ST/SOA/Series A
Population Studies, No. 10

Manuals on methods of estimating population

MANUAL I

Methods of Estimating Total Population for Current Dates

UNITED NATIONS

POPULATION STUDIES, No. 10

Manuals on methods of estimating population

MANUAL I

Methods of Estimating Total Population for Current Dates

UNITED NATIONS
Department of Social Affairs
Population Division
New York

ST/SOA/Series A. POPULATION STUDIES, No. 10

List of reports in this series to date:

Reports on the population of Trust Territories

No. 1. The population of Western Samoa

No. 2. The population of Tanganyika

Reports on population estimates

No. 3. World population trends, 1920-1947

No. 10. Methods of estimating population
Manual I: Methods of estimating total population for current dates

Reports on methods of population statistics

No. 4. Population census methods

No. 6. Fertility data in population censuses

No. 7. Methods of using census statistics for the calculation of life tables and other demographic measures. With application to the population of Brazil, by Giorgio Mortara

No. 8. Data on urban and rural population in recent censuses

No. 9. Application of international standards to census data on the economically active population

Reports of demographic aspects of migration

No. 5. Problems of migration statistics

No. 11. Sex and age of international migrants: statistics for 1918-1947

No. 12. Economic characteristics of international migrants: statistics for selected countries, 1918-1950

Reports on legislation relative to population

No. 13. Survey of demographic legislation

UNITED NATIONS PUBLICATION
Sales No.: 1952. XIII. 5

Price: $U.S. 1.50
(or equivalent in other currencies)

ST/SOA/SERIES.A/
Population Studies No. 10

MANUAL ON METHODS OF ESTIMATING POPULATION

MANUAL I

METHODS OF ESTIMATING TOTAL POPULATION FOR CURRENT DATES

Corrigendum No. 1

Page 28, second column, line 6, under section 2.

For *Switzerland* read *Swaziland*

FOREWORD

This is the first of a series of manuals on methods of estimating population, which the Population Division of the United Nations Department of Social Affairs has undertaken to prepare in accordance with a recommendation of the Population Commission (E/CN.9/88, paragraph 25). This Manual deals with methods of estimating the total population of a country or territory for current dates. In later manuals it is intended to take up other problems in the field of population estimates, including estimates of past and future population and of population characteristics.

The primary purpose of these manuals is to assist governments in improving the quality of official population estimates. In addition, it is expected that they will be useful as materials for national and international training courses and university teaching in the field of demographic methods, and as reference works for individual technicians engaged in demographic research.

The present Manual was prepared by the Population Division in consultation with the Statistical Office of the United Nations.

CONTENTS

INTRODUCTION

1. The need for current population estimates

Information about the size of a country's population has been sought since ancient times. In some of the old empires, efforts were made to ascertain approximate numbers of inhabitants so as to estimate the military forces which could be recruited, the numbers of slaves who could be impressed, and the numbers of persons from whom taxes could be collected. Similarly, the headmen of tribal societies have wanted to know the numbers of families in tribal divisions in order to estimate the areas which were required for seasonal pastures, or the number of representatives who would appear at a gathering of the tribal council.

In modern times, the functions of government have become increasingly complex. In addition to the preservation of law and order, government has become more and more entrusted with the planning of economic and social programmes; this requires a fairly accurate knowledge of the size of the country's population, its rate of growth, its distribution among the various towns and provinces, its composition by sex, age, ethnic and educational groups, and the extent to which it is engaged in, or depends on, various branches of economic activity. The further a country has progressed in its technical development, the greater is the need for accurate and detailed information.

Information regarding the size of the population and its growth can be secured by periodic census enumerations, records of births, deaths and migration, and, at least in a few countries, by a continuous population register. Such information can be supplemented in some areas by records of school attendance, rationing registration, occasional sample enumerations, taxation lists, records of military conscription, social insurance records, voting registers, etc. However, facts of this kind are not available in all countries, nor are they always up-to-date even in those countries where they have been gathered.

In most countries in which technical development has advanced relatively little, censuses have been taken sporadically, if at all, and have yielded only a limited amount of information. The cost of periodic detailed censuses is large in relation to the financial resources of the governments in such countries. Where the level of education is low, moreover, it is difficult to assemble a sufficient staff of trained enumerators for a simultaneous counting of the entire population, and difficult to get accurate answers to census questions from the people. Obstacles to the accurate recording of births, deaths and migratory movements appear even greater; they are considerable even in countries which are much richer and which have a more educated population.

In some of the most advanced countries, the social and economic tasks of government have reached such high complexity that detailed and accurate statistics regarding their population and its characteristics are needed. Even in these countries, however, there are limits to the frequency with which censuses can be taken, to the number of questions which can be asked, and to the numbers of answers which can be tabulated on the occasion of any one census. Moreover, the final results of an enumeration are not entirely up-to-date by the time they become available. Estimates are therefore required if changes since the census, or items not included in the census, are to be taken into account.

2. Recent progress in population estimates

Much progress was made in the last century in collecting statistics concerning the size, structure and rates of change of the population in various countries of the world. The periodic enumeration of the population and of its main characteristics became an established practice in most Western countries during the nineteenth century. Towards the end of the century, censuses were taken in India and the Russian Empire, while the size of the Japanese population was determined by periodic reviews of the population registers. Statistics of births and deaths, though not always complete, were being collected in many parts of the world. A close study of population movements and up-to-date knowledge of the population required the development of various estimating techniques, which led to the establishment, in the census offices of many countries, of special sections concerned with population estimates.

Progress during the present century has been even more rapid as the recognized need for demographic knowledge has increased both in extent and in scope. Scientific censuses have been taken for the first time in many more countries, and the characteristics of the population investigated in much greater detail and with increased precision. The registration of births and deaths has been begun in new areas, while in older areas it has been made more complete and detailed. Still more recently, significant investigations have been made regarding the quality of the statistics collected. The improvement of scientific sampling techniques in recent years has provided useful tools for the collection of additional data and the evaluation of the accuracy of censuses and vital statistics.

In the countries where census-taking and the collection of vital statistics has become a long-established practice, elaborate techniques are used for estimating a wide range of population aspects. In other countries, where basic statistics are relatively new and more limited in scope, recent estimates are less refined and detailed than in the former countries, but are markedly superior to those estimates which previously were produced. There is still a considerable number of countries and territories in which there has been no census and where statistics relating to population trends are either absent or very scanty. Even in those countries, however, increasing efforts are being made at estimating the population by various other means, some of them requiring considerable ingenuity. Although these estimates, at best, can be only rough approximations, they can serve a variety of purposes if their limitations are kept in mind.

The advances made in preparing current estimates of total population are reflected in international statistical compendia. The International Statistical Institute, prior to the Second World War, did not find it feasible to publish reasonable population estimates for several countries and territories of the world. Between the two world wars, the League of Nations published current estimates for *every country* of the world in its *Statistical Yearbook,* though many of these estimates were known to be unreliable, and the reliability of many others was unknown since the method of obtaining them could not be ascertained. The United Nations, in its *Demographic Yearbook,* is now in a position not merely to publish current estimates for every country, but also to specify the manner in which most of these estimates have been derived and thereby to indicate — at least roughly — how reliable these estimates are.

As a result of rapid progress in obtaining basic data in some countries and less rapid progress in others, the best way of making population estimates varies more widely than was formerly the case. In many countries, the estimates that are published currently are probably the best which can be made in view of existing statistical information. However, there are still many countries and territories in which the basic information, whether good or poor, is not utilized to best advantage.

3. Scope and purpose of this Manual

This Manual is confined to the problem of estimating the total population of a country at a current date. There are many other related problems, such as estimating the past and future population, its geographic distribution, its composition by sex and age, and its structure according to various other characteristics. It is planned that some of these problems will be dealt with in future manuals.

Although this Manual is intended for use in all countries of the world, it is concerned especially with the preparation of population estimates in those countries where the available statistical information is rather limited. The Manual shows that current estimates of population can be produced even in countries having very few basic statistics, that existing information often permits the improvement of the estimates which are now being made, and that additional data, which would permit the making of superior estimates, can often be obtained at small cost. The Manual also shows that in countries with highly developed statistics, there are still various problems regarding the comparability and consistency of estimates.

The Manual is concerned not only with the methods of constructing population estimates, but also with the problem of evaluating their quality.[1] In some countries, official population estimates are published with comments relating to their nature and accuracy, but these countries are still in a minority at the present time. The appraisal of the quality and reliability of estimates is a subject which should concern all countries of the world — those with highly developed statistical systems as well as those where statistics are more rudimentary. Mediocre statistics, whose nature and reliability are not indicated, have often proved to be more dangerous and have led to more incorrect conclusions, than much poorer statistics which are clearly shown to be such.

The information necessary for an accurate evaluation of the reliability of population estimates can often be obtained rather cheaply, and it can enhance greatly the usefulness of results obtained from base data which are secured at great expense. Sampling offers a cheap and effective means both of checking the accuracy of existing information and of obtaining new data. Some of the ways in which sampling and other techniques can be used for these purposes are discussed in this Manual.

Population estimates are used to serve various purposes in various countries. For this reason the estimates now being made often differ from country to country as regards definition, coverage and time reference. Some countries, for instance, define their population in terms of habitual residence, while others define it in terms of actual presence within the territory. Categories such as temporary migrants and travellers, nomadic tribes, merchant seamen, foreign troops stationed inside the country or national troops stationed abroad, certain cultural or racial groups, migrant workers and their dependants, or displaced persons, are sometimes included and sometimes excluded. Similarly the estimates may refer to the middle of the year or to some other date.

While these differences may be well justified because of specific national needs, they lessen the comparability of estimates from country to country which is highly desirable from an international point of view. The Population Commission and the Statistical Commission of the United Nations have therefore established certain

[1] One requirement for a full evaluation of the quality of population estimates is the appraisal of the quality of the basic statistics; this subject, which is somewhat specialized, is scheduled for separate treatment in a future manual.

standards for demographic statistics. In preparing estimates which are useful from a national point of view, it is highly desirable that every country consider carefully the modifications which would facilitate comparisons with those of other countries. To achieve better international comparability, some of the estimates now being made in various countries need to be adjusted in certain ways, which are discussed in this Manual. Where conformity to international standards is not possible, it is desirable that the deviations from the standards be clearly indicated, or that appropriate sub-totals be given in order that the desired figure may be derived. Information relevant to international comparability is still lacking in the statistical publications of many countries.

4. Types of methods of estimating total population for current dates

Because there is great variety in the amount and type of information relative to population size and population growth available in the various countries of the world, the methods used in making population estimates must vary accordingly.

At one extreme are the few countries which have a current bookkeeping system, known as a "continuous population register". In such a register a new entry is made for each birth and for each arrival of an immigrant, while a deletion is made for each death and for each departure of an emigrant, the actual events being recorded within a very brief interval of time. By balancing the books at convenient intervals, such as once or twice a year, a highly accurate estimate of the population can be obtained.

In other countries which are well advanced statistically, no such system is in operation, but information on each of the component items of population change is available from separate sources. The population total is known for the date of the last census, the number of births and deaths can be obtained from registers of vital statistics and the number of immigrants and emigrants from the records of migration across the frontiers. Adding to the last census the total excess of births over deaths and the net immigration (or subtracting the net emigration) gives an estimate of the current population. This procedure takes into account every component of population change. If the census and the statistics of births, deaths and migration are complete, such an estimate is highly accurate. Even if there are minor defects in the vital statistics and inadequate migration records, quite accurate estimates of total population can often be obtained.

There are many countries where censuses have been taken, but where birth and death registration either does not exist or is known to be very inaccurate. In such countries, population estimates for current dates can be produced by means of mathematical extrapolation. This method is less reliable than the foregoing, since it is unlikely that population change will conform precisely to a mathematical formula. It may, however, be the best method if vital statistics are markedly defective. No mathematical methods can be used in countries where the results of only one census are available, or where censuses have been of doubtful and variable accuracy.

In some countries, and many territories, there has so far been only one census, while in others there have been only certain types of enumeration (e.g., registration for food rationing or a count of the number of households) which are not comparable to a census. Population figures of this type can be brought up to date only by the use of certain assumptions — some of them quite arbitrary — concerning rates of population growth. Some non-censal counts, however, are repeated annually, which facilitates obtaining current population estimates.

There are, finally, a considerable number of countries and territories in which the population has never been counted in any way. In such areas population estimates can be based only on conjectures, or "reasoned guesses" Conjectures can be based on comparisons of certain conditions in these areas with those in other areas where population size or density is known. There is scope for the use of much ingenuity in making the best possible conjecture.

Even where a high degree of reliability cannot be attained in view of the limited information at hand, it is still important to obtain at least some kind of a population estimate. A high degree of precision is not needed for every purpose. If an accurate estimate cannot be made under given conditions, this should not stop efforts to produce as good an estimate as circumstances permit. It is important, however, for the reasons stated above, to indicate the methods used and the degree of reliability of the result.

In this Manual, the problems which arise in the making of population estimates and the possible ways of improving them are discussed separately with reference to each of the types of estimating procedures mentioned above. The discussion begins with the estimates having the least adequate foundation: the conjectural estimates. Succeeding chapters relate to estimates based on defective censuses and non-censal counts, estimates made by mathematical extrapolation and similar procedures from the results of one or more censuses, and finally, estimates based on the results of censuses and current vital and migration statistics, and on continuous population registers. The discussion of the various methods is preceded by a chapter dealing generally with the problem of assessing the quality of population estimates.

I. CONSIDERATIONS OF QUALITY IN POPULATION ESTIMATES

1. Determination of the accuracy of estimates

Population estimates are like tools, some of which can properly be used for rough work only while others are refined instruments. Accurate estimates can serve a great variety of purposes; approximate estimates have more limited uses and cannot safely be employed where precision is important. The user of population estimates needs to know how much he can rely on their accuracy. It is the responsibility of the producer of the estimates to state as definitely as possible the magnitude of the possible errors, and thus to guard against misuse of the figures.

In some instances the margins of error of a population estimate can be determined with considerable precision on the basis of adequate investigation and experimentation. Where such precise indications are not possible, at least an approximate indication of the possible amount of error can usually be furnished. As a rule, the determination of the margins of error requires an appraisal of the accuracy of the data on which the estimates are based; for example, an investigation of the completeness of census enumeration and of birth and death registration. The methods of appraising various types of data used for population estimates are to be discussed in a later publication. Knowledge of the accuracy of these data being given, it remains to determine the consequent margins of error in the population estimates, and this problem is treated in the present Manual, with reference to each of the major types of estimates.

Most current estimates of population have two components: (*a*) a "base figure", that is, a count or estimate of the population at a previous date, and (*b*) a "time adjustment", that is, an allowance for population increase or decrease since the previous date. The accuracy of the estimate, of course, depends on the accuracy of both components.

In the case of conjectural estimates and of certain estimates based on non-censal counts or defective census enumerations, even the base figure is an estimate. In making estimates of these types, the base figure is often established by counting or estimating a certain element or category of the population, or some other quantity which bears a numerical relation to population, and multiplying the result by a factor in order to reach the total population figure. In that case, the accuracy of the base figure has to be considered in view of the possible error, not only in the statistics used but also in the multiplier.

In most cases, however, the time adjustment is considerably less accurate than the base figure. The relative importance of the error in the time adjustment depends on the length of the period between the date of the base figure and the current date. If the base figure is of recent date, the reliability of the current estimate is almost entirely controlled by the accuracy of the base figure. Population usually does not change abruptly within a short period of time, and although some error is introduced by the time adjustment, this error is likely to be smaller than that in the base figure. On the other hand, if the base figure refers to a date in the distant past, the error in the time adjustment may be the more important.

Example: A count resulted in a population figure of, say, 5 million. In the following year, it is estimated that the population has increased by 50,000. If the total of the count was subject to an error of 2 per cent, and the estimated increase was subject to an error of 20 per cent, the population may have increased from 4,900,000 to 4,940,000, on a minimum assumption, whereas on a maximum assumption it may have increased from 5,100,000 to 5,160,000. An intermediate estimate of the population in the following year is 5,050,000 subject to an error of 110,000, or only slightly over 2 per cent.

After the lapse of twenty years, however, the situation is changed. If an annual increase by 50,000, subject to a relative error of 20 per cent, is assumed for the entire period, the population may have increased by as little as 800,000, or by as much as 1,200,000. Under extreme assumptions, it may have grown to a total of 5,700,000 as the minimum, or 6,300,000 as the maximum. The estimate may then be put at an intermediate value of 6 million, subject to an error of 5 per cent.

In the foregoing example it is assumed that the probable direction of the errors both in the base figure and in the estimated increase, is unknown; the figures may equally well be too large or too small. In some cases, an assessment of the quality of the data used will give evidence of an error in one direction, and in those cases the data should be corrected before the estimate is made. If the amount of the correction required remains in doubt, an effort should be made to determine reasonable upper and lower limits for the correction, and these limits should be considered in stating the margins of error of the resulting population estimate.

Example: In a census, 5 million persons were enumerated. Tests of the completeness of enumeration were carried out in two sets of sample areas, indicating, in one set of areas, an under-enumeration of 4 per cent, and in the other, an under-enumeration of 2 per cent. There is no *a priori* reason to suppose that either test was more valid than the other. The correction for the census count may be estimated at 3 per cent, and its upper and lower limits may be placed, after due consideration of all the relevant facts, at, say, 5 per cent and 1 per cent, respectively. The corrected census figure is then 5,150,000. A population estimate for the next year, using vital statistics which

show an increase of 50,000 subject to an error of 20 per cent in either direction, would be 5,200,000. The possible error of the estimate may be put at about 110,000, or slightly more than 2 per cent, in either direction.

In considering the possible errors in statistical time series which may be used as a basis for time adjustments in population estimates (notably, statistics of births, deaths and migration) it is useful to distinguish what may be called biased, random and self-correcting errors. Biased errors are those which tend systematically either to minimize or to exaggerate the population increase over a period of years; an example is under-registration of either births or deaths. Random errors, in the sense intended here, may tend toward either an underestimate or an overestimate of the increase during any year; an example is the error in an estimated correction for under-registration of births or deaths. Self-correcting errors are those, the occurrence of which in one year creates a tendency toward a compensating error in subsequent years; an example is the error due to late registration of births or deaths. It is the biased errors which present the most serious problem. Self-correcting errors do not accumulate in the long run, and random errors accumulate only slowly, but biased errors accumulate in direct relation to the number of years over which the time adjustment is continued.

Where time adjustments are made by mathematical extrapolation of population increases observed in the past (for example, by extrapolation of the increases between successive census dates) it is difficult to find the basis for an objective evaluation of the possible errors. Certain biases are inherent in each method of extrapolation, and the selection of a particular method depends on an individual's judgment, which may also be biased. Nevertheless, there are means, which will be discussed in a later chapter of this Manual, of arriving at an indication of the possible errors in such time adjustments.

Where information on the population is scant and the levels of birth rates, death rates and rates of increase are unknown, it is necessary to make time adjustments rather arbitrarily, by assuming some plausible rate of population increase, or by merely holding a figure, derived at some time in the past, at a constant value. Possible errors arising from such assumptions with respect to population increase can be evaluated, though only very roughly. There are upper and lower limits for the rates of population change (increase or decrease) which, in a given situation, can be regarded as within reason. With the passing of time, the error resulting from either holding a previous figure constant, or from adjusting it at an arbitrary rate, is likely to increase cumulatively.

The accuracy of the time adjustment has an especially important bearing on the usefulness of population estimates as indicators of population growth, or of changes from one time to another in per capita measures such as income and production or consumption of various items per person. For such comparisons over time, population estimates are adequate if the time adjustment is nearly accurate, even though the base figure is unreliable and the magnitude of the population in any year is therefore uncertain. Conversely, comparisons over time are vitiated if the time adjustments are inaccurate, even though the base figures may be of the highest quality. Since one of the most frequent uses of current population estimates is comparison over time, it is important that the method of adjustment be known.

Adjustments using statistics of births, deaths and migration of at least tolerable accuracy may be regarded as adequate for this purpose since they reflect, more or less faithfully, the true variations in population size from year to year. Adjustments depending on an assumed rate of increase, on the other hand, can tell us nothing new about recent population changes, since they merely reflect those assumptions which were made beforehand; a time comparison of estimates of this type leads, so far as population size is concerned, to foregone conclusions. The same is also true of current estimates adjusted by mathematical extrapolation which reflect an assumption that recent rates of population change are a continuation of changes observed in the past.

One way of determining margins of error of an estimate, particularly of one based on certain assumptions (conjectures, estimates involving the use of a multiplier, and extrapolations) is to make several independent estimates using different assumptions, all of which are within reason, and then to compare the results. The extreme values obtained may indicate approximately the limits of the range which should be stated, and the figure derived by what is believed to be the most reliable method may be taken as the best estimate. A critical comparison of the several values which are derived by different methods may, however, lead to the conclusion that the best estimate is some other figure than that derived by the method that was at first believed to be most reliable, or perhaps some intermediate value between the figures resulting from estimates of similar reliability. In the case of conjectural estimates, in particular, it may even be desirable to compare several estimates made by a number of fairly competent persons independently of one another, to consider the range within which their estimates are contained, to determine some intermediate figures as the "best estimate" subject to an error represented by the extreme estimates.

In some cases, when the result of a census was of dubious quality, or when vital statistics of dubious quality have been used in an adjustment over a long series of years, some of the above processes may also be helpful in determining a reasonable margin of error for the estimate or in suggesting some upward or downward revision.

2. Methods of stating the degree of reliability

Where the information available permits a quantitative statement of the margins of error in a population

estimate, it is desirable to publish this statement, perhaps in the form of a percentage of possible error in each direction, together with the estimated figure. The foregoing discussion, however, makes it clear that in many cases the evidence regarding possible errors is not sufficient to permit such a definite statement. In many cases the best that can be done is to make an informal guess at the possible extent of errors, based on knowledge of the statistical procedures in the country and of the extreme values which can be assumed for data and assumptions relating to population size and population change, combined with critical judgment. The margins may be indicated in such cases by a statement such as "figure believed correct within about 5 per cent", or "approximate estimate; believed correct within about 10 per cent", or "figure believed correct to the nearest 100,000". The margins stated should be such that it is unlikely, though not necessarily impossible, that the error exceeds the stated amount.

If it is felt that the margin of error of a figure cannot be safely expressed as a percentage of that figure, it should be indicated by qualifications such as "estimate believed to be fairly accurate", or "approximate estimate", or "very approximate estimate, possibly subject to a large error".

