

POPULATION STUDIES, No. 23

*Manuals on methods of estimating population*

MANUAL II

Methods of Appraisal of Quality of  
Basic Data for Population Estimates



UNITED NATIONS

Department of Economic and Social Affairs  
Population Branch  
New York, 1955

The previous reports prepared by the Population Branch of the Bureau of Social Affairs are listed below:

**ST/SOA/Series A. POPULATION STUDIES**

**Reports on interrelationships between population growth and economic and social changes**

- No. 17. Determinants and consequences of population trends  
No. 20. Population growth and the standard of living in under-developed countries

**Reports on fertility and mortality**

- No. 13. Foetal, infant and early childhood mortality. Vol. 1. The statistics  
No. 13/Add.1 Vol. 2. Biological, social and economic factors  
No. 22. Patterns of human mortality

**Reports on migration**

- No. 5. Problems of migration statistics  
No. 11. Sex and age of international migrants: statistics for 1918-1947

**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  
No. 16. The population of Central America (inc. Mexico), 1950-1980

- No. 21. The population of South America, 1950-1980

**Reports on methods of population statistics and analysis**

- 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  
No. 18. Training in techniques of demographic analysis

**Reports on the population of Trust Territories**

- No. 1. The population of Western Samoa  
No. 2. The population of Tanganyika  
No. 14. Additional information on the population of Tanganyika  
No. 15. The population of Ruanda-Urundi

**Demographic Dictionary**

- No. 19. Multilingual demographic dictionary—Provisional edition

**ST/SOA/Series N. POPULATION BULLETIN**

- No. 1. December 1951  
No. 2. October 1952

- No. 3. October 1953  
No. 4. December 1954

ST/SOA/Series A/23

October 1955

UNITED NATIONS PUBLICATION

Sales No.: 56. XIII. 2

Price: \$U.S. 2.00  
(or equivalent in other currencies)

## FOREWORD

This is the second of a series of manuals on methods of estimating population which the Population Branch of the United Nations Bureau of Social Affairs has undertaken to prepare in accordance with a recommendation of the Population Commission (E/CN.9/88, paragraph 25(C)). The first manual in this series dealt with methods of estimating a country's total population for current dates.

The present manual is concerned with procedures for appraising the accuracy of census enumerations of the total population and sex and age groups, and of statistics of births, deaths and migration. These are the principal types of statistics used in the preparation of both current population estimates and future population projections. An appraisal of their quality, leading to necessary corrections, is an essential step in making reliable estimates and projections, and in determining the limits of confidence in the estimates which can be made.

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 material 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 with the co-operation of Dr. Abram J. Jaffe of Columbia University, Bureau of Applied Social Research.



## TABLE OF CONTENTS

	<i>Page</i>
<b>INTRODUCTION</b>	
The need for appraisal of demographic statistics.....	1
Purpose of this manual.....	1
Types of testing procedures.....	2
Nature of illustrative examples.....	3
Organization of this manual.....	3
<b>CHAPTER I</b>	
<b>The accuracy of census totals</b>	
<b>A. INTRODUCTION</b> .....	4
Factors affecting the completeness of a census count.....	4
<b>B. APPRAISING THE HEAD COUNT OF A SINGLE CENSUS</b> .....	5
Review of census procedures and performance.....	5
Relating the census count to non-census counts.....	5
Relating the census counts for various areas to known characteristics of the areas.....	7
Comparing the head and household counts.....	7
Comparison of urban and rural areas.....	7
Comparison of political subdivisions.....	8
Comparison of results of different tests.....	9
<b>C. CONSISTENCY OF TOTALS AT SUCCESSIVE CENSUSES</b> .....	9
Analysing the rate of change between two census dates.....	9
Comparing rates of change during intervals between three or more censuses .....	10
Population totals for subdivisions of a country.....	10
Comparing successive censuses.....	10
Comparing population changes with social and economic conditions....	11
<b>D. CONSISTENCY OF CENSUS TOTALS WITH VITAL STATISTICS AND MIGRATION STATISTICS</b>	
The balancing equation.....	12
Definition and limitations.....	12
Treatment of migration.....	13
Examples of the use of the balancing equation.....	13
<b>E. DIRECT CHECKS ON THE ACCURACY OF CENSUS TOTALS</b> .....	15
Timing of a direct check.....	15
Selection of areas to be re-surveyed.....	15
Comparing census records with lists of names.....	16
Example of partial re-enumeration in Costa Rica .....	16
Example of the post-enumeration survey in the United States.....	16
Sample verification of the census count in India.....	17

## CHAPTER II

### The completeness of vital statistics

	<i>Page</i>
A. INTRODUCTION .....	18
Effects of errors in vital statistics on population estimates .....	18
Coverage .....	18
Time and place of occurrence .....	19
Characteristics recorded on the certificates .....	19
Factors affecting the completeness of birth registration .....	20
Factors affecting death registration .....	20
Relative completeness of birth and death registration .....	20
B. INTERNAL CONSISTENCY OF VITAL STATISTICS .....	20
Numbers of reported births, deaths and infant deaths .....	21
Levels of rates .....	21
Birth rates .....	22
Consistency of crude birth rates with age structure .....	22
Death rates .....	22
Infant mortality rates .....	23
Trends in crude rates .....	23
Patterns of death rates, by age and sex .....	24
C. THE USE OF BALANCING EQUATIONS .....	25
Appraisal of death statistics by means of balancing equations .....	27
Appraisal of birth statistics by means of balancing equations .....	28
Application to the subdivisions of a country .....	29
D. DIRECT CHECKS ON COMPLETENESS OF VITAL STATISTICS REGISTRATION ..	30
Birth registration test in Puerto Rico, 1950 .....	30

## CHAPTER III

### The accuracy of age and sex statistics

A. EVALUATION OF CENSUS RESULTS CLASSIFIED BY SEX ONLY, OR BY SEX AND BROAD AGE GROUPS .....	31
B. EXAMINATION OF DETAILED AGE CLASSIFICATIONS OF THE POPULATION AT A SINGLE CENSUS DATE .....	33
Inspection of the data .....	33
Comparison of data with an expected configuration .....	36
Analysis of ratios computed from the data .....	39
Measurement of age-accuracy by means of an index .....	40
C. EXAMINATION OF DETAILED AGE STATISTICS FROM TWO OR MORE CENSUSES	43
Computation of survival rates from age 5 and over to age 15 and over ...	49
Use of the balancing equation .....	49
D. EVALUATION OF THE ACCURACY OF STATISTICS OF DEATHS BY SEX AND AGE GROUPS .....	53
E. DIRECT CHECKS .....	53

## CHAPTER IV

### Adequacy of migration statistics

A. ADEQUACY OF INTERNATIONAL MIGRATION STATISTICS .....	55
Use of the balancing equation .....	55

	<i>Page</i>
International comparison .....	56
Direct checks .....	57
B. ADEQUACY OF STATISTICS ON INTERNAL MIGRATION .....	57
Continuous population registers .....	57
Census statistics relating to internal migration .....	57
Tests of data on place of birth .....	57
Tests of data on place of residence at a previous date .....	58
Direct checks .....	58

#### APPENDIX A

##### The Haitian census of 1918/19

Appraising the church estimates .....	59
Comparing the church estimate and the 1918/19 census count with the 1950 census count .....	61
Comparison of rate of growth with that of other countries .....	61
Estimation of 1920 population from 1950 census data .....	61

#### APPENDIX B

##### The reliability of census data for Libya 63

#### APPENDIX C

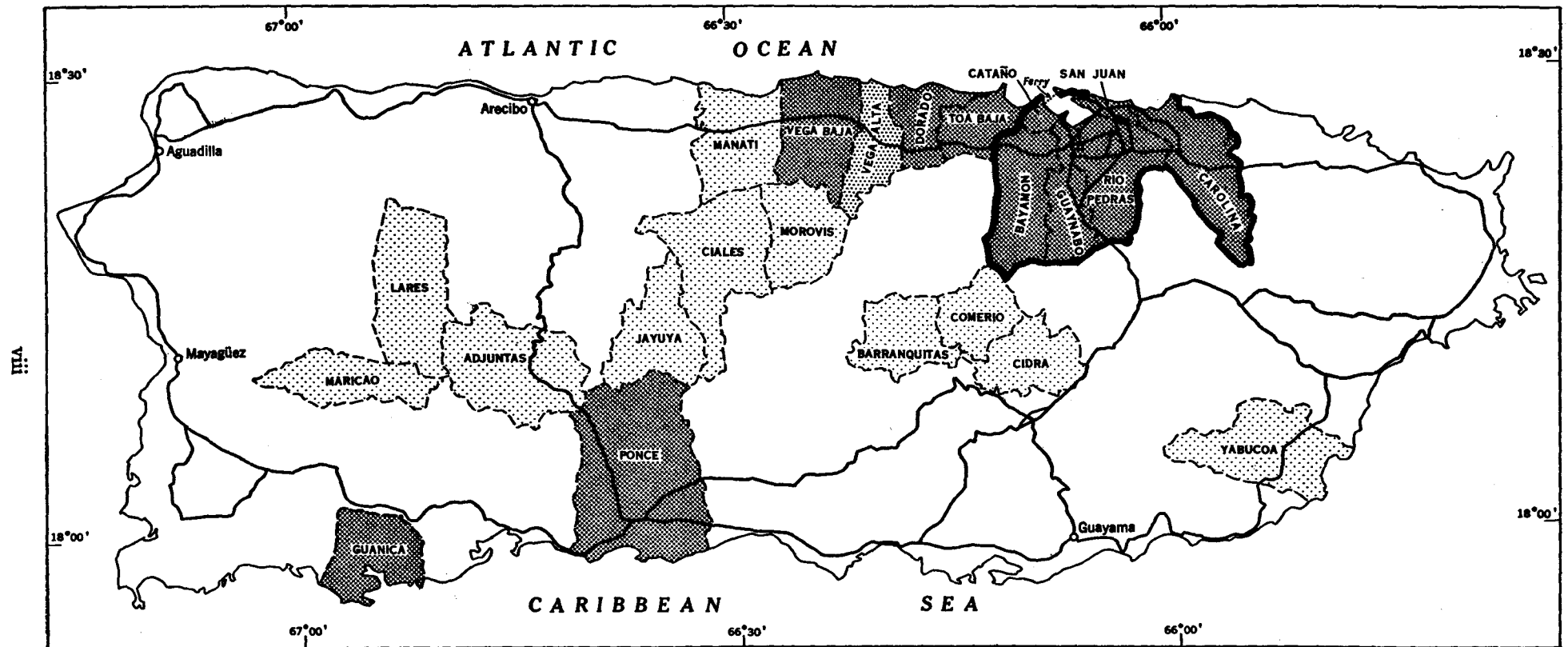
##### The censuses of Honduras

The consistency of the census totals .....	65
Consistency of census results with vital statistics .....	66

#### FIGURES

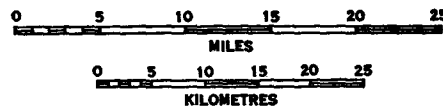
<i>Map.</i> Puerto Rico. Municipalities with given rates of population increase, 1940-1950 .....	viii
<i>Figure 1.</i> Population of Turkey, 1945, by sex, by single years of age and 5-year age groups, according to census .....	34
<i>Figure 2.</i> Population of Turkey, 1945, by sex and 5-year age groups. Comparison of census figures with a stable population .....	38
<i>Figure 3.</i> Ten-year survival rates for male and female cohorts computed from Egyptian census data for 1937 and 1947 .....	46
<i>Figure 4.</i> Ten-year survival rates for males and females computed from censuses of Honduras, the Philippines, Portugal and Turkey .....	48
<i>Figure A1.</i> Annual population estimates, 1900-1944, for Haiti and departments, and census results, 1918/19 and 1950 .....	60

# PUERTO RICO SELECTED MUNICIPALITIES WITH GIVEN RATES OF POPULATION CHANGE 1940-1950



MAP NO. 765 UNITED NATIONS  
JANUARY 1956

- Limit of municipality
- San Juan metropolitan area
- Major highways



- Increase of 20% or more
- ▨ Increase of 15% to 19%
- ░ Increase of less than 5% or decrease



## INTRODUCTION

### The need for appraisal of demographic statistics

Population statistics, like all other statistics, whether they are obtained by enumeration, registration, or other means, are affected by errors. The errors may be large or small, depending on the obstacles to accurate recording which are present in the area concerned, the methods used in compiling the data, and the relative efficiency with which the methods are applied. The importance of the errors, given their magnitude, depends on the uses to which the data are put. Some applications are valid even if the statistics are subject to large errors; other uses require more accurate data. When dealing with any given problem, it is important to know whether the data are accurate enough to provide a significant answer.

Until quite recent times, Governments paid little attention to an appraisal of the accuracy of demographic statistics. It is now becoming widely recognized that an effective statistical programme in this as well as other fields involves not only the compilation of the needed statistics but also an adequate measurement of their reliability. Hence, efforts have been made recently in various countries to ascertain the probable extent of errors in the statistics being obtained. This practice, however, has so far been mainly confined to those countries where statistical activities have been most fully developed. The need for investigation of the quality of the data collected is at least as great in the countries where little work in this direction has so far been done.

In connexion with population estimates, the evaluation of the census or registration statistics on which the estimates are based has a double importance. In the first place, an investigation of the accuracy of the base data is a prerequisite to any attempt at determining the reliability of the estimates. Errors of estimation result both from inaccuracies in the basic population statistics and from errors in the assumptions involved in deriving the estimates (for example, in the assumed population changes between the date of the latest statistics and the date to which the estimate applies); and both sources of error must be taken into account if the degree of confidence that may be placed in the estimate is to be known. Second, where an investigation of reliability of the base data has revealed errors, the direction and magnitude of which can be estimated, it is possible to make compensating adjustments, as the estimates of population are prepared, and thus to avoid a compounding of errors. In some instances, where difficulties are involved in the publication of adjusted census or registration statistics, estimates of population may, nevertheless, be made which are more reliable than the published official base statistics themselves. An adequate programme of investigating the accuracy of basic population statistics is therefore indispensable for the work of making good population estimates.

### Purpose of this manual

The purpose of this manual is to describe certain methods for appraising the accuracy of those types of statistics which are most commonly used as a basis for current population estimates and future population projections, namely, census figures on total population and on age groups, registration data on births and deaths, and statistics of migration. It is assumed that at least one census has been taken in the country concerned and that the results have been compiled. The statisticians of the country are faced with the problem of determining the accuracy of the census and other population data, but are not in a position to re-enumerate the whole population or repeat any major part of the census undertaking. With respect to the statistics of births, deaths and migration, the same assumption is made. The procedures used by the country for collecting these data are established, and the statisticians desire to appraise their accuracy in conjunction with the census results.<sup>1</sup>

Procedures for appraising the accuracy of certain population data have been described in various United Nations publications. With reference to vital statistics, such procedures are discussed in the *Handbook of Vital Statistics*.<sup>2</sup> In *The Population of Tanganyika*<sup>3</sup> there is an appendix on the problem of population statistics and in *Problems of Migration Statistics*,<sup>4</sup> there is a "Note on the relationship between excess of arrivals over departures and net migration". A few publications have been concerned exclusively with appraisal procedures as, for example, "Accuracy Tests for Census Age Distributions Tabulated in Five-year and Ten-year Groups".<sup>5</sup> Procedures that have been described in these other publications are referred to in this manual and only in rare instances are reproduced in summary form. In using this manual, therefore, it is desirable to have available for reference those publications mentioned above. As this manual is intended to supplement in certain respects the preceding manual on methods of estimating total popu-

<sup>1</sup> For information on the procedures of census enumeration and the collection of vital and migration statistics, see the following United Nations publications: *Population census methods*, document ST/SOA/Ser.A, Population Studies, No. 4, 1949; *Handbook of population census methods*, document ST/STAT/Ser.F, Studies and Methods, No. 5, 1954; *Handbook of vital statistics method*, document ST/STAT/Ser.F/7, in press, 1955; *Principles for a vital statistics system*, document ST/STAT/Ser.M, No. 19, 1953; *Informe final del primer Seminario Interamericano de Registro Civil, Santiago de Chile, 1954*, document ST/STAT/Ser.M/23, 1955; *Problems of migration statistics*, document ST/SOA/Ser.A, Population Studies, No. 5, 1949.

<sup>2</sup> *Op. cit.*, especially chapter XIV.

<sup>3</sup> United Nations, Population Studies, No. 2, 1949.

<sup>4</sup> United Nations, Population Studies, No. 5, 1949.

<sup>5</sup> *Population Bulletin of the United Nations*, No. 2, October 1952.

lation,<sup>6</sup> study of the latter will in some instances make clearer the meaning or purpose of some of the procedures described here.

