersey).

58/ Ibid.

*Programme Management and Development — Report of a Regional Seminar* (E/CN.11/1225): 1976

No.25 *Report and Selected Papers of the Expert Group Meeting on Training in Family Planning Programme Administration* (E/CN.11/1226): 1976

No.26 *Report and Papers of the Expert Group Meeting on Social and Psychological Aspects of Fertility Behaviour* (E/CN.11/1231): 1974

No.27 *Report and Selected Papers of the Regional Project on Pre-testing and Evaluation of Educational Materials Used in Family Planning Programmes* (E/CN.11/1159): 1975

No.28 *Population Strategy in Asia — Report, Declaration and Selected Papers of the Second Asian Population Conference, November 1972* (E/CN.11/1152): 1975

No.29 *Report of the Comparative Study on the Administration of Family Planning Programmes in the ESCAP Region: Organizational Determinants of Performances in Family Planning Services in Selected ESCAP Countries* (under preparation)

No.30 *Report of the Comparative Study on the Input-Output Relationships in Family Planning Programmes in Selected Countries of the ESCAP Region* (under preparation)

No.31 *Report of the Multinational Study in Methodologies for Setting Family Planning Targets in the ESCAP Region* (under preparation)

No.32 *Guidelines for Preparing Subnational Population Projections* (E/CN.11/1230): 1975

No.33 *Report of the Expert Group Meeting on Population Projections* (E/CN.11/1250): 1976

No.34 *Report and Selected Papers of the Expert Group Meeting on Translation of Population Materials* (under preparation)

No.35 *Comparative Study of Population Aspects of Manpower and Employment* (mimeographed): 1976

**B. Country Monograph Series**

No.1 *Population of Hong Kong* (E/CN.11/1120): 1974
No.2 *Population of the Republic of Korea* (E/CN.11/1241)
No.3 *Population of Thailand* (under preparation)
No.4 *Population of Sri Lanka* (under preparation)
No.5 *Population of the Philippines* (under preparation)
No.6 *Population of Japan* (under preparation)

**C. Miscellaneous publications**

1. *Report of the Expert Working Group on Problems of Internal Migration and Urbanization, and Selected Papers* (SA/Dem/EGIM/L.22): 1967

2. *Projections of Populations of Subnational Areas — Report of a Working Group* (E/CN.11/897): 1969

3. *Directory of Key Personnel and Periodicals in the Field of Population in the ECAFE Region* (E/CN.11/898; United Nations publication, Sales No. E.70.II.F.14): 1970

**D. The Second Asian Population Conference, Tokyo, 1972**

Official report (refer *APSS* No.28)
Available documents

**E. Asian Population Programme News**

A quarterly regional newsletter, published since 1971. Issues to date comprise Vol. 1, Nos. 1, 2, 3; Vol. 2, Nos. 1, and 2 and 3 (combined); Vol. 3, Nos. 1 and 2 (combined), Nos. 3 and 4 (combined); Vol. 4, Nos. 1, 2 and 3 (combined), No. 4; Vol. 5, Nos. 1, 2; and occasional special issues.

**F. Population Headliners**

A monthly mimeographed news-sheet. Circulation is limited to ESCAP Population Correspondents and Member Government decision-makers.

**G. ADOPT (Asian and World-wide Documents On Population Topics)**

A monthly listing of publications with in-depth subject analysis for population libraries and reference centres in the ESCAP region. Available to institutions, professionals, libraries and reference centres on request.

## HOW TO OBTAIN ESCAP PUBLICATIONS

*ESCAP publications for sale may be obtained from:*

(1) ESCAP Secretariat, United Nations Building, Rajdamnern Avenue, Bangkok 2, Thailand

(2) Sales Section, United Nations, New York, N.Y. 10017, USA

(3) Sales Section, European Office of the United Nations, Geneva, Switzerland

*ESCAP free publications may be obtained only from ESCAP at address (1) above.*