It is unfortunate that the practice of indicating margins of error in population estimates, as well as many other types of statistics, is not more generally followed in the national statistical offices. The failure to indicate the approximate nature of certain statistics may sometimes be motivated by an expectation that figures not so annotated will give an impression of being exact, or at least nearly so. However, it is a well-known fact that in many countries, in view of very obvious difficulties, it is next to impossible to obtain highly accurate statistics. On the other hand, it is the practice of many countries with well developed and highly accurate statistics to indicate errors, however small, which are present in the figures. Indications of inaccuracies are an almost unfailing sign that efforts are made to appraise the quality of the statistical information, and hence to improve it. Far from detracting from the value of the figures, indications of the degree of their reliability actually increase their usefulness.

An indication of the methods used in making the estimates is also very valuable, as has been shown above, to the user of the statistics. Such information, together with an assessment of the probable accuracy of the results, not only helps to prevent unjustified uses of the estimates, but also to encourage those uses for which the estimates are adequate.

In certain countries, the population has been estimated arbitrarily by an authority guided only by the intention of producing a desired figure, for purposes of propaganda or to increase the prestige of the country or its government; such estimates are, of course, useless. Cautious users of population estimates may be led unjustly to suspect that figures for some other countries have been arrived at equally arbitrarily. Statements of the methods used can dispel such suspicions where they are not warranted.

3. The rounding of figures

It is common practice to round large figures for one of two reasons. In the first place, figures running into many digits are cumbersome to handle in computations, while little is gained in their usefulness by presenting all the digits; for this reason it is often found more expedient to abbreviate them, showing only the first few significant digits, perhaps the nearest 100 or the nearest 1,000, as the case may be. On the other hand, it is often felt that a figure is not sufficiently reliable to justify showing it to any large number of significant digits, and it is therefore preferred to show only the first two or three digits, as the case may be, in order to indicate that they are approximate.

However, the mere rounding of a figure is not a sufficient indication that it is inexact. Thus, if the population of a country is stated to be 7 million without any qualifying remark, there is no way of telling whether this figure can be subject to an error of 20 per cent or one-twentieth of 1 per cent. On 21 July 1946, the population of Austria was reported to be 7,000,003, an estimate which was probably fairly accurate, though not to the last digit. The population of Afghanistan, on the other hand, was estimated very roughly for the years 1927 to 1939 by the League of Nations at 7 million, a figure which may have been in error by several million.

The greatest shortcoming of rounded figures is that they vitiate comparisons of estimates over time. The population of a country may have been estimated very roughly at 7 million at some time in the past and may be assumed to have increased by 5 per cent in each of two subsequent periods of time. Accepting 7 million as the initial figure, we should have to estimate the population at 7,350,000 at the end of the first, and at 7,717,500 at the end of the second period. Rounded to the nearest million, the figure would remain at 7 million at the end of the first, but would rise to 8 million at the end of the second. A comparison of these rounded figures would suggest no increase in the first, by an increase by 14.3 per cent in the second period. The rounding of figures in such a case may obscure valuable information regarding possible population changes.

The rounding of figures is, therefore, only a poor substitute for other indications regarding their reliability, and should not be used for that purpose. Some rounding may be desirable for purposes of abbreviation with the intent of saving space and labour in computing where a higher degree of precision is unnecessary. Also, in an estimate of a low order of reliability it is clearly absurd to show all digits. As a rule, estimates should be shown to at least as many figures as are significant in view of the method of estimating (e.g., from the assumption of certain rates of increase, etc.), while indications of the approximate

nature of these figures are given in annotations or separate statements.

4. The problem of internal consistency

An estimate of population can hardly be reliable unless its components are at least approximately consistent in regard to the definition of population and the area covered. Where estimates are made from census statistics adjusted to current dates by means of vital statistics and migration statistics, the area and population covered by all these types of statistics should be the same. It is particularly important where results of several censuses are used in making estimates by mathematical extrapolation that the coverage and the completeness of enumeration in the series of censuses should be constant.[1] Likewise, the coverage of series of vital statistics and migration statistics used in making population estimates should be the same from year to year. Where estimates of the total population of an area are built up with data for component parts of the area, the figures used for the parts should be consistent as regards not only the definition of population but also the dates to which they refer, and to the procedures of counting employed. If the conditions of consistency are not satisfied, it is desirable to make corrections in the various figures, in order to bring them into line.

Consistency is important not only in the components of each single estimate, but also in a series of estimates. If the coverage of the estimates for one period of years differs considerably from that for another period, the usefulness of the series is obviously much impaired. The problems of establishing consistency in a historical series of population estimates are to be taken up in a later publication.

A fairly common source of major inconsistencies is changing national boundaries, which obviously require corrections in series of demographic statistics which may be used in making population estimates. Another source of inconsistencies which often assume major dimensions is the failure to cover some parts of the national territory or of its population. Vital statistics are sometimes confined to "registration areas" which comprise more or less large fractions of the areas covered by the censuses. Records of migration are sometimes taken only at certain points of entry or exit, for example at the ports but not at land frontiers. In such cases the partial statistics of births, deaths and migration may have to be amplified before they can be used for reliable adjustments of census figures to current dates.

The exclusion of certain population categories, such as native population in various African territories or tribal aborigines in some Latin-American countries, may create problems of consistency. Such categories may be covered by one census, but not by another; if covered by the census they may be excluded from the vital statistics and from records of migration. In such cases, difficult problems of estimation may be involved in establishing a consistent series of figures relating to the whole population.

Such groups as prisoners of war, displaced persons and armed forces at home or abroad may be excluded from some of the statistical series and included in others, and their coverage may change from time to time within the same series. In many cases vital statistics relate only to the births and deaths of residents, and migration statistics to persons entering to take residence in the country or leaving to take residence elsewhere, whereas census statistics refer to all persons present in the country at the time of enumeration. It may be safe to ignore these inconsistencies in many instances because the changes in population indicated by the vital and migration statistics may be nearly the same as those which would be shown by data defined in a manner consistent with the census definition. In other instances a "correction", either in the census total or in the vital and migration statistics, may be necessary in order to avoid substantial distortion of the estimates.

Inconsistencies may be created by changes in the degree of completeness of enumeration from one census date to another, or by changes in the completeness of registration of births and deaths, or the completeness of recording of migration, over a period of time. The elimination of such inconsistencies is, of course, part of the problem of evaluating the accuracy of the statistics and correcting them for use in population estimates.

If important adjustments or corrections in the data are made, this fact should be indicated in the publications which contain the estimates of population. For example, a census enumeration limited to the settled population may have given a figure of 800,000, to which an estimate of the number of nomads, say 50,000 at the time of enumeration, is added to arrive at the total population. Population estimates for subsequent dates, based on this total for the census date, should be accompanied by a statement such as, "Includes nomads estimated at 50,000 at the time of the census". If there is no further information regarding the nomads, adjustment for population changes in time may be carried out on the assumption that the nomadic population increases at the same rate as the original population. If, on the other hand, there is reason to believe that the nomadic population is either stationary or increases at a rate different from that of the rest of the population, adjustments for change in time should be made separately for the enumerated population and for the estimated number of nomads. This should then be recorded in a note saying "estimate adjusted to include nomads estimated at a fixed number of 50,000", or "estimate adjusted to include nomads, estimated at 50,000 at the time of the enumeration, and believed to be increasing at a rate of . . ."

If known inconsistencies in the components of an estimate cannot be eliminated by correcting and adjusting the figures, the nature of these inconsistencies should be stated in the publications presenting the estimates.

[1] See chapter V, section C.

5. The problem of international comparability

Consistency is desirable, not only among the components of each population estimate and within each time-series of estimates, but also between these estimates and other statistical series, in conjunction with which they are often used. An important aspect of consistency in this broader sense is consistency between the estimates for one country and those for another — that is, international comparability.

It is highly desirable that estimates be internationally comparable, so that the density or growth of population, and various per capita measures for a nation can be compared with those of other nations. It is often desirable to obtain data for a group of countries, and this cannot be done satisfactorily by adding together several estimates which are not strictly comparable. It is also advisable that internationally comparable data be available for identical dates so that they may be added up to give simultaneous data for entire regions.

It is difficult, if not impossible, to apply identical census definitions, or identical definitions in vital statistics and migration statistics, in all countries. The conditions under which enumerations are made in different countries vary greatly, and public expenditure is always directed toward obtaining those results which are of the most immediate interest to the administrative requirements of the particular country and which are relatively less difficult to obtain. Differences in definitions in the statistics collected in various countries are, therefore, likely to persist.

It has been recommended by the United Nations that census statistics be obtained in accordance with certain standard definitions wherever possible, without prejudice to obtaining those statistics which are of immediate interest to the country concerned. These recommendations favour a "modified *de facto*" definition of the total population including all persons present in the country, with the exception of foreign armed forces stationed in the area, but including national armed forces located abroad.[2] It is, of course, understood that for many purposes, both national and international, this population may not be relevant; but from the point of view of ensuring complete world coverage, free of double counting, it is the most desirable type of figure to obtain. Recommendations for standards in vital statistics and migration statistics are also being prepared by the United Nations.

While it is not possible in many countries to collect all statistical data in conformity with such definitions, it should always be possible to form at least those estimates of total population size which would conform to the standard, or to "correct" a figure conforming to a different definition by the amount by which it is estimated to differ from the standard definition. If, for different purposes, a different definition of the population is also required, it is desirable to provide separate figures for those population categories (e.g., foreigners temporarily present, nationals temporarily abroad, etc.) which are involved in the change of definition. If appreciable differences obtain, it is also desirable to estimate separately numbers of births, deaths and migrants, according to different definitions of the population.

In some instances, it may be felt that "corrections" made for purposes of conformity with standards may result in much loss of accuracy while national statistics, although diverging from international standards, are fairly accurate. If this is the case, it may be preferable to give a non-comparable national estimate rather than one conforming to standards. In all such cases, however, indications should be given to show in what manner a national population estimate deviates from international standards in order to provide the means, whereby users of statistics can make some alternative estimate conforming to international standards.

With regard to time reference, it has become a practice of most countries with a well-established statistical office to provide mid-year figures for every year. An annual mid-year figure, dated 30 June or 1 July, may therefore be regarded as the international standard. Theoretically, for the computation of birth and death rates or other per capita rates, the best possible figure for a year would be the figure for the "mean population", i.e., the average population for the entire year. In a few countries, such a "mean population" is computed by averaging the twelve monthly estimates.[3] This, however, is a great refinement which can make only a very small difference, particularly if the accuracy of population statistics is not very great. The difference between the mid-year and the "mean population" of a year can be disregarded in most cases.

If, however, population figures refer to some other time of the year, this circumstance is a considerable handicap to international comparability. Such figures are also less useful than mid-year estimates for many national purposes, since they differ to a greater extent from the "mean population" and introduce some errors into the computation of vital rates and other per capita ratios calculated on a calendar-year basis.

In some countries, population estimates are made for the first or the last day of every year. An arithmetic average, or some other interpolation, between two subsequent end-of-the-year (or beginning-of-the-year) figures may then be suitably used instead of a mid-year figure. In fact, while such interpolation usually results in a mid-year figure which differs slightly from the true mid-year figure, it may differ no more from the annual "mean population" than does the true mid-year figure.

In some countries, a mid-year estimate is not given for a year during which a census was taken, even though the date of the census was not 1 July. The reason may be that the census figure is considered more reliable

[2] United Nations. *Population Census Methods* (Population Studies No. 4). Lake Success, November 1949.

[3] In some countries, averages of national estimates for the beginning and the end of each year are published as "mean population" estimates.

than an estimate for some other date. However, unless the census was taken exactly at mid-year, the figure does not meet standards of comparability and is not appropriate for certain uses.

6. Conclusions

The discussions in the present chapter lead to the following principal conclusions:

1. It is desirable to appraise the possible errors in population estimates, at least in broad terms, and to publish a statement of the approximate degree of reliability in order to minimize the dangers of misuse and to encourage legitimate uses.

2. Estimates should ordinarily be presented with at least as many digits as are useful, in view of the degree of reliability for purposes of indicating both population size and population change.

3. Where there are important inconsistencies in the data used for population estimates, the data should be rendered consistent by adjustments or "corrections". These adjustments or corrections should be indicated in the publications containing the estimates, and any remaining inconsistencies should be pointed out.

4. In general, estimates are internationally comparable if they conform to the "modified *de facto*" definition. Deviations from this definition should be indicated.

5. Estimates should be made for the middle of each year, both for the sake of international comparability and for their greater general usefulness in computing annual rates.

II. CONJECTURAL ESTIMATES

1. The nature of conjectural estimates of total population

A population estimate may be described as "conjectural" if it is not based on numerical data relating to the population itself. Conjectural estimates depend mainly on quantitative and qualitative information concerning one or more factors which are related to population size. Among such factors are land area, types of settlement and the population density which, under given conditions, may appear feasible, or the total production or consumption of a staple commodity and estimated per capita rates of production and consumption.

Estimates based on numerical data pertaining more directly to part or all of the population, for example a count of houses, huts or tents, numbers of tax-payers, voters, or recipients of rations, are not conjectural; though not constituting actual population enumerations, they are based on counting procedures which must be regarded as "non-censal".

In some countries and territories, estimates of population have been derived, in part or entirely, from local population estimates, such as those supplied in reports of local administrators to the central administration. The basis of some or all of the local estimates is often unknown, and the quality of many is highly dubious. Owing to these uncertainties regarding the local figures, estimates of total population made by compiling the reports of local administrators should also be considered as conjectural, though they are probably superior to direct conjectures on total population size.

Sometimes, a small part of a country's population may have to be estimated by conjecture, the rest being estimated by superior methods. This is frequently the case with aboriginals in remote areas of a country, about which information is scarce. In such cases it is important, for the sake of a proper indication of the reliability of the total estimate, to state the fact that it includes a conjectural estimate of part of the population.

The first distinction to be made among estimates based on conjectures is that between conjectural "base figures" and estimates adjusted for a current date. A conjectural estimate made in the past is very often regarded as so unreliable that, whether or not population has changed in the meantime, it is still considered as the best estimate of current population size. The population of Ethiopia, for instance, has been estimated at a round 15 million, it being understood that this figure represents only a rough approximation. In the absence of further information, this same figure is also used as an estimate of probable population size in the past, and is likely to be retained for several years in the future unless new attempts are made to obtain a more reliable estimate. Nevertheless, it should be realized that the retention of the same figure over a series of years creates the impression that population size has, in fact, remained nearly constant. It is not advisable where there is reason to believe that population is either increasing or decreasing.

Thus although the population of Liberia was estimated by a rough conjecture in 1947 at the round number of 1,600,000 it was estimated in 1949 at 1,648,000, on the assumption that population is increasing at a rate of approximately 1.5 per cent per annum.

In a conjectural base figure estimate, two elements must be distinguished: the *circumstantial data* (e.g., total land area, or any other measure to which population may be thought to be in a certain relation), and the *multiplier* (e.g., the population density, i.e., number of persons per unit of land area, or any other ratio expressing the assumed relation between population and the circumstantial measure). Errors in the estimate can arise from both these components, each of which must be derived separately. The greatest difficulty usually attaches to the selection of a plausible multiplier, and this is the main reason why conjectural estimates are usually of a low order of reliability.

There are considerable areas of the world where, at the present time, population size cannot be determined on any firmer basis. A conjectural estimate, however, is always possible, and it is useful to make the best conjecture that can be made under given circumstances.

2. Conjectural estimates made by explorers and travellers

In 1885, travelling through the remote inner regions of China, the Russian explorer Potanin encountered a branch of sedentarized Mongols on the upper reaches of the Yellow River, near the present borders of Kansu and Chinghai Province, whom he called the Shirongol-Mongols. He estimated their numbers, by conjecture, as follows:[1]

> In San-chuan, i.e., in the area between Gyango-gol and Unchzhagol, there are 1,200 households; assuming five souls of either sex per household, we obtain 6,000 souls for San-chuan. If we allow for the population of Bouchzha-aral, Itel-gol, and Sombra with Badu-ol a figure of 2,000 souls, then the total population of this section amounts to 8,000. In Tun-syan one

[1] G. N. Potanin. *Tangutsko-tibetskaya okraina Kitaya i tsentralnayo Mongoliya.* Moscow, 1950, p. 377.

reports thirty-six imyks, i.e., villages; estimating that there are 100 households per village, one may put these at 18,000 inhabitants; however, assuming that the figure thirty-six is exaggerated, we may reduce the estimate for the population of Tun-syan to 10,000. The Shirongolian population of the lower part of Sinin-gol and of Day-tong-gol may also be accepted at 10,000. In the surroundings of U-yan-bu and Mubayshintu there can hardly be more than 20,000 souls. Around Bou-nan there are, probably, no more than 2,000 souls. These estimates result in the following total:

	Persons
San-chuan	8,000
Day-tong-gol	10,000
Tun-syan	10,000
U-yan-bu	20,000
Bou-nan	2,000
TOTAL	50,000

The author himself had travelled only through San-chuan and Bou-nan and had an opportunity to estimate these parts of the Shirongolian population from personal observations. Utilizing his personal knowledge regarding the average size of households, average numbers of households per village, and possible population densities, he proceeded to make an inference on possible numbers of population in neighbouring areas which he had not visited but the approximate extent of which he knew from local reports.

Stanley, travelling through the region of Uganda in 1878, estimated the population of "Uganda proper" at 750,000, and that of the entire empire of Uganda at 2,775,000, with this remark: "But it is to be understood that it is only a rough estimate, made by a traveller who has had to compile his figures by merely taking into consideration the number of the army assembled at Nakaranga, and enumerating districts and villages along the line of his travels."[2]

The population of French Equatorial Africa was estimated at 9 million in 1911 and at 5 million in 1914. These high estimates probably rested on an assumption that the population of remote and little travelled areas was comparable in density to that of better known areas in which French administration was already well established. The count of 1921 (largely an estimate based on a compilation of local reports) resulted in a figure of only 2,850,000.[3]

Vastly divergent conjectures were made in early years regarding the population of Madagascar:

Benyowski, judging by the uninhabited forests of the Bay of Antogil, estimated the total population of the island at less than 300,000 inhabitants. Jean Laborde, who was mostly familiar with Imerina, raised this figure to 8 million. Grandidier, who had traversed the entire island, made a more exact computation by placing, after 1868, the figure at 3 million. The first enumeration in 1900, which was still rather deficient, resulted in 2,500,000.[4]

Of special interest is Czekanowski's method of estimating the population of Ruanda on the basis of the observations which he made in the course of an exploration expedition in 1907-08 — one of the earliest explorations in this part of Africa. His method is described in detail in the case of an estimate for part of Mpororo, an adjoining region where, on the occasion of a trip of 25 kilometres in length, he had counted 100 huts. Assuming that his range of observation on the trip included a strip of territory of an average width of about one kilometre, he regarded these 100 huts as representing the population in an area of 25 square kilometres. Since it had been observed that the average population per hut was about four persons, he computed a population density of sixteen persons to the square kilometre. Applying this computed density to the area of the region, estimated from a map, he obtained a population estimate for the entire region which he believed to be "fairly close to the truth".[5]

Using similar methods for estimating the population of each separate region of Ruanda, he arrived at an estimate for the total population of Ruanda, within an area of 28,900 square kilometres, of 1,710,000.[6] This estimate, though based on extremely rough methods, is surprisingly close to recent estimates, placing the population of Ruanda, within a somewhat smaller area of 24,306 square kilometres, at 1,752,000 (official estimate for 1947).

In this connexion it may be pointed out that the use of airplanes in modern times may offer an opportunity to make far more reliable estimates by using similar methods. In particular, it is most uncertain that the area under observation on Czekanowski's trips (supposedly contained within a strip one kilometre in width "in the open country", and of different width in different terrain) could be estimated with any degree of exactness. On the other hand, it is very likely that the itineraries were largely confined to more densely inhabited areas, while avoiding the more difficult terrains where population was relatively sparse. Both of these sources of error are eliminated if huts are counted from aerial photographs taken in various places irrespective of the difficulties of travel along the surface and with an unlimited field of vision.

Conjectural estimates made by these early explorers were often wide of the mark. The observations of travellers led, in many cases, to overestimates because these persons travelled along the most convenient routes, and these routes were likely to be more heavily populated or more favourable to a dense population than other parts of those countries. Thus, Czekanowski's estimate was soon revised downward when it was found that some parts of Ruanda-Urundi were not at all favourable to dense settlement and that settlement was indeed the densest along his route of travel. Similarly, when the islands of Western Samoa were visited by the French navigator, La Perouse, in 1787, judging by the high density of settlement along the coastline, he esti-

[2] Kuczynski, R. *Demographic Survey of the British Colonial Empire*, vol. II. Oxford University Press, 1949, p. 235.

[3] Bunle, H. "Notes Statistiques sur la Demographie des Colonies Françaises", *Metron*, vol. IV, no. 3-4. June 1925, pp. 605-805, ref. to p. 658.

[4] Deschamps, H. *Madagascar, Comores, Terres australes.* l'Union Française, 1951.

[5] Wissenschaftliche Ergebnisse der deutschen Zentral-Afrika-Expedition 1907-1908, Band VI, Erster Teil. Dr. Jan Czekanowski, editor, Leipzig, 1917.

[6] *Ibid.*, pp. 111-115.

mated the population at 80,000 persons. Half a century later, when a number of missionaries had become active on the islands and it had become known that there was very little settlement in their interior, the estimate was reduced to 47,000 (in 1839) and to 32,000 (in 1849).[7]

The population of Swaziland was estimated in 1890 on the basis of its estimated fighting strength, as follows: "The numbers of the Swazie nation may be roughly estimated at 63,000. The calculation is made by taking the fighting men at 9,000 strong and multiplying by 7."[8]

3. Composite basic estimates compiled by an administration

Considerable gains in accuracy can be expected if population estimates are made piecemeal, separately for each part of a country, and added together. It is then possible that many of the errors of individual estimates, some upward and others downward, will be compensated and result in a smaller error in the total. It may be of interest to recount here the history of early official estimates of the population of the Gold Coast Colony.

In his report on the Blue Book for the year 1846, Lieutenant-Governor Winniett stated:

> There has not been any census of the native population of this colony ever been attempted to be taken; from certain data, however it may be safely assumed that the aggregate number of the population of those districts which acknowledge and are amenable to the jurisdiction of this Government is not less than 275,000, scattered over a territory of about 6,000 square miles.