The nature and quality of the demographic data existing in different countries vary greatly. Population censuses have been taken with varying frequency and available series of birth and death statistics contain different kinds of detail. Migration across national boundaries may be relatively important or not. Consequently, different methods have to be employed in different situations for the appraisal of the accuracy of statistics. It has not been possible in the present manual to consider all the possible situations in which it may be desired to test the existing statistics. Nor has it been possible to consider all the detailed tests to which every conceivable kind of data on the subjects covered here can be submitted. The methods presented here may, therefore, not always be directly applicable to a specific problem; modifications must be worked out by the statisticians in the country to suit their particular requirements.

The best appraisal of a country's statistics can be made only if all available statistics, published or unpublished, are at hand and if various other information concerning a country's economic and social conditions is also utilized. It has not been possible in the preparation of this manual to consider in every detail all the possible information which, in a given country, can be utilized for an appraisal of its demographic data.

The results of the tests described in this manual are of various kinds. Sometimes, a test will reveal only that statistics are either "probably reasonably accurate" or "suspect"; if they are "suspect", further intensive investigation is required before a definite judgement can be made. Other tests will not only indicate that errors are present, but also lead to an estimate of the direction and probable extent of the error. In the latter case, it is desirable to adjust or correct the faulty statistics and to revise the estimates based on them. The description of procedures to be used in the revision of estimates, however, is outside the scope of this manual.

### Types of testing procedures

Whether one is dealing with census data, vital statistics, or records of migration, the same basic types of testing procedures are applicable. This similarity results from the fact that demographic phenomena are interrelated both among themselves and with other social and economic phenomena. Some of these relationships are direct and necessary. For example, the increase in population during a given interval is precisely determined by the numbers of births and deaths, and the migratory movements, occurring in that interval. Other relationships are less precise and less definite. For example, in some countries, an economic depression is likely to result in a declining, and prosperity in a rising, birth rate, but the exact amount by which the birth rate will change cannot be inferred even from detailed knowledge of the economic situation.

The basic types of possible testing procedures can be summarized as follows:

- (a) Comparison of observed data with a theoretically expected configuration;
- (b) Comparison of data observed in one country with those observed elsewhere;
- (c) Comparison with similar data obtained for non-demographic purposes;
- (d) Balancing equation of directly interrelated data; and
- (e) Direct checks (re-enumeration of samples of the population et cetera).

The first two types of tests are similar. The demographic changes observed in some other country where conditions are presumably similar can sometimes be substituted for a theoretically expected configuration. When the observed data are compared with those which are theoretically expected, the two sets of figures will always be found to differ, whether by a large or a small amount. The essence of the test then rests on the answer to the question: Can the difference between the observed and expected values readily be explained by historical events or current conditions in the country, the data of which are being tested? If not, then it must be concluded that the observed data are "suspect" and may be in error. Further investigation may yield an explanation of the difference, or it may furnish clear indications that the "suspect" data are indeed in error. Very often this kind of method is to be applied merely as a preliminary step, to suggest along what lines further testing should be undertaken.

In many countries, other data are available which can be compared with the demographic statistics. For example, young men may be registered for the purposes of military recruitment. Estimates of the whole or parts of the population are often obtained from tax lists, voters' registers, school statistics, housing censuses et cetera. If such estimates differ from the population census data, the question arises whether there is a satisfactory explanation for the difference. This is the essence of the third type of tests.

"Balancing equations" can be applied to test the consistency of the increase in population shown by two enumerations at different dates, with the increase shown by statistics of the various elements of population change—births, deaths, and migration—during the interval. If all the data were accurate, the two measures of increase (or decrease) should evidently be balanced. Aside from population totals, the test can also be applied to sex and age groups and other categories of population that are identifiable in the statistics. Furthermore, by rearranging and re-defining the components of this equation, separate appraisals can be made regarding the accuracy of birth, death and migration statistics.

Direct checks involve a field investigation, such as a re-enumeration of a population sample, or a recount of births. The advantage of a direct check consists in the fact that the individual persons enumerated, or the individual events registered, can be identified, so that not only the consistency of totals, but the specific errors of omission or double-counting come to light.

It will be noted that the first four types of testing procedures only give an indication of *relative accuracy*, as it appears from the comparison of two sets of data both of which are subject to error. If several testing procedures are applied, or if there is a strong presumption that one set of data used in the comparison is highly

<sup>6</sup> United Nations, *Manuals on methods of estimating population*. Manual I: *Methods of Estimating Total Population for Current Dates*, document ST/SOA/Ser.A, Population Studies, No. 10, 1952.

accurate, the evidence so secured can have considerable weight, amounting to a practical certainty. In other instances, the comparison may reveal little more than that at least one, if not both, sets of data are in error.

### Nature of illustrative examples

Wherever possible, specific examples are included showing how various tests can be applied to the statistical data of certain countries. For the most part, the data for these examples have been drawn from the United Nations *Demographic Yearbook*. It is important to note that only a fraction of the data and knowledge which must be available in each country was actually used in working out these examples. Many more data, some of them not published anywhere, exist in these countries. In addition, the statistician's knowledge of the country, its people and their economic and social condition, and of the operations of the census and vital statistics system, is relevant to an appraisal of the demographic statistics. Hence, the examples presented here should be regarded merely as illustrations of methods and the results should not be taken as definitive evaluations of the quality of the particular data employed.

### Organization of this manual

The procedures are presented in four chapters with supplementary illustrations in three appendices. Each of the four chapters is concerned with methods of testing as applied to a particular kind of statistics, namely census totals, statistics of births and deaths, statistics of population by age and sex, and statistics of migration. The five basic types of tests are presented, in so far as applicable, in each of these chapters. In each case, the problem of testing is considered both on the assumption that there has been only one census and that there have been two or more censuses in the past. The application of testing methods to statistics of a country's political subdivisions, such as provinces, departments, counties,

municipalities, cities et cetera, as well as to statistics for the entire country, is considered wherever possible.

The first chapter, on the accuracy of census totals, is concerned with testing only the accuracy of the "head count", that is to say the total number of persons enumerated, without reference to the characteristics of the population. In the second chapter, emphasis is placed on methods for evaluating the completeness of the registration of births and deaths. The procedures described in those two chapters are carried into further detail in the third chapter, which deals with the accuracy of age and sex reporting, a subject of considerable importance in making population estimates and in analysing mortality, fertility, economic activities of the people et cetera. The last chapter, on the adequacy of migration statistics, is mainly concerned with statistics of international migration, but some attention is also given to internal population movements.

Three appendices are included to show how various tests, described separately in the several chapters, can be combined and brought to bear on a particular problem of appraisal. Appendix A and appendix C deal with census statistics of Haiti and Honduras, respectively. It is emphasized that the calculations presented there have been made without detailed study of the sources of information or the conditions existing in these two countries; the results are therefore only illustrative and far from a definitive appraisal. The purpose is only to show how several methods can be combined in an attempt at appraising a particular set of data. Appendix B reproduces some tests which were applied by C. L. Pan<sup>7</sup> to the Libyan censuses of 1931 and 1936, in an effort to appraise their accuracy. It is not expected that any other country will have exactly the same problems as these three, nor is it expected that the tests can be applied in other countries in exactly the same manner. The adaptation of these tests to meet the unique conditions of any specific country must be worked out by the statisticians of the country concerned.

<sup>7</sup> Reproduced from Chia-lin Pan, "The Population of Libya", *Population Studies*, Vol. III, No. 1, June 1949.

## CHAPTER I. THE ACCURACY OF CENSUS TOTALS

### A. Introduction

The degree of accuracy in a count of the total number of people in a country is a function of the accuracy with which the entire census operation was conducted. The head count may be either more or less accurate than the enumeration of some of the parts, such as age or marital-status groups, but if all the census procedures are of poor quality and the characteristics of the population have not been accurately determined there is little likelihood that the head count will be correct. Indeed, one of the ways of appraising the quality of the head count consists of analysing the accuracy of data on various characteristics of the population. This analysis may not only reveal evidence of inaccurate classification of the individuals enumerated, which is likely to be associated with a faulty head count, but also may reveal a tendency to omit certain categories of the population, which of course has a direct effect on the accuracy of the total count.<sup>1</sup>

Proper evaluation of the head count requires consideration not only of the census figures for components of the population and other information bearing on the quality of the work, but also of any collateral vital statistics, migration statistics, and other relevant data, and the reliability of these collateral data. Ideally all the materials should be examined, and evaluated simultaneously, but in practice this is impossible. In this manual the ways of analysing the different items are taken up separately.

#### FACTORS AFFECTING THE COMPLETENESS OF A CENSUS COUNT

The factors which contribute to the completeness of a census count will be but briefly mentioned here. These factors have already been considered in greater detail in other publications dealing with census procedures.<sup>2</sup> The main function served in reviewing them here is to help direct the analyses of the accuracy of census details. For example, social customs sometimes operate against reporting certain segments of the population in certain countries; accordingly, if it is known that among a certain population group the custom of not talking in public about baby boys is prevalent, then special attention can be directed toward evaluating the completeness of census enumeration of male children among that group.

It is assumed that administrative and technical procedures at the central census office have been arranged as well as possible so that clerical errors or other mistakes in handling the census returns are reduced to a minimum. The factors affecting the completeness of the census count can then be listed as follows:

<sup>1</sup> For the analysis of age and sex classifications as clues to the accuracy of the head count, see chapter III.

<sup>2</sup> See: United Nations, *Population Census Methods*, Population Studies, No. 4; and *Handbook of Population Census Methods*.

1. *Obtaining full public co-operation.* Sometimes the general public may not fully comprehend the significance of a census because of ignorance resulting from illiteracy, lack of communication facilities et cetera. In the past in some areas censuses have been connected with taxation, military service, rationing, or other benefits or obligations; where the public believe that such a connexion exists it is difficult to obtain their full and unbiased co-operation. Superstition and various social customs sometimes are important; if it is believed that "bad luck" will follow if one gives out personal information about the members of his family, there will be trouble in obtaining such information. In some cases enumerators may have to be recruited from an educated class which does not enjoy the full confidence of less educated segments of the population.

In general, but not always, this type of problem is most acute in the case of poor and illiterate populations, among those groups sometimes characterized as "under-developed".

2. *Problems of geography.* In an area which is difficult of access because of geographical barriers such as mountains, swamps, deserts, bodies of water et cetera, and where the road and transportation system is inadequate, considerable trouble will be had in locating the population. Often, where such geographical conditions are found, there is also an absence of adequate maps, which complicates the problem.

3. *House identification.* The intent of every census is to relate the population to specific areas; this involves being able to identify the dwelling or other place where each person is to be counted. Furthermore, it is necessary to identify dwellings so that coverage can be assured; the enumerator must know that he has or has not as yet enumerated a given person. Among some populations, such identification is difficult for various reasons: there are houses which lack exact addresses, persons who dwell on boats or barges, nomads or seminomads who have mobile dwellings, or persons who have no fixed place of residence.

4. *Obtaining good enumerators.* If there are few educated and literate people in a given area, it may be difficult to obtain enough persons qualified to do a good job of enumeration. Sometimes also, even when there are enough qualified people in an area, it may be difficult to recruit them for census work. There may also not have been enough opportunity to provide enumerators with the necessary training to enable them to interpret the census instructions correctly.

Other problems affecting the accuracy of census enumerations may be mentioned. Some such problems are peculiar to one country only. It is useful, before appraising the accuracy of the census results, to review all the problems which may have been encountered and to consider what defects in the enumeration may have resulted. Special efforts should be made to appraise the

completeness of the census counts in those areas or among those population groups which are known to be subject to conditions unfavourable for census taking.

### B. Appraising the head count of a single census

The problem considered in this section of the chapter is that of evaluating the accuracy of a single census without reference to demographic data other than those of the census itself. This step should be taken in the appraisal of any census. If the data of a previous census, records on births, deaths and migration, or detailed statistics on population characteristics also exist, this step is only a part of a more extensive analysis. If the results of this one census, without detailed breakdown by population characteristics, are the only demographic data available, methods of appraisal are of necessity confined to those described in this section, unless direct checks, considered in section E of this chapter, can be undertaken.

Even in the case of those censuses which yield no other statistics than the head count, the results are invariably available for various political subdivisions of the country. In many cases, but not invariably, information is also available on the numbers of households (or families) enumerated. The enumeration procedures are generally conducted on a household basis (that is to say the information for all persons living in one household is grouped together on the census schedule and the household is separately identified). Procedures for using the figures for subdivisions of the country and the information on households, if available, for helping to appraise the total head count, are accordingly introduced here.

In the circumstances stated above, only a direct check such as described in section E below, can give a definite measure of the extent of error in the head count. However, without a direct check, the following four tests may show whether or not an error of considerable importance is probably present, and in some cases may yield an estimate of the amount of errors.

- (1) Review of the enumeration procedures and information on the quality of performance.
- (2) Comparing the census figures with any available data from non-demographic sources which relate to the numbers of the population or parts thereof.
- (3) Relating the population distribution as revealed by the census findings to known characteristics of the subdivisions.
- (4) Comparing the head and household counts.

Additional tests, some of which are described subsequently in the present section, are also recommended, depending upon the circumstances of the case.

### REVIEW OF CENSUS PROCEDURES AND PERFORMANCE

With respect to the first test, there is little that can be said other than that all the factors involved in the census undertaking which may bear on the accuracy of results should be examined. Such an examination may very well be regarded as an integral part of the census work rather than as a separate testing procedure. The enumerators' maps should be checked to see whether the entire country was covered, and the returns should be checked to see whether every known city, town, village and hamlet was included. Available reports of the

census supervisors for different areas should be searched for observation on the quality of the personnel engaged and of their performance. The completed schedules for each area should be scanned for an impression of the degree of care taken in filling them, and of the extent to which the instructions appear to have been understood and followed.

The prescribed census procedures themselves should be studied with attention to the possibilities of error which may be inherent in them. For example, if it is found that in some or all areas the enumeration was made by convoking the population of each village, the members of each tribe et cetera, at an appointed time and place instead of visiting their dwellings individually, the possibility of a substantial error may be envisaged. If the enumeration was conducted over a very long period of time, errors may be expected as a result of population flux during the interval.

An example of the use of some of the methods described above, together with others, to verify the probable extent of omissions is found in the Brazilian census of 1940. Persons connected with government and business organizations were required to report any known case of census omission on a special questionnaire; in one area, postmasters were required to identify persons in their respective districts who had not been reported at the census; students were offered rewards for information on persons not enumerated; local judicial authorities were asked to express their opinion on the probable efficiency with which enumeration was carried out in their respective areas. Direct checks were also employed. By comparing the information obtained from these several sources, it was estimated that about 1.7 per cent of the population of the country had not been enumerated, this percentage varying greatly from area to area, depending on whether population was dense or scattered and whether transportation was easy or difficult.<sup>3</sup>

### RELATING THE CENSUS COUNT TO NON-CENSUS COUNTS

Often, some estimate of population can be obtained from such sources as police records, church records, tax lists, counts of persons eligible for military service, school attendance et cetera. Information relevant to population, or to households, farms et cetera, may also be obtained from censuses of housing and agriculture. In some cases such information may purport to be total counts of the population living in each area; in other instances the information may relate only to selected population segments or other relevant data which can be converted into estimates of the total population.<sup>4</sup> These figures are then compared with the census counts for the whole country and each region. Where there is relatively "close" agreement the presumption is that both counts are reasonably correct; at least such agreement does not prove that the census count is incorrect. If there is considerable disagreement, it is necessary to check further in an effort to appraise the two counts.

<sup>3</sup> International Statistical Conference, Washington, D.C., 1947. *Note on the Completeness of the 1940 Brazilian Population Census*, by José Carneiro Felipe and Octavio Alexander de Moraes.

<sup>4</sup> Methods of deriving population estimates from non-censal statistics have been described in the preceding manual in this series, Manual I: *Methods of Estimating Total Population for Current Dates*, Population Studies, No. 10, especially chapter III.

Of course, even when both counts are in agreement, the evidence is not conclusive that the census count is correct. Both counts may be incorrect by the same amount, as for example, by the omission of the population living in some remote area. Further checks, including the conducting of sample re-enumerations, are necessary in order to provide positive proof of the degree of inclusiveness in the census. If the non-census counts and estimates are believed to be nearly accurate, however, the comparison yields at least an approximate estimate of the error in the census figure for the area concerned. Also, if there is reason to believe, for example, that the non-census figure is below the true number, and if the census figure is still lower, the inference can be drawn that the census figure is too low by at least the amount of the difference.

The application of this test, that is to say, comparing census results with non-census estimates—is illustrated with data for Haiti, Northern Rhodesia and Southern Rhodesia in table 1 below. A census was taken in Haiti in 1918/19; its results are compared with the estimates derived from church membership data for the same period. The census count of 1,631,000 population is 30 per cent below the non-census estimate of 2,121,000. Among the several regions of the country the differences between the two counts are quite variable. In the Artibonite the two figures are very close; in the North, however, the non-census estimate is about 71 per cent greater than the census count. In order to evaluate these data each set of counts has to be explored at considerable length and other evidence introduced. This problem is illustrated at further length in appendix A.