The Blue Book for 1849 said:

> A Census of the Population of this Settlement was attempted to be taken this year but from the Suspicious and jealous eye with which the Natives view giving any information to Government Officials especially respecting numbers caused this important measure to fail.
>
> The population is however rapidly increasing and the peaceful and continued prosperity of the Settlement, which in the absence of Statistical information, renders it impossible to form an accurate estimate of, may be Stated at fully 5 per Cent above that Assumed to have been the Population in 1846, Viz. 275,000 (as Stated in the Blue Book of that Year) . . .

In his report on this Blue Book, Acting Lieutenant-Governor Fitzpatrick wrote:

> . . . I find the population is estimated at 288,500. I have no means of corroborating or correcting this estimate, but I apprehend it can scarce be an exaggeration, as, with the exception of a few sea-side towns, the vast district extending from Assinee to Pram Pram and back to Ashantee, is all under the jurisdiction of the British authorities.

But in the following year Lieutenant-Governor Bannerman wrote to Earl Grey:

> Upon the subject of population, where no census has been taken, and especially throughout such an extensive country, it would be impossible to state anything with certainty regarding actual numbers. My own opinion is that there has been exaggeration upon this point, as the country is far from being thickly populated; although, taking into account the immense space over which our jurisdiction extends, even without exaggeration the number must be great . . . Since the last Report to your Lordship, the territory formerly under the Danish flag has been added to our rule. This has nearly doubled the amount of population claiming English protection . . .

In his report on the Blue Book for 1851 Governor Hill took account of the increased area and population.

> . . . taking into consideration the opinion of men who have resided many years in this country, and travelled much, I am led to conclude that the total number may be put down as at least amounting to 400,000 under British protection, occupying about 8,000 square miles of country.

This report, dated 26 April 1852, was the first to put the population of the Gold Coast at 400,000, a figure which, as we shall see presently, became the standard estimate of the Administration for a whole generation . . .

In the years 1852 to 1883, various estimates were also made of the population of the Gold Coast Colony on the basis of numbers of persons found subject to the poll tax. However, since there were doubts regarding the completeness of the tax census, and regarding the numbers of dependants per person subject to tax, there was much controversy regarding the reliability of those figures. A different type of conjectural population estimate for the Gold Coast Colony was made in 1883.[10] It was estimated that in the Central Districts there were 72,000 fighting men. This number was multiplied by six "as allowing for every fighting man the existence of one woman and 4 old people and young children", resulting in a total of 432,000. For the district of Wassau, it was estimated that there were 10,000 fighting men which, this time multiplied by five, gave a total of 50,000. Estimates for other tribes and districts, derived in different ways, amounted to 169,000, giving a grand total of 651,000 as the estimated population of Gold Coast Colony in 1883.

All these estimates for Gold Coast Colony, although derived by various and unreliable methods, do not appear entirely unreasonable in the light of results of the censuses of 1891 and 1901. Both of these censuses consisted in group enumerations of most of the population, augmented by supplementary estimates. They resulted in totals of 895,350 and 1,043,350 respectively.

It may be assumed that in many cases the conjectural estimates of travellers, missionaries, administrators and military persons in particular areas were taken into consideration and were either added up, or were taken as a basis for estimating the population in those areas where this kind of information was not forthcoming. An interesting attempt at systematizing local conjectures for the purpose of estimating the population of a large area was made by the administration of Southern Nigeria:[11]

The early official reports state that it is impossible to estimate the native population of Southern Nigeria. On 11 January 1904 instructions were issued for the "Collection of Intelligence respecting Districts" which dealt also with population estimates:

[7] United Nations. *The Population of Western Samoa.* Lake Success, 17 January 1948.

[8] Kuczynski, R. *Demographic Survey of the British Colonial Empire,* vol. II, p. 25.

[9] Kuczynski, R. *Demographic Survey of the British Colonial Empire,* vol. I, pp. 404-405.

[10] *Op. cit.,* pp. 410-412.

[11] *Op. cit.,* pp. 582-586.

II. In estimating the population of towns an Officer should estimate the average number of persons in a house and the average number of houses in a compound; the number of compounds is easily ascertainable and consequently an approximate estimate can be made of the population. When the population of one or more towns has thus been ascertained it is easy for an Officer to make a "preliminary estimate" of the population in a place through which he travels for the first time; the entries in the Intelligence Book should always show whether the population is based upon a "preliminary estimate" or the reverse.

Probably, on the basis of such estimates for some districts, the total population was put for 1904 at 2,000,000. But on 17 February 1905 the Acting Secretary published a detailed estimate which yielded a population of over 3 millions.

The following rough estimate of the Population of Southern Nigeria, is published for general information. The statistics have been compiled by District Officers and are estimates of the towns and villages known to them. Many parts of the Protectorate have not yet been visited, and the returns are necessarily only very approximate.

The "estimates by some District Officers" were very rough indeed. Thus, the officer for Ikot-Ekpene put the number of women (111,786) exactly at twice the number of men (55,893), and the number of children (224,141) at almost exactly twice the number of women. Some other officers likewise overstated the number of children while others evidently understated it.

This attempt was evidently a failure since, already in 1907, the population estimate for Southern Nigeria was raised to 6 million, and in 1911 to nearly 8 million. This failure was probably due not so much to local misreporting as to uncritical compilation of those reports that were available. Instead, a critical examination of available local reports might have yielded better estimates for those areas concerning which local reports were clearly faulty or deficient, and an improved estimate might have been made for the country as a whole.

The procedure attempted by the Southern Nigeria administration, if carried out scientifically, would have some resemblance to a sampling procedure. In selected localities, precise averages of sizes of households, numbers of households per compound, numbers of compounds per settlement, and numbers of settlements, could have been determined; these averages could then have been applied to larger areas in which conditions could be assumed to be similar. From these estimates in larger areas, a total for the whole country might have been derived with a certain margin of error.

4. Other types of conjectural estimates

The foregoing examples have shown that conjectural estimates can be based on considerations of population density, of the average size of households and numbers of households per village or town, or the sizes of armed forces. This list by no means exhausts the possibilities of population conjectures. The following examples show that it is possible to derive conjectural population estimates also by other means.

The population of parts of Bengal and Bihar was estimated, in 1807-1814, by Dr. Francis Buchanan, who obtained a figure not far different from that found sixty years later at the first census of India (in 1871-72).

The mode which he adopted was to ascertain the extent of cultivation, and, allowing five or six acres (according to the character of the district) to each plough, which he assumed to represent five persons of all ages, to calculate the aggregate agricultural population, whence, by consulting the most intelligent inhabitants as to the proportion which the agriculturists bore to other classes in that district, he arrived at the total number. This rough estimate was in some cases checked by ascertaining the aggregate agricultural produce, and, after abatement for exports, calculating the number of mouths for which the remainder would suffice. The result of Dr. Buchanan's survey was that . . . he reckoned the population to be 15,443,220 . . . The population of this tract by the last census was 14,926,331.[12]

A less elaborate conjecture of population size was also made in another area of India:

In the year 1813, Mr. Butterworth Bayley, at that time the Judge and Magistrate of Burdwan, endeavoured to ascertain the population of his district. By inquiries among the Native proprietors of estates and of European residents, he satisfied himself that an average of 5½ persons should be allowed for each dwelling, and that the number of houses might be taken at 262,634, which gave a population of 1,444,487. The territory as comprised in the district as then constituted appears from the recent census to contain 322,830 houses, with a population of 1,305,316 souls, or 4½ to each house.[13]

However, much simpler devices have also been used to estimate population. Thus, the population of Hong Kong used to be estimated by the amount of nightsoil which had to be disposed of by a certain contractor, the estimate being derived with the use of some suitable multiplier.[14]

In former French Indochina, "censuses" were made every five years by means of group enumeration. The heads of villages received forms which they had to fill out in detail, but it is doubtful whether they knew enough of the required detail or whether they took the trouble actually to count the persons in their villages. For Annam, a figure of 4,183,000 was returned, but this was declared to be understatement. On the grounds that the returns of Cochinchina were relatively accurate, a ratio between the population figure of Cochinchina and its salt consumption (subject to tax) was computed; on the assumption that the same ratio would apply also in Annam, the known salt consumption (subject to tax) was multiplied by this ratio, and a "corrected" figure of 4,933,000 was obtained. However, no trouble was taken to ascertain whether a fixed ratio of salt consumption (subject to tax) to population could be regarded as a valid assumption.[15]

Many other methods than those described would be applicable. Much depends on conditions in a particular area and the amount of available knowledge. It is, however, always desirable to seek confirmation of a vague conjecture by making alternative conjectures, based

[12] *Memorandum on the Census of British India of 1871-72.* London, 1875, p. 9.

[13] Loc. cit.

[14] Edge, G. *Vital Statistics and Public Health Work in the Tropics,* p. 56.

[15] Congrès International de la Population, Paris 1937. Vol. VI, *Démographie de la France d'outre-mer.* T. Smolski, "Les statistique de la population indochinoise", and Pierre Gourou, "La densité de la population dans le delta du Tonkin".

on different information, as was done by Dr. Buchanan in the example quoted above.

Often, for instance, there is fairly reliable information on the size of district capitals, though the remainder of the population of most districts may never have been counted. If there is reason to believe that, on an average, a more or less fixed percentage of each district population resides in the district capital, it is possible to use a multiplier by which the country's total population is estimated from that of the district capitals.

Example: In a certain district, which is fairly typical for the country, the population is known to be 120,000, while the population in the chief town is 15,000. The multiplier may then be assumed in the neighbourhood of eight. If the sum of the population of all district capitals is, say, 225,000, the population of the entire country may be estimated at 1,800,000. This, however, must be recognized as an extremely unreliable figure.

In a country with varied physical characteristics, population density often varies greatly; and it may not be feasible to estimate an average density. However, the density of agricultural settlement is likely to vary to a smaller extent. If the amount of cultivable, or cultivated, land is known, the use of agricultural density as a multiplier is preferable in estimating a chiefly agricultural population, or in estimating the agricultural segment of a total population. The relatively small variation in the density of agricultural settlement can be demonstrated by means of statistics of mainly agricultural countries where over-all population density varies widely. Thus, the general population density (number of inhabitants per square kilometre of total territory) of Puerto Rico in 1949 was 246, that of the Philippines in 1939 was fifty-three, that of Thailand in 1937 was twenty-eight, that of Mexico in 1940 was ten, and that of Venezuela in 1941 was four. The number of males actively engaged in agriculture per 100 hectares (i.e. one square kilometre) of arable land in these countries was as follows: Puerto Rico: fifty-nine; Philippines: thirty-six; Thailand: sixty-four; Mexico: thirty-eight; and Venezuela: thirty-nine.[16] The smaller variation in the latter ratio suggests that agricultural density provides probably a safer multiplier than over-all population density. In some cases it may be used to advantage provided the total agricultural area can be estimated within safe limits.

In a country with a nomadic population, the numbers of tribes and tribal subdivisions are usually known. Some investigation may provide a clue as to the average size of tribal subdivisions of each major tribe or in each major region. This should make it possible to estimate the nomadic population. Confirmation of such an estimate should be sought by considering land areas needed for grazing, or actually in use during certain seasons, in order to maintain certain sizes of herds, and by considering the sizes of herds required for the needs of an average family.

There are undoubtedly many other, hitherto unexplored, possibilities for making conjectural estimates. Thus, the consumption of tobacco or salt (particularly if such commodities are subject to tax, and the amount on which tax was collected is known) may furnish the basis for a plausible population estimate. Depending on local conditions, other useful indications may also be found. The making of conjectural population estimates is a subject which has not yet been fully explored and, with ingenuity, many new devices may be discovered which will make it possible to make improved conjectures.

[16] Figures computed from *Yearbook of Food and Agricultural Statistics, 1950,* vol. IV, part I. Food and Agriculture Organization of the United Nations, Washington, D. C., 1951.

5. The reliability of a conjectural estimate

As already indicated, the reliability of a ***simple*** conjectural estimate depends on the margins of error of the data (area, number of villages, size of cities, volume of salt consumption, etc.) as well as those of the multiplier (population density, average size of villages, ratio of urban to total population, per capita consumption of salt, etc.). The margin of error of the population estimate has to be determined by considering the extreme values of the two components. Expressed as a percentage, it is usually the approximate sum of the percentage errors in each component:

Example: The area of a country is estimated at 100,000 square kilometres, subject to an error of plus or minus 5 per cent. The population density is estimated at twenty persons per square kilometre, subject to an error of plus or minus 15 per cent. The total population of the country may then be estimated at no less than 95,000 times 17, i.e., 1,615,000, and no more than 105,000 times 23, i.e., 2,415,000. The figure for total population may then be expressed as approximately 2,000,000, subject to an error of plus or minus 20 per cent.

The number of villages may be estimated in a country at 6,000 (with an error of plus or minus 10 per cent), and the average size of a village at 500 inhabitants (with an error of plus or minus 15 per cent). The total population may be no less than 5,400 times 425, i.e., 2,295,000, and no more than 6,600 times 575, i.e., 3,795,000 or at an intermediate value of about 3 million, subject to an error of about 25 per cent.

In *composite* conjectural estimates (i.e., those derived by compilation of local estimates), the margin of error may be greatly reduced, on condition that errors in local estimates are not all in the same direction and may in part cancel each other. If local estimates have been made by closely similar methods or by the same person, there is some likelihood that most of them err in the same direction. In this case, the estimates may be described as *systematically biased,* and the estimate of total population, derived by addition of the local estimates, may be almost as much in error as if it had been derived directly by a simple conjecture for the entire country. If, however, local estimates have been made in various ways and by different persons, it is likely that some local populations have been overestimated and others have been underestimated, and that the error in the total is consequently reduced. In an actual situation, it may be difficult to decide to what extent the local estimates are likely to be systematically biased, since the methods used in many cases are somewhat similar, or multipliers may have been bor-

rowed from one local area and applied to another. The usual situation is probably intermediate between an assumption of systematic bias and an assumption of absence of bias.

If bias is completely absent, and the local estimates apply to populations of comparable sizes, the probable percentage error in the total, according to laws of probability, should be equal to the mean percentage error in local estimates, divided by the square root of the number of local estimates. If, on the other hand, bias is completely systematic, the percentage error in the total will be the same as the mean error in local estimates.

Example. In a country of sixteen provinces, the population of each province has been estimated by conjecture, these provincial estimates being subject, on an average, to an error of about 30 per cent.

If it is assumed that bias is totally absent, and that consequently some of the errors are likely to compensate each other in the total, the probable error in the total may be considered as only 8 per cent (i.e., 30 per cent divided by the square root of sixteen).

If, on the other hand, provincial estimates are systematically biased, the error in the total would also be about 30 per cent.

If, after realistic consideration of the possible nature of errors in provincial estimates, it is believed that estimates are partly, but not systematically, biased, the probable error in the total may be set at some intermediate value, possibly 15 or 20 per cent.

The error in a sum of unbiased estimates is, however, not reduced in the same manner, if the estimates are for component areas of greatly different population size. Thus, if a large proportion of the country's total population lives in one province, the error in the estimate for that one province has a strong effect on the error in the total, despite compensating errors in some of the smaller provinces. In such a case, the error in a total of unbiased estimates has to be computed after some of the local estimates have been grouped together to form units of more nearly equal population size.

So far, margins of error have been expressed from the mid-value of the range within which an estimate may be supposed to fall. Most conjectural estimates may, in fact, be adequately expressed as the mid-point of such a range. There are, however, instances where it is not desirable to do so. If, for instance, several independent conjectures are made, the estimate resulting from the probably most reliable method should be regarded as the best possible estimate, but the maximum estimate may deviate from this value by a greater or smaller amount than the minimum estimate. Similarly, if the margin of error of a conjectural estimate is almost as great as the estimate itself, it is more realistic to place the estimate at a value below the middle of the range and to allow for a larger possible error in the upward direction then in the downward direction.

Example. In commenting on the official estimate of the population of British Somaliland, which is a very rough figure, Kuczynski stated: "The native population of British Somaliland has been estimated in the Blue Books for many years at 344,700. There is no evidence as to how this figure was obtained, and it should, I think, be treated at best as a reasoned guess with a margin of error of + 200,000 / — 100,000."[17] It follows that the population may be estimated anywhere between the extremes of about 150,000 and 550,000.

6. Time adjustment of conjectural estimates

As has already been mentioned, it is frequent practice to retain a conjectural estimate, once made, over a long series of years without adjustment for possible changes in the population size, although retention of the figure implies that population is regarded as nearly constant. Such unadjusted estimates have no precise time reference, since they apply over a long period of time.

For purposes of international comparability, however, it is desirable to state a time reference. This may be taken as the date, or the middle of the period, as of which the observations were made, or to which the circumstantial data used in the estimate referred.

There can be little advantage in publishing the result of a new conjecture every year without appropriate revision of previously published figures. Conjectures made by different methods lead to divergent results which in no way express real changes in population size. It is, however, obvious that a conjectural estimate cannot be retained indefinitely since, after a long period, population may indeed have changed considerably, and an old estimate may no longer be the best which can be made.

If an old estimate is replaced by a new one, the result is usually an abrupt change in estimated population size from one year to the next. Such a result is unrealistic since most changes in population size occur gradually and continuously. It is, therefore, preferable to adjust a conjectural estimate, once made, year by year, by an amount which reflects the presumed change in population size.[18] In addition, it may be desirable to utilize such further information as is available. The use of refined methods of time adjustment in the case of conjectural estimates cannot be recommended, but it would appear useful at least to assume a rate of population growth which in the light of existing knowledge appears plausible.

Example. The population of a country was estimated by conjecture at 1 million in 1940. It is believed that under normal conditions the population increases at a rate of possibly 1 per

[17] Kuczynski, R. *Demographic Survey of the British Colonial Empire,* vol. II, p. 97.

[18] "Any method which attempts to estimate merely the *increase* of the population possesses an advantage over those which aim at estimating directly the *actual* population. For since the *increase* generally bears only a small ratio to the *actual* population, a large error in the computation of the former may lead to a comparatively small one in the latter. Any reasonable method of estimating the *increase* of population should, therefore, be worth trying." (*Source:* Snow, E. C. "The application of the method of multiple correlation to the estimation of post-censal populations", *Journal of the Royal Statistical Society,* May 1911, vol. LXXIV, part VI, p. 586).

cent per annum. Aside from this "normal" assumption, it is believed that the population was further augmented in 1941 by an unusual immigration of 50,000 individuals, but that in 1949 some 5 per cent of the total population perished in a severe epidemic. The conjecture may then be adjusted for each of the years 1940-50 as follows:

Year	*Estimate*	*Rounded nearest 10,000*
1940	1,000,000	1,000
1941	(1,010,000 + 50,000)	
	1,060,000	1,060
1942	1,070,600	1,070
1943	1,081,306	1,080
1944	1,092,119	1,090
1945	1,103,040	1,100
1946	1,114,070	1,110
1947	1,125,211	1,130
1948	1,136,463	1,140
1949	(1,147,828 — 57,391)	
	1,090,437	1,090
1950	1,101,341	1,100

Severe *rounding* of figures may be desirable in the estimate for the base date. This estimate may be so rough that it may not be justifiable to show more digits than the first one or two. However, to express rates of change, subsequent adjusted estimates cannot be rounded to the same extent. The degree of rounding in each case has to be determined in such a way that no significant information is lost with regard either to the size of the population or its increase.

The assumption of a rate of population increase raises a special problem in a country with no censuses or national vital statistics. Among the considerations which may be useful in estimating a rate of increase for such a country are the following:

1. There may be visible evidence of population growth or decline. The desertion of former settlements may bear witness to depopulation, whereas the rise of new settlements, clearing of new lands, or increased parcellization of land holdings may indicate growth of numbers.[19] Unless these signs of change are studied very carefully, however, they may be misleading. The abandonment of lands in one part of the country may be balanced by an extension of settlement elsewhere, or by increased density of settlement in other lands, or by movement to non-agricultural areas. Likewise, changes in the production or consumption of a given staple commodity may often reflect shifts in tastes, technology, productivity, etc., rather than population change.

2. There may be indications of the order of magnitude of the birth rate, the death rate, and the inflow or outflow of migrants, even though no records are kept. Available information relating to marriage customs, frequency of separation, taboos on sexual intercourse, abortion, the use of contraceptive devices, and observation of sizes of families may permit an inference regarding the birth rate.[20] Signs of a high or a low death rate may be found in observations relating to sanitary conditions, the prevalence of various diseases, malnutrition, etc.

3. In certain localities registration of births and deaths may be in force, or statistical studies of fertility and mortality may have been made by other means. The results of such registration or local studies, together with available comparative information regarding the relevant conditions in these areas and in the rest of the country, may be used to estimate the rate of growth of the whole population.

4. A knowledge of events such as migratory movements, internal warfare, poor harvests, floods, droughts and epidemics, may be taken into account.

5. Statistics of births, deaths, migration and population growth may be available for neighbouring countries or even for distant countries where conditions are similar to those in the country where conjectural estimates are being made. These statistics interpreted in the light of all available information as to the similarities or differences in relevant circumstances may be helpful in arriving at an estimate.

The publishing of new conjectures within brief intervals of time is not ordinarily useful unless a new conjecture can be regarded as a real improvement over an old one. It is, however, advisable, constantly to check published conjectures by making new conjectures whenever existing information makes this possible. Eventually a new conjecture should replace an old one for purposes of publication, either because it is believed that a new conjecture is based on more reliable information, or because with the lapse of time the old conjecture can no longer be regarded as representing current conditions.

7. Standards of comparability for conjectural estimates

Owing to the very crude nature of most conjectures, little can be gained by further refinement for purposes of international comparability. However, attention should be given to the following points:

1. Conjectural estimates should be accompanied by a time reference.

2. If at all possible, they should be adjusted for every current year.

3. The nature of the estimate should be indicated by cautionary remarks. Thus, it should be pointed out

[19] "A large extent of fresh ground has been brought under cultivation, and the population as well as the livestock has been increased by the arrival of a good many Basutos returning with their earnings from the Free State and Cape Colony" (Cape of Good Hope, *Blue-Book on Native Affairs 1874*, p. 35, reporting on Basutoland). Quoted in Kuczynski, R., *Demographic Survey of the British Colonial Empire*, vol. II, p. 17.