The data for Northern Rhodesia show much closer agreement than those for Haiti. Here the census count of 1,837,000 is 9 per cent greater than the non-census count of 1,674,000 derived from tax registers. In four of

the six regions the differences are 5 per cent or less, and may be considered almost negligible for many practical purposes. In two of the regions, the Eastern and Northern, the differences are quite substantial. Part of the difference between the census and non-census counts may be due to the fact that the census count is based on a sample and therefore has some sampling variance; sampling variance, however, does not account for all of the differences between the two sets of counts.

Detailed pursuance of this problem is not within the scope of this manual. It can be pointed out, however, that the procedures for obtaining non-census counts such as those of Northern Rhodesia can lead to some error. Generally, these procedures consist in obtaining figures on the numbers of taxable heads of households, collected from every village by visiting officials, with subsidiary estimates and adjustments to make estimates of total population. The figures on taxpayers may be incomplete or out of date and the estimates of their dependants, mostly women and children, may be quite inaccurate. Actually, in the present instance, the two counts for adult males are quite close, the totals differing by less than 2 per cent. For women and children, however, the non-census counts are about 12 per cent below the census counts; it is quite possible that most of this difference is due to error in the non-census figures. Furthermore, non-census figures for some villages were out of date, having been kept without change since the time of the last visit.<sup>5</sup> In short, there are sufficient discrepancies within the non-census count to disqualify it as a positive check on the census count. It can be concluded only that in four of the regions the agreement is quite good, and in two regions there is much less agreement. It must be concluded, however, that the results of the two counts in Northern Rhodesia substantiate each other much more closely than they do in the case of Haiti.

In Southern Rhodesia the agreement between the census and non-census counts appears somewhat better than in Northern Rhodesia. In none of the regions is there exact agreement. The maximum difference, however, is only 9 per cent; in some regions the census count is higher and in others the non-census count is higher. As in the case of Northern Rhodesia, the census count is based on a sample and hence has some sampling variance; the non-census counts also are in error to some degree.<sup>6</sup>

Practically speaking, the agreement between the census and non-census counts is probably about as close as can be expected.

The observations on these three countries can be summarized as follows:

1. For Southern Rhodesia, the comparison with the non-censal estimates gives support to an inference that the census count is not subject to great error.

2. For Northern Rhodesia, the result of the test is less conclusive, but it does not suggest a large error in the census figures.

Table 1

COMPARISON OF CENSUS COUNTS AND NON-CENSAL POPULATION ESTIMATES, FOR SELECTED AREAS  
(numbers in thousands)

Area and date	Census count (a)	Non-censal estimate (b)	Difference	
			Number (a)-(b)	Per cent of census count
<i>Haiti, 1918/19</i>	1,631	2,121	-490	-30
West .....	671	780	-109	-16
North .....	256	437	-181	-71
Artibonite .....	241	255	-14	-6
South .....	385	541	-156	-41
Northwest .....	78	108	-30	-39
<i>Northern Rhodesia, 1950</i>	1,837	1,674	+163	+9
Barotse .....	278	268	+10	+4
Central .....	176	175	+1	+
Eastern .....	376	322	+54	+14
Northern .....	410	338	+72	+18
Southern .....	212	204	+8	+4
Western .....	385	367	+18	+5
<i>Southern Rhodesia, 1948</i>	1,619	1,631	-12	-1
Salisbury .....	408	436	-28	-7
Gwelo .....	366	334	+32	+9
Victoria .....	284	298	-14	-5
Bulawayo .....	313	335	-22	-7
Umtali .....	248	228	+20	+8

\*Difference negligible.

<sup>5</sup> Report on the 1950 Demographic Sample Survey of the African Population of Northern Rhodesia, Central African Statistical Office, Salisbury, Southern Rhodesia, April 1952, pp. 10 and 11.

<sup>6</sup> Report on the Demographic Survey of the African Population of Southern Rhodesia, Central African Statistical Office, Salisbury, Southern Rhodesia.

3. For Haiti, the magnitude of the difference calls for detailed investigation of sources of error in both sets of figures; pending the result of such an investigation the possibility of a major error in the census count must be admitted.

#### RELATING THE CENSUS COUNTS FOR VARIOUS AREAS TO KNOWN CHARACTERISTICS OF THE AREAS

Generally the statistical agencies of every country have intimate knowledge about the geographical, social, and economic characteristics of the country's subdivisions. Such knowledge is sometimes quantitative and perhaps very often of a qualitative nature. In either event such details provide clues as to the possible population of the various parts of a country. Such knowledge does not in itself prove that the census count is correct or deficient, but it serves as a background against which to appraise the census count. For example, many countries have detailed land use maps and/or maps showing the agricultural areas or comprehensive aerial photographs. Such maps or photographs can never prove how many people should have been enumerated in each segment of the country, but they can suggest whether there may be many more, or many fewer people than the census counted. In this connexion use of density figures (population per square kilometre, for example) according to land use in the various subdivisions of the country may be the most illuminating procedure. For example, an area devoted to coffee production is likely to have a lower population density than one devoted to sugar cane or tobacco, since coffee requires perhaps only one-third as much labour per unit of land; an area devoted to cattle raising should be more sparsely settled than one devoted to dairying.

This type of appraisal can be illustrated with data for Puerto Rico. A group of municipalities were selected in which sugar-cane raising was very important; they were all located on coastal plains. The average density shown by the census in these municipalities was 470 persons per square mile; the range in density was from 341 to 706 persons. Another group of municipalities located in the mountains, in which coffee raising was very important, had an average density of 332 persons per square mile. The range in density among these municipalities was from 200 to 483 persons per square mile. Clearly, the two types of regions have different typical population densities, but for some of the sugar areas the census shows no higher density than for certain coffee areas. If it were desired to check the census counts, this comparison would suggest paying special attention to those sugar municipalities which reported low population densities, and perhaps those coffee municipalities which reported very high densities.

Other information can be brought to bear in accordance with the known conditions of the country. Perhaps a given part of a country contains extensive public lands which are largely closed to settlement; few people should be enumerated as living there. The results of censuses of agriculture, industry, or housing, if available, are likely to be especially useful. All such types of information cannot be listed here; the statisticians in each country can seek the information appropriate to their purpose.

#### COMPARING THE HEAD AND HOUSEHOLD COUNTS

Comparison of these two counts is done most easily by computing the numbers of persons per household. Several such measures are available: (a) the arithmetic average number of persons per household; (b) the median number of persons per household; (c) households classified by number of members, such as one-person households, two-person households et cetera. For the sake of simplicity the tests illustrated below are based on the arithmetic average number of persons per household (or family).

#### COMPARISON OF URBAN AND RURAL AREAS

This test involves calculating the average number of persons per household for urban and rural areas, not necessarily by political subdivisions. In most countries the censuses classify the population as urban and rural; hence it is possible to tabulate the census results so as to obtain counts of the numbers of persons living in each type of area, and the numbers of households in each type. Such data without reference to political subdivisions will be used in these illustrations.

As a general rule, the average size of household in rural areas should be larger than in urban areas. The absence of such a relationship constitutes a warning that the census count may be deficient and that an investigation of the reasons for the observed differences is required. This type of test can be illustrated with the following data for Puerto Rico and Costa Rica. (The choice of these illustrations, of course, does not imply that the census figures for these countries are thought to be particularly unreliable.)

##### *Persons per household*

##### *Puerto Rico (1950)*

Total country .....	5.07
Metropolitan areas (containing central cities of 50,000 population and over).....	4.69
Central cities .....	4.55
Population living outside central cities....	5.16
Cities of 25,000 to 49,000 population.....	4.40
Cities of 10,000 to 24,000 population.....	4.56
Small cities and rural areas.....	5.35

##### *Costa Rica (1950)*

Total country .....	5.59
San José metropolitan area.....	5.40
San José city.....	5.31
Population living outside San José city..	5.72
Other cities .....	5.26
Rural areas .....	5.74

It will be noted that in both of these countries the number of persons per household shown by the census is larger in rural areas than in the cities. Within the urban areas of these countries there seems to be little, if any, relationship between city size and number of persons per household. In some countries the average size of household may become progressively smaller as the size of the city increases; this condition, however, is not as commonly found as the urban-rural difference mentioned above. As regards the figures for Puerto Rico and Costa Rica, the conclusion from this test is that it has produced no evidence of a possible error which would require investigation.



# COMPARISON OF POLITICAL SUBDIVISIONS

This test consists of calculating the number of persons per household by urban and rural areas, for the various subdivisions of a country. If the average size of household is very unusual in any one area—either very large or very small—it serves as an indication that there may be errors in the census count which ought to be investigated further. This test can be illustrated with data for Costa Rica as follows:

Province	Persons per household, 1950		
	Total	Urban	Rural
Total country .....	5.59	5.33	5.74
San José .....	5.58	5.40	5.80
Alajuela .....	5.80	5.30	5.94
Cartago .....	5.86	5.54	5.98
Heredia .....	5.79	5.57	5.91
Guanacaste .....	6.31	5.69	6.42
Puntarenas .....	5.34	5.21	5.39
Limón .....	3.98	4.08	3.94

Examination of these data on average size of household reveals some variation among the provinces, and shows that the province of Limón has the smallest households. Furthermore, within each province the rural area has the larger average size household, with the exception of Limón. Indeed, within the urban areas there is comparatively little variation among the provinces, excluding Limón; the maximum variation is less than one-half a person (compare Guanacaste and Puntarenas). Within the rural areas the variation is somewhat higher than among the urban areas, the maximum variation, excluding Limón, being about one person (again, compare Guanacaste and Puntarenas). Limón appears so unusual by this test as to raise the possibility that there was an undercount of population there and further investigation seems called for. The possibility to be investigated is that many persons were not counted in the census although their households were enumerated. Actually, as the following analysis shows, other plausible explanations can be found for the Limón figures.

Further analysis can be based on data relating to characteristics of the population as shown by the census.

Although the evaluation of such data is to be taken up in a later chapter, it seems convenient to introduce here, for illustration, the use of figures such as those assembled in table 2, on this page. Inspection of lines 2 to 4 inclusive reveals that Limón has an unusually high percentage of one-person households, 18.3 per cent as compared with 4.8 per cent in the total country. This could have occurred if, in some households, the wives had been omitted from the census count.

Hence, the next step is to examine the sex composition. Limón has an excess of males of all ages (line 6), in comparison with the total country; only Puntarenas has a larger proportion of males. Turning to the population aged 15 and over (line 7), it is noted again that Limón has a larger proportion of males than the country as a whole.

This high percentage of men might be a further indication that many women were missed. Accordingly, marital status composition can be studied; if many wives had been missed then there ought to be a preponderance of married men in this province. Among the married population, however, there is but a very slight preponderance of men (line 9). In this respect Limón differs but very little from the entire country. Among the single population there is a very large excess of men (line 8); in Limón 63.3 per cent of the single population are men as compared with but 52.5 per cent in the total country. Only Puntarenas has a somewhat higher proportion of single men. These figures do not support the hypothesis that wives were omitted from the census in Limón in any unusually large numbers.

Might an unusually large number of children have been omitted from the census count in Limón? In line 5 is presented the number of children under 5 years of age per 100 women aged 15 to 44 years inclusive. If the great majority of the women in the province were included, but if a large number of children were omitted, this ratio ought to be relatively small in Limón. In fact, however, this ratio is exactly the same for Limón as for the total country (line 5). Hence, the presumption is that unusually large numbers of children were not omitted from the census count in this province.

Table 2

## SELECTED POPULATION CHARACTERISTICS FOR COSTA RICA, BY PROVINCES, 1950

Line No.	Entire country	San José	Alajuela	Cartago	Heredia	Guanacaste	Puntarenas	Limón
1. Population (in thousands).....	800.9	281.8	148.9	100.7	51.8	88.2	88.2	41.4
2. Number of households (in thousands) .....	143.2	50.5	25.7	17.2	8.9	14.0	16.5	10.4
3. Number of households containing only one person (in thousands).....	6.9	1.9	0.8	0.6	0.3	0.5	0.9	1.9
4. One-person households as percentage of all households.....	4.8	3.8	3.0	3.6	3.3	3.4	5.5	18.3
5. Children under 5 years per 100 women aged 15-44.....	74	65	81	79	68	89	80	74
Percentage of males:								
6. In total population.....	49.9	47.9	49.9	50.5	48.9	51.2	53.7	53.0
7. In population aged 15 and over.....	49.3	46.4	49.1	49.7	47.9	51.3	55.5	54.3
8. In single population aged 15 and over	52.5	47.3	51.9	53.3	49.3	55.2	65.3	63.3
9. In married population aged 15 and over <sup>a</sup> .....	49.9	49.5	49.7	49.9	49.6	50.0	50.8	50.9

<sup>a</sup> Including consensually married persons.



In summary, the small average size of household in the province of Limón seems consistent with a census count, as complete there as in the remainder of Costa Rica. Further checks can be made, of course, not only by re-enumerating the province or a sample of it, but also by investigating other aspects of its social, economic and ethnological structure, factors which could not be included in this brief example.

#### COMPARISON OF RESULTS OF DIFFERENT TESTS

As the examples given above imply, the results of the types of tests described in the present section are often not sufficient in themselves to prove that the census count is either correct or in error, or to give an estimate of the amount of error. For the most part, they serve as indicators of possible errors that deserve investigation by means of more rigorous checks. It is often possible, however, to get more definite indications by comparing the results of different tests of the type described. For example, if a review of the census methods and evidence on the quality of performance gives rise to a suspicion that the results for a given administrative area may have been inaccurate, and if a comparison with non-census data or an analysis of the household figures suggests a possible deficiency in the count for the same area, the inference to be drawn from the two tests combined is far more positive than the result of either test considered alone.

#### C. Consistency of totals at successive censuses

When the total population of a country is known for several successive censuses, better estimates of the probable accuracy of each census, including the most recent one, are possible than if data were available for only one census.

The guiding principle to be followed in comparing the results from two or more successive censuses, is that population changes normally proceed in an orderly manner. This is to say that in the absence of unusual events the rate of increase for the whole country and for each of its parts, can be expected to change only gradually in successive inter-censal periods, and to follow a fairly constant trend. When such an orderly pattern is not observed, the deviations should be explainable in terms of known events, such as the curtailment of immigration, the presence of famine, or any other event. Deviations from the pattern which cannot be so explained constitute a warning of possible errors; and the presumption of error is greatly strengthened if the results of other tests, however inconclusive in themselves, are found to point in the same direction.

#### ANALYSING THE RATE OF CHANGE BETWEEN TWO CENSUS DATES

When the total population of a country is known for two successive census dates a test which can be applied consists of comparing the observed rate of change with the rates of change observed in other countries. For this purpose comparisons should be made with such other countries as seem to have approximately comparable demographic, social and economic conditions. Actually, of course, no two countries have identical conditions; nevertheless, countries which have fairly

similar social and economic conditions should have fairly similar rates of population growth.

This procedure can be illustrated with data for Angola and neighbouring areas:

Area	Dates	Average annual rate of increase <sup>1</sup>
Angola	1940/1950	0.96
Mozambique	1940/1950	1.18
South Africa, excl. Europeans	1946/1951	2.03
Bechuanaland	1936/1946	1.02
Swaziland	1936/1946	1.68

Although the average annual rate of population increase in Angola is lower than that in any of the other areas with which it is compared, its rate is not so extremely different as to lead to suspicion of the census counts. The result of the comparison is compatible with an accurate census both in Angola and in the other areas.

A second illustration is provided by data for Martinique and other island territories in the lesser Antilles. The population of Martinique, according to censuses, increased from 1910 to 1921 at an annual geometric rate of 2.66 per cent. Such a rate of population growth is high but is known to have occurred in some areas. Population changes during the same period in other islands of the region were as follows, according to censuses of 1911 and 1921:

	Per cent per annum
Barbados	-0.93
Leeward Islands	-0.90
Bahamas	-0.53
Turks and Caicos Islands	-0.17
Windward Islands	+0.42
Guadeloupe	+0.79
Trinidad and Tobago	+0.92

In some of the territories, population decreases were registered while in others the increases were only moderate. The decreases, or low rates of increase, may have been caused by the great influenza epidemic following the First World War, which took a heavy toll of lives in this region. The epidemic may have affected some islands more than others. Differences in population growth may also have arisen through migration, both between these islands, and from these islands to other regions. However that may be, the average rate of population change did not exceed one per cent (plus or minus) per annum in any territory except Martinique, where an exceedingly high rate of growth was observed. How could this be? Did Martinique completely escape the influenza epidemic? Was there a large immigration into Martinique? If there is no such explanation, the conclusion must be either that the 1910 census of Martinique was incomplete or that the 1921 figure was inflated.

In some cases it may not be helpful to compare the rate of population growth in the country in question with that of other countries. It may be that conditions are so unusual as to preclude any useful comparison. Or it may be that no census figures are available for other countries with approximately similar social and economic conditions. Even so, it is possible to set probable upper and lower limits for the average annual rate of increase.

If the country's population changes only through natural increase, it is very unlikely to have an average

<sup>1</sup> Geometric rate of increase, per annum.

annual rate of growth exceeding 3 per cent. This would be the result of a high birth rate (say 40 or more per 1,000) and a very low death rate (say 10 or less). On the other hand, it is only in unusual circumstances that the population would be likely to decline without heavy emigration. In fact, nearly all observed rates of natural increase in the various countries of the world in modern times have been in the range from zero to 3 per cent. If, in any given country, the rate of population change approaches or exceeds these limits without large-scale immigration or emigration, the question must be raised as to whether there is some explanation for such an unusual rate, or whether the census counts were in error.

An illustration can be found in data for Honduras. According to the census counts of 1926 and 1930, the population increased 5.8 per cent per year. Such a rate could not have been attained by the excess of births over deaths. Only an extraordinary immigration into the country could have caused it. If there was no such immigration, it is clear that the census figures are erratic. Possibly the earlier census was incomplete and the later one complete, or the first was correct and the second unduly inflated, or more likely, both were in error. The inconsistencies in census data for Honduras are treated in more detail in appendix C.

With some information, however approximate, regarding the conditions of mortality and fertility in the country, the limits of the rate of growth may be defined more closely. For example, if there is reason to believe, even without reliable vital statistics, that the birth rate is normally at least 30 per 1,000 (corresponding to moderately high fertility) and that the death rate does not normally exceed 20 per 1,000 (corresponding to fairly good control of major infectious diseases), then, in the absence of large-scale emigration, any average rate of growth below one per cent per annum shown by the censuses gives a strong presumption of error.

#### COMPARING RATES OF CHANGE DURING INTERVALS BETWEEN THREE OR MORE CENSUSES

If population counts are available for three or more successive censuses it becomes possible to make a more accurate evaluation by comparing the successive rates of growth. Again, the same principle is followed, namely, that the pattern of population growth should be regular except in so far as it can be shown that changes in the circumstances may have led to departures from the pattern.

An example of this test is afforded by Guatemalan census figures. The average annual rates of increase were as follows:

1880 to 1893.....	0.84
1893 to 1921.....	1.36
1921 to 1940.....	2.67
1940 to 1950.....	-1.63

The rate of increase in the period 1921 to 1940 was improbably high in comparison with the earlier period; furthermore, it approaches the upper limit of about 3 per cent per year, which was previously noted. Finally, the census counts show a decrease in population in the next decade, 1940 to 1950. The latter also seems improbable, since, as was mentioned previously, declines are unusual. These observations strongly suggest that the census count of 1940 was exaggerated; or, which is less probable, that both the 1921 and 1950 counts, and per-

haps also the 1893 enumeration, were grossly deficient. No evidence is available to show that there were such changes in social and other conditions, or such large migrations, as could have led to these diverse rates of change.

As a further check, the average annual rate of change for the period 1921 to 1950 can be calculated. This rate is 1.16 per cent. The fact that it is below the rate for 1893 to 1921 is difficult to explain except on the hypothesis that the census figures for 1893, 1921, or 1950, as well as 1940, are in error (some exaggeration in 1921, or some deficiency in 1893, 1950, or both).

Another example of this test is afforded by data for the Indian peninsula (the combined areas of India and Pakistan). The average annual rates of increase for the inter-censal periods were as follows:

	<i>Per cent</i>
1901 to 1911.....	0.65
1911 to 1921.....	0.09
1921 to 1931.....	1.02
1931 to 1941.....	1.41

Inspection of these rates suggests, at first glance, a serious under-count at the 1921 census. It is known, however, that the Indian peninsula suffered an enormous number of deaths as a result of the influenza epidemic following the First World War, and this is a sufficient explanation for the abnormally low rate of population increase during the decade 1911 to 1921.

#### POPULATION TOTALS FOR SUBDIVISIONS OF A COUNTRY

The type of tests previously described, for analysing the changes in a nation's total population, can be applied in principle to the analysis of changes in the various subdivisions. Here too, we expect to find "orderly" patterns of population change, both within the same subdivision in successive inter-censal periods, and among different subdivisions in any period. Any dissimilarities should be explainable in terms of known conditions. As a practical matter it is well known that there is considerable diversity in the rates of population change among the various parts of any nation. Accordingly, the problem becomes one of trying to distinguish between changes which are explainable in terms other than errors in the statistics and those which are not. It should be noted that although these procedures may reveal the presence of errors and in some cases indicate their order of magnitude, they provide no basis for exact estimates of the size of the errors. Definite estimates must be made by other methods, some of which are described in chapters II and III.

#### COMPARING SUCCESSIVE CENSUSES

With this method alone, final evaluation of one census count on the basis of preceding counts cannot ordinarily be made until one or more subsequent enumerations are available for purposes of comparative analysis. Preliminary analysis can be made, however, without waiting for subsequent counts. Actually, it is a part of normal census procedure, as soon as an area has been enumerated, to calculate the percentage change in its population since the previous census date and compare it with the changes in surrounding areas and during previous inter-censal periods. Any unusual rate of change should be questioned immediately, and if

necessary, the areas should be re-enumerated, either on a complete or sample basis.

Let us turn now to an example of how the rates of change can be analysed for several successive censuses. In the United States there was serious under-counting in the Southern States at the time of the 1870 census. This deficiency became evident when the percentage increase in population for the decade 1860 to 1870 was compared with the changes during the preceding and following decades. The percentage increases in population for the South as compared with the remainder of the country were as follows:

Period	South	Remainder of United States	Total United States
1850 to 1860.....	23.2	43.0	35.4
1860 to 1870.....	10.8	29.6	23.0
1870 to 1880.....	34.1	28.0	30.0
1880 to 1890.....	21.2	27.3	25.3

The rate of increase in the South in the decade 1860 to 1870, which encompassed the Civil War and the ensuing reconstruction period, seems much too low in comparison with the preceding decade. The rate of increase for the period 1870 to 1880 seems abnormally high in comparison with the preceding and following decades. The population changes in the remainder of the United States seem much smoother, in contrast. The differences are of such magnitude that, even with no other information, it would appear highly probable that the enumeration of 1870 in the South was seriously incomplete. Taken together with the knowledge that conditions in the South in 1870, during the aftermath of the Civil War, were not favourable to accurate census work, the figures give convincing evidence of a major error. This under-counting was not recognized as an established fact, however, until after the 1890 census had been taken.<sup>8</sup>

#### COMPARING POPULATION CHANGES WITH SOCIAL AND ECONOMIC CONDITIONS

Procedures for evaluating diverse rates of change in terms of differences in the social and economic conditions in a country can be illustrated with selected census data for Puerto Rico. No attempt is being made here to evaluate the censuses of 1940 and 1950 for the whole island; rather examples of a few selected municipalities have been chosen to show how such comparisons can be made and conclusions drawn regarding the probable completeness of the census counts.

The San Juan metropolitan area (see map on page viii) increased in population by about 62 per cent in the decade 1940 to 1950, as compared with a rate of growth of 18.3 per cent for the entire island. Such a large increase in this area is consistent with the economic and other developments which occurred there since 1940. These included: (a) about a doubling in the volume of manufacturing activity (as measured by the increase in the number of production workers); (b) substantial increases in commercial and transportation activities, both directly and indirectly related to the growth of manufacturing; (c) substantial increases in governmental activities and numbers of employees; this increase was in part related to the needs of the

Second World War and in part came about when Puerto Rico achieved self-government and the types and amounts of government services were greatly expanded.

Another factor which was probably important was the building of the new highway along the northern coast during this decade (see map). As a result the San Juan area became a focal point for a much larger volume of traffic. Furthermore increased employment opportunities became available in connexion with the road—in its construction and maintenance and in the operation of roadside garages and refreshment stands et cetera. Even more important as a source of employment and population support was the fact that this new road made it possible for commercial establishments and factories to be located near the San Juan port, but on the outskirts of the metropolitan area. In the absence of improved transportation some of these establishments probably would never have been opened. Other new roads were put in south of the city of San Juan—into Rio Piedras, Guaynabo and Bayamon, and eastward into the northern part of Carolina. At present the entire metropolitan area is served by excellent transportation facilities, much superior to those available in 1940. Catano, across the harbour from San Juan, has excellent ferry service, and now contains large residential areas.

Proceeding westward from metropolitan San Juan, the census figures show that Toa Baja and Dorado both grew substantially, by 38 and 24 per cent, respectively. Both of these municipalities probably benefited from the developments in the San Juan area. Distances are short enough and transportation facilities good enough so that a person may live near the main highway in either of these municipalities and have his employment in metropolitan San Juan; from Bayamon to Dorado, for example is hardly more than fifteen minutes by private or public automotive transportation, the total distance being hardly ten miles. That the highway is an important factor influencing population growth in these two municipalities is suggested by the fact that those *barrios* through which the main highway passes, grew the most rapidly, as follows:

	Per cent increase in population, 1940-1950
Toa Baja:	
Total .....	38.1
Highway area .....	41.6
Remainder .....	36.6
Dorado:	
Total .....	23.9
Highway area .....	34.4
Remainder .....	21.6

Continuing westward along the northern coast the next three municipalities and their rates of population growth in the decade 1940 to 1950 are:

	Per cent
Vega Alta .....	15.3
Vega Baja .....	25.2
Manati .....	3.7

In all three of these municipalities those parts through which the main highway passes grew more rapidly than the other parts. All three municipalities are in part within the zone of influence of the San Juan metropolitan area; numbers of people travel to work regularly between Manati and the metropolitan area. It can then be asked why were the differences in population growth so large, and is there a possibility of an under-

<sup>8</sup> 1950 U.S. Census of Population, U.S. Summary, Bulletin P-C 1, p. xxviii, Washington, 1953.

count in the 1950 census in Vega Alta, and more particularly in Manati? Both areas produce sugar and pineapples, and neither has much manufacturing. In both areas there were very significant increases in the production of both these crops during the decade. Presumably, then, both areas experienced increases in employment opportunities and should have grown at about the same rate. Why they grew at such different rates, the present analysis does not reveal. Further investigation, including consideration of earlier census data, the returns for smaller subdivisions of these municipalities, and any available information regarding the performance of enumerators in this area in both 1940 and 1950 is required.

Another municipality which had a relatively high rate of population growth during the decade, was Ponce on the southern coast. This grew by 20.6 per cent. This growth seems possible since there was considerable growth in manufacturing, and large increases in sugar and coffee production. Furthermore, Ponce is advantageously situated with respect to transportation, having both main highways and a seaport.

Guanica, to the west of Ponce, increased in population by 23.2 per cent during this decade. No reason for such an increase is readily apparent. This municipality is largely dependent on raising and processing sugar cane; there was but a moderate increase in the volume of such activity, however, and not enough to account for the growth in population. Further analysis of the possibility that either the 1940 or 1950 census counts may have been in error, is indicated.

Turning now to municipalities which lost population, or gained but very little, it is found that the majority are in the interior mountainous areas where the main economic activities consist of raising coffee and/or tobacco. This group includes: Maricao, Lares, Adjuntas, Jayuya, Ciales, Morovis, Barranquitas, Comerio and Cidra. The development of these activities has not been sufficient to support large population increases; by 1940 or earlier these municipalities already had enough people to supply all the labour needed for these two crops. Furthermore, these municipalities are all poorly supplied with roads. Of this group, Adjuntas is the only one through which a main highway passes; however, since the area is mountainous the highway is a difficult one to travel. This lack of transportation, in turn, is probably one of the factors contributing to the lack of population growth and, hence, to the lack of population growth.

There is no apparent explanation for the low rate of population growth in Yabucoa, in the southeast corner (see map). This municipality's economy is based largely on sugar and tobacco; the production of both these crops increased considerably during the decade. Yabucoa is also advantageously situated with respect to highway transportation.

This procedure of examining each municipality in considerable detail and relating the changes in population size as shown by the census enumerations to social and economic changes can be continued indefinitely. Data on all the various types of agricultural production, on the opening of new factories, on the building of new roads and power stations and dams et cetera, together with vital statistics, the tabulations of the census data on population characteristics, the evidence of the census

schedules, and the results of earlier censuses, can all be introduced into the final evaluation. It is apparent that this procedure calls for detailed knowledge of the country and its parts as well as of the census materials and cognate statistics, if a proper evaluation of the adequacy of census returns is to be made.

It should be understood in making such comparisons that internal migration plays an important part in bringing about the differences in rates of population growth in the subdivisions of an area such as Puerto Rico. The expansion of economic opportunity in the form of manufacturing, cane-sugar production, coffee raising, dairy farming, trade et cetera, was uneven in different parts of the island. As job openings appeared in a specific area, workers and their families migrated into the area, and the inter-censal rate of population growth increased. Differences in natural increase, the excess of births over deaths, may also be an important factor, in some circumstances, producing regional variations in the rates of growth within a country, but natural increase is unlikely to respond so greatly and so quickly to changes in economic opportunity as to produce such large differences in inter-censal rates of change as are observed in Puerto Rico.

#### D. Consistency of census totals with vital statistics and migration statistics

##### THE BALANCING EQUATION

The *balancing equation*, which is to be considered in this section, is a procedure which, if properly used, permits a more precise quantitative evaluation of the accuracy of a census count than can usually be made by such methods as described in the preceding sections. This procedure, employing vital and migration statistics, brings much more evidence to bear and provides a much more definitive and quantitative answer on just how accurate and complete a given census enumeration probably is. In this section only the completeness of the head count is considered; in subsequent chapters the accuracy of data on age and sex composition and of the vital and migration statistics will also be included. However, in practice a final determination of the completeness and accuracy of a census count, using the balancing equation, cannot be made until all of the components—population by age and sex, births, deaths and migration—have been tested.

##### DEFINITION AND LIMITATIONS

The principle of the balancing equation is essentially very simple. In any interval of time, such as between two censuses, the population of a country can increase or decrease only as a result of births and deaths and movements across the country's boundaries. Births and immigration add to the population, and deaths and emigration subtract from it. Accordingly, if data are available from two censuses, and the numbers of births, deaths and in- and out-migrants are known, then the equation must balance exactly, if all the data are perfectly accurate. The equation is very simple:

the population at the second census ( $P_1$ ) equals:

the population at the first census ( $P_0$ )

plus the number of births during the inter-censal period (B)

plus the number of immigrants in the inter-censal period (I)  
 minus the number of deaths during the inter-censal period (D)  
 minus the number of emigrants during the inter-censal period (E)

$$P_1 = P_0 + B + I - D - E$$

Completely accurate census, vital and migration statistics have never been attained in any country, and never will be. Accordingly, this equation never balances out exactly; some unaccounted-for residual always remains. If the residual is large, at least one of the components contains a sizeable error, and further investigation has to be conducted in an effort to determine where the error lies. Sometimes the major part of the error may be found in one set of data; generally it is found that there are errors of varying size in all of the components of the equation. By testing each of the components separately by means of the balancing equation, the approximate size of the error in each of the components can be determined.

Even if the residual is small, however, there is no definite proof, until all of the components have been separately evaluated, that the censuses, vital and migration statistics are correct. This is because of *compensating errors*. For example, under-registration of births may be compensated for by under-registration of deaths; incompleteness of enumeration at one census may be compensated for by incompleteness at the other; incompleteness of census enumeration may be compensated for by under-registration of births or deaths, or by incomplete recording of migrants.

One further caution is required. Compensating errors may sometimes be of such a nature that, even after studying the various components as thoroughly as possible by means of the balancing equation and finding that the data are all in good agreement, it may still not be justified to conclude that they are all reasonably correct in absolute terms. For example, the application of various tests may make it appear that the birth statistics are consistent with the census enumeration of young children and therefore probably correct, that the migration statistics tally with the census returns on the foreign-born population and thus seem to be correct, and that the entire equation balances with but a very small residual, implying that the death statistics and the census enumerations were also correct. Yet, the censuses, vital statistics and migration data may all fail to cover some part of the population—for example, persons with no settled abode, or primitive tribes, possibly not under effective administration, or the population living in a frontier region which is difficult of access, or an outcast or criminal segment which defies census enumeration and reports none of its activities to the authorities. Generally, however, such groups constitute but a small proportion of any country's population.

#### TREATMENT OF MIGRATION

The element of migration in the balancing equation often presents a difficulty because many countries do not have comprehensive and reliable statistics on the flow of migrants across their borders, and migratory movements within the country are even less likely to be

adequately recorded. However, even though such statistics are lacking or seriously defective, the available sources of information will be sufficient as a rule to show whether or not migration has taken place on a large enough scale to have a significant effect. If not, the balancing equation may usefully be applied without regard to migration.

It is primarily in the case of statistics for subdivisions of countries that the lack of adequate migration statistics interferes with the use of balancing equations. In most countries, the volume of internal movements is so large, at least for the majority of their subdivisions, that in the absence of a proper measure of it the balancing equation is inapplicable.

#### EXAMPLES OF THE USE OF THE BALANCING EQUATION

The procedures and examples used in this chapter are dictated both by the purpose and by the nature of the available data. The purpose is to evaluate the total count of two successive censuses by means of the balancing equation; evaluation of the vital and migration data and evaluation by age and sex are reserved for subsequent chapters. Because the results of the balancing equation in the form used here cannot be interpreted until the remaining statistics are also evaluated separately, no definite conclusions can be drawn in these examples regarding the accuracy of any particular components. Some tentative inferences can be made, however, in certain cases.