[20] "The reason for the great overcrowding to-day, to my mind, is that the last fourteen or fifteen years have seen a tremendous change in native custom as it affects birth and population. Formerly, no Kikuyu woman was allowed to conceive a second child until the first child had stopped suckling, which was usually not until after the end of the second year, so there were generally intervals of about three years between the children. That has been broken down entirely. It used to be considered unlucky, but now they have discovered that is not true, and children are being born now — according to figures from the Kabete Mission — about one every one-and-a-half years." (Kenya Land Commission, 1932, *Evidence and Memoranda*, vol. I, p. 676. Quoted in Kuczynski, R., *Demographic Survey of the British Colonial Empire*, vol. II, p. 216.

that the estimate is conjectural and whether or not it has been adjusted for population change in time. If the estimate is composite, being a compilation of local estimates made by unknown methods, this should also be indicated.

4. In rare cases, e.g., if the territory consists largely of a trading port or a military establishment, attention should be paid to the definition of the population, i.e., resident or present population, with or without inclusion of merchant seamen, armed forces, temporary travellers, etc.

8. Improvement and appraisal of conjectural estimates through full utilization of existing knowledge

No conjectural estimate can be regarded as an adequate substitute for population statistics such as are obtained from actual enumeration and registration of vital events. However, where the latter cannot be supplied it is worth while to take care that the estimates made are the best possible under the circumstances.

The first step in making a conjecture as to population size should be a survey of all available information which has any bearing on the question. Much relevant information may be found in reports made by administrators, missionaries, health officers, travellers and military personnel. These should be surveyed systematically. Pertinent information may also be found for other countries where conditions are similar.

Full use should be made of available maps. A study of a good map may be helpful in determining whether conditions are likely to be similar or variable in various parts of the country. An aerial survey can be even more useful than maps. Aerial photographs suitably assembled may give a concrete picture of the densities of settlement in various parts of the country.

The available information has to be examined by a person sufficiently acquainted with the geography and culture to judge what is typical or atypical of conditions in various parts of the country. For reasons already discussed, it is preferable to estimate the population piecemeal, area by area, rather than for the country as a unit.

There is an advantage in making several alternative estimates on the basis of different types of information whenever the opportunity to do so exists. In this way it may be possible not only to arrive at a better estimate of the most probable size of the population, but also to get a clearer indication of the range of error than could be obtained from an estimate by one method only.

In many cases, the reliability of conjectural estimates can be greatly improved by adding to the information which is available through special studies or investigations of actual conditions in the whole country or in certain areas. For example, it may be possible to ascertain more exactly the average size of villages or tribes, or the variations of population density, by small-scale field studies in certain localities, even though a comprehensive census or even a full-fledged sample enumeration is not practicable. Likewise, it may be feasible by small-scale studies to obtain an improved estimate of per capita production or consumption of certain staples where the absolute volume of such production or consumption is being used as a basis for population estimates.

The quality of a conjectural estimate depends in the last resort on the qualifications and objectivity of the person making it, his resourcefulness, his familiarity with the country, and his knowledge of factors which influence population size and growth.

III. ESTIMATES BASED ON THE RESULTS OF INCOMPLETE CENSUSES AND NON-CENSAL COUNTS

1. Types of incomplete enumerations and non-censal counts

This chapter refers to population estimates derived from two types of basic data: (*a*) the results of population censuses which failed to achieve a complete, or nearly complete, enumeration of the people; and (*b*) the results of non-censal counts, that is, counting operations pertaining to individuals, groups of individuals, or categories of the population, which are not equivalent in procedure or in the validity of results to real censuses of the population. These two types of data are considered in general to be of a similar order of reliability as bases for population estimates. Estimates based on such data are clearly more reliable than conjectures, for they are at least substantiated by some quantitative observation relating to the population itself. On the other hand, they do not have the same validity as estimates based on actual population censuses, which may be defined as nearly successful attempts at counting the whole population, person by person, within a single year.

Estimates based on the results of sample censuses of the population are not included in the category discussed in this chapter. Although such censuses do not aim at complete coverage of the population, they can, under favourable conditions, give results as accurate as those of a complete census. The problems of making current population estimates with the use of sample census figures are therefore similar to the problems encountered in putting the results of complete censuses to the same use; they are discussed in a later chapter of this Manual.

Some counting procedures have often been called "censuses", but must really be regarded as non-censal counts:

"The existence of a census is something of a prestige item indicative of a certain progress. Consequently there has been a tendency to promote rough head counts and even much cruder estimates to the categories of censuses in official publications ... Many of the so-called 'censuses' listed as such in international compendia are in fact censuses only in the loosest meaning of the word.

"In Africa, almost all of the so-called censuses of colonial dependencies are true enumerations only for the infinitesimal European minority. Figures for the natives are often based on such evidence as the reports of local chiefs or administrators, whose returns may be much influenced by their purpose to which they assume the figures to be put. Thus, in their minds, enumeration may be associated with taxation, road service, or other levies. In this circumstance, there is a very understandable inclination to return a low estimate. Conversely, when population is used as a basis for determining administrative prestige and distribution of funds, etc., there is an equally understandable inclination to exaggerate. Even when the local administrators make a sincere effort to count people under their jurisdiction, this is normally in the form of some rough approximation such as counting the number of huts and estimating the average number of inhabitants."[1]

For the present purpose, the following three major types of "non-censal counts" may be distinguished:

1. Enumeration of individuals by registration boards.

2. Enumeration of groups rather than of individuals.

3. Enumeration of individuals with specific qualifications.

There are also enumerations which are not primarily censuses of population although they incidentally furnish statistics relating to either the whole population or some parts of it. Examples are agricultural censuses, industrial censuses, housing censuses and school censuses. With respect to the determination of population size these must be regarded as counts of the second or the third type above.

In many of the countries where true population censuses have been taken, various other counting procedures are also frequently practised. The results of the latter need not be utilized for estimates of population because the results of the population censuses are, in nearly all cases, far more reliable. Non-censal counts may, however, serve temporarily as a basis for estimates in countries where important events, such as a major war, have rendered the results of previous censuses obsolete for the purpose of current estimates. This was the case in some European countries in the years immediately following the Second World War, until new censuses were taken. In cases like these, the reliability of non-censal counts is usually fairly high and can also be evaluated with the aid of corroborating evidence.

2. Incomplete or defective census enumerations

An enumeration by actual census methods is sometimes carried out successfully in only a part of a country. An incomplete census enumeration of this kind can lead to an estimate of total population by means of supplementary estimates (conjectural, or based on non-censal counts) for the remainder of the country. For example, the 1946 census of South West Africa was excuted only in the Police Zone, where 176,000 persons were enumerated; the population in the remainder of the territory was estimated in that year at 185,000.

[1] Dudley Kirk, "Problems of Collection and Comparability of International Population Statistics", in Milbank Memorial Fund, *Problems in the Collection and Comparability of International Statistics,* New York, 1949, p. 22.

An incomplete census may, incidentally, help to improve estimates for those parts of the country where the population was not enumerated. Thus, a comparison between estimates made previously and the census results in areas where the census enumeration took place can show whether the estimates made by previous methods tended to be too large or too small. If population estimates in the remainder of the country are made by the same or similar methods, there is some reason to believe that these are subject to a similar error, and they may be "corrected" accordingly. This latter assumption cannot be made, however, if the population enumerated in the census is not similar to the non-enumerated population. For example, if a census taken only in urban areas shows in what manner previous estimates for these urban areas tended to err, it cannot be assumed that the estimates previously made by the same methods in rural areas erred in the same direction or to the same extent.

A census enumeration may also have been undertaken in the entire country but may have failed as a result of poor organization, faulty procedures, or obstacles encountered in the process. For instance, this is believed to have been the case with the census of Haiti in 1918-19, which did not yield complete returns for the whole country. In such a case, it is not necessary to reject entirely the results of the census, even though they are defective, because by careful examination of the completeness of enumeration in each area and detailed comparisons of local census results with other estimates for the same localities, an improved population estimate can be constructed for the whole country.

Observation and local reports may provide important clues for identifying those areas where enumeration was most nearly complete; in those areas the census results may be regarded as the best population estimates available for the census date and should properly form the basis for subsequent estimates of local population. In some of these areas, a comparison of census results with the results of previous methods of estimating the population may furnish an indication of the nature and extent of the error arising from other estimating procedures; population estimates, made by such procedures, can be "corrected" accordingly in those areas where the census enumeration was unsuccessful. Some areas may also be found where fairly reliable estimates can be made by other procedures, but where the census results were very incomplete. For these areas a coefficient of incompleteness of the census enumeration can be estimated, and this coefficient may be applied to other areas where census results were probably incomplete to a similar extent but where no adequate estimates can be formed by alternative methods.

3. Enumeration of individuals by registration boards

Attempts at enumerating every individual in the population are sometimes made by non-censal methods for specific purposes. Some examples of this type of enumeration are afforded by the establishment of a ration register in French Morocco prior to the count in 1947, and the development of the central register of identity cards in Iran in the course of several years up to 1941. Such "civil registers" are usually kept up-to-date, so far as possible, by adding the reported births and subtracting the reported deaths. However, they must not be confused with continuous population registers, since internal checks for accuracy are insufficient and, in particular, since there are no adequate provisions for the transfer of records in cases of change of residence. Nor is the establishment of such a register equivalent to the taking of a census, because of the great difference in procedure.

The Moroccan registers were originally established for purposes of food rationing, a purpose which is conducive to over-registration, since there is a tendency not to report deaths or departures. In 1947, the Moroccan population was enumerated from these rationing lists, and the result was a total of 8,293,000 (excluding French and foreigners). However, on the basis of sampling investigations, it was believed that this figure was an overestimate of at least 4 per cent, and it was consequently replaced by an estimate of 7,900,000 Moroccans.[2]

After the partition of Palestine, the population of territories surrounding the new State of Israel was swelled by large numbers of Arab refugees. The United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWAPRNE) attempted to enumerate the refugees who were on relief, in May 1950, and arrived at a total of 957,000, including many double registrations. After repeated revisions of the relief rolls, the total number was found to be about 876,000 in June 1951,[3] a figure which was believed still to be slightly excessive.

The establishment of a civil register of Iran took many years. It was originally intended to serve the purposes of military conscription, but no accurate registration could be obtained. Later, it became a central register of identity cards issued, cards being added for each birth and withdrawn for each death. By 1940, the registration of live persons was believed sufficiently complete to serve as a basis for estimates of the country's total population. By the end of 1950, the population total, according to this register, amounted to 18,952,000 persons. It was, however, believed at that time that, if an accurate count of the population could be taken for the whole country, it would result in a total of about 20 million.[4]

4. Group enumeration

Group enumerations of the population are often carried out for purposes other than that of determining

[2] *Annuaire Statistique de l'Union française Outre-Mer,* 1939-46, Chapitre B — Territoire et Population. Paris, 1948, p. 83.

[3] United Nations. *Report of the Director of the United Nations Relief and Works Agency for Palestine Refugees in the Near East, Supplement No. 6* (A/1905). Paris, 1951, p. 3.

[4] Iran, Ministère de l'intérieur. *Bulletin de statistique,* No. 4. Ordibéhechte. 1330. (21 avril — 21 mai 1951) pp. 2-3.

the population size. For example, in a housing census, the primary object of the enumeration is to establish the numbers of dwelling units of different types and characteristics; since one of the important characteristics of a dwelling is the number of persons living in it, a housing census leads incidentally to an enumeration of the great majority of the population, on a household basis. Group enumerations of this type are frequent in countries where there are also true population censuses; in those countries, the results of the group enumerations ordinarily need not be employed for the purpose of estimating total population, particularly because the results of the population censuses are nearly always more reliable.

In some countries, however, group enumerations are made specifically for the purpose of determining the size of the population. This method is used especially in countries where there are serious obstacles to individual enumeration. In many parts of Africa it has not been found possible to take a real census of the population because a sufficiently large staff of literate, co-operative and well-instructed enumerators to make a person-by-person enumeration could not be rapidly assembled, or because funds for hiring and training such staff were insufficient. A small number of enumerators, therefore, have to cover large territories within a brief space of time. Generally this is done by asking village headmen or chiefs of tribal divisions and subdivisions to report on the numbers of persons under their jurisdiction.

Theoretically, group enumerations may give results which are no less accurate than those obtained by individual enumeration. In practice, this is seldom the case because errors are introduced by lack of complete knowledge of the group on the part of the individual who reports it, or by the inattention or carelessness of that individual or his unwillingness to give full and accurate reports. The danger of error is especially large when the contact between the local populace and the central administration of the country is a rather loose one. A certain amount of diplomacy is often required to elicit satisfactory answers from tribal or village headmen. There may be frequent misunderstandings, both as regards the purpose of the enumeration and the precise kind of answers which are required. Local dignitaries may sometimes also lack the necessary authority to determine the numbers of persons in their respective territories. In some instances they may not take the trouble to ascertain exact numbers, while in other instances they may be motivated either to exaggerate or to minimize the number of people. In some areas there is a reluctance to declare the numbers of women, while in others, small children are not considered as "persons" that should be counted. There may also be varying practices regarding the inclusion or exclusion of persons who have originated in another village or tribal group, or of persons who have left the particular group but are still loosely associated with it. A great drawback of group enumerations is that it is usually impossible to say whether they have resulted in an under-count or an over-count.

Thus, a "census" by group enumeration was taken in Nyasaland in 1945. On this occasion, it was frankly admitted in the official publication that neither the procedure nor the accuracy of the results was comparable to that of a real census. "As on previous occasions, it was not possible to deal with individual Africans separately. Each village was treated as a separate entity and a return entered for that village."[5]

". . . the tables of statistics should be only treated as a useful and, in the aggregate, a fairly accurate estimate of the African population, based on a count".[6]

Some group enumerations have been performed in such cursory fashion that they differed little in practice from estimates derived from tax lists, described in the following section. Thus, in the 1911 "census" of Sierra Leone, "It was recognized that to take such a detailed Census of the aboriginal natives [as of the non-natives] of the Protectorate would be quite impracticable . . . and accordingly it was decided that the Census should be limited to obtaining such information as could by approximation be furnished by the District Commissioners."[7] However, it turned out that such information was derived largely from tax lists, and that population figures for most districts were returned with varying arbitrary proportions of the population reported as men, women and children (20 per cent men, 30 per cent women, 50 per cent children, in some districts; 30 per cent men, 40 per cent women, and 30 per cent children in other districts; various other percentages were also used).

5. Enumerations of individuals with specific qualifications

Examples of this type of enumeration are tax lists, where only taxable persons (e.g., able-bodied males, heads of households) or taxable units (huts, tents, households) are enumerated; voting lists, where the numbers of persons eligible to vote are registered; conscription lists, recording the numbers of persons liable for military service; school censuses, in which the number of children of pre-school and school age are counted; and agricultural censuses with incidental enumeration of the agricultural population.

Whereas the types of counts previously discussed can lead to estimates of total population size either directly or by the application of correction factors for under-enumeration or double reporting, enumerations which are restricted to special categories of the population can only lead to a figure of total population with the application of a multiplier. Estimates based on this type of count are sometimes not much better than conjectures, since they are subject to a double error, one pertaining to the accuracy of the count itself, the other to the selection of a representative multiplier. The accuracy of such estimates is usually much more

[5] *Nyasaland Census Report*, 1945, pp. 1-2.
[6] *Ibid.*, p. 11.
[7] Kuczynski, R. *Demographic Survey of the British Colonial Empire*, vol. I, p. 33.

affected by the selection of a suitable multiplier than by the accuracy of the count itself. In most cases, however, the basis of estimating is likely to be improved by the fact that such counts are usually repeated periodically, leading to improved procedures and increased experience. In some cases, special efforts are made to arrive at a representative multiplier.

Enumerations of the population of the Cameroons were made under German administration. It was reported about a particular district:

> The number of men has been ascertained through direct enumeration, but it will be, of course, too low, since there are chiefs who will, either intentionally or unintentionally, make wrong statements. The number of women and children has been calculated by means of a table which, to judge from sample tests, may be considered as more or less accurate.
>
> This table provided that the number of women and the number of children were to be calculated by multiplying the number of men for four tribes by 1.25 and by 1.5 respectively, and for the other three tribes by 1.5 and 2 respectively.[8]

A special problem in enumerations of this type arises where there is difficulty in defining sharply the category of the population which is to be counted. Thus, in the territory of Tanganyika, population estimates were formerly based on tax counts, but "the details of the system have varied. The principal features of the system have been as follows: A hut tax has to be paid for every dwelling owned by an African. Where an African has more than one wife living with him in a hut, a tax is levied in respect of each additional wife. A poll tax has to be paid by every adult male not liable to hut tax. An adult male was formerly defined as a male who appears to be over the age of 16. Since 1934 this age limit has been 18 years".[9] In cases like these, where the definition of the enumerated category is somewhat complex or necessarily vague (since for many Africans the exact age is not known), it is particularly difficult to find multipliers which can express adequately the ratio between the numbers counted and the total size of the population. Moreover, some persons may escape registration while others are exempt from tax payment. Therefore, "changes in numbers of taxpayers may reflect changes in population, or they may reflect changing fiscal legislation, changing numbers of exemptions, or differences in the procedures developed for allocating migrant men to specific areas".[10]

In the years 1943-47, the total population of Tanganyika was derived from tax reports by multiplying the number of taxable males by 3.5. Similar methods were also used in earlier years. Whether as a result of inaccurate counting, or because of an important error in the multiplier, the estimates were far too low. Whereas the estimate for 1947 was 5,838,000, the census of the following year, which is believed to have been the first fairly accurate enumeration of the population, resulted in a total of no less than 7,478,000. It appears that population estimates of Tanganyika derived by tax counts fell short of the true population size by about 20 per cent.

In Kenya, the numbers of native huts were repeatedly enumerated for purposes of a hut tax, but much difficulty was experienced in selecting suitable multipliers. In 1913-14, population figures were estimated by assuming an average of three persons per hut, and entering a round figure which approximated to the results obtained. In 1924, instructions were issued which stipulated:

> In Districts where Officers are not satisfied as to the accuracy of their census of native children, it is recommended that their figures be compiled upon the principle of taking the adult population as 63% and the child population as 37% of the general total.[11]

The number of adult females was first determined by assuming a ratio of forty-seven adult males to fifty-three adult females, and the number of children was then added according to instructions stated above. Neither of these ratios had any statistical foundation whatever, and it is probably largely owing to the use of improper ratios that the population estimate for Kenya, which stood at 4,200,000 in 1947, had to be raised on the basis of the census of 1948 to 5,382,000 in the following year. Population estimates of Kenya must have been some 20 per cent short of the true population size.

In 1932, a special inquiry was conducted in the Digo District on the Kenya coast, where a population of 26,000 persons was enumerated with distinction of age and sex. It was found that the ratio of adult males to adult females was, in fact, roughly forty-seven to fifty-three (which may be a mere coincidence), but that of the total population 52 per cent — not 37 per cent — were children.[12] The Digo District may not have been at all representative of Kenya as a whole. Nevertheless, if the findings in this inquiry had been utilized, the population estimate for Kenya in 1947, would have amounted to some 5.5 million, a figure not greatly different from the census result. Although this may be only a coincidence, it is very probable that if the assumed multipliers had been based on some actual observation, like that made in the Digo District, population estimates would have been much more accurate than they were.

6. Determining improved multipliers

It appears from the foregoing discussion that the accuracy of estimates based on enumerations of specific categories in the population, such as estimates based on tax lists, depends very largely on the accuracy of the multiplier. It is therefore surprising that, despite the expense and labour involved in counting the households or special categories in the population, efforts to establish a reasonable multiplier have often been quite insufficient. True, for purposes of taxation, a knowledge of actual population size is not necessary so long as

[8] Kuczynski. *The Cameroons and Togoland*, p. 21.
[9] United Nations. *The Population of Tanganyika* (ST/SOA/Series A, No. 2), Lake Success, 1948, p. 71.
[10] *Ibid*, p. 87.
[11] *Quarterly Bulletin of Statistical Research for British East Africa*, vol. I, part I, p. 21 (quoted in Kuczynski, R., *Demographic Survey of the British Colonial Empire*, vol. II, p. 136).
[12] Kuczynski, R. *Demographic Survey of the British Colonial Empire*, vol. II, p. 155.

the number of taxable units is known with a fair amount of accuracy, but knowledge of the size and composition of the population may be regarded as an important by-product of such non-censal counts, where the obstacles to real census-taking seem insuperable.

A means of determining the ratios which should be used in converting, say, numbers of taxable males into total population estimates is to actually enumerate the population, with distinction of the special categories, in selected localities. The ratio found in such an enumeration, which may be confined to a relatively small part of the population, may then be assumed to be approximately true for the rest of the population, on condition that the enumerated localities can be regarded as typical.

Reference has already been made to the enumeration in the Digo District in Kenya, which might have been utilized to establish multipliers better than the ratios actually used. It is, however, doubtful whether the Digo District could have been regarded as typical of the country as a whole, and some variations in this ratio should have been allowed for to take account of varying conditions in different parts of the country. An example of such practice can be found in the case of Togoland under German administration, where, in 1912-13, some estimates were made in the following fashion:

> In Lomé-Land the vaccination of the population in the entire District, started by the vaccination physician, offered an opportunity for counting the huts and the natives in a large part of the District. This showed that on an average 4.4 inhabitants have to be reckoned for one person subject to taxation. By applying this ratio also to that part of the District in which an enumeration could not yet be effected, a population of 136,400 inhabitants was calculated for the whole District, a figure which exceeds the former estimates — 119,000 — not inconsiderably. In the Anecho District the population was ascertained on the basis of the tax rolls by multiplying the number of persons taxed by a coefficient which according to the local conditions varies between 4 and 4.5. The coefficient itself was ascertained through enumeration in various localities. The population figure computed by means of this procedure showed 123,237 native inhabitants while the population formerly had been estimated at 112,000.[13]

However, such methods of obtaining improved multipliers are not a sufficient safeguard against errors. In the first place, there can be no certainty that conditions in the enumerated localities can be regarded as representative of those in the districts or territories to which these multipliers are applied. Secondly, the accuracy and completeness of the actual counting of taxable individuals or households remain in doubt, even if the multiplier is accurate. Finally, there are probably considerable variations in the precision with which taxable subjects are defined in various localities, as was shown in the case of Tanganyika (see p. 21).

An adequate procedure would require an enumeration, independent of the non-censal count of the total population in suitably selected localities, and the determination of the ratio between the inexact count of taxable subjects and the relatively exact enumeration of total population. This ratio may, of course, differ somewhat from the true multiplier (i.e., from the true ratio if taxable subjects were counted accurately), since it also contains an element of correction (i.e., a factor of adjustment for inaccuracy in the count of taxable subjects).