Since migration statistics for all countries are not available, examples of the use of the balancing equation are presented for some countries having such data, and for some not having them. For all of the examples shown, the equations were computed without migration statistics; and for those countries having such data, additional computations were made which included the migration balance.

The specific procedures employed are as follows:

- (1) By subtraction of successive census totals, the apparent inter-censal increase is determined.
- (2) The apparent natural increase—or excess of births over deaths—for the inter-censal period is then determined on the basis of the reported birth and death statistics. (Where necessary, adjustments for a fraction of a year are made.)
- (3) Subtraction of the apparent natural increase from the apparent inter-censal increase provides a measure of the *discrepancy*.

For countries with migration statistics, the following additional steps are involved:

- (4) The migratory balance is computed for the inter-censal period by subtracting recorded departures from recorded arrivals.
- (5) This migratory balance is added to the apparent natural increase to determine that population increase which can be accounted for by means of available vital and migration statistics.
- (6) *The residual* is obtained by subtracting the figure obtained in step 5, from the apparent inter-censal increase obtained in step 1.

The above steps are illustrated in detail for the following areas:

### 1. Puerto Rico

Step (1)	Population according to census of 1 April, 1950 .....	2,211,000
	Population according to census of 1 April, 1940 .....	1,869,000
	Apparent inter-censal increase.....	342,000
Step (2)	Number of births reported in inter-censal period .....	830,000
	Number of deaths reported in inter-censal period .....	289,000
	Apparent natural increase.....	541,000
Step (3)	Apparent inter-censal increase.....	342,000
	Apparent natural increase.....	541,000
	Discrepancy .....	-199,000
Step (4)	Number of recorded arrivals in inter-censal period.....	627,000
	Number of recorded departures in inter-censal period.....	808,000
	Migratory balance.....	-181,000
Step (5)	Apparent natural increase.....	541,000
	Migratory balance.....	-181,000
	Population increase accounted for.....	360,000
Step (6)	Apparent inter-censal increase.....	342,000
	Population increase accounted for.....	360,000
	Residual .....	-18,000

The residual of 18,000, or the amount by which the equation fails to balance, is relatively small in this case, and amounts to only 0.8 per cent of the 1950 population as recorded in the census enumeration. It should be noted that some of this residual is explained by the fact that deaths to members of the armed forces are excluded from the reported death statistics. Hence, the amount of natural increase as recorded in step 2 should be somewhat smaller than the figure shown. This, in turn, would decrease the size of the "Population increase accounted for" in step 5, and lead to a smaller residual in step 6.

The residual, expressed as a percentage of the recorded inter-censal increase, amounts to 5.3 per cent. As will be seen from the other examples to be presented, the equation for Puerto Rico balances out fairly well—better than in some countries but not as well as in some others. Without further analysis of the components of the balancing equation, and without applying the balancing equation to specific age-sex cohorts, it can be concluded that the balancing equation gives no evidence of major errors.

### 2. Thailand

Censuses were conducted in 1937 and 1947. Vital statistics are available for these years, together with an estimate of the net immigration. The balancing equation for the period 1937 to 1947 then is:

Census count, 1947.....	17,443,000
Census count, 1937.....	14,464,000
Apparent inter-censal increase.....	2,979,000
Recorded natural increase in inter-censal period....	2,569,000
Migratory balance.....	100,000
Population increase accounted for.....	2,669,000
Residual (excess of census increase):	
Number .....	310,000
Per cent of inter-censal increase.....	10.4
Per cent of 1947 population.....	1.8

The relatively small residual suggests that the components of the equation are not subject to very large errors; in anticipation of material to be presented in the chapter on vital statistics, however, it should be pointed out that there were actually rather large compensating errors in the reports of births and deaths. Furthermore, the data on migratory balance are meagre and unsatisfying; the

figure of 100,000 presented above for net in-movement is an estimate pieced together from incomplete records.

### 3. Dominican Republic

Censuses were conducted in 1920, 1935 and 1950. A complete series of vital statistics covering these 30 years was not found; birth records were missing for the years 1920 to 1924, 1929, and 1931 to 1934, while death records were missing for the years 1920 to 1923 and 1929 to 1934. Data for these missing years were estimated. No statistics on migration were found. The balancing equation for the period 1920 to 1935 is:

Census count, 1935.....	1,479,417
Census count, 1920.....	894,665
Apparent inter-censal increase.....	584,752
Recorded natural increase in inter-censal period....	426,384
Discrepancy:	
Number .....	158,368
Per cent of inter-censal increase.....	27.1
Per cent of 1935 census count.....	10.7

The balancing equation for the period 1935 to 1950 is:

Census count, 1950.....	2,135,872
Census count, 1935.....	1,479,417
Apparent inter-censal increase.....	656,455
Recorded natural increase in inter-censal period....	706,442
Discrepancy:	
Number .....	-49,987
Per cent of inter-censal increase.....	-7.6
Per cent of 1950 census count.....	-2.3

Any migratory balance during either of these two periods was probably small in comparison with the discrepancies. In order to determine whether the discrepancies are due chiefly to errors in the census enumerations or the vital statistics, it is necessary to test the accuracy of the latter.

### 4. Fiji Islands

Censuses were taken in 1936 and 1946, and vital statistics were reported for the entire decade. Data on migration, however, were found only for the years 1936 to 1938 and 1945. The balancing equation for this decade is:

Census count, 1946.....	259,638
Census count, 1936.....	198,379
Apparent inter-censal increase.....	61,259
Recorded natural increase in inter-censal period....	58,578
Migratory balance.....	2,007
Population increase accounted for.....	60,585
Residual (excess of census increase)	
Number .....	674
Per cent of inter-censal increase.....	1.2
Per cent of 1946 census count.....	0.3

Since complete data on the migratory balance were not found, it is difficult to determine the probable degree of accuracy of the statistics. The residual is so small, however, that it does not seem likely that the censuses contained any very large errors. Even with an error of several thousand in the migratory balance, the equation would still balance out fairly well. To be sure, compensating errors may be present.

### 5. Mexico

Although a number of censuses have been taken in Mexico, the equation will be applied only to the period 1940 to 1950. The result is as follows:

Census count, 1950 (preliminary figure).....	25,781,000
Census count, 1940.....	19,654,000
Apparent inter-censal increase.....	6,127,000
Recorded natural increase in inter-censal period....	5,770,323



<b>Discrepancy:</b>	357,000
Number .....	5.8
Per cent of inter-censal increase.....	1.4
Per cent of 1950 population.....	— 122,000
Recorded migratory balance.....	5,648,000
Population increase accounted for.....	
<b>Residual:</b>	479,000
Number .....	7.8
Per cent of inter-censal increase.....	1.9
Per cent of 1950 population.....	

The discrepancy (if migration statistics are not taken into account) is not very large in proportion to the population. If the recorded migratory balance is included in the computations, a somewhat larger residual is obtained. Whether or not the migration statistics are accurate, it is very probable that during this period more persons left the country than entered it. The analysis shown here does not yield any clues as to whether the residual should be mainly attributed to defects in the vital statistics or to differences in the completeness of the two census enumerations. However, further tests of the Mexican statistics (to be presented in a later chapter) suggest that the statistics of births and deaths are fairly accurate, so that probably no very large residual would have resulted from inaccuracies in the vital statistics.

## 6. Egypt

Although several censuses have been taken in this country, the balancing equation will be applied only to the period 1937 to 1947. For this period vital statistics are available, but no migration data were found. The balancing equation is:

Census count, 1947.....	18,967,000
Census count, 1937.....	15,921,000
Apparent inter-censal increase.....	3,046,000
Recorded natural increase in inter-censal period....	2,543,000
<b>Discrepancy:</b>	503,000
Number .....	16.5
Per cent of inter-censal increase.....	2.7
Per cent of 1947 population.....	

In the absence of migration data, the balancing equation is incomplete. However, unless there were a net in-migration of some 500,000 persons, the equation would not balance. For fuller investigation of the data for this country, it would be necessary to obtain some estimates of the probable volume and direction of migration, and to analyse the vital statistics component of the balancing equation separately. Furthermore, the balancing equation should be applied separately to the various age-sex groups as explained in chapter III. It would then be possible to reach a conclusion as to probable accuracy of the census counts.

## E. Direct checks on the accuracy of census totals

Direct checks consist primarily in re-enumeration of at least a part of the population, in such a way as to provide a measure of the numbers of persons who were omitted from or erroneously included in the census count. The list of persons recorded in the re-enumeration is compared with the record of the census, and any discrepancies are investigated in order to determine whether or not the persons concerned should have been included in the census. Some of the ways in which such a check can be made are described below. The examples presented here refer to entire countries, but the same

procedures are evidently applicable to particular subdivisions of a country, and it may be efficient, in some instances, to limit their use to those subdivisions in which errors are suspected on the basis of other tests.

An important consideration, and one reason why such checks have not often been made, is the question of cost. The cost of a properly planned check, however, can be very modest in comparison with the cost of the census itself, and it can greatly enhance the value of the census results, either by confirming their accuracy or by providing the basis for a correction.

## TIMING OF A DIRECT CHECK

Because of the necessity of name-by-name checking, direct checks are carried out most efficiently directly after a census enumeration. The mechanics of checking the names obtained by the re-questioning with the names on the census schedule become quite complicated and laborious, particularly if people migrate. Ideally, immediately after a small area has been enumerated—as for example one block in a city, or one small village—the procedures for a direct check should be inaugurated in that area; thus the problem of searching the census records for the persons found in the check enumeration is greatly facilitated. The longer the time allowed to lapse before the check is instituted, the more the problems of identifying persons and deciding who should have been counted is complicated by births and deaths, by changes in address, by changes of name on the part of women marrying et cetera.

## SELECTION OF AREAS TO BE RE-SURVEYED

There is no need to re-enumerate the entire population of the country in order to check the census count. Depending on the precise objective of the check, the re-enumeration can be carried out for a selected sample of the entire country, for one or several special areas of the country, or for samples of such special areas. Any combination of these procedures can also be made.

Re-enumeration of a sample of the entire country is desirable if the chief objective is to establish some measure of the over-all accuracy of the census count. Scientific sampling techniques make it possible to select a relatively small sample in such a manner that the most representative results are obtained at smallest cost. A sample of the entire country is also sufficient if there is no special reason to suspect that the census enumeration may have been particularly inaccurate in some special part of the country, or among some special segment of its population.

The various tests described in sections A and B of this chapter are relevant to determining whether or not the census results for particular areas should be questioned. By these or other tests, areas may be selected for re-enumeration. Because of the desirability of conducting checks as soon after the census as possible, it is important that census results should be analysed from this point of view within the shortest possible time.

It should be noted that direct checks conducted in selected areas only, whether on a sample basis or by complete re-enumeration, do not provide an accurate measure of the accuracy of the census for the entire country. In the "suspected" areas, the relative magnitude of errors in the census is likely to be greater than

in areas not so suspected. A combination of intensive checks in suspected areas with checks of small samples for the remainder of the country is perhaps the most effective method of checking over-all accuracy.

#### COMPARING CENSUS RECORDS WITH LISTS OF NAMES

A method of direct checking which does not require re-enumeration and which can be undertaken even after a considerable lapse of time is the matching of census records with other lists of persons who were living in the area at the time of the census, or with samples from such lists. The kinds of lists that may be used for this purpose include the names of pupils enrolled in school, lists of taxpayers, registers of voters et cetera. The comparison may be made at any time after the census, even if it is already too late to proceed to a sample re-enumeration. It is important, however, to make every effort to ascertain that the persons on such lists actually lived in the specified area at the time of the census.

This method of checking, though it may be less expensive, is normally less satisfactory than re-enumeration, for a variety of reasons. It is not always possible to be certain that the persons named on the lists were actually living in the specified area at the time. Such lists usually refer only to portions of the population living in the area; for example, persons who were not on the taxpayers' list are not represented. The extent of coverage of the list may not always be known; for example, the list of schoolchildren may include only a part of all the children enrolled in school. A list of names generally does not provide any information about the families of these individuals, nor very much information about their personal characteristics. A re-enumeration, on the other hand, attempts to locate all persons (or a correct sample of all persons) who were in the area at the time of the census enumeration; in addition to verification of the head count, it also provides an opportunity to determine whether their personal characteristics (age, occupation et cetera) have been correctly reported.

For these reasons, it is generally preferable to check the accuracy of a census by sample re-enumeration rather than by reference to existing lists of persons, unless too much time has elapsed to make such re-enumeration still useful for the purpose of checking.

#### EXAMPLE OF PARTIAL RE-ENUMERATION IN COSTA RICA

At the time of the 1950 Costa Rican census, efforts were made to resurvey those parts of the country where the enumeration was deemed suspect, for one reason or another. Included in this group were those areas which lacked good maps, and those for which there were any doubts as to the quality of the returns. For census purposes Costa Rica was divided into 316 districts; a post-enumeration check was made in 135 of these. This check was not carried out uniformly throughout the country; the Province of Guanacaste was resurveyed in its entirety whereas in the Province of Cartago only 9 of the 39 districts were revisited, and in Limón, only 2 of the 10 districts.

There were 6,878 persons located in the post-enumeration check who had not been recorded at the census. These comprised 0.86 per cent of the total population of the country. It was calculated that if the entire

country had been resurveyed the number of omissions which would have been uncovered would have amounted to but 2.01 per cent of the total population. This number was so small that no efforts were made to revise the census count to include an estimate of omitted persons in the country as a whole. Instead the final census count includes only the 793,997 persons originally enumerated plus the 6,878 omissions actually discovered in the post-enumeration check, for a total of 800,875. For all of these people the required data on personal and family characteristics were available so that they could be included in the detailed tabulations; it was not necessary to adjust the tabulations so as to include estimates of other persons presumably omitted, classified by their characteristics.<sup>9</sup>

#### EXAMPLE OF THE POST-ENUMERATION SURVEY IN THE UNITED STATES

Following the 1950 census, the United States Census Bureau undertook a re-enumeration on a sample basis. This sample was designed to provide an estimate of the number of households completely missed, and the number of individuals within enumerated households who had been missed or enumerated erroneously. A sample of 3,500 small areas representing the entire United States was carefully chosen, and re-enumerated meticulously by interviewers who were especially selected and given more intensive training than had been received by the average census enumerator. In these areas the enumerators obtained the addresses of all households that could be located and the names and characteristics of their members. These data were checked against the original census schedules for those 3,500 areas. The post-enumeration check covered a total of 22,000 households.

The check indicated a net under-enumeration of about 1.4 per cent in the original 1950 census count. This was the net result of erroneous omissions and erroneous inclusions, as follows:

erroneously omitted,	3,400,000 persons, or 2.3 per cent of the census count
erroneously included,	1,309,000 persons, or 0.9 per cent of the census count
net error.....	2,091,000 persons, or 1.4 per cent of the census count

The great majority of those missed were the members of households that were entirely omitted:

erroneously omitted.....	3,400,000 persons
in missed households.....	2,416,000 persons
in enumerated households.....	984,000 persons

Of the persons erroneously included, the great majority should have been enumerated in another enumeration district:

erroneously included.....	1,309,000 persons
should not have been enumerated anywhere .....	198,000 persons
should have been enumerated in another enumeration district.....	1,111,000 persons

Even this elaborate post-enumeration check, however, probably failed to obtain a complete estimate of the number and proportion of persons missed, since the

<sup>9</sup> Costa Rica, Dirección General de Estadística y Censos, *Censo de Población de Costa Rica*, San José, 1953, pp. 2, 3.



check itself contains some errors. A small-scale field check was carried out following the post-enumeration check, and indicated that the original post-enumeration check contained errors which were, in general, in the direction of underestimating the number of erroneously omitted persons. Some of the errors found in the original post-enumerating check were of the type commonly found in any census enumeration. For example, in some cases, it was not possible to locate a person thought to be living in a given household; in other cases, although the person was located, no interview could be obtained. Infants, who are always under-enumerated in the census, were also under-enumerated in the original post-enumeration check, as other data indicated. Although the magnitude of the errors in the check is largely unknown, those errors which could be identified were all in the direction of under-estimating the number of persons erroneously omitted. Altogether, these various checks indicate that the estimated net deficiency of 1.4 per cent in the original census enumeration is the minimum.<sup>10</sup>

In short, the United States experience indicates that, as repeated checks are employed, the size of the errors in the enumeration is progressively reduced, but that perfection is never attained. Some people are always missed no matter how carefully a census enumeration is conducted. Since any post-enumeration check survey is somewhat expensive, it is evident that a point is eventually reached at which additional cost of an additional improvement in accuracy exceeds the value of the improvement.

#### SAMPLE VERIFICATION OF THE CENSUS COUNT IN INDIA

After the Indian census of 1 March 1951, the Government of India "requested all State Governments to carry out a sample verification of the 1951 Census Count according to a scheme framed by the Registrar Gen-

eral."<sup>11</sup> The re-enumeration of samples of the population was carried out by most Indian States, in the majority of cases during the latter part of the year 1951. In general, the scheme provided for re-enumeration of one-tenth of one per cent of occupied houses in every census tract, by verification officers specially appointed from among qualified civil officials. The purpose of the operation was to discover, by visits to carefully selected households, any cases of "clear omissions", "fictitious entry", or "erroneous count of visitors and absentees", as well as any evidence of omission of entire households in the vicinity of households selected. The selection of households was to be made from census lists and had to conform rigidly to specified procedures.

The details of re-enumeration procedures were not uniform in every State. Enumerators were given special training and their attention was directed particularly to the kinds of census errors which they might be expected to encounter. In at least one State (Mysore), the verification officers were also informed that the lists given to them, which had been copied from census returns, contained some "ghost" entries, that is to say, entries of fictitious persons purposely added to the list and made to look entirely plausible. These "ghosts" were added to the lists in order to increase the vigilance of the verification officers, and most "ghosts" were indeed recognized as such when the field visits were completed.

The area in which the verification was carried out contained 81 per cent of India's population. Within this area, clear evidence was found of omissions amounting to 1.3 per cent of the population, and of faulty inclusion of persons amounting to 0.4 per cent of the population, in addition to various minor errors consisting of omission or double counting of temporary visitors and absentees. As a net result of all these errors, the census count appeared to have fallen short of the truth by 1.1 per cent for all of India (1.0 per cent in rural, and 1.4 per cent in urban areas). The percentage varied among the several regions.

<sup>10</sup> United States, Bureau of the Census, 1950 Census of Population, *United States Summary*, Bulletin P-C 1, pp. xxix, ff., Washington, 1953.

<sup>11</sup> India, Registrar General, *Sample Verification of the 1951 Census Count* (Census of India, Paper No. 1, 1953), New Delhi, 1953, p. 3.

## CHAPTER II. THE COMPLETENESS OF VITAL STATISTICS

### A. Introduction

The term "vital statistics" is often used to include births, deaths, marriages and divorces. For the purposes of population estimates, it is mainly the data on births and deaths that are important. Accordingly, in this chapter methods for appraising births and deaths only will be considered. Some of the methods to be described here are adaptable, however, to tests of the completeness of marriage and divorce reporting.

#### EFFECTS OF ERRORS IN VITAL STATISTICS ON POPULATION ESTIMATES

There is an important difference between errors in vital statistics and errors in census enumerations, as regards their effect on population estimates. The difference can be illustrated by referring to those annual estimates of population for the years following a census which are made by adding to the census total the recorded annual excesses of births over deaths. If the census figure is in error—due to under-enumeration, for example—the annual estimates will reflect the same error, but the error will be a constant. The whole series of estimates will be comparable with the census and the picture of population growth since the date of the census will not be invalidated. Errors in the vital statistics, on the other hand, may have a cumulative effect. For example, if there is an under-registration of births every year, so that the recorded annual natural increase is too small, the consequent error in the population estimates will grow larger in each successive year. By the end of a ten-year period, both the estimated size of the population and the estimated amount of its growth since the census date may be considerably below the truth, even though the amount of error in the birth statistics for any one year is not very great.<sup>1</sup>

It is also important to note that the methods of collecting vital statistics are essentially different from those involved in collecting population data by means of a census. A census is taken at infrequent intervals, generally once in ten years. The data are usually collected by a staff of enumerators especially selected for the purpose, who visit the households and obtain the information by direct interview. Vital statistics, on the other hand, are usually recorded continuously by officials who carry this responsibility over a period of years; these officials are not ordinarily expected to visit the people in their areas and make direct inquiries, but only to receive and record the reports of vital events which, by law, are to be brought to their attention. The fact that vital statistics are collected continuously is an advantage in appraising them, since certain tests can be applied which are not possible for a one-time operation such as the census. On the other hand, per-

haps the main problem in obtaining complete reporting of vital events is the fact that some one other than the registrar must make a positive effort to report the event.

#### COVERAGE

In many countries, registration of births and deaths is not compulsory, or is compulsory for only a part of the population (for example, for "Europeans" in certain African territories), or for certain areas within the country. Obviously, in such cases registration is not complete for the entire country even though it may be nearly so for those parts of the country or the population to which it applies. If the areas or segments not covered by registration are relatively small, then it is possible to estimate how many births and deaths would have been reported if registration were applied to the entire country. Such estimates are sometimes needed if a balancing equation is made. It is then necessary to distinguish between the terms "completeness" and "coverage"; the former means the extent to which all births and deaths are reported where the registration system applies, whereas the latter term means the extent to which the registration system applies to the entire population.

In some areas there is a system of "voluntary" registration. Such a system results in the registration of only a fraction of all births and deaths. There is no known instance of voluntary registration of vital events that has even remotely approached completeness.

#### TIME AND PLACE OF OCCURRENCE

When vital statistics are used for making population estimates, or for the construction of a balancing equation to test the accuracy of demographic data, it is important that they should reflect accurately the numbers of births and deaths occurring in a given time interval. Many countries tabulate and publish their vital statistics according to the time of registration rather than according to the time in which the event occurred. Where there is a very short time lag between time of occurrence and time of registration the effect on annual totals is negligible; but in some countries a lag of two years is not uncommon.<sup>2</sup> Where the number of events is changing significantly from year to year or where an appreciable time elapsed between occurrence and registration, the results may be misleading as to the number of vital events which actually occurred during the specified year. This consideration is particularly important in relation to statistics of births; deaths are ordinarily more promptly reported. Therefore, as a first step in the proper appraisal of vital statistics where tabulations are by date of registration

<sup>1</sup> United Nations, *Manuals on methods of estimating population. Manual I: Methods of Estimating Total Population for Current Dates*, Population Studies No. 10, pp. 38, 39.

<sup>2</sup> See for example, Ricardo, Jimenez J., *Comparative Study of the 1950 Population Census Results and the 1950 Population Estimate of Costa Rica*, San José, Costa Rica, December 1951.

instead of date of occurrence, the factor of delayed registration should be examined.

Vital statistics are often tabulated according to the areas in which the events occurred rather than the areas in which the persons concerned were residents. Very often hospitals serve persons living in a wide radius outside the city in which the hospital is located. Many births and deaths occur in such hospitals, to residents of areas outside the city in which the hospital is located. On the other hand, when vital statistics are used as materials for population estimates, which are usually on a *de jure* basis, that is they refer to the population *resident* in a given area, it is desirable that they should reflect the births and deaths to residents of the areas for which estimates are made. For this purpose, tabulations are required, in which the births and deaths of non-residents are re-allocated to the areas where they reside. The difference between tabulations by place of occurrence and place of residence is especially important where calculations are being made separately for urban and rural areas. In the statistics for an entire country, of course, place of occurrence and place of usual residence will almost always coincide, but the same may not be true of data for component regions.

#### CHARACTERISTICS RECORDED ON THE CERTIFICATES

Almost universally when births are registered the sex of the infant and the age of the mother are recorded, and when deaths are registered the sex and age of the deceased are noted. Information on various other personal characteristics is often also contained in birth and death records. For the purposes of population estimates and tests of the accuracy of vital statistics and census counts, reliable data on sex and age are of particular importance.

Sex is probably very rarely mis-stated, either at the census or in vital registers. Mis-statement of sex is perhaps most frequent in the case of birth records, but even in this case is probably rare. If the ratio of reported numbers of male and female births is unusual, the reason is probably not so much mis-statement of sex as a less complete reporting for one sex than for the other.

The ages of parents of a new-born child, or the age at death of a deceased person, may be frequently mis-stated. Considerations relating to the accuracy of age statements are presented in chapter III. If age mis-statements are similar in the vital records and census enumerations, then the censuses and vital statistics are still comparable. It is sometimes believed that ages are more accurately reported in the registration of vital events than in censuses. If the frequency and nature of age mis-statements in vital records differ greatly from those at the census, then the two sets of statistics are to that extent inconsistent and caution is necessary if they are used jointly.

It is thought that other characteristics, such as occupation, for example, may be returned more correctly on the census schedule than on the vital certificate.

#### FACTORS AFFECTING THE COMPLETENESS OF BIRTH REGISTRATION

As already pointed out, responsibility for the reporting of vital events rests with the public, usually with

the relatives of new-born infants or deceased persons. Hence, more than in a census enumeration, co-operation of the public is decisive in achieving a complete record of births and deaths. The imposition of fees for the registration of births may tend to discourage registration. This may be particularly important among poor people, if the registration fee amounts to a significant proportion of their daily earnings. In countries imposing such fees temporary removal of the registration fee in order to encourage registration may result in a sharp increase in the number of recorded births during the period in which no fees are charged.

Social customs sometimes may lead to non-registration. In areas where a stigma is attached to illegitimate births, there may be a deliberate attempt on the part of the mothers to avoid reporting them, or to report them as legitimate. If the births are reported as legitimate, of course, population estimates and the balancing equation are not affected, but any analyses of marital fertility will be somewhat in error. In some countries also, children of one sex may not be considered quite as important as those of the other sex, in which case there may be a tendency to report less completely births of the less favoured sex. The fear that boys will be conscripted for military service may sometimes lead to the non-registration or late registration of their births.

A rural population thinly dispersed over a wide area may find it physically difficult to reach the registrar's office in order to report a birth. The consequent tendency to neglect registration may be aggravated if the means of transportation are deficient, if illiteracy is prevalent, if the people see no positive inducement to register the births, and if the births most frequently occur at home without a physician in attendance.

The population living in an urban area or within easy access to the registrar's office is more likely to register births. Indeed, registration is probably most complete among urban, literate populations in which most of the births occur in a hospital and practically none occur without a physician in attendance. It is much easier to ensure that a few hospitals or physicians will report all births than it is to ensure that every individual family will report on its own initiative.

Any system of family allowances based on the number of children should increase the completeness of registration since the payment of the allowance with respect to any child can be conditioned upon the registration of the birth. Any requirement that a person give documentary proof of his age, his parentage, or that he was born in the country or is a citizen of the country for inheritance or other legal purposes will also act to improve the completeness of registration.

The rationing of consumers' goods is another factor tending towards complete registration. Indeed, it may even lead to over-registration. This may be particularly true in those situations where an infant is entitled to as much of the rationed item as an older person, or where the rationed item is particularly needed by children. For example, sugar may be rationed at so many units per person irrespective of age, or milk may be limited exclusively to children; in such a case it may be advantageous to the family to have extra births registered. If the rationed item is for the consumption of adults only, as tobacco or

coffee for example, there may be no incentive to over-register births.

Over-registration may also occur, in some cases, for other reasons; for example, duplicate reporting by the parents through ignorance of the fact that the birth has already been registered without their participation or re-registration when an original certificate of registration is lost. Normal precautions such as birth registration indexes will minimize such duplicate registration, but in some cases the precaution may not be adequate.

#### FACTORS AFFECTING DEATH REGISTRATION

Some of the same factors which operate to reduce the completeness of birth registration may also tend towards incomplete death registration. In particular, among a rural population thinly dispersed over a wide area, in which there are few doctors available so that many of the deaths occur without a physician in attendance, registration is likely to be incomplete. Under such circumstances it is comparatively easy to bury the body privately; if the burial has been accompanied by whatever religious and social ceremonies are deemed important by the people concerned, they may feel no need to register the death.

The requirement that a death must be registered in order to obtain a burial permit is a very important aid to complete death registration, particularly in areas where most bodies are attended to by a relatively small number of undertakers, and where most burials take place in but a very few cemeteries. Death registration may be more complete in urban than in rural areas if for no other reason than that it is likely to be easier to control burials in urban areas.

In areas where most deaths are attended by a physician, complete registration is likely to be more easily achieved than elsewhere. From this point of view also, the situation in urban areas is likely to be more conducive to good registration than that in rural areas. At least the causes of death will certainly be reported more accurately if a physician is in attendance.

Certain legal factors may be conducive to complete registration. For instance, inheritance or other legal claims involving succession or the right of a widow to re-marry may require proof of a death which can be supplied by a death certificate. The State is also interested in deaths in which a crime may be involved. Thus there is some inducement to register deaths in order to avoid entanglements with the law which might result from non-registration.

The rationing of consumers' goods, on the other hand, may have a tendency to encourage non-registration or delayed registration of deaths. If the family of the deceased can continue to collect his rations by failing to report the death, there is obviously some incentive to avoid reporting.

#### RELATIVE COMPLETENESS OF BIRTH AND DEATH REGISTRATION

For the purposes of population analysis and particularly population estimates, it is highly desirable to know whether the births or the deaths in a given country are more completely reported. If births are

reported more fully than deaths, the computed natural increase will be too high, and if the reverse occurs it will be too low. On the other hand, if both are registered with about the same degree of completeness, the recorded natural increase may not be greatly in error even though there is some omission of both births and deaths.<sup>3</sup>

In spite of the importance of this question, the statistical reports of few countries give any indication as to which of the two events are believed to be reported more fully. An indication to the effect that both births and deaths are reported "fairly completely", or that the registration of both events is "markedly deficient", does not suffice to resolve this question.

#### B. Internal consistency of vital statistics

Some clues as to the possible degree of completeness of reporting of births and deaths are afforded by examination of the data themselves. The use of the procedures to be described in this section will not, by themselves, establish the quality of the registration; they may serve, however, to reveal possibilities of error which deserve further investigation. Procedures for measuring the completeness of registration more exactly will be presented further on in this chapter.

#### NUMBERS OF REPORTED BIRTHS, DEATHS AND INFANT DEATHS

In some areas, birth rates or death rates cannot be computed because sufficiently reliable population estimates are not available. Sometimes this is the case for an entire country; more frequently, it applies to a country's geographical or civil divisions. Valid comparison can then be made only of the various absolute numbers of vital events. Even where population estimates are realistic enough for a computation of valid birth or death rates, a comparison of the mere numbers of reported births and deaths is a very useful first step in an examination of the accuracy of the statistics.

In countries where birth registration is known to be accurate, male births invariably outnumber female births by a slight margin. For biological reasons, the sex-ratio of births can vary only within rather narrow limits. Usually, there are very nearly 105 male births for every 100 female births, though some variation in this ratio is possible. A ratio lower than 102, or higher than 107, while not impossible, must be regarded as rather unusual. Any significant deviation from these limits should be regarded with suspicion. In unusual circumstances, a slight deviation from this pattern might occur, but if the deviation is at all large it is practically certain that the births of one sex are less completely reported than those of the opposite sex and, very probably, births of both sexes are incompletely reported. If, on the other hand, the ratio of male to female births happens to be very near 105 per 100, birth reporting may still be deficient for both sexes.

Ordinarily, when a population consists of more or less equal numbers of men and women, the deaths of males slightly outnumber those of females. The exact ratio of male to female deaths depends on a variety of

<sup>3</sup> See: United Nations, Population Division, Manual I: *Methods of Estimating Total Population for Current Dates*, Population Studies, No. 10, chapter VI.

factors which cannot be fully considered for the purpose of such a simple comparison. Nevertheless, one should expect that variations in the ratio of male to female deaths among a country's administrative divisions should tend to conform to a systematic pattern; if variations are unsystematic, or if the ratio in some particular areas deviates greatly from the norm without any apparent reason, then the accuracy of death registration is open to suspicion. The comparison of numbers of male and female deaths may also be extended to individual age groups. As a general rule, more males than females die in infancy, childhood, and adolescence, but more females than males die in old age and, at least in some countries, also during the child bearing age.<sup>4</sup> Proceeding from one age group to another, one should expect the ratios of male to female deaths to change gradually and never abruptly. Sharp fluctuations in this ratio between one age group and another may indicate either incomplete recording of deaths of certain sex-age groups, or mis-statement of ages of the deceased (see also chapter III).

Among infants, male deaths almost invariably exceed female deaths. This is not only because slightly more boys than girls are born, but because the boys are generally subject to a higher rate of mortality. The exact ratio of male to female infant deaths may vary in accordance with the circumstances. In countries where infant mortality is high, this ratio is usually, though not necessarily, within the range of 110 to 120 male per 100 female deaths. In countries with low infant mortality, the ratio is often of the order of 130 to 140 male per 100 female deaths. These are only rough standards, but they may help to determine in some cases whether inaccuracy of infant death registration should be suspected. Large and unsystematic differences in this ratio among various subdivisions of a country likewise serve as a warning that registration may be inaccurate at least in some areas.

Other numerical relationships can also be utilized. For example, the ratio of infant deaths to the total of deaths at all ages usually exhibits a considerable stability, at least so far as the various segments of any one country are concerned. This ratio will be high—of the order of one in four, or even one in three—in countries of high fertility and high mortality, and low—about one in ten, or lower still—in countries of low fertility and low mortality but, so far as the various segments of a country or consecutive years are concerned, the ratio should not fluctuate unreasonably. Finally, one may compare the ratio of all births to all deaths. This ratio, which has sometimes been used for purposes of demographic analysis, can vary rather widely, but may also serve to detect irregularities so far as particular areas within a country are concerned.