An example of this method is furnished by the practice in the Belgian Congo. "Sample centres, as far as possible typical of a large area, are chosen where counts are made each year; by 1935, studies of this kind covered one-sixteenth of the estimated total population. From these detailed studies, a coefficient is calculated for each region to four decimals, and the estimated population for the whole region is obtained by multiplying the number of tax-payers by this figure."[14]

7. Time adjustment of estimates derived from non-censal counts

Some non-censal counts, particularly tax counts, are performed at regular intervals such as once a year, and therefore result in current estimates without need for time adjustment. An annual series of such figures has important defects. There may be year-to-year fluctuations in the figures which, though reflecting in part actual changes in population size, are probably mainly the result of the varying errors to which annual figures are subject. An imprudent user of statistics, in glancing at a series of such annual estimates, is apt to conclude that population has increased or decreased markedly during particular years, although such a conclusion is entirely unfounded in view of the varying errors of each separate estimate. A comparison of figures separated by a long time interval can be likewise misleading since the general accuracy of the estimates could have improved with the progress of time.

It is therefore necessary, where figures are secured annually, to give sufficient explanation of the manner in which these figures were derived, whether they can be strictly compared from one year to another, and also whether it is believed that estimates have become progressively more accurate. Without such indications, the comparison of consecutive current estimates can be misleading, even though each current estimate represents the best estimate which was possible at the particular time. On the whole, it would seem that an annual repetition of non-censal counts is likely to lead to greater and greater accuracy, with increased experience and avoidance of past mistakes on the part of the enumerating officials and greater responsiveness on the part of the population; but experience has shown that this is not necessarily true.

Example. The difficulty of deriving a plausible rate of population increase from a series of population estimates of varying accuracy can be illustrated with the annual estimates for the African population in the indigenous sultanates of Ruanda-Urundi. In that territory, non-censal counts of able-bodied males used to be taken annually and estimates of total population were derived by means of special investigations designed to establish the ratios between enumerated persons and total

[13] Kuczynski, R. *The Cameroons and Togoland.* Oxford University Press, London, New York, Toronto. Issued under the auspices of the Royal Institute of International Affairs, 1939, p. 375.

[14] Hailey. *An African Survey,* p. 121.

population. The following results were obtained (numbers in thousands):

1932	3,451	1941	3,832
1933	3,244	1942	3,840
1934	3,165	1943	3,803
1935	3,340	1944	3,577
		1945	3,386
1936	3,449		
1937	3,603	1946	3,495
1938	3,725	1947	3,662
1939	3,766	1948	3,761
1940	3,798	1949	3,807
		1950	3,863

The figures show a rapid decline for 1932-34, then an increase, at first rapid, then gradual, up to 1942, a sharp decline to 1945, a rapid increase to 1948 and a more gradual increase to 1950. It is believed that in 1932-34 population may in fact have declined (though not as drastically as suggested by the figures) as a result of famine conditions; part of the decline in the figures may, however, be due to the disorganization in the collection of statistics occurring under these conditions; similarly, part of the rapid increase in subsequent years may be a result of statistical reorganization. In the years from 1941 to 1946, budgetary restrictions impeded the collection of accurate statistics, but in the years 1946 to 1948, more accurate statistical procedures were re-established.[15]

The figures yield the following geometric rates of increase for the periods before 1950:

1948-1950	1.3% per annum
1945-1950	2.7% per annum
1941-1950	0.1% per annum
1934-1950	1.3% per annum
1932-1950	0.6% per annum

As in the case of mere conjectures, it is not advisable to hold an estimate derived from a non-censal count at some time in the past at a constant figure, unless there is reason to believe that the population has in fact remained nearly stationary. Unless non-censal counts are repeated annually, it is desirable to apply adjustments for population change up to the current date. The methods of time adjustment which may be used include assumed rates of increase, extrapolation and the use of vital statistics.

The use of assumed rates of increase has been discussed in the preceding chapter (chapter II, p. 16). Even if the assumptions relating to population growth are considerably in error, this error can form only a small part of the total error in the current population estimate because the base figures are not highly reliable. The error in estimated growth can contribute substantially to the total error in current estimates only if many years have passed since the last non-censal count (for an illustration, see the example in chapter I, p. 4).

Estimates by extrapolation, in the case of non-censal counts, can only be made with greatest caution. In particular, full consideration must be given to the fact that successive counts are usually not comparable with regard to completeness or accuracy. In many cases, it is probable that recent counts are more reliable than those taken in the more remote past. An extrapolation from counts of varying degrees of accuracy is apt to carry forward some of the efforts contained in the less accurate counts. However, provided the accuracy of successive counts is somewhat comparable, extrapolation may be better than assumed rates.

Example: In a non-censal count, taken twenty years ago, the population was found to be 1,000,000. A second count, taken ten years later, resulted in a population figure of 1,200,000. For the current year, an estimate by means of arithmetic extrapolation would be 1,400,000.

If the first of the two counts was incomplete, and the population was underestimated by some 100,000, the method of extrapolation used would imply an exaggerated rate of population growth, and the current estimate should be lowered accordingly, to about 1,300,000.

If, on the other hand, the first count was relatively accurate but the second count omitted about 100,000, the rate of growth implied in the extrapolation is understated, and the current estimate should be raised to about 1,600,000.

Extrapolation of totals based on non-censal counts should never be carried over a very long period. This fact was appreciated with regard to population estimates in Northern Rhodesia:

"The native population of the Territory is estimated on the basis of the ratio of increase (or decrease) over the last five years . . . This system of estimating the population is an innovation, and cannot be used indefinitely, as it cannot reasonably be applied to those districts that attract immigrants from adjoining Territories." (Northern Rhodesia. *Report upon Native Affairs, 1928,* p. 5).[16]

Methods of extrapolation will be discussed more fully in chapter V.

There are only a few countries using non-censal counts where extensive vital statistics are available. Where this is the case, the accuracy of statistics on births and deaths is probably doubtful. The use of statistics of births and deaths appropriately corrected is, however, greatly preferable to an assumed rate of increase, particularly since the statistics reflect, even if inaccurately, the real year-to-year fluctuations in rates of growth. Where non-censal counts exist in the form of statistics obtained by registration boards (other than "continuous population registers", discussed in chapter VII), there is reason to assume that registration of births and deaths is subject to the same biases as the original population count. If the purpose of registration is rationing or the allocation of benefits, it is probable that births are fully registered but that many deaths are not reported. If the registers serve purposes of military conscription or the imposition of other levies, there may be a tendency not to report births. Therefore, wherever vital statistics are used in adjusting estimates derived from non-censal counts, their probable accuracy should be carefully examined. It may be found preferable, for instance, to substitute for the recorded data some estimated numbers of births and deaths based on assumptions with regard to the probable completeness of registration.

8. Comparability and consistency of estimates based on non-censal counts

Sometimes non-censal counts cannot be made simultaneously for an entire country, but may extend over a

[15] Figures are taken from United Nations, *The Population of Ruanda-Urundi* (in press 1952).

[16] Kuczynski, R. *Demographic Survey of the British Colonial Empire,* vol. II, p. 402.

certain period of time. This means some loss of accuracy in the total results, particularly if numbers in various localities are affected by seasonal movements. It is therefore desirable to indicate in the publication of results the entire time-period during which the count was taken. In some cases, a count may be taken for one section of the population or for one part of the country at one time, and for the remainder at another time. In such cases, it is desirable to adjust the results of one sectional count so that it conforms to the date of the other and explain this adjustment.

The observations regarding inclusion or exclusion of specific categories of the population which were made in chapter I apply in the case of non-censal counts as well as census enumerations. The standard definition (modified *de facto* population) should be compared carefully with the categories included in the count, and deviations from the standard should be noted by indications, or further estimates should be made to allow for adjustments necessary in order to make the estimate conform with the standard definition. Counts based on voters' lists or tax lists are particularly likely to omit temporary residents or the dependants of migrant workers, and care should be taken to determine whether the estimate includes these or not. Similarly, refugees, displaced persons or prisoners of war in camps may not figure in a count of ration-holders, voters or taxpayers, but they should be included in a country's total *de facto* population.

If an estimate, based on non-censal counting procedures, has been derived from a combination of various counts pertaining to various sections of the population and differing considerably in accuracy, figures for each section should be shown separately, since the figures resulting from estimates of higher accuracy may be amenable to certain statistical uses from which the figure for the total population must be excluded owing to its low reliability. If an estimate is based on a non-censal count covering the majority of the population, but estimates of special categories (e.g., displaced persons in camps) from other sources are added to the total, this fact and the relevant figures should also be stated.

9. Improvement of estimates based on non-censal counts by full utilization of existing knowledge

In a number of countries, non-censal counting procedures may be in operation though they have never been utilized for population estimates. There should, therefore, be a survey of all the statistical information available in the country. It is quite possible that certain counting procedures will be found well suited for the purpose of estimating the population.

In some cases, population estimates are based on only one type of non-censal count, although the results of other counts are also available. The possibility that the latter will yield population estimates of the same or greater reliability should be fully explored. A comparison of estimates, each derived independently and on a different basis, may be very helpful in confirming the result of any particular method or in determining and narrowing its possible margins of error.

Full advantage should be taken of every indication bearing on the completeness and exactitude of the non-censal counts. Reports of enumerators should include their opinion as to whether figures collected locally are fairly exact or whether they are likely to be exaggerated or understated, and to what extent. If such reports exist, they should be fully utilized in formulating a judgment concerning the reliability of the total figure and the possible direction and magnitude of its bias.

If the enumeration is of the kind which results in a total of certain units of population (e.g., households, families), or certain of its categories (e.g., voters, heads of households, taxable subjects, children of school age), it is important to utilize fully all available information which may be helpful in determining the multiplier. Estimates of sizes of households, of the proportion of persons in certain ages, of the average size of a family, etc., may already have been made, if not for the given country, perhaps for other countries where conditions are presumably comparable. The utilization of such knowledge may help greatly in improving the accuracy of multipliers. Of particular value are investigations carried out on a representative sample basis. Sampling investigations may be suitably combined with the counting procedure itself. However, if it is intended further to "correct" the results of counts for errors in counting, sampling investigations must be pursued independently of the count, by separate enumerations in selected localities.

There may also be unutilized information regarding the accuracy of the counts. The sifting of such evidence is a prerequisite to the use of extrapolation of the results of successive counts, since it is essential that the accuracy of the extrapolated figures be at least somewhat comparable.

In some cases, where the reliability of the results of a count is very low, a conjectural estimate may be made independently of the count for comparison.

The utilization of the results of a partial or defective census has been discussed in section 2 of this chapter. Similar methods may be used where censuses or superior counts have been taken only in small parts of the country or in particular localities. A comparison of the results of the latter with the results, for the same localities of the country-wide count, can furnish the basis for evaluating the accuracy of the count and for applying some "corrections", if needed.

A patient search may yield various pieces of information relevant to rates of population increase. The great variety of background knowledge which is helpful in making reasonable assumptions for population growth has been indicated in section 6 of chapter II. Perhaps the best adjustments for population change, where vital registration is not in effect, can be made by estimating vital rates on the basis of sampling investigations. A brief outline of the uses of sampling procedures for such purposes will be found in appendix B of this Manual.

IV. ESTIMATES BASED ON ONE CENSUS ONLY

1. Importance of a census for accurate population estimates

A census is defined for the present purpose as an individual enumeration of the whole population, or at least a great majority of the population, during a period of not more than one year. Though censuses vary greatly in accuracy, as a rule they give much more reliable measures of population size than non-censal counts. The main reason for this is that the determination of population size is one of the primary objectives of a census; the census procedure is designed to realize this objective as accurately as possible.

The first census taken in a country is usually not as accurate as subsequent censuses can be, since successful census-taking under the conditions prevailing in any country is in part a matter of local experience. It is highly important for purposes of population estimates to evaluate the accuracy of a census enumeration and to "correct" the census result accordingly if it is found that there have been omissions or double counting. The accuracy of a first census is, however, more difficult to appraise than that of subsequent enumerations, which provide the means of a comparison of results for two or more dates.

The accuracy of population estimates based on a census is, of course, greatly affected by the length of the time interval which has elapsed since that census was taken. We shall, therefore, give separate consideration to estimates based on a recent census and to those based on a census taken in a more distant past, assuming in both cases that this was the country's only census.

2. Estimates based on one recent census

For the date of the census, the census result can be regarded as the best possible estimate of population size, provided there is no reason to believe that it is above or below the mark. If there is evidence of over-enumeration or under-enumeration, the census figure should be "corrected" and it should be stated, in the publication giving the subsequent estimates of population, that this correction has been made.

Time adjustment for population changes since the census date[1] can be made in a number of ways, some of which are discussed elsewhere in this Manual. The use of birth, death and migration statistics for this purpose is considered in chapter VI,[2] and the assumption of a plausible rate of increase, where no better method can be found, is taken up in chapter II.[3] The following paragraphs refer to two methods which are particularly likely to be useful in a country where a census has been taken for the first time: (*a*) the calculation of rates of increase from results of non-censal counts, and (*b*) the estimation of rates of increase from data obtained in the census itself.

If non-censal counts were used to estimate the population before the census was taken, these may permit an estimate of the increase after the census date.

Example. In the census of Tanganyika as of 2 August 1948, a population of 7,477,677 was returned. There was no reason to suppose that this figure was too high or too low. Comparison of the census total with previous estimates based largely on tax lists and group enumerations showed that the latter were too low. The estimate for 1947 was 5,838,000 and that for 1928 was 4,741,000. On the assumption that the extent of under-estimation was the same in 1947 as in 1928, the average annual increase during the interval was calculated at 1.1 per cent. One method of estimating the mid-year 1949 population (though not the one used in making the official estimate) is to add eleven-twelfths of 1.1 per cent to the 1948 census total, which gives an estimate of 7,550,000. (The fraction, eleven-twelfths, is used because the 1948 census date was one month after the middle of the year.)

The derivation of an assumed rate of increase from non-censal counts rests on an assumption that the non-censal counts for various years have erred in the same direction and to the same extent. This assumption may not be tenable in some cases. Non-censal counts performed after the taking of the census may differ in accuracy from those performed before the census, particularly if the staff employed for the non-censal counts take advantage of the information and experience acquired at the census. If counts of unequal accuracy are compared, the derivation of an estimated rate of increase is still possible by arbitrarily selecting a value somewhere between extreme values resulting from the comparison of several counts.

It is usually not safe to derive a rate of increase directly from a comparison of a non-censal count with the census result, as the latter is probably more accurate and not subject to the same kinds of error.

Given some knowledge of the mortality rates, the rate of population increase may sometimes be estimated,

[1] A time adjustment is needed even in making an estimate for the year of the census, unless the census was taken near the middle of the year. The census result should be increased or diminished by the estimated amount of population change during the fraction of the year between the census date and 1 July, in order to ensure comparability with other mid-year estimates.

[2] Comprehensive data on these subjects are likely not to be available in countries where the first census has recently been taken.

[3] See chapter II, section 6.

at least very roughly, from the results of the census itself, classified by age groups. A possible way of estimating the birth rate is to use the number of infants aged under one year who were enumerated at the census, with suitable allowance for infant mortality, to arrive at the number of births in the preceding year from which these infants have survived. The method, of course, requires some information about infant mortality. Moreover, few censuses succeed in an accurate enumeration of infants. Considerable proportions of infants under one year of age are often omitted from the enumeration, while some of those enumerated may be reported as more than one year old. A comparison of the numbers of infants reported as under one year of age with the numbers reported at ages of one, two, three, or four years may provide some basis for correcting the figure for the age under one year, and thus lead to an improved estimate of the birth rate. By similar methods, estimates of births over a period of several years may be made from the returns on children of various ages.

Example. Mortara, utilizing the results of the Brazilian census of 1920, estimated the birth rate of Brazil during the years 1910-20 on the assumption of certain coefficients of survival and inaccuracies in age statements.

That ages were reported inaccurately could be seen from the fact that more children were reported at ages of 2 or 3 years than at ages of one or under one year. However, despite this inaccuracy, Mortara applied survival coefficients to these numbers to arrive at a first approximation, as follows:

Years of age	*Persons enumerated*	*Coefficient of survival*	*Number of births (first estimate)*	*Year preceding census*
1	828,000	0.883	934,000	1
2	776,000	0.780	996,000	2
3	1,015,000	0.736	1,378,000	3
4	1,006,000	0.717	1,404,000	4
5	950,000	0.706	1,345,000	5
6-10	4,537,000	....	6,576,000	6-10
TOTAL	9,112,000	....	12,636,000	

Assuming that a total of about 12,500,000 births had indeed occurred during the ten years preceding the census, but allowing for the fact that, in an increasing population, the annual number of births increases under a fairly constant birth rate, he recomputed the series with corrected estimates. According to these second estimates, the number of births in the year preceding the census was about 1,363,000, and the estimated birth rate was 47.2 per 1,000.[4]

Census data on population by age groups can be used also to estimate the death rate, but in this case the results of at least two censuses are ordinarily needed for accurate estimation.

Information regarding fertility and mortality can also be obtained from special questions asked in a census enumeration, either of the whole population or a sample. On the occasion of the censuses of Tanganyika, Kenya and Uganda, in 1948, questions relating to the numbers of births and deaths which had occurred in the preceding year were asked among 10 per cent of the enumerated population. The resulting information made it possible to estimate rates of births, deaths and natural increase prevailing in these territories. These estimated rates are being used in making estimates of current population size. Fertility and mortality rates can be estimated also by means of sampling investigations independent of the census. This is being done at present in a selected area of India, in the course of a field study being conducted jointly by the United Nations and the Indian Government. The experience obtained in that study is expected to be helpful as a guide for similar investigations in other countries where reliable vital statistics are lacking.

3. Estimates based on a census taken in the more distant past

The usefulness of a census result for current population estimates diminishes with the passage of time, for two reasons. On the one hand, cumulative errors due to inaccurate assumptions with respect to population growth tend to remove current population estimates more and more from reality the more time has elapsed since the last census. On the other hand, in most countries where there has been a census at some time in the past, other statistical series have also been established and their accuracy has been improved. Eventually reliance on other statistical information leads to better estimates than reference to an old census.

Example. A census has been taken, the result of which is believed to have been in error by no more than 3 per cent. The error in annual estimates of population increase may be, on the average, 0.5 per cent of the population. The cumulative effect of the latter error can lead, after twenty years, to a total error in estimates amounting to 3 per cent plus 20 times 0.5, i.e., to 13 per cent of the population.

Supposing also that non-censal counts, which are taken regularly, can be used to make population estimates subject to an error of no more than 10 per cent. It appears that, after twenty years, it is better to rely on the non-censal counts for current estimates than to continue basing them on the census result.

This does not mean that the census loses all its usefulness, because it can still be utilized in an appraisal and improvement of current statistics. A census, even though taken in the remote past, may also be very helpful in the design of samples for current investigations of the population.[5]

In some countries, no national census has been taken for a long time, but there have been provincial censuses from time to time. A comparison of the results of such provincial censuses with those of the previous national census may be used for estimating the increase in the population of the country as a whole since the date of the national census, with appropriate allowance for differences in the rates of increase in various parts of the country.

In Argentina, for instance, no national census was taken during the interval from 1914 to 1947. However, censuses were taken in the cities of Buenos Aires (1936) and Santa Fé (1923), in the provinces of

[4] United Nations. *Methods of using census statistics for the calculation of life tables and other demographic measures* (with application to the population of Brazil). By Giorgio Mortara. Lake Success, November 1949, pp. 14-16.

[5] See appendix B.

Buenos Aires (1938) and Mendoza (1942), in Chaco Territory (1934), Pampa Territory (1935), and National Territories (1920). Censuses of children of school age were taken in Santa Fé province (1912, 1918, 1925, and 1937) and in Rosario city (1934).[6] The results of these local censuses, together with the current registration statistics of births, deaths and migration, probably permitted fairly adequate provincial population estimates to be maintained during the long period between the two national population censuses.

There may be reason to believe that in the period since the last census the completeness of birth and death registration has improved. If so, current estimates may eventually have to be reconstructed: earlier estimates in post-censal years, when birth and death registration was presumably less accurate, may have to be revised before an improved series of post-censal estimates can be carried up to the current date.

4. Evaluation and improvement of the quality of estimates based on one census only

Where the results of one census are available, the most obvious method of improving current population estimates is to taken another census. Owing to the fact that the taking of a census involves important considerations of planning and expense, frequent census taking is not always practicable. But after a long time the results of a census lose most of their usefulness especially where the registration of births and deaths and the records of migration are not very accurate, or are lacking entirely. Under those conditions, after twenty years or so, the state of information about the population is little better than if no census had ever been taken. If it is intended to maintain a standard of statistical information above that of countries in which there has been no census, the census must be repeated.

As an alternative, though not an entirely adequate substitute, a sample census may be taken. Even after a considerable number of years has passed, the results of a previous census can provide a framework for a highly efficient sample census, an opportunity which should always be kept in mind.

The rapidity with which the results of a previous census lose their usefulness depends on the accuracy of information with regard to rates of population change. If only vague assumptions regarding rates of increase can be made, errors of estimates accumulate very rapidly. If, on the other hand, tolerably accurate vital statistics are at hand, a longer interval of time may pass before time adjustments fall greatly out of line with actual population changes. Hence a great interest attaches to the availability of vital statistics, by means of which the value of the census results can be preserved over a longer period of time. Sampling for estimates of birth and death rates can be very helpful where it is not found practicable to establish a well-functioning registration system within any short time. Sampling, however, is at best a poor substitute for a complete system of vital registers.

In the absence of statistics of births and deaths, the considerations which should govern the assumption of rates of increase, and the kind of information which may be helpful in making such assumptions, have been discussed in preceding chapters and should also apply where a census has been taken. However, as was shown in the present chapter, special methods for the derivation of possible rates of increase can also be used where a census result is available. A combination of all possible information is likely to lead to more realistic assumptions than the reliance on one method only.

Information on the accuracy of a country's first census can best be obtained in quality checks by means of samples;[7] the census itself also provides various internal evidence of its possible degree of accuracy; discussion of this topic is, however, reserved for a later manual.

The appraisal of the quality of non-censal statistics is often possible in the light of the census, particularly if non-censal counts were taken near the date of the census. Such knowledge is of importance if the census is rather old and if more reliance must be placed on non-censal data for purposes of current estimates.

For all estimates based on a census, care should be taken that the definition of the population (*de jure, de facto,* etc.) corresponds if possible to international standards. If necessary, census figures may be adjusted by means of estimated numbers of persons in certain categories, in order to conform to the standard definition of the total population. The definition of the population, if deviating from international standards, should be indicated.