#### LEVELS OF RATES

Vital rates<sup>5</sup> are known to be functions of the health conditions and other social and economic conditions of

<sup>4</sup> The reason for more numerous female deaths in old age is not a higher female mortality at such ages, but the survival of larger numbers of women than men up to these ages.

<sup>5</sup> The "vital rates" considered here are crude birth rates (live births per 1,000 inhabitants per year), crude death rates (deaths per 1,000 inhabitants per year), and infant mortality rates (infants dying at less than one year of age per 1,000 births during the same year).

the population.<sup>6</sup> Accordingly, the rates for a given country should be examined in the light of the conditions known to exist there, and of the results of researches on the relationships between these conditions and vital rates. Such an examination can indicate very roughly the ranges within which the rates should probably be contained, and point the way to a more precise investigation of possible errors if the recorded rates seem to differ too much from those that are expected. It must be borne in mind in this connexion that erratic rates may be produced by error either in the numbers of births and deaths reported, or in the population figures which form the bases for the rates, or both.

#### BIRTH RATES

Crude birth rates (number of live births per 1,000 population) much lower than 15 or higher than 50 have rarely been observed where data were known to be accurate. If the observed rate falls below 15, the completeness of registration is open to question; if it is over 50, the two possibilities should be considered: that there was over-reporting of births or that the population figure was understated.

The most probable range of birth rates, given certain economic and social conditions, can be defined more narrowly. Birth rates in the range between 35 and 45 per 1,000 are usually considered as "high", those in the range between 25 and 35 as "moderate", and those between 15 and 25 as "low". These limits should, of course, be interpreted somewhat liberally. High rates are characteristic of most areas of the world where social customs and relationships for the great majority of the population have been little affected by the changes associated with industrialization according to the modern pattern. Low rates are generally found in those countries which have undergone a profound social transformation as a result of urbanization, industrialization, a rising level of popular education, and related factors. Moderate rates tend to prevail in areas where the types of conditions mentioned are intermediate between the two extremes.

This generalized statement, however, is over-simplified; in each case other factors must be considered in determining within what range the birth rate should be expected to fall.

The components of the birth rate, as well as the crude rate for the entire country, should be examined in detail. For example, almost universally among those countries having fairly complete birth registration, it has been found that the birth rate in urban areas is lower than that in rural areas.<sup>7</sup> Therefore, the rates for the urban and rural areas of a country should be studied separately. If the rural rate turns out to be lower than the urban rate (even after allowing for possible differences in age composition and marital status by means of standardizing techniques), the completeness of registration in the rural areas is open to question. In this connexion it will be noted that it is very important that births be

<sup>6</sup> United Nations, *The Determinants and Consequences of Population Trends*, Population Studies, No. 17, especially chapters 3 and 4.

<sup>7</sup> For the results of studies on the various relationships mentioned in this and succeeding paragraphs, the reader is referred to the United Nations report mentioned above, *The Determinants and Consequences of Population Trends*, especially chapters 3 and 4.

tabulated according to the usual place of residence of the mother. If many rural women go to urban areas for confinement, and if their births are credited to the urban areas, the urban rate may be higher than the rural even though nearly all births to rural women have been registered.

The birth rates in the various subdivisions of a country should be examined in detail. After they are standardized for possible differences in age composition, the rates in the various subdivisions should be consistent with the social, economic and public health conditions in the various areas.

The birth rates for important ethnic groups or other components of the population (aborigines, nomads, immigrant groups of distinctive culture and living conditions et cetera) should also be studied in the same manner.

It is also useful to compare the level of the birth rate with the census statistics on age composition of the population. In general, a high birth rate is reflected by a high percentage of children in the population and a low rate by a low percentage of children, unless the relation is distorted by major migratory movements.

#### CONSISTENCY OF CRUDE BIRTH RATES WITH AGE STRUCTURE

It might appear at first thought that the birth rate could easily be checked directly against the number of infants enumerated in the census. Upon closer examination, however, this matter presents considerable complications. Infant enumeration is incomplete in most censuses. Furthermore, the number of reported "infants" (that is to say, those aged less than 12 months) can be much affected by slight mis-statements of age. Finally, only in countries having excellent registration systems is the rate of infant mortality known exactly.

A useful test, called the "forty per cent test", has been devised by Wertheim.<sup>8</sup> As demonstrated by its author, a population of which 40 per cent or more are aged less than 15 years is most likely to have a birth rate of at least 40 per 1,000. The birth rate might be as low as 39, with moderate infant mortality. As is generally known, infant mortality is usually high in populations with very high birth rates.

This test is applicable where only crude age statistics are available. Experience has shown that in most countries where the censuses show categories such as "adults" and "children" but not the exact ages of the populations, the mean age at which persons begin to be reported as "adults" is usually in the vicinity of 15 years (though, as a rule, slightly lower for females and slightly higher for males). Many such countries have, in fact, birth rates of the order of 40 per 1,000; variations around this level are probably reflected—however approximately—by variations in the percentage of persons aged less than 15, at least among population groups which are similar in other respects.

#### DEATH RATES

Crude death rates based on accurate statistics are seldom much less than 8 per 1,000 population, except

<sup>8</sup> W. F. Wertheim, "La population de l'Indonésie et le test des 40%", *Population*, 1954, No. 4, pp. 655-674.

among populations of abnormal age composition. This is not to say that a lower rate cannot be correct, but rather that if the rate is lower than 8 and if this fact cannot be explained by an abnormality of age structure, the completeness of registration or the accuracy of the population base figures should be investigated.

In areas where the statistics are accurate, death rates considerably in excess of 30 per 1,000 in normal years are unusual. Normal death rates as high as 35 per 1,000, or even higher, are possible if fertility is so high that the population can be reproduced despite such heavy losses. In years of calamity (for example, famine, epidemic, or mass destruction by war), the death rate may, of course, rise much higher. Except in the case of such catastrophes, a recorded death rate in excess of 35 per 1,000 suggests that the statistics should be examined for possible errors.

As in the case of birth rates, the range of expected death rates in a given area can be specified within narrower limits when the conditions affecting public health in that area are taken into consideration.

One further observation is relevant. High birth rates occur in conjunction with either high or low death rates. Low birth rates, on the other hand, are normally found only in association with rather low death rates. Hence, if the recorded birth rate is found to be of the order of, say, 15 to 25 per 1,000, the death rate would not normally be expected to be much lower than 10 per 1,000 nor higher than perhaps 15 per 1,000. If the recorded birth rate is low and the death rate either extremely low or higher than about 15 per 1,000, the accuracy of the statistics should be examined.

Tests for the completeness of death reporting are substantially similar to those described above for testing birth reporting. The data can be analysed for the various subdivisions of a country, or the various population groups, and the rates related to the social, economic and health conditions of these several segments.

In relating the death rates to these other factors it is necessary to take into consideration the age composition. If a population contains relatively few older people it will have a lower death rate, all other factors being the same, than a population containing relatively many older persons. This can be illustrated by comparing data for Israel and Norway. The crude death rate in the former country was 6.4 (in 1951) which might seem suspiciously low. In Norway (1950) the rate was 9.1. The population of Norway, however, was older than that of Israel. It is possible to standardize the Israel age-sex specific death rates on the Norwegian population so as to find out what the death rate in Israel would be if its population had the same age distribution as that of Norway. Standardizing in this manner yields a death rate of 10.3. In short, the difference in the crude death rate between these two countries was accounted for to a considerable extent by differences in age structure.

#### INFANT MORTALITY RATES

The level of the infant mortality rate, in general, is related to the level of the total death rate. Countries having high general mortality rates also have high rates of infant mortality. It is impossible, however, to assign any specific value to the infant mortality rate which



should accompany a given crude death rate. A preliminary appraisal of the infant mortality rate can be made by considering its level in the light of the social, economic and public health conditions of the country. If a rate seems unusual in relation to the known conditions, the accuracy of the vital data can be considered suspect.

The infant mortality rate is calculated by relating the number of infant deaths, generally under one year of age, to the number of births. Hence, it is affected by whatever errors there may be in the registration of both these events.

In this connexion it is well to note that in some countries infants who do not survive a minimum length of time, as one day or two days or three days, are not included in either the birth or death registration data. The effect of such a procedure is to lower the reported infant mortality rate, and make appraisal of it somewhat more difficult. For example, let us assume that 1,000 births are reported of which 100 died in infancy. The infant mortality rate is then calculated as 100 (100 deaths per 1,000 live births). If there were 20 infants born who died during the first day and were not registered, so that 20 should be added to both the numerator and the denominator, the rate becomes 117.6 per 1,000. Fortunately, for the purpose of population estimates, this error is of little importance since the additional number of births is almost instantly cancelled by the same additional number of infant deaths. In an investigation of birth rates and infant death rates, however, this circumstance can produce misleading results.

The infant mortality rate may also be affected if the age of the infant is not correctly reported. It is possible that the deaths of some children who are not quite one year old may be reported as deaths above the age of one year, but it is also possible that some children who have died after their first birthday may still be reported as infant deaths. If either of these two errors is more frequent than the other, the number of infant deaths is accordingly under-reported or over-reported.

A useful test consists in an examination of numbers of infant deaths by months of age, if such statistics are available. Deaths are by far most frequent during the first month, and especially the first few days, of life; in subsequent months mortality declines continuously. Ordinarily, one should expect the number of deaths in the first month to exceed significantly the number in the ensuing five months which again should be substantially larger than the number of deaths between 6 and 12 months of age. Beginning with the second month, the progression should be a fairly smooth one.

The sources of error and ambiguity in statistics relating to infant mortality, and methods of appraising the quality of the data, have been described in another publication.<sup>9</sup>

#### TRENDS IN CRUDE RATES

It is expected that there will be year-to-year variations in the birth, death and infant mortality rates. Where these yearly variations are very large, however,

<sup>9</sup> United Nations, Population Branch, *Foetal, Infant and Early Childhood Mortality*. Volume I. *The Statistics*. ST/SOA/Series A/13, New York, 1954.

there is reason to suspect the accuracy of the statistics. In Bolivia, for example, the recorded birth rate rose from 14.7 in 1940 to 28.5 in 1941 and has fluctuated around 30 per 1,000 ever since; the death rate as recorded rose from 4.9 in 1940 to 19.4 in 1941 and has since been within the range of 15 to 20 for most years. Such extreme variations normally do not occur; in this case they are explained by the fact that Bolivia introduced compulsory registration in 1940. It can be deduced that there was considerable under-reporting prior to 1941; how complete the reporting has been from 1941 to date has yet to be determined.

Sometimes, however, a rather sharp change in a vital rate may occur as a result of a major change in the relevant conditions of life. An example is found in the statistics of Japan. In 1946 the reported crude birth rate was 25.3 and in 1947, 34.3. The Japanese birth rate is known to have fallen during the years of the Second World War, and in particular during those years when many of the men were in the armed forces outside the country. With the termination of the war in 1945 and the repatriation of the armed forces as well as of many other Japanese men who had been abroad, the birth rate rose rapidly.

Another example is found in the Netherlands, where the death rate was 11.8 in 1944, 15.3 in 1945, and 8.5 in 1946. In this case, mortality had been rising gradually during the years of the Second World War, as living conditions deteriorated. At the end of 1944 and the beginning of 1945, the Netherlands were the scene of military operations which had devastating effects, causing a considerable temporary increase in mortality. By 1946, conditions were again relatively normal and the death rate returned to its pre-war level.

It should be noted that where the births and deaths are tabulated by year of registration rather than year of occurrence, large fluctuations in the rates may result from variations in the numbers of delayed registrations. One further point should be noted. The obtaining of fairly complete reporting of all births and deaths which occur is a slow process. After a compulsory registration law is introduced which applies to all the inhabitants, and even after the necessary administrative mechanism is instituted, several years may be required before satisfactory reporting of all the events is ensured. A steadily rising trend in crude birth or death rates over a period of several years is then likely to indicate improvement in coverage rather than a rise in fertility or mortality.<sup>10</sup>

#### PATTERNS OF DEATH RATES, BY AGE AND SEX

Another clue as to the possible completeness of registration of deaths is afforded by examination of the patterns of death rates by age and sex. In those countries in which all the evidence has proven quite conclusively that registration of deaths is virtually complete, there appears an almost standard pattern of rates by age and sex. This typical pattern is illustrated with data for Sweden (see table 3 below). The essential points to note are as follows:

(1) The rates for males, in each age group, are higher than those for females;

<sup>10</sup> See also United Nations, *Demographic Year Book*, 1952, p. 27.

(2) The rates are quite high in the age group under one year, after which they rapidly decrease to a low point after age 10. From this minimum, the rates rise slowly at first and at the oldest ages increase very rapidly. Thus there is a *steady and smooth progression* describing a U-curve.

In reviewing the age-sex pattern of any country, if deviations from the above typical pattern are observed they should be explained in terms of known peculiarities; otherwise errors in the statistics are to be suspected. For example, in some populations the death rate for women is higher in the reproductive ages than it is for men in the same age groups, because of unusually high maternal mortality associated with a very high birth rate. In some countries, tuberculosis affects

the survival of adolescents and young adults to such an extent that the death rate rises fairly rapidly until about age 20, to remain at a fairly constant level until about age 30. Any deviations that are not explainable imply either that the death reporting (number of deaths and/or age of deaths) is deficient, or that the age data obtained from the population census (or population estimates) are in error. Tests for attempting to determine the accuracy of the age data will be shown in chapter III.

Some of the kinds of deviations from the normal pattern of sex-age specific mortality rates which may be found in the statistics for various countries, are illustrated by the figures for Egypt and the Moslem populations of pre-partition Palestine in table 3.

Table 3  
DEATH RATES BY AGE AND SEX, FOR SWEDEN (1948), EGYPT (1937) AND PALESTINE MOSLEMS (1944)  
(Deaths for 1,000 persons in specified age-sex class)

Age	Sweden		Egypt		Palestine Moslems	
	Males	Females	Males	Females	Males	Females
All ages .....	10.0	9.7	29.5	25.0	17.3	16.8
Under 1 year.....	27.4	19.4	257.6	211.2	112.8	108.5
1 to 4 years.....	1.5	1.2	85.5	72.7	33.0	38.3
5 to 9 years.....	0.7	0.5	8.5	6.7	4.6	3.6
10 to 14 years.....	0.6	0.4	5.1	3.9	3.4	2.3
15 to 19 years.....	1.3	0.7	5.8	4.2	3.6	2.5
20 to 24 years.....	1.8	1.1	8.1	4.5	4.3	4.7
25 to 29 years.....	1.7	1.1	8.8	5.4	8.0	7.6
30 to 34 years.....	2.0	1.5	10.5	7.3	6.8	5.8
35 to 39 years.....	2.4	2.1	10.5	7.2	6.5	6.5
40 to 44 years.....	3.2	2.5	13.7	9.6	7.1	4.8
45 to 49 years.....	4.9	3.9	14.0	8.2	10.0	5.2
50 to 54 years.....	8.0	6.2	21.9	13.9	8.8	5.8
55 to 59 years.....	12.2	9.1	22.1	12.5	10.9	9.3
60 to 64 years.....	19.1	15.2	37.9	22.1	22.1	13.5
65 to 69 years.....	30.8	25.9	45.3	25.3	30.2	24.3
70 to 74 years.....	48.7	44.5	76.4	50.5	44.4	37.8
75 to 79 years.....	81.1	74.4	103.7	74.7	—	—
80 to 84 years.....	136.3	129.5	192.6	144.9	—	—
85 and over.....	235.8	222.5	566.4	578.1	—	—
75 and over.....	—	—	—	—	103.9	119.7

The death rate for all ages in Egypt is about double that of Sweden. As in Sweden, the rates for males are higher than those for females, age for age, except in the very oldest age group. The rates for males in the successive age groups exhibit no deviations from the expected steady and smooth progression of a U-curve. In the case of female rates, minor deviations are noted; between the ages of 30 and 65 the progression is not smooth. Whereas the expected pattern is that of a steadily rising rate, the Egyptian data show:

Age	Rate
30 to 34 years .....	7.3
35 to 39 years .....	7.2
40 to 44 years .....	9.6
45 to 49 years .....	8.2
50 to 54 years .....	13.9
55 to 59 years .....	12.5
60 to 64 years .....	22.1

Intensive testing is necessary to determine whether the dips at ages 35 to 39, 45 to 49 and 55 to 59 reflect primarily errors in the death statistics or in the population figures by age used for computing the rates.

The data for the Palestine Moslems (table 3 above) reveal greater discrepancies. The male rates are lower than those for females at three age groups—1 to 4 years, 20 to 24 years, and 75 years and over; only the deviation at the 20 to 24 year group might be explainable in terms of high maternal mortality. The patterns by age, for both males and females, are much more irregular than those observed in Egypt. Among both sexes the patterns between ages 25 and 59 are highly irregular. For example, it is most implausible that women aged 45 to 54 years have lower mortality than women aged 25 to 29 years, or that men aged 40 to 44 years have lower mortality than men aged 25 to 29 years. The age-sex rates for this population strongly suggest the need for intensive investigation into the completeness of death reporting, as well as the accuracy of the population figures.