Some censuses, although conducted in the greater part of the country, or for a majority of its population, do not include special parts of the population (e.g., tribal Indians, displaced persons, merchant seamen aboard vessels, etc.). To obtain estimates of total population, special estimates must be made for these categories and added to the census total, with indication that this was done. If the numbers involved in the adjustment are large, it is desirable that they be also shown separately with statement of the method by which they were estimated.

[6] United States Bureau of the Census. *Argentina. Summary of Biostatistics.* Washington, February 1945, pp. 7-9.

[7] See appendix B.

V. ESTIMATES DERIVED BY EXTRAPOLATION OF CENSUS RESULTS

A. METHODS REQUIRING TWO CENSUSES ONLY

1. The utilization of the results of two censuses

Estimates based on the results of two or more censuses are more reliable, as a rule, than those based on one census only. As already pointed out, the census data are likely to improve in quality as more experience in census taking is gained, and the availability of a series of census results makes possible a better appraisal of the accuracy of enumeration than is possible with a single census. A comparison of the results of two or more censuses provides, furthermore, a means of estimating the increase of population since the last census.

Extrapolation — that is, the use of the assumption that the mode of increase observed between two dates in the past has continued since the time of the last enumeration — often gives better results than any other method which can be applied in a country where current data on population change are lacking or of poor quality. Extrapolation, however, must not be applied mechanically wherever the results of two or more censuses are available. There may be good reasons to believe that the rate of increase since the last census has differed considerably from the rate during the inter-censal interval, either because of unusual events in one or both of these periods, or because of a general tendency towards acceleration or deceleration in these rates. In some of these cases, a modification of the extrapolation is still preferable to other less well-founded assumptions, though less reliable than time adjustments by means of adequate statistics of births, deaths and migration. Extrapolation depends, strictly speaking, on an assumption that the population change in recent times has not deviated from past trends and that it proceeds, year by year, according to a constant progression. Such precise conformity of biological phenomena to a mathematical formula never occurs in reality. Even though deviations from previously established trends are not always large, new factors of change arise constantly.

In some countries, existing statistics of births, deaths and migration are too inadequate to reveal the true changes in population size. Inadequate migration statistics are almost the rule, but so long as migatory movements are not very large this defect in the statistics is not of major importance. However, if births, deaths or large migratory movements are inadequately recorded, extrapolation may lead to better current estimates than the use of such defective data.

If the results of two censuses only are available, there is a choice between two simple methods of extrapolation: the use of an arithmetic, or a geometric rate of increase. The choice of method should depend on a knowledge of the given situation. For special purposes, more elaborate methods can also be used.[1] The greater variety of methods which can be used if three or more censuses have been taken are discussed in part B of this chapter. The selection of the most suitable method of extrapolation in each case requires good judgment, and no general rule can be stated as to the method which will give the most reliable results in any particular instance.

2. Extrapolation by means of arithmetic rates of increase

The simplest method of extrapolation is to compute the average annual number by which the population has increased from one census to the next, and to add an equal number for every year which has elapsed since the last census.

Example. In Switzerland, 156,715 persons were enumerated in 1936, and 185,215 in 1946. Current estimates are made in the following manner:

The increase observed in the census interval of ten years was 28,500 persons, giving an average annual increase of 2,850. This figure is added to the 1946 census total for each year which has elapsed since that date. The estimate for 1950 was, therefore, 185,215 plus four times 2,850, i.e., 196,615 or, as a rounded figure, 197,000.

It should be noted that a constant numerical increase in a growing population really implies a diminishing rate of growth relative to population size. The use of an arithmetic rate is often justified for some of the following reasons:

1. It is easy to compute an arithmetic rate. If it is doubtful that very reliable estimates can be obtained by any more complicated method of extrapolation, there can be no objection to using the simplest one.

2. An arithmetic rate is less likely to give absurd results than a constant geometric rate of increase if the rate is high and the period of time is long. A population of 10 million increasing at a geometric rate of 2 per cent per annum would attain 79,900,000 within 100 years and 638,470,000 within 200 years. Except in an initially very sparsely settled country, it would appear

[1] For a description of several methods and bibliographic reference pertaining to many other methods of extrapolation, the reader is referred to R. H. Wolfenden, *Population Statistics and their Compilation* (Actuarial Society of America, New York, 1925).

inconceivable that population growth could continue at such a rate for a very long time. An arithmetic rate in such a case gives much more conservative results. Thus, an initial population of 10 million, increasing by 200,000 per annum, would attain 30 million after 100 years and 50 million after 200 years.

3. There are grounds for supposing in some instances that the rate of population growth tends to slacken as time passes. For example, if the birth rate is believed to be declining and the death rate to be nearly constant or declining more slowly, a gradually decelerating rate of growth such as is implied by a constant numerical addition to the population is more reasonable as an assumption than a constant geometric rate. Likewise, if immigration is the source of a major part of the increase and if there is no reason to suppose that the annual number of immigrants increases as the population grows larger, an arithmetic rate may be more reasonable than a geometric one.

4. There is often good reason to assume that enumeration in the earlier census was less accurate than it was in the later one. More often than not a first census results in a considerable under-enumeration which is partially eliminated in the second census. In extrapolating the apparent increase between a first and a second census it may be preferable to be conservative and to use an arithmetic rather than a geometric rate.

5. An advantage of the use of arithmetic rates, which will be discussed again in section 4, arises from the fact that the sum of arithmetically extrapolated sub-totals is always equal to the arithmetically extrapolated total, whereas extrapolation of sub-totals and totals by any other method leads to results which are more or less inconsistent and require further adjustment.

3. Extrapolation by means of geometric rates of increase

Geometric extrapolation corresponds to the assumption that a population increases constantly by numbers proportionate to its changing size. In this case, the computation is carried out in the same manner as a computation of compound interest. A population increasing by a geometric rate obeys the formula

$$Pn = Po\,(1 + r)^{t} \quad \ldots\ldots\ldots\ldots (1)$$

where Po is the population at the beginning of a period,

t is the period of time, in years,
r is the annual rate of increase, and
Pn is the population at the end of the period.

If the rate of increase is to be determined from a comparison of census results, it can be found by the formula

$$(1+r) = \sqrt[t_1]{\frac{P_2}{P_1}} \quad \ldots\ldots\ldots\ldots (2)$$

where P_1 and P_2 are the populations according to the first and the second census respectively, and t_1 is the inter-censal time interval, in years.

Thus, for a geometric extrapolation, from the date of the last census onward, of the rate of increase indicated by a comparison of the last census figure with an earlier census figure, the formula can be written:

$$Pn = P_2\left(\sqrt[t_1]{\frac{P_1}{P_2}}\right)^{t_2} \quad \ldots\ldots\ldots\ldots (3)$$

where t_2 is the post-censal time interval (up to current date).

The application of this formula requires calculating aids (logarithms, a slide-rule, or a compound-interest table). Actually, with a geometric rate of increase, the logarithms of the population increase at an arithmetic rate.

Example. The population of Spain amounted to 23,563,867 according to the census of 31 December 1930, and to 25,877,971 according to the census of 31 December 1940. For the years following 1940, current estimates were made by means of a geometric extrapolation of these two census results.

Using formula (2), the rate of increase can be determined as follows:

$$(1 + r) = \sqrt[10]{\frac{25,877,971}{23,563,867}} = 1.0094\,;\ r = 0.0094.$$

Substituting this rate of increase in formula (1), the following estimate can be obtained for midyear 1950, i.e., after a lapse of 9½ years:

$$P_{1950} = 25,877,971\ (1.0094)^{9.5} = 28,286,533,$$

which may be rounded to 28,287,000. (This estimate may have to be revised in view of the fact that the preliminary result of the census taken on 31 December 1950 totalled 27,861,000).

The same result can also be obtained more directly by formula (3):

$$P_{1950} = 25,877,971\left(\sqrt[10]{\frac{25,877,971}{23,563,867}}\right)^{9.5} = 28,286,533.$$

Geometric extrapolation always leads to a higher estimate of population size than arithmetic extrapolation whether a population is increasing or declining. Thus, in the above example, if arithmetical extrapolation had been used it would have been found that the population of Spain had increased in the ten years from the end of 1930 to the end of 1940 by 2,314,104, i.e., by 231,410 on an annual average; assuming this arithmetic annual increase to have continued for another 9½ years, up to mid-year 1950, the estimate for the latter date would have amounted to 28,076,366 only as compared with 28,286,533 resulting from geometric extrapolation.

One drawback of geometric extrapolation is that the sum of extrapolated sub-totals is not equal to the extrapolated total.

Example. Let it be assumed that separate geometric extrapolations are made for the urban, the rural and the total population of Canada, up to 1950, using the results of the national census of 1931 and 1941.

In 1931, the urban population was found to be 5,572,058, the rural 4,804,728, and the total 10,376,786. In 1941, the urban population was 6,252,416, the rural 5,254,239, and the total 11,506,655.

Geometric extrapolations up to 1950 would bring the urban population to 6,935,000, the rural population to 5,695,000,

and the total population to 12,628,000. The sum of the extrapolated urban and rural populations, however, amounts to 12,630,000, which is slightly more than the extrapolated total population.

While the assumption of a constant geometric rate of increase may be plausible in many circumstances, it cannot be relevant in every situation. Such a rate of increase would occur in a population not affected by migration if the birth and death rates were constant or if both birth and death rates were rising or falling by equal amounts each year. In a population being increased by immigration or depleted by emigration, the rate of growth would be constant only if the net rate of immigration or emigration as well as the birth and death rates were constant in proportion to the population, or if changes in the birth and death rates were exactly counterbalanced by changes in migration.

No country's population can be expected to grow for an indefinite period of time at a constant rate for eventually it would become so large that further increases would hardly be possible. Many countries have experienced in recent times a drop in death rates followed, a generation or so afterwards by a drop in birth rates; as a result, population increased rapidly for a time, but at a slackening pace subsequently; a strictly parallel movement of birth and death rates has hardly been observed over any long time period. It is true that during the period of rapid population growth many countries of Europe sent large numbers of migrants to other continents; this did, however, not entail any large migrations into Europe during the subsequent period of slackening growth.

It is, therefore, desirable to confine geometric extrapolation to relatively short time intervals. For short periods it may be plausible to assume that a given population increases approximately according to a geometric rate and that one of the above-stated combinations of conditions is satisfied to some degree of approximation. If several decades have elapsed since the taking of the last census geometric extrapolation becomes increasingly unreliable and may lead to a cumulative exaggeration of the estimated population, especially if the projected rate of increase is high. It may then be preferable to use arithmetic extrapolation which gives more conservative results. Geometric extrapolation is hazardous also if the accuracy of the basic census figures is subject to considerable doubt; the accumulation of increases which are apparent (only because of varying completeness of enumeration) may lead to greater and greater error in the extrapolation as the years pass.

In some populations growth is for a time accelerated because the death rate is declining while the birth rate remains nearly constant. In such a situation even the geometric rate of increase will underestimate the growth of population and an arithmetic rate is definitely inappropriate. In this event the use of a harmonic rate of increase may be suggested (see section 4, below).

Geometric rates are preferable to arithmetic rates for the extrapolation of decreases in population over a series of years, for an arithmetic extrapolation of a negative rate of increase would eventually result in a negative population estimate, which is obviously absurd. Geometric extrapolation, on the other hand, would show a diminishing numerical decline and would always result in positive population estimates no matter how long the time period. In a decreasing population geometric extrapolation is the more conservative of the two methods. For the same reason geometric extrapolation is probably preferable in countries with a large emigration.

4. Modified methods of extrapolation

As already pointed out, the computation of a geometric rate of increase cannot be carried out without a logarithmic table or other calculating aids. If it is desired to confine the computation to an arithmetical operation, but to obtain a result which approximates that of a geometric rate, the following method can be used.

Instead of keeping the annual increase at a constant figure, one may express the average annual numerical increase between the two census dates as a percentage of the arithmetic average of the totals from the two censuses. This procedure yields a simple approximation to the geometric rate for the purpose of arithmetical computations. This rate can then be applied, year by year, to estimates of a changing population.

Example. Using the data for Spain (see example on p. 29 of this chapter), extrapolation, by arithmetical approximation to a geometric rate of increase, can be performed as follows:

The arithmetic mean of the census results for the end of 1930 and the end of 1940 amounts to 24,720,919. Since average annual increase in this period amounts to 231,410, a rate of 231,410 in 24,720,919, i.e., of 0.936 per cent, can be computed, this being an approximation of the geometric rate. Applying this rate to each successive figure, we obtain the following series:

Date	*Period (years)*	*Population*	*Increase*
31 December 1940.....	½	25,877,971	121,117
1 July 1941.....	1	25,999,088	243,351
1 July 1942.....	1	26,242,439	245,629
1 July 1943.....	1	26,488,068	247,928
1 July 1944.....	1	26,735,996	250,249
1 July 1945.....	1	26,986,245	253,591
1 July 1946.....	1	27,238,836	254,956
1 July 1947.....	1	27,493,792	257,342
1 July 1948.....	1	27,751,134	259,751
1 July 1949.....	1	28,010,885	262,182
1 July 1950.....		28,273,067	

The result, 28,273,000, is a close approximation to the result of the actual geometric extrapolation, which was 28,287,000, and may be compared with 28,076,000, the result of arithmetic extrapolation.

This method of extrapolation always results in figures somewhat higher than those obtained by arithmetic extrapolation and differing slightly from those obtained by geometric extrapolation. Like geometric extrapolation, it has the disadvantage that sub-totals do not add up to totals extrapolated by the same method. This method is to be recommended only where it is desired to compute, by approximation, a geometric rate of increase without the use of computing aids.

In some situations the two alternatives of arithmetic and geometric extrapolation may seem to offer too narrow a choice. This is particularly so in the case, already referred to, where population increase is temporarily accelerated. If one wishes to assume an accelerating rate of increase relative to population size, the extrapolation can be carried out with respect to the reciprocals of the numbers involved. A rate of growth accelerating in this manner may be described as a harmonic rate.

Example. Employing again the data for Spain (see example on p. 29 of this chapter), harmonic extrapolation can be carried out as follows:

The reciprocal of the 1930 census figure multiplied by 100 million is 4.243786; and 100 million times the reciprocal of the 1940 census figure is 3.864291. The reciprocal value, therefore, has declined by 0.379495 in ten years, or by 0.03795 per annum. To arrive at an estimate for mid-year 1950, this decline must be extrapolated for 9½ years beyond the value for the end of 1940, and a reciprocal value of 3.864291 minus 0.36052, i.e., 3.503771, is obtained, which corresponds to a figure of 28,540,700 as the estimate for 1950, on the assumption of an accelerated rate of increase. This figure is markedly higher than the result of geometric extrapolation.

If there is reason to believe that the rate of population growth is either accelerating or slowing down, a reasonable assumption as to the rate of acceleration or deceleration can also be made and a formula for modified extrapolation with a changing arithmetic or geometric rate can be derived.

Various methods have been devised for reconciling the sum of geometrically extrapolated sub-totals with a geometrically extrapolated total. In most of these methods the geometric extrapolation for the total is retained and adjustments are applied to sub-totals to make them agree with the total. One method consists in extrapolating sub-totals by means of an arithmetic progression and then multiplying each value by the ratio of the total population derived from geometric extrapolation to the total derived from arithmetic extrapolation. This method has the drawback that the sub-total for an area which had a stationary population in the inter-censal interval would show a decline in the post-censal period if the total population were increasing. Another method which was developed by A. C. Waters consists in an arithmetic extrapolation of the ratio of the sub-total to the total. The disadvantage of this method is that the sub-total for an area with stationary inter-censal population would show a post-censal increase if the total population were increasing. Other methods have also been developed but they require more complex mathematical operations.

B. METHODS REQUIRING THREE OR MORE PREVIOUS CENSUSES

5. Extrapolation by means of parabolas

The question as to whether the rate of population growth tends to be constant or to accelerate or slow down cannot be answered with much assurance by comparing successive census totals unless the results of at least three censuses are available. But even the results of three censuses are insufficient for determining whether an observed acceleration or deceleration of increase appears to be temporary or a long-term trend, whether it is fairly recent and may still gather momentum, or whether it has apparently run its course. To answer such questions with any assurance a minimum of four census totals is required.

It is, of course, not necessary to employ the results of more than two censuses in an extrapolation even though more are available. If the results of a series of censuses suggest that population growth has proceeded at a fairly even pace, and if there is reason to suppose that it may continue to do so, arithmetic or geometric extrapolation is quite sufficient for most purposes, at least if the post-censal period is not too long. Similarly, if one or several of the census results are of doubtful accuracy, those results may be selected for purposes of extrapolation which are believed to be most nearly comparable. Extrapolation by methods more refined than arithmetic and geometric rates is hardly justified if a basic figure is unreliable.

The most widely used method of extrapolation which employs the results of three or more enumerations is extrapolation by parabolas of the second or the third degree. A parabola of the second degree can be computed from the results of three censuses; this type of curve is sensitive not only to average rates of growth but also to the observed acceleration or deceleration in these rates. A parabola of the third degree, which can be computed from the results of four censuses can take account not merely of the acceleration or deceleration in rates of growth, but also of a changing momentum in the acceleration or deceleration.

If the results of three censuses are plotted on a chart where the horizontal axis measures intervals of time and the vertical axis the population enumerated at each of the three dates, these points can be connected by a parabolic curve of the second degree which, if continued beyond the date of the last census, indicates the population which may be estimated for each subsequent date. This parabolic curve is defined by the formula

$$Px = a + bx + cx^2,$$

where x is the time interval, in years, measured from any fixed date, such as perhaps one of the census dates.

Px is the population size to be expected at a time of x years after the fixed date,

any fixed date, such as perhaps one of the census dates.

solving the above formula for each of the three census dates.

Example. Let it be assumed that, according to a census of 1936, the population was 40,000; according to a census of 1942, it was 44,000, and according to a census of 1946, it was 47,000. (For the sake of simplicity, exact calendar years and round population figures are used in this example.) How is the population in 1950 to be estimated, using parabolic extrapolation?

Taking the year 1936 as the fixed date, $x = 6$ for 1942, 10 for 1946, and 14 for 1950. The equations can be set up as follows:

(1) for 1936: $40{,}000 = a + 0.b + 0.c$ ..... $(x = 0)$
(2) for 1942: $44{,}000 = a + 6b + 36c$ ..... $(x = 6)$
(3) for 1946: $47{,}000 = a + 10b + 100c$ ..... $(x = 10)$

The solution of these simultaneous equations gives the following values:

$$a = 40{,}000$$
$$b = 616.7$$
$$c = 8.33$$

The formula, $P_x = a + bx + cx^2$ may now be expressed as:

$$P_x = 40{,}000 + 616.7x + 8.33x^2.$$

Applying this formula for 1950 ($x = 14$), we obtain:

$$P_{14} = 50{,}267.$$

Computations are simpler if the time intervals are equal. In that event it is convenient to take the date of the second of the three censuses as the fixed date.

Extrapolation of this type has been employed for current population estimates of the Dominican Republic in the years 1920-49. Current estimates of Peru in the years after 1940, on the other hand, were made by the use of a parabola of the third degree, based on the censuses of 1836, 1850, 1876 and 1940.

A parabolic curve of the third degree is defined by the formula

$$Px = a + bx + cx^2 + dx^3.$$

The computation of these constants requires four separate equations, which must be solved for the four census dates.

Example. Censuses were taken in a country in 1910, 1920, 1930, and 1940, resulting in population totals of 250,000, 300,000, 320,000, and 360,000, respectively. How is the 1950 population to be estimated, using extrapolation by a parabola of the third degree?

(1) $250{,}000 = a + 0b + 0c + 0d = a$
(2) $300{,}000 = a + 10b + 100c + 1{,}000d$
(3) $320{,}000 = a + 20b + 400c + 8{,}000d$
(4) $360{,}000 = a + 30b + 900c + 27{,}000d$

Solving these four simultaneous equations, we obtain the following values:

$$a = 250{,}000$$
$$b = 8{,}166.7$$
$$c = -400.$$
$$d = 8.333$$

To obtain the estimate for 1950, we have to substitute a, b, c, and d, for $x = 40$, in the general formula of the third-degree parabola.

$$P_{40} = a + 40b + 1{,}600c + 64{,}000d, \text{ i.e., } 470{,}000.$$

(If a second-degree parabola had been used in conjunction with the census figures of 1920, 1930 and 1940, the population estimate for 1950 would have amounted to 420,000. The difference in the results is due to the fact that, in the present example, the second-degree parabola would have expressed some acceleration of population growth, whereas the third-degree parabola also expresses an increasing momentum of acceleration.)

The selection of a second-degree or third-degree parabola for extrapolation presupposes, of course, that the form of growth expressed by such a curve appears plausible in view of what is known about the trends of births, deaths and migration. It may seem advisable to use a parabola of the second degree if, for instance, the birth rate is believed to be rising while the death rate falls, so that the rate of population growth is accelerating; or if there is reason to suppose that the rate of growth is slackening because of a falling birth rate and a more nearly constant death rate. A parabola of the third degree can take into account changes in the speed at which the gap between birth and death rates is widening or narrowing; for instance, when death rates fall rapidly at first, but less rapidly as time progresses. An increasing or decreasing volume of migration can also lend support to the assumption of a parabola.

Compared to arithmetic and geometric rates, parabolas have much more flexibility. It would, however, be fallacious to expect that population growth will continue to conform to some parabolic trend, merely because it has done so in the past. Too much refinement may also lead to an illusion of increased accuracy in the estimates although estimates must always be less reliable than the basic data.

A very important defect of parabolas, if they are extrapolated over a considerable length of time, is that they eventually move more and more rapidly either in an upward or a downward direction. This tendency is more pronounced in third-degree curves than in those of the second degree, owing to their greater flexibility. Although, for a time, population growth may indeed be accelerating, there is an upper limit to the rate of growth which can ultimately be attained. There is also a lower limit to population size, which can never become negative, whereas some parabolas may eventually fall below zero.

In most situations, a parabola of the second degree should be preferred to one of the third degree. The additional refinement of the latter is usually not warranted by the available data, nor can it be very realistic to expect population growth to conform to a function of such complexity. In the particular case, however, the use of a third-degree parabola is sometimes suggested when it appears that it leads to an extrapolation with less exaggeration than a second-degree parabola.

6. Parabolic extrapolation of transformed data

The tendency of parabolas after several years of projection to show a steeper and steeper rise or fall has already been mentioned. In many cases, this defect can be modified by applying parabolic extrapolation to the logarithms of the figures instead of to the figures themselves. The extrapolation of logarithms implies a projection of changing rates of growth rather than of changing absolute figures.

Example. Using the data in example on pp. 31-32, a second-degree parabola can be computed on logarithms as follows:

Date	*Number*	*Log.*	*Equation*
1936	40,000	4.60206	log. $P_0 = a + 0b + 0c = 4.60206$
1942	44,000	4.64345	log. $P_6 = a + 6b + 36c = 4.64345$
1946	47,000	4.67210	log. $P_{10} = a + 10b + 100c = 4.67210$

Solving these equations, we obtain:

$$a = 4.60206$$
$$b = .006742$$
$$c = .000026$$

To obtain the logarithm for the 1950 estimate, these values must be substituted in the equation.

$$\log. P_{14} = a + 14b + 196c$$

The logarithm for the 1950 estimate, therefore, is 4.701544, and the 1950 estimate is 50,297. This result of logarithmic extrapolation compares with 50,267, resulting from direct extrapolation.

Another transformation of data which can sometimes be used with advantage consists in using reciprocals. As has already been mentioned, extrapolation of reciprocals implies an assumption of accelerating growth, which is realistic in some instances. Parabolic extrapolation of reciprocals may be pertinent where estimates derived by direct parabolic extrapolation tend to be too conservative or result in rapidly decreasing figures.

7. Extrapolation by means of growth curves

For purposes of estimating the population at future dates or at dates in a distant past when no adequate information was given, highly elaborate methods of extrapolation are often employed. Most important among these is the logistic curve of growth, first discovered by P. F. Verhulst and later developed by R. Pearl and L. I. Reed.[2]

The logistic curve shows a long-term growth cycle, usually extending over a century or longer, which contains the following phases: (1) a gradual transition from almost stationary conditions to a noticeable increase in the population, (2) an acceleration in the rate of increase until the rate approaches a maximum, (3) a slowing down of the rate of increase, and (4) a gradual transition towards almost stationary conditions. It has the advantage that it can fit any set of data showing at least one of these phases of growth.

It has been demonstrated that population growth over periods of a hundred years or more in many countries has approximated fairly closely a logistic curve of growth. It has also been shown that logistic curves possess a certain predictive value and that future estimates made by means of logistic extrapolations have in many cases been approximately confirmed by actual observation as censuses were taken subsequently.

Studies of these kinds have led some scientists to the belief that the logistic curve expresses some universal law of population growth. Experiments were made with cultures of yeast-cells or with a colony of fruit-flies in a bottle and it was found that biological organisms, multiplying within a limited or closed environment, showed changing phases of growth closely resembling those of the logistic. It does not follow, however, that the same applies to human populations since human beings, as distinguished from flies in a bottle, have both the capacity to change their environment and to control their rates of reproduction.

The Gompertz curve has also sometimes been used as a curve of population growth.[3] Its main difference from the logistic curve consists in the fact that it expresses a more rapid acceleration in the early phases of growth and a more gradual slowing down in the late phases.

Growth curves are used only very seldom for purposes of current estimates. A logistic curve was computed by the Banco Nacional of Cuba,[4] to estimate the population of Cuba for the years 1937-49 because estimates made at various times during that period by the Cuban census office, using defective statistics of births and deaths, were found inconsistent.

Although a simple logistic or a Gompertz curve can be fitted to the results of three censuses, and a modified logistic curve to the results of four censuses — provided these show a somewhat regular procession of growth — this is rarely done. Growth curves are usually computed from a long series of census figures in such a manner that they do not coincide exactly with any one of the census results, but approximate each and all of them. The reason for this practice is that individual census results are affected not only by the long-term trend in population growth, but also by temporary departures from this trend, whereas growth curves are intended to portray the long-term trend only. This makes it inadvisable to employ growth curves in an extrapolation of the results of three or four censuses only. Any particular census result may deviate more or less from the over-all trend and an extrapolation of such temporary deviations from the trend would lead to unrealistic results and defeat the purposes for which these elaborate curves are intended. Nearly all countries which have a long series of census results also possess adequate systems of birth and death registration, making extrapolation unnecessary for the purpose of current estimates.

C. GENERAL OBSERVATIONS PERTAINING TO CURRENT ESTIMATES DERIVED BY EXTRAPOLATION

8. Elements of reliability in estimates derived by extrapolation

The following are the main factors affecting the reliability or reasonableness of extrapolation:

1. The accuracy of the census data;
2. The relative accuracy, or comparability, of the data from the different censuses;
3. The length of inter-censal intervals and of the post-censal period; and
4. The relevancy of the method of extrapolation selected.

[2] Hagood, M. J. *Statistics for Sociologists.* Henry Holt & Company, New York, 1951, pp. 272-282.

[3] Croxton, F. E., and Cowden, D. J. *Applied General Statistics.* New York, Prentice-Hall, Inc., 1950, p. 448.

[4] Cuba. *Banco Nacional de Cuba.* Memoria 1949-1950, La Habana, Cuba, p. 155.

The accuracy of the census results is important because a current estimate, made by extrapolation, cannot be more accurate than the base data. If the probable direction and possible extent of errors in the census figures (due to over- or under-enumeration) are known, the census data should be "corrected" before they are used in the extrapolation, as this will reduce the margin of error in the resulting estimates.

Differences in the accuracy of various census results can affect extrapolation even to a greater extent than constant errors running through a series of census enumerations. For example, if in two successive censuses there was under-enumeration of about the same amount, a comparison of the results leads to an estimated rate of population increase which is fairly realistic. Although the results of an extrapolation of these census totals are likely to be underestimates, this component of error is not greatly increased by extrapolation. However, if two successive censuses have differed with regard to the extent of under-enumeration or over-enumeration, or if they have erred in opposite directions, a comparison of the totals can lead to a computed rate of increase which is much too high or too low. Extrapolation of such an erroneous rate of increase will probably lead to even greater error in the resulting estimates. This is also true if one of two censuses was fairly accurate, but the other was not. Thus, a rate which is extrapolated from an earlier census which was notably incomplete and a later census which was more accurate is likely to lead to overestimates of the current population; this error increases every year. The possible errors in the successive censuses should therefore be given careful consideration, and if there is reason to believe that they have varied in accuracy, the totals should be "corrected" before extrapolation.

Consideration should be given also to the comparability of the census results with regard to the definition of the population, the inclusion or exclusion of special categories, and the territory covered. It would clearly be wrong to extrapolate census totals without adjustment if one census was taken within a larger or a smaller territory than the other. Similar errors may result if the censuses used in extrapolation differ in coverage of particular categories of the population.

The reliability of an extrapolation is affected by the length of inter-censal intervals. A long interval can be favourable for two reasons: year-to-year fluctuations in population increase tend to be compensated over a longer period and true trends are reflected with more weight in a long period than incidental differences in completeness of census enumerations. A short interval, on the other hand, has the advantage that it refers to a less remote past and may therefore give a somewhat better basis for estimating trends in the most recent years, and that census results obtained within a relatively short succession are likely to be more nearly comparable.

In some instances it may be desirable to select two or three censuses from a series of censuses in such a manner that extrapolation can be made on the basis of inter-censal intervals of a desirable length. Where population trends are apt to change rapidly, more reliable extrapolations can be obtained if they can be based on rather short intervals. Longer intervals should be selected, if possible, where population trends are fairly constant or subject to gradual changes only.

The reliability of an extrapolation depends directly on, and is in inverse proportion to, the length of the post-censal period, i.e., the period between the last census and the date for which an estimate is desired. As the post-censal period becomes longer, errors may accumulate more and more rapidly, because recent trends deviate more and more from extrapolated trends; this is of necessity true in the case of parabolic extrapolation. In particular, it is undesirable to continue extrapolating over a post-censal period which is as long as, or even longer than, the inter-censal intervals. If a long time has elapsed since the last census was taken, it may be desirable either to base extrapolation on longer inter-censal intervals (provided a sufficiently long series of previous census results is available), or to abandon the method of extrapolation and to make current estimates by other methods. In particular, parabolic extrapolation should not be carried over a long period (in view of the tendency of parabolas to rise or fall more and more steeply after a certain time has passed), unless it is based on very long inter-censal intervals. Thus, with the passage of time, parabolic extrapolation may have to be replaced, first, by extrapolation of arithmetic or geometric rates and, eventually, by different types of estimates, such as those based on the results of more recent non-censal counts, or even conjectures.

A consideration of all the factors discussed in this section (accuracy and comparability of census results, length of inter-censal and post-censal periods) as well as any knowledge regarding probable recent developments affecting the countries' population must influence the choice of the method of extrapolation. A method can be regarded as appropriate if it leads to estimated post-censal increases in the population which do not appear in any way unrealistic. The selection of a method of extrapolation depends, therefore, largely on sound judgment and a knowledge of the situation. If extrapolation is made by an unrealistic method, errors in estimates are likely to increase more and more rapidly in the course of years.

9. Improvement of estimates derived by extrapolation, and standards of comparability

Using the same methods of extrapolation, estimates can be improved by either selecting the most useful censuses if a sufficient number of census results are available, or by "correcting" census results where this seems necessary. Censuses may be so selected that time intervals which are too short are avoided and census

figures that are not comparable are omitted. If the last census interval was unusual with respect to rates of increase (e.g., owing to temporary disturbance of normal trends, or owing to a war), an arithmetic or geometric rate of increase observed in some previous census interval may be applied to the total figure obtained at the last census.

Fresh consideration should continuously be given to the selection of the method of extrapolation, particularly if a long time has passed since the last census. In particular, parabolic extrapolation may have to be abandoned after several years and some simpler method substituted. Eventually, any method of extrapolation must be abandoned and estimates have to be formed on some new basis rather than by relying on remote census results.

If it is believed that recent developments have caused the growth of population to depart from previous trends, all knowledge regarding the possible influence of such developments on birth and death rates or on the volume of migration must be taken into account in selecting the most suitable method of extrapolation or in modifying the results of an extrapolation. If unusual events have occurred in a particular year only (e.g., a famine resulting in much excess mortality, the arrival of large numbers of refugees, etc.) the increase or decrease of population will have to be estimated separately for that year by some other means, but extrapolated rates of increase may again be applied for subsequent years.

The best way of improving estimates derived by extrapolation is, of course, by taking a new census. So long as no new census is taken, it is important to indicate, so far as possible, the reliability of estimates. One way of ascertaining approximate margins of error is to perform several independent extrapolations, one by assuming a rate of increase which can be regarded as a maximum, and one by assuming a rate which would seem to be a minimum.

If the results of the last census had to be modified in order to derive comparable figures for purposes of extrapolation, this fact should be indicated. Current population estimates are frequently used in conjunction with the results of the last census, but a comparison of these two figures leads to fallacious conclusions unless it is made clear how the extrapolated figures differ from those of the census. If a census result was increased to allow for an estimated amount of under-enumeration, or for a category of the population which had not been enumerated, this should be explained.

10. Comparative results obtained by various methods of extrapolation

In order to illustrate the results obtained by various methods, estimates of population at the date of the last census have been made, for several countries, by extrapolating in different ways the trends shown by earlier censuses. Comparison between these estimates and the enumerated figures for the latest census dates gives some insight into the types of errors produced by the different methods of extrapolation under various conditions. The countries selected for these illustrative calculations differ greatly as regards population trends, length of inter-censal intervals, and presumable accuracy of census enumeration.

The following types of extrapolation were carried out:

(1) Arithmetic rate of increase between two preceding census dates;

(2) Geometric rate of increase between the two preceding census dates;

(3) Parabola of the second degree fitted to the totals from three preceding censuses;

(4) Parabola of the third degree, fitted to the totals from three preceding censuses;

(5) Second-degree parabola computed on logarithms of the totals from three preceding censuses;

(6) Third-degree parabola computed on logarithms of the totals from three preceding censuses.

For British Honduras in 1946, for example, the following results were obtained:

Population according to the 1946 census:	59,220		
Result of extrapolation (1):	60,342, error:	+	1.9 per cent
Result of extrapolation (2):	61,865, error:	+	4.5 per cent
Result of extrapolation (3):	62,588, error:	+	5.7 per cent
Result of extrapolation (4):	70,342, error:	+	18.8 per cent
Result of extrapolation (5):	63,280, error:	+	6.9 per cent
Result of extrapolation (6):	59,856, error:	+	1.1 per cent

The population of British Honduras has been growing from census to census at first at a rising, but more recently at a declining rate. It grew from 37,479 in 1901 to 40,458 in 1911, to 45,317 in 1921, to 51,347 in 1931, to 59,220 in 1946. Hence the tendency of all methods of projection was to exaggerate the result for 1946. In this case, method (6), a third-degree parabola computed on the logarithms of the data furnished the best result but method (4), a third-degree parabola computed on the data themselves, furnished the worst result.

Errors resulting from extrapolations by the same six methods for fifteen countries are shown in table I.

Table I

ERRORS RESULTING FROM EXTRAPOLATION OF PREVIOUS CENSUS RESULTS TO DATE OF LAST CENSUS, BY SIX METHODS OF EXTRAPOLATION[a]

Country and date of last census[b]	Interval since last previous census (years)	Per cent error					
		Method (1)	Method (2)	Method (3)	Method (4)	Method (5)	Method (6)
Egypt, 1947	10	— 7.9	— 6.9	— 5.7	—12.5	— 5.4	— 2.9
Mauritius, 1944	13	— 1.4	— 1.2	+ 1.9	+ 0.7	— 0.2	+ 0.8
Canada, 1941	10	+ 4.0	+ 6.5	+ 4.1	+ 6.3	+ 3.1	+ 8.9
Alaska, 1950	11	—32.8	—30.3	—24.2	—21.8	—18.6	—26.5
British Honduras, 1946	15	+ 1.9	+ 4.5	+ 5.7	+18.8	+ 6.9	+ 1.1
Puerto Rico, 1950	10	— 0.5	+ 1.7	+ 2.7	—16.9	+ 4.1	—20.5
Ceylon, 1946	15	—14.1	—13.2	+ 9.6	+28.2	+11.7	+53.9
Thailand, 1947	10	+ 4.6	+11.1	+12.1	+13.8	+17.6	+14.4
Denmark, 1940	5	— 1.6	— 1.5	— 1.4	— 2.2	— 1.4	— 2.3
France, 1946	10	+ 2.4	+ 2.5	+ 1.4	—12.4	— 6.8	—11.5
Ireland, 1946	10	+ 0.3	+ 0.3	+ 3.3	+ 2.1	+ 9.8	+ 7.1
Portugal, 1940	10	— 1.3	+ 0.01	+ 8.4	+22.5	+12.3	+34.7
Switzerland, 1950	9	— 6.1	— 5.9	— 6.2	— 7.3	— 6.2	— 7.5
New Zealand, 1951	6	— 8.5	— 8.3	—11.5	— 5.5	— 8.6	— 3.7
Western Samoa, 1945	9	+ 0.8	+ 7.4	+16.1	+31.5	+24.4	+ 0.1
Mean error, irrespective of sign		± 5.9	± 6.8	± 7.6	± 13.5	± 9.1	± 13.1

[a] The methods are described on page 35. Data from *Demographic Yearbook*, 1951.
[b] Last census for which final results were available.

Only a few conclusions can be drawn from this comparison. It appears that projections of any kind are liable to err greatly in those countries where population growth has been irregular in the past. It also appears that extrapolations over a time period about equal to preceding census intervals can lead to large errors, particularly when third-degree parabolas are employed. However, parabolas of the higher order may be more reliable than some of the simpler methods if only a short time period is employed for projection from the last census. This seems to be the case in Ireland, where the last time interval was five years, and in New Zealand, where it was six years, while in most other cases the last time interval was ten years or more.

Method (1) (arithmetic rate) furnished the best results for Puerto Rico, Thailand and Ireland, but the poorest result for Alaska; method (2) (geometric rate) furnished the best results for Portugal and Switzerland, and in no case the poorest; method (3) (second-degree parabola) gave the best results for Ceylon, Denmark and France, and the poorest for Mauritius and New Zealand; method (4) (third-degree parabola) gave the poorest results for Egypt, British Honduras, Denmark, France and Western Samoa, and in no case the best; method (5) (second-degree parabola on logarithms) gave the best results for Mauritius, Canada and Alaska, but the poorest for Thailand and France; method (6) (third-degree parabola on logarithms) gave the best results for Egypt, British Honduras, New Zealand and Western Samoa, but the poorest for Canada, Puerto Rico, Ceylon, Portugal and Switzerland.

For Egypt, Alaska and New Zealand, all the methods used result in underestimates because population growth was accelerated in the most recent period. In each of these three cases, a third-degree parabola computed on the reciprocals of the data furnished a better result than any of the above methods, with an error of + 2.8 per cent for Egypt, — 6.3 per cent for Alaska, and — 2.3 per cent for New Zealand. In the cases of Denmark and Switzerland, where the above methods of extrapolation also resulted consistently in underestimates, no advantage was gained by an extrapolation of reciprocals.

The only general conclusion which can be drawn from this comparison is that a method which can give the best result in one situation may give the worst result in another situation. Good judgment must be used in selecting the most suitable method of extrapolation for the purpose of obtaining current population estimates in any given situation.

11. Short-period extrapolations for the purpose of making provisional estimates

In countries where births and deaths are regularly recorded it is sometimes necessary to obtain a preliminary population estimate for immediate purposes before complete statistics of births and deaths for the preceding year or some other brief recent interval have become available. Such a provisional estimate may be made by means of extrapolation.

The length of the period over which extrapolation must be extended in such cases depends on the regularity and promptness with which complete statistics of

births and deaths are obtained. Usually, this period is very short, perhaps a quarter or half a year. Much refinement is not necessary because the error resulting from extrapolation over such a short period can hardly be great. The choice of method is therefore not very important and arithmetic rates, being the simplest for computation, are ordinarily satisfactory, unless recent population changes have been of an unusual character.

If there is reason to believe that trends are fairly stable, extrapolation may be based on the changes during a period of a few recent years. If trends have been changing, it is preferable to use some other time interval during which population changes were of the kind which may have occurred during the most recent period. Thus it may be best to base such extrapolations on the last available census result and on the last available population estimate (such as, e.g., the official estimate for the middle, or the end, of the last year for which an estimate has been made by means of vital statistics).

If it is preferred for any reason to use a different method, such as geometric extrapolation, it is better to extrapolate numbers or rates of births, deaths, immigration and emigration, than to extrapolate population size.

A provisional estimate derived in such manner should be indicated as "provisional" in order to avoid confusion when it is replaced by a firmer estimate.

VI. ESTIMATES BASED ON CENSUSES, VITAL REGISTERS AND RECORDS OF MIGRATION

1. Advantages of the use of vital statistics and migration statistics

In countries where censuses are taken periodically and where there are accurate records of births, deaths and migration, the best method of making current population estimates is obviously to add the recorded natural increase and migratory changes to the population at the last census date. If the census figures and statistics of births, deaths and migration were perfectly accurate and consistent, this method would result not merely in estimates, but in precise figures of population size.

In practice, of course, the statistics are never perfect. In many countries where censuses have attained a fairly high degree of accuracy, vital statistics are markedly defective, although there are also a few countries where vital statistics are more accurate than census enumerations. Vital statistics in some countries are so imperfect that for purposes of current estimates, extrapolation is preferable. There are few countries in which international migration is recorded fully and with sufficient accuracy for all practical purposes. The more or less serious defects of migration statistics in most countries preclude the drawing of a precise migratory balance for any given time interval. The effect of these defects on current estimates of population depends both on the importance of the gaps in the statistics and on the actual volume of migration.

In some countries where the volume of international migration is small and where vital statistics are imperfect, errors in these statistics are usually greater than the net migratory movement to and from the country. In that case, even the omission of the migration factor from the computation of current population estimates can only result in errors which are small compared to errors arising from other sources. If, in addition, the statistically recorded movements account for only a small part of all migration to or from a country, it may be considered unnecessary or even unwise to take migration into account when making current population estimates. If the migration statistics are biased — that is, if the movements in one direction (e.g., immigration) are overstated in relation to the movements in the other direction — the estimates may be more accurate if these data are ignored in making the computations. However, the complete omission of migration statistics is ordinarily not advisable since the use of these statistics will usually tend to lessen the total error in the estimate, even though only by a rather small amount.

In countries with highly accurate vital statistics, on the other hand, accurate accounting of migratory movements, even if these are rather small, is important. In many countries, where other statistics are highly accurate, errors in current estimates are ascribed chiefly, if not entirely, to defective migration statistics, and further improvement of the estimates depends chiefly on any improvement in these statistics.

2. Components of error in estimates based on censuses and vital statistics with or without migration statistics

Current estimates derived from an inaccurate census with the use of fairly accurate vital and migration statistics may be considerably in error from the date of the census onward, but the error does not increase with the passing of time. On the other hand, current estimates based on an accurate census and inaccurate vital statistics may be of high quality immediately after the census, but will deteriorate as time passes. It is true that errors can compensate one another (e.g., the census may have been an under-enumeration while vital statistics exaggerate the natural increase because of under-registration of deaths, or under-registration of births may be partly compensated by a failure to record all immigrants, etc.), but there is no assurance that they will balance. Consequently, estimates based on components that are subject to error are unreliable even though some of the errors may occasionally happen to compensate.

Since the natural increase of a population is merely the numerical difference between births and deaths, errors in the computed natural increase arise only from differences in the completeness of registration of births and deaths. If birth registration is more nearly complete than death registration, the error in the computed natural increase is positive; if death registration is the more nearly complete, the error is negative.[1] While it is widely believed that death registration is usually more nearly complete than birth registration, there is evidence that in a number of countries the

[1] This statement requires some modification. In a country with a high birth rate and a low death rate, the natural increase of the population may be understated even though the registration of deaths is less complete than that of births, because a smaller proportion of unregistered births may involve larger numbers than a larger proportion of unregistered deaths. Suppose that, during some interval, 1 million births and 400,000 deaths have occurred, resulting in a natural increase of 600,000, but that birth registration is 90 per cent complete and death registration is 80 per cent complete. Then, despite greater completeness of birth registration, only 900,000 births and 320,000 deaths have been registered, resulting in an "apparent natural increase" of 580,000, which is less than the true natural increase.

opposite is true. It is unfortunate that estimates of the completeness of birth and death registration have been made only in a few countries.

Statistics of births and deaths are tabulated in some countries according to the time of their registration, in others according to the time of their occurrence. In the long run, the statistics tabulated in these two ways must give approximately the same totals, but over short periods the results may differ considerably if the delay between the occurrence of the events and the registration is variable. In that case, tabulations according to time of registration may involve considerable errors in the calculation of natural increase during various short periods of time. In some countries, delayed registration may result in some re-registration and double counting of births and deaths.

Where migration statistics are used in making current population estimates, errors may result from a more nearly complete recording of arrivals than of departures, or vice versa. For example, if the migration statistics are derived from passport records, they may include a larger proportion of all arrivals than of all departures, since passports are in many cases not required upon departure.

Migration statistics, however, present many other problems.[2] In many countries such statistics are kept only to serve specific requirements in the enforcement of certain kinds of legislation. In such cases, the statistics comprise only certain types of migrants, such as those entering or leaving at specific ports or other points on the boundary, or using specific modes of transportation. They may or may not include persons intending to change their residence for a short or long duration, and may make various provisions for temporary travellers, visitors, tourists, students, etc. Special migration statistics may exist for particular categories, such as persons entering or leaving on labour contracts, or persons of military ages. The definitions used for compiling the migration statistics may not be at all suitable to indicate the changes in population that result from movements across a country's boundaries. A particularly serious problem arises where statistics of immigration are not at all comparable with statistics of emigration.

3. Methods of improving estimates using vital statistics and migration statistics

The main ways of improving current population estimates of the type discussed in this chapter are to take censuses more frequently; to improve the accuracy of census statistics, vital statistics and migration statistics; to devise more adequate means of estimating the magnitudes of the errors in these statistics, and of "correcting" them; and to obtain better information on the "corrections" required for consistency in the various series. Sampling methods have many applications in these fields. The means of identifying errors and inconsistencies and of estimating the "corrections" are to be discussed in a later publication.

One important consideration in the use of vital statistics and migration statistics is their consistency with the census statistics. If statistics of births and deaths do not refer to the same territory or the same categories of the population as does the census enumeration, errors in the current population estimates may result.

To be strictly consistent with a *de facto* census definition, vital statistics should refer to the place of occurrence of each vital event, as they usually do. If the census definition is a *de jure* one, vital statistics should be employed, if available, which refer, in the case of deaths, to the place of residence of decedants, and in the case of births, to the residence of the mother or father. However, differences betwen the two kinds of vital statistics are usually negligible since rather few births and deaths occur in any one country to the residents of another country. Further difficulties arise from the special treatment of the military population in a census, not only in times of war, but also of peace. Deaths of military personnel and such births as occur to them may or may not find their way into civil registers.

If perfectly consistent vital statistics cannot be obtained, the data can be rendered approximately consistent with census definitions by means of estimates of the differences arising from the inconsistencies. In countries where there is a "vital statistics registration area" which does not cover the entire country, estimates have to be made for births and deaths occurring outside the registration area. Estimates are also needed where births and deaths of certain minorities, such as those of the aborigines of Australia or tribal Indians of some Latin-American countries, are not registered.

Consistency is usually a major problem in the case of migration statistics. To be fully consistent with *de facto* census figures, migration statistics should record every arrival and departure, whether permanent or only for a short duration. Migration statistics that are mainly confined to movements of persons who intend to change their residence for a considerable length of time are more nearly consistent with a *de jure* census definition. Migration statistics of the latter type, however, are often particularly unsatisfactory since it cannot always be determined whether a particular person crossing a country's boundary will turn out to be a real migrant or merely a temporary traveller. Over a long series of years, the day-to-day arrivals and departures of temporary travellers tend to balance, if the records are accurate, and the residual number represents, in fair approximation, the number of persons who have actually changed their residence. However, if there are any omissions in the records, this method may give a seriously biased measure of migration even over long periods, especially where temporary travellers are far more numerous than true migrants. In that case, in order to avoid large possible errors, it is preferable to use statistics of "migrants" proper, even if it is intended to estimate the *de facto* population.

[2] A detailed discussion of the inadequacies of migration statistics in many countries may be found in United Nations, *Problems of Migration Statistics*. Lake Success, November 1949.

If vital statistics or migration data are not yet available for the latest year, a provisional estimate of the increase of population during that year may be made by a short-period extrapolation, as explained in section 11 of chapter V. Similarly, if data on births, deaths and migratory movements are incomplete or not available for particular years owing to some temporary defect in the system of compiling these data, suitable figures for the missing years may be found by interpolation. For example, the missing number of births in a particular year may perhaps be assumed to equal the arithmetic average between reported births in the preceding and in the following year.

Auxiliary data may also be used to improve the estimates. In countries with a planned economy, various registration systems are in effect for specific purposes. These include registration for identity cards, labour permits, factory enrolment, residential registers, registers of school children, registers for family and old-age benefits, and various other types of registers. A coordination of the various separate compilations by provincial or central authorities may result in current population estimates which, though perhaps inferior to the estimates from census and vital statistics, may nevertheless serve to verify their accuracy. Particularly where a long time has elapsed since the taking of the last census, such registration data (which have already been referred to as results of "non-censal counts") may become important alternative means of estimating population.

The centralization of various registration procedures, including the registration of births, deaths and migration, can eventually result in the establishment of a continuous population register. In countries where registration is of relatively recent origin, it will usually be found difficult to establish an effective continuous population register within a short time.

4. Comparison of recorded natural increase with increase shown by successive censuses

In order to illustrate the errors which can result from the use of censuses and vital statistics over a series of years for current population estimates, the population of several countries has been estimated for the date of the latest census for which final results are available by adding to the results of the previous census the numbers of births and subtracting the numbers of deaths reported for the intervening period. Care was taken to use data of births and deaths for the precise period, from the exact date of the penultimate census to the exact date of the last census. The resulting estimates were then compared with the final results of the latter census.

For example, in Cyprus 347,959 persons were enumerated at the census of 28 April 1931 and 450,114 at the census of 10 November 1946; the population increase recorded by these two censuses amounted to 102,155. During the same time interval, 173,478 births and 76,961 deaths were registered, so that the apparent natural increase was 96,517 — that is less by 5,638 or 5.5 per cent, than the recorded increase between the two censuses. If the population at the date of the latter census were to be estimated from the former census with addition of births and subtraction of deaths in the intervening period, the estimate would amount to 444,476, or 1.3 per cent less than the census result. This error may be attributed to one or several of the following causes: an under-registration of births, and excessive registration of deaths, under-enumeration in 1931, over-enumeration in 1946, or net emigration. Similar computations have been made for the countries shown in table II.

The computed excesses or deficiencies shown in this table may be ascribed to varying completeness of census enumeration, relative inaccuracies in statistics of births and deaths, or failure to take migration into account. In countries of immigration, such as the United States during 1940-50, part or all of the deficiencies would disappear if migration statistics for the interval are added to the balance. In certain countries, the discrepancies must be attributed to the errors in some or all of the statistics used in the computations.

A discussion of the use of such comparisons as are illustrated in table II, in ascertaining the adequacy of available statistics, is reserved for a future publication.

Table II

COMPARISON OF RECORDED NATURAL INCREASES DURING INTER-CENSAL PERIODS WITH INCREASES SHOWN BY THE CENSUSES, FOR TEN COUNTRIES[a]

Country	*Census interval*	*Apparent increases*		*Excess (+) or deficiency (—) of apparent increase according to birth and death*		
		According to censuses	*According to recorded births and deaths*	*Numbers*	*In per cent of last census total*	*In per cent of apparent increase according to censuses*
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Tunisia	1936-46	622,639	377,072	— 245,567	—7.6	—39.4
Canada[b]	1931-41	1,129,869	1,245,431	+ 115,562	+1.0	+10.2
Cuba	1931-43	816,239	429,292	— 386,947	—8.1	—46.2
Dominican Republic	1935-50	641,666	706,442	+ 64,776	+3.1	+ 9.2
United States	1940-50	19,028,086	16,316,734	—2,711,352	—1.8	+14.2
Venezuela	1941-50	1,134,945	953,078	— 181,867	—3.6	—16.0
Cyprus	1931-46	102,155	96,517	— 5,638	—1.3	— 5.5
Norway	1930-46	342,756	304,449	— 38,307	—1.2	—11.2
Australia	1933-47	949,519	886,970	— 62,549	—0.8	— 6.6
Fiji	1936-46	61,259	55,005	— 6,254	—2.4	—10.2

[a] Data from *Demographic Yearbook*, 1951.
[b] Excluding Newfoundland.

VII. BALANCES DERIVED FROM CONTINUOUS POPULATION REGISTERS

1. The nature of a continuous population register

A few countries, notably those of Northern Europe, have developed in the course of time a system of human bookkeeping which may be described as a continuous population register. For the present purpose, such a register is defined as a system of registration which maintains an individual record for each person in the population, a record being added for each birth, a record being removed for each death, and provision being made for the transfer of records with changes of residence. This system must be distinguished from that of vital registers, where records of births, deaths, marriages and divorces are kept at the locality of their occurrence (or of the residence of the persons concerned), these records, once made, being final. It must likewise be distinguished from registers for specific purposes, under the control of registration boards, where provisions for entry, deletion or transfer are not sufficient to avoid omissions or double entries.

The distinguishing feature of all these registers is the provision of internal checks to avoid duplicate entries, particularly in the case of persons who change their residence within the country. It is also usual for population registers to be checked and corrected by a periodic census, once every five or ten years, or by comparison with the returns from tax schedules, rationing registers, school registers, etc.

The individual records of a continuous population register thus follow each person, from his birth, through the major events in his life, such as marriage, divorce, and the birth of children, or changes of residence, until death. Addition of the number of live records, at convenient time intervals, results in a population count almost equivalent to a census; and if the register is well kept the population figures which it yields may be superior in quality to those obtained from the average census.

Aside from providing at convenient intervals estimates of the total population, continuous registers serve also a variety of other purposes. They can combine the functions of various non-censal counts, such as counts of persons liable for taxation, persons subject to military service, persons eligible to vote, children of school age, etc., they can supply records for the identification of persons for legal purposes, and they fulfil the functions of registers of births, marriages and deaths. Continuous population registers can be particularly useful in determining the volume of internal migration within a country, a subject on which it is usually not possible to obtain satisfactory statistics from any other source. They are therefore of great value in providing current estimates of population within a country's administrative subdivisions.

2. The comparability of estimates derived from a continuous population register

It can be generally assumed that estimates drawn from continuous registers are highly accurate. Continuous registers have, however, two main drawbacks. In the first place, they record the resident population according to the manner in which residence is defined by the registration enactments. The definition of this "resident population" may correspond, more or less, to that of *de jure* population used in various censuses, but it differs from a *de facto* definition by excluding persons temporarily present and including persons temporarily absent. Estimates drawn from population registers are, therefore, not strictly comparable with the population estimates made in the majority of countries. It is true that the numerical difference between the *de jure* and the *de facto* population is usually rather small. However, in view of the generally high accuracy of estimates derived from registers, a greater precision of the definition of the population and a greater conformity with international standards appear desirable. For purposes of international comparability, therefore, countries having continuous population registers should endeavour to obtain *de facto* population figures by means of suitable estimates of the numbers of temporary absentees and temporary visitors from abroad.

A second drawback of continuous population registers arises from the difficulty of providing internal checks in the case of persons who have left the country and have either died abroad or established permanent residence in a foreign country. Entries of persons who have left the country without declaring themselves as emigrants but have not returned may therefore remain in the registers for a long time before being removed. A periodic check with census results or the establishment of a special register of persons whose whereabouts could not be ascertained will usually help to remedy this defect, but the remedy is not always perfect. Whereas continuous registers can record migrations within the country to a high point of precision, they are not always successful in recording the true numbers of emigrants.

Appendix A

A NOTE ON THE "MEAN POPULATION"

In the present Manual it has been recommended to present, wherever possible, estimates of the mid-year population, i.e., of the population as of 30 June or 1 July. The mid-year population appears to be the most suitable standard for comparability with regard to time reference, and it is in most cases a practically sufficient approximation for the computation of annual rates, such as birth, death and marriage rates, annual per capita income, production and consumption. The mid-year population suffers from a slight defect for some uses because it is not necessarily the exact average number of people living in the area during the entire year. In a country where the population is subject to considerable seasonal fluctuations, the mid-year population may sometimes differ by a significant amount from this mean population. Where this is the case, there is some theoretical advantage in basing the computation of various per capita rates on the mean rather than on the mid-year population. However, in nearly all cases this difference is practically negligible. It need not be taken into account except in countries in which statistics are highly accurate (such as balances from population registers derived quarterly or even monthly), or where seasonal fluctuations are of major importance (e.g., owing to the presence of seasonal foreign workers at the time of the harvest, or tourist travel into or out of the country during the holiday season).

A mean population can be computed by averaging the population figures for each month, or each quarter, of the year, provided that estimates at such intervals are available. The average of the population at the beginning and at the end of one year is often used as an approximation of the mean, or the mid-year population.

Monthly or quarterly population figures may be obtained where population registers are balanced at such intervals, or where data on births, deaths and migration are available for such intervals. If population is estimated by means of assumed rates of increase or by means of extrapolation, an estimate can of course be computed for any date whatever.

If an arithmetic rate of increase is either assumed or extrapolated, the mean and the mid-year population coincide exactly. A slight difference arises, however, if the assumed rate or the extrapolation proceeds geometrically or according to a curve.

Example. A population increases geometrically at a rate of 2 per cent per annum, from 1 million at the beginning of a year to 1,020,000 at the end of the year. What are the mean and mid-year estimates?

The mid-year population is computed by letting the population increase at an annual geometric rate of 2 per cent for half a year. The precise result of this computation is *1,009,950.*

The mean population is computed by adding the population for the beginning of each of the twelve months and for the end of the last month, and dividing the sum by 13. The resulting mean population is *1,009,969.*

It can be seen that the difference is, in this case, exceedingly slight. If the mean or mid-year population had been computed by simply averaging the population at the beginning and at the end of the year, i.e., by assuming an arithmetic rate of increase during the year, it would have amounted to *1,010,000.*

Appendix B

THE USES OF SAMPLING IN THE IMPROVEMENT OF POPULATION ESTIMATES

1. The principles of sampling

Statistical sampling is a procedure by which a number of units is selected from a group of persons or objects in such a way that the examination of the sample will permit valid inferences with regard to characteristics of the whole group. This procedure, which has been used for a long time in commerce — e.g., in determining the quality, weight, etc., of shipments of merchandise by inspection of sample batches — has been greatly elaborated in recent times through the development of scientific techniques of statistical sampling.

In order to permit valid inferences, i.e., inferences of measurable reliability, regarding the group from which a sample is drawn, the sample must be selected according to prescribed procedures. For example, when a sample of individual persons is to be drawn from a population, each person in the population should be given the same chance of being selected. One of the cardinal principles in drawing a "representative" sample, therefore, is *random* selection. If this criterion is fully satisfied, the inferences drawn from an examination of the sample will be free of bias and the errors of sampling can be calculated by means of standard formulae based on the laws of probability. It is not always easy to devise a method of truly random selection, and methods which appear at first to be suitable are often found, after careful consideration of the characteristics of the group, to involve serious biases. Hence, a scientific sampling investigation must be carried out under expert direction.

It is also important that the sampling procedure should be *efficient;* that is, the size of the sample and the method of selecting it should be such as to yield the minimum errors of sampling that can be achieved at a given expense, or to minimize the expense of achieving an acceptable standard of accuracy. Given some knowledge or some previous experience of the degree of homogeneity of the group with regard to the characteristics being investigated, a sampling expert can estimate in advance the sampling errors that will result from a given procedure and, by considering the costs of alternative procedures, he can choose the one which is most efficient.

Errors of sampling do not arise in an investigation covering all cases, as in a census enumeration of the entire population, or a comprehensive birth and death registration. Nevertheless, other errors occur in complete enumerations, owing to mistakes on the part of the enumerators, inaccuracies in the information supplied by respondents, and clerical errors in the process of compiling the results. Similar errors, of course, may also occur in sample studies, in addition to errors of sampling. The scale of a sampling investigation is, however, much smaller than that of a complete enumeration, and a relatively small staff of well-trained and highly qualified enumerators who can be easily supervised can perform the task with greater precision than can be attained in a larger-scale operation. Under favourable conditions, the total error in estimates based on samples — that is, the sum of the error due to inaccurate returns and the sampling error — may be smaller than the error due only to inaccurate returns in a complete enumeration. In some situations, therefore, a sample can provide more accurate results than a census, and do so at considerably less expense, in a shorter time, and with the use of a much smaller staff. Sampling can also be repeated at more frequent intervals than a census, in which case the efficiency of sampling is greatly increased. Sampling is sometimes the only possible means of obtaining census information under conditions where complete enumeration is impracticable.

2. Summary of possible applications of sampling in the improvement of population estimates

At many points in this Manual it has been suggested that sampling could be used to advantage in obtaining information on certain topics relevant to the improvement of estimates of total population — topics on which information is defective or entirely lacking. The uses of sampling which have been mentioned can be summarized under the following headings:

(*a*) To obtain basic data on population size by the enumeration of samples, in lieu of a complete population

[1] For further information see also:

United Nations, Statistical Office. *Population Census Handbook* (provisional edition), Lake Success, October 1949, pp. 46-51.

United Nations, Statistical Commission, Sub-Commission on Statistical Sampling. *A brief statement on the uses of sampling in censuses of population, agriculture, public health, and commerce.* Lake Success, February 1948; *Reports of Sessions,* 1-5 (E/CN.3/37, 52, 83, 114, and 140); and *Statistical Papers, Series C,* No. 1 (revised; "The preparation of sampling survey reports") and Nos. 2-5 ("Sampling surveys of current interest"), Lake Success and New York, February 1949 to March 1952.

A reference book on sampling methods is F. Yates, *Sampling methods for censuses and surveys,* New York, 1949.

census. There are many ways of doing this which have been described in other reports;[2] the choice of a method will depend largely on the type of information that is available as a framework for the selection of a sample. The best sampling frames for human populations are, of course, a previous census or, at least, a fairly complete list of individuals or households. Lists of villages, administrative divisions (in some cases also tribal divisions) can likewise be utilized. Where information of the latter type is lacking, samples can also be derived on the basis of maps or aerial surveys. In most cases, sampling for population size is not as efficient as sampling for population characteristics. If, however, the sampling process is repeated, the efficiency of sampling can be greatly increased.

(*b*) To estimate the multipliers which should be applied to certain known quantities in order to arrive at conjectural estimates of the population; for example, to estimate the average number of persons per tribe or village by studying a sample of tribes or villages, to estimate the average density of population in certain areas by examining a sample of areas, or to estimate the average per capita consumption or production of a staple commodity by determining consumption and production on the part of a sample of the population.

(*c*) To estimate appropriate multipliers for the results of non-censal counts relating to portions of the population, such as average numbers of non-taxpayers per taxpayer, non-voters per voter, women and children per adult male, persons of other ages per child of school age, or persons per dwelling. Such multipliers may be estimated by enumerating samples of taxpayers, voters, adult males, school children, dwellings, etc., and persons attached to them.

Where it is doubtful that the basic lists (of taxpayers, voters, households, etc.) are complete, a sample may also be selected from a different framework, and complete enumerations may be carried out in some selected localities. The ratio between the population enumerated in such localities and the numbers of individuals appearing on the basic lists corresponding to the same localities then determines a multiplier which also contains a correction factor for incompleteness of the lists. The latter type of multiplier, if applied to the total of listed individuals, tends to result generally in an efficient estimate of the total population.

(*d*) To estimate birth rates, death rates, rates of immigration or emigration, and rates of population increase, where adequate statistics on these subjects are lacking. The types of sampling studies that are appropriate for this purpose are various. If the results of a previous census, even a fairly old one, are available, rates of increase can be ascertained quite efficiently even by small samples. A selection of households may be canvassed and asked to report on numbers of births, deaths, arrivals or departures during a stated period, e.g., in the course of the past year; such a canvas of selected households may be performed concurrently with a census. Vital registers may be established in selected localities. Migrants, or samples of groups containing migrants, may be counted at selected points of entry or departure.

(*e*) Sampling is, in all countries, an almost indispensable device for the checking of the quality of statistical information. Thus, the completeness of the enumeration of children may be checked, at least approximately, by matching census returns of children with a sample of recent birth records. On the other hand, an approximate check for the completeness of birth registration can be provided by matching birth records with a sample of census returns. Similar checks may be made by comparing census returns on foreigners with recent records of immigration. The accuracy of enumeration or registration can also be checked by obtaining new measures of the same magnitudes, with the same or different methods and the use of specially trained enumerators in sample portions of the territory.

Sampling also finds important applications in controlling the accuracy of the tabulating process in a census.

(*f*) The population of certain areas or certain categories of the population not covered by the census or other basic data can be estimated by sampling. The procedures that are appropriate for these purposes are like those which may be used for sampling censuses, except that they may be limited to the particular areas or population groups on which special information is desired.

* * *

The present summary is intended merely to suggest the potentialities of sampling procedure. More detailed information regarding actual sampling procedures may be found in the sources referred to at the beginning of this appendix. The service of a sampling expert is indispensable if an actual sample investigation is to be carried out.

[2] United Nations, Statistical Commission. *Statistical Papers, Series C*, Nos. 2-5, "Sampling surveys of current interest".

HOW TO OBTAIN UNITED NATIONS PUBLICATIONS

United Nations publications may be obtained from bookstores and distributors throughout the world. Consult your bookstore or write to: United Nations, Sales Section, New York or Geneva.

COMMENT SE PROCURER LES PUBLICATIONS DES NATIONS UNIES

Les publications des Nations Unies sont en vente dans les librairies et les agences dépositaires du monde entier. Informez-vous auprès de votre librairie ou adressez-vous à: Nations Unies, Section des ventes, New York ou Genève.

КАК ПОЛУЧИТЬ ИЗДАНИЯ ОРГАНИЗАЦИИ ОБЪЕДИНЕННЫХ НАЦИЙ

Издания Организации Объединенных Наций можно купить в книжных магазинах и агентствах во всех районах мира. Наводите справки об изданиях в вашем книжном магазине или пишите по адресу: Организация Объединенных Наций, Секция по продаже изданий, Нью-Йорк или Женева.

COMO CONSEGUIR PUBLICACIONES DE LAS NACIONES UNIDAS

Las publicaciones de las Naciones Unidas están en venta en librerías y casas distribuidoras en todas partes del mundo. Consulte a su librero o diríjase a: Naciones Unidas, Sección de Ventas, Nueva York o Ginebra.

Litho in U.N. - 28690-Nov.1961-1,200
Reprinted in U.N. - 09032-May 1971-2,300

Price: $U.S. 1.50
(or equivalent in other currencies)

United Nations publication
Sales No.: 52.XIII.5