### C. The use of balancing equations

The use of balancing equations has been described in chapter I. It was noted that the balancing equation alone does not permit a final appraisal of any one of



the sets of statistics involved, that is, census counts and statistics of births, deaths and migration.

In the present chapter, balancing equations are made for certain segments of a population, thereby concentrating attention on the errors which arise from inaccuracies in the vital statistics. Again it is to be noted that any one such equation, by itself, is insufficient to determine exactly from which set of statistics the major part of the discrepancy originates. Nevertheless, as the terms of the equation are changed, additional evidence is gathered and the conclusions, which were at first tentative, may be either confirmed or disproved, at least to a certain extent. Every such appraisal is relative only, and the degree of accuracy of each set of statistics can in no case be determined with absolute certainty by these equations alone. Absolute appraisal would be possible only if at least one set of statistics were known to be perfectly accurate or its error were known precisely. Even so, balancing equations can often determine with sufficient assurance whether a given set of statistics is probably reasonably accurate or not.

#### APPRAISAL OF DEATH STATISTICS BY MEANS OF BALANCING EQUATIONS

It is obvious that the population above a certain age enumerated at a recent census must have already been alive at the time when the census previous to it was taken. This part of the population would have aged by the number of years which have passed between the two census dates and, during the interval, would have been diminished by a certain number of deaths and otherwise modified as a result of migration. Births which occurred during the interval do not affect this segment of the population.

For reasons to be explained in chapter III, the comparison of the two populations is likely to be more precise if children under 5 years of age at the earlier census are not included. Let it be assumed that censuses have been taken in 1940 and 1950. The balancing equation then becomes:

$$P_1 = P_0 - D + I - E$$

$P_1$  = population 15 years of age and over in 1950,  
 $P_0$  = population 5 years of age and over in 1940,  $D$  = number of deaths in 10 year period to persons who were 5 years and over in 1940,  $I$  = number of immigrants who were 5 years of age and over in 1940,  $E$  = number of emigrants who were 5 years of age and over in 1940.

By setting up the equation in this way, the factor of births is eliminated so that the balancing equation can be used to appraise the death statistics, the migration data, and the census enumerations at ages above the stated limits. Thus a better appraisal of the death reporting is obtained than could be had if the balancing equation also included births.

It would be preferable also to eliminate the migration factor and this is possible for a country in which the census and vital statistics are presented separately for the native population, if it is known that few of the native-born population emigrate; in such a case the equation becomes:

$$P_1 = P_0 - D$$

where  $P_1$ ,  $P_0$  and  $D$  refer to native-born individuals only.

In some countries it may be known that net migration of both native and foreign-born is very small. In this case it is not necessary to tabulate the data by nativity. The balancing equation can be applied using only the total count of persons 5 years of age and over in the earlier census, the number 15 years of age and over in the following census, and the total number of deaths during the decade to persons who were 5 years of age or over at the time of the earlier census.

The number of deaths to persons aged 5 and over at the time of the earlier census is obtained by adding together the deaths at ages 5 and over in the census year, those at ages 6 and over in the following year, those at ages 7 and over in the subsequent year, and so forth. A refinement of the calculation is possible if the exact date of the census is taken into account. If the census was taken at the beginning of the year, then some of the persons dying at age 5 in that year were less than 5 years old at the date of the census. If the census was taken near the end of the year, only a small portion of the deaths of that year could have occurred after the date of the census.

Great refinement of the computations, however, is unnecessary, unless the statistics are very nearly accurate. Otherwise, the slight modification in the estimated number of deaths which results from a refinement of the calculation can only be negligible as compared to the probable magnitude of the residual in the balancing equation. For most practical purposes, it is sufficient to utilize death statistics tabulated by 5-year age groups. It is enough to assume that all deaths occurring at ages 5 to 9 during the first year after the census was taken were the deaths of persons aged 5 to 9 at the census date; for the next year, it may be assumed that four-fifths of the deaths in this age group were to persons aged 5 to 9 at the census; two years later three-fifths; three years later, two-fifths; and four years later only one-fifth. In subsequent years, the deaths of persons aged 5 to 9 are no longer to be considered, but of the 10 to 14 year age group all deaths should be taken into account five years after the census, four-fifths of the deaths six years after, and so forth.

#### Example of Thailand

Age data are available from the censuses of 1937 and 1947, and deaths by age for the inter-censal period together with an estimate of the net volume of migration for the period, are also available. These data can be used to illustrate the above-described form of the balancing equation as follows:

Population 15 years of age and over according to census of 1947.....	10,068,000
Population 5 years of age and over according to census of 1937.....	12,027,000
Apparent inter-censal decrease.....	1,959,000
Number of reported deaths of persons who were 5 years of age and over in 1937.....	1,530,000
Migratory inward balance of persons who were 5 years of age and over in 1937.....	100,000
Total expected decrease (deaths minus immigrants)	1,430,000
Apparent inter-censal decrease.....	1,959,000
Population decrease accounted for.....	1,430,000
Residual unaccounted for.....	529,000

The residual amounts to about 5.3 per cent of the population 15 years of age and over in 1947, and 27.0

per cent of the apparent inter-censal decrease. This amount is large enough to suggest that further investigation into the accuracy of the component parts is required.

If the residual were to be attributed to any one factor, this single factor could have been a greatly excessive enumeration in 1937, a greatly deficient enumeration in 1947, a large emigration not accounted for, or a considerable deficiency in death registration.

Excessive enumeration at one of the censuses does not seem very probable. Census enumerations may have been deficient, but there is no obvious reason why the 1947 enumeration should have been much less complete than the earlier one. In this connexion, it may be stated that examination of the census data by age and sex (see chapter III) yields no clues with regard to any major irregularity in the census enumeration. Whether, despite the available estimate of an inward migratory balance, there has in fact been a large emigration, cannot be decided in the absence of further evidence, but it would seem unlikely. Therefore—although some of these errors may have been present—it is probable that the major part of the error should be attributed to incomplete registration of deaths.

One estimate of the completeness of death registration is arrived at by assuming that the two censuses were either complete, or deficient to the same extent, and that the migration balance is substantially correct. In this way the total number of deaths which occurred is estimated as the number reported plus the residual; or a total of 2,059,700, of which 1,530,000, or 74 per cent were reported.

On this basis, the average crude death rate of Thailand for the period 1937-1947 inclusive should be raised from the recorded level of 16.6 to 22.4. This rate is comparable with an average death rate of 21.0 in Malaya during 1932-1940, when death registration is believed to have been fairly accurate.

On the assumption that the censuses differed with respect to completeness of enumeration, other estimates of the completeness of death registration would be obtained. If it were assumed that there was a larger under-enumeration in the 1947 census, the residual attributed to under-registration of deaths would be reduced, and the conclusion would be that death registration was somewhat more than 74 per cent complete. On the other hand, if there were thought to have been a larger amount of under-enumeration in 1937 than in 1947, the residual due to incomplete death registration would be larger and the estimate of completeness of registration would be less than 74 per cent. By making various assumptions as to the completeness of the two censuses, based on whatever information can be obtained on that score, it is possible to arrive at maximum and minimum estimates of the completeness of death registration.

#### *Example of Puerto Rico*

The procedures can also be illustrated with data for Puerto Rico as follows. Censuses were taken on 1 April, 1940, and 1 April, 1950. For the intervening years data are available on deaths and migrants.

Population 15 years of age and over according to census of 1 April 1950.....	1,255,300
Population 5 years of age and over according to census of 1 April 1940.....	1,588,900
Apparent inter-censal decrease.....	333,600

Reported deaths of persons who were 5 years and over in 1940.....	157,800
Migratory outward balance of persons who were 5 years of age and over in 1940.....	172,000
Expected decrease (deaths plus emigration).....	329,800
Apparent inter-censal decrease.....	333,600
Decrease accounted for.....	329,800
<i>Residual</i> .....	3,800

The residual would be less than 3,800 if deaths which occurred among men in the armed forces had not been excluded from the death statistics employed here. Even the residual of 3,800 amounts to only 0.3 per cent of the enumerated 1950 population 15 years of age and over.

These results suggest that the death statistics, as well as the two population censuses and the migration data, are fairly accurate. It is possible, of course, that an under-reporting of deaths was compensated by faulty migration data. This result could have been produced by an over-count of emigrants or an under-count of immigrants. However, since the migratory balance is computed from statistics of total civilian arrivals and departures (so that no problem of distinguishing between migrants and other travellers, such as visitors, is involved), neither of these possible errors is likely to be important. Puerto Rico is an island, and most of the migration occurs by air via the one central airport; it is unlikely that many persons leave from or arrive at the island without being recorded, and the possibility of duplication of records in any large number of cases is remote.

With respect to the remaining two elements in the balancing equation, the census counts and the death statistics, the fact that the equation does balance out with but a very small residual suggests that both these elements are quite accurate also. This is not proven beyond doubt, since there could be compensating errors among these two sets of data. If the size and direction of any errors in the census were the same at both dates, it follows that the reporting of deaths was quite complete. If the 1950 census was more complete than that of 1940, it follows that deaths were under-reported by a corresponding amount. This possibility could be checked in part, by making the same comparisons for the 1940 and 1930 censuses, and for the 1950 and 1960 censuses when the results of the latter become available. If the equation should balance out as well for the other two inter-censal periods as it did for the decade 1940 to 1950, the evidence would be quite conclusive that the censuses and death reporting were nearly complete.

#### *Example of Portugal*

The censuses of 1940 and 1950, together with mortality data and estimates of net migration for the inter-censal decade, permit the construction of this version of the balancing equation for Portugal:

Population 15 years of age and over according to census of 1950.....	5,953,100
Population 5 years of age and over according to census of 1940.....	6,890,500
Apparent inter-censal decrease.....	937,400
Reported deaths of persons who were 5 years and over in 1940.....	787,900
Migratory outward balance of persons who were 5 years of age and over in 1940.....	64,000

Expected decrease (deaths plus emigration).....	851,900
Apparent inter-censal decrease.....	937,400
Decrease accounted for.....	851,900
Residual .....	85,500

In this case the residual amounts to 1.4 per cent of the population 15 years of age and over enumerated in 1950, and 9.1 per cent of the inter-censal decrease. This residual is smaller than that observed in Thailand and greater than that found in Puerto Rico. Further investigation of the components parts is required before any definitive statement can be made about the completeness of death reporting. Perhaps the migration data are in error; perhaps the 1950 census was more nearly complete than the 1940 census; perhaps death reporting was incomplete.

If it is assumed that the two censuses were equally complete and the migration data substantially correct, the residual is accounted for by unreported deaths. On these assumptions, the true number of deaths which occurred is estimated at 873,400, of which 787,900 or about 90 per cent, were reported. Other estimates can be obtained in the same manner as was shown for Thailand previously.

#### APPRAISAL OF BIRTH STATISTICS BY MEANS OF BALANCING EQUATIONS

Another form of the balancing equation is particularly suited to direct attention to an evaluation of birth statistics. In this instance, the procedure consists essentially in a comparison of the number of children below a certain age enumerated at the census with the number of births during the years immediately preceding the census, allowance being made for the deaths and migratory movements by which this child population has been affected prior to the census. The balancing equation can then be stated in the form:

$$P_1 = B - D + I - E$$

where  $P_1$  = number of children enumerated, under a specified age (for example, under 5 years),  $B$  = number of births reported during an equivalent number of years prior to the census,  $D$  = number of deaths among children born during these years, up to the time of the census,  $I$  = number of immigrants (children born during the specified period),  $E$  = number of emigrants (children born during the specified period).

In computing the number of deaths which have occurred to children born during the given period, statistics of deaths by age are required. It is of some importance to refine this calculation because deaths are far more frequent to children in their first year of life than in subsequent years. Usually more than two-thirds of the deaths under one year of age occur during the first six months of life, and a disproportionately high percentage of these deaths occur to children aged less than one month. This consideration has a bearing on the ratio which should be selected in order to estimate that number of infant deaths which should be attributed to children born in the earliest year included in the calculation.

#### Example of Thailand

The above balancing equation can be applied to Thailand with the following results:

Population under 5 years of age according to census of 1947.....	2,644,000
Births reported in 5 years preceding census date....	2,513,000
Deaths reported to children born during these 5 years	386,000

Expected survivors under 5 years of age to be enumerated in the 1947 census.....	2,127,000
Number actually enumerated.....	2,644,000
Residual .....	517,000

This calculation indicates that there was under-reporting of births. It is possible, but very unlikely, that the census count exaggerated the number of children. In fact, in almost every census, young children are under-enumerated at least to some extent, and if the Thailand census figure for this age group was in error, it almost certainly erred on the low side. It is hardly possible, on the other hand, that several hundred thousand young children immigrated; some of the residual, but probably only a small part, could be accounted for by migration.

How incomplete might the birth reporting have been? If it is assumed that the number of children reported by the census was correct, and the number of child deaths was fully reported, one estimate of the completeness of birth reporting can be computed as follows: to the number of births as reported, is added the amount of the residual, giving an estimated total of 3,030,000 births, and an estimate of 83 per cent for the completeness of birth registration.

Another estimate can be arrived at in the following manner: The number of children under 5 years of age reported in the census is probably too small; let it be assumed, purely for illustrative purposes, that 90 per cent of the children were enumerated. Then the true number of children under 5 living in the country at the census date would be 2,930,000. With reference to deaths, the calculations earlier in this section of the chapter suggested as one possibility that death reporting was 74 per cent complete. If it is assumed that the same percentage applies to the deaths of children under 5 years of age, the true number of deaths for this age group is estimated at 521,000 (386,000 divided by 0.74). Adding the estimated number of children to the estimated number of deaths gives 3,451,000 as an estimate of the number of births, which implies that birth registration was 73 per cent complete (2,513,000 divided by 3,451,000 = 0.73).

On this basis, the average crude birth rate for Thailand for 1942-1946 would be raised from the recorded 30.7 to an estimated 42.1. The latter rate is comparable with that for Malaya during 1947-1951, namely 42.6, where birth registration is believed to have been very nearly complete.

This result cannot be taken as a reliable estimate of the completeness of birth reporting. The calculations were made only to illustrate the methods; better estimates might be obtained by investigating other available information bearing on the completeness of the census enumeration and of death registration, or by applying direct checks. The probability is very high, however, that not all children under 5 years of age were enumerated in the census of 1947, and that not all deaths to infants born in the 5 years preceding the census were registered. Therefore it is not likely that birth regis-

tration was as much as 83 per cent complete—the figure first derived.

### *Example of Puerto Rico*

For illustrative purposes, let us use the balancing equation to test the completeness of registration of births in the year preceding the census date. The calculations are:

Population under 1 year of age according to census of 1 April 1950.....	71,000
Births reported during preceding 12 months.....	85,500
Deaths reported to children born during this period....	4,600
Expected surviving children under 1 year of age to be enumerated in the census.....	80,900
Number actually enumerated.....	71,000
<i>Residual</i> .....	9,900

It should be noted that data on migration are missing from this balancing equation. It is known that there was considerable net outward migration from Puerto Rico, but the number of babies born in the period 1 April, 1949 to 1 April, 1950 who emigrated before the latter date is unknown; it is probably reasonable to assume, however, that this number was very small and could have accounted for but a very small portion of the residual of 9,900.

There might have been a significant under-reporting of infant deaths. This balancing equation proves nothing about the quality of death statistics. However, the results of the balancing equation used earlier for estimating the completeness of death reporting suggested that deaths were reported quite completely for the population which was 5 years of age and over in 1940. The presumption is therefore strong that reporting of infant deaths was also at least fairly complete.

The above reasoning leads to the conclusion that the census enumeration of infants was probably deficient, and that birth registration may very well have been nearly complete. The census deficiency could have taken two forms: complete omission of some infants from the head count and enumeration of some as one year of age or more. In either case, the result is an understatement of the number of infants. The question becomes: how much under-enumeration was there? If the under-enumeration were of about the size of the residual, it would follow that birth reporting was very nearly complete. However, if the amount of under-enumeration were considerably greater than the residual, it would be clear that a significant proportion of the births had not been registered. The under-enumeration of infants under one year of age in Puerto Rico in the 1950 census can be estimated, when the 1960 census has been taken, by the methods to be explained in chapter III.

### APPLICATION TO THE SUBDIVISIONS OF A COUNTRY

As has been noted in chapter I, balancing equations may be applied to the population of parts of a country as well as to the population of the entire country. The usefulness of balancing equations is diminished, however, by the fact that migration between the various areas of a country is usually important and that satisfactory statistics on internal migration can rarely be found. Hence, in many circumstances, the balancing equations for parts of a country provide insufficient

evidence, particularly on the quality of death statistics. As a matter of fact, in countries where the census figures and birth and death statistics are supposed to be of fairly high quality, such equations are sometimes used for the purpose of estimating the balance of migration rather than detecting errors in the vital and census statistics. On the assumption that the errors are negligible, the discrepancies are ascribed to the factor of migration.

The limitations of the use of the equation are not so strict if the census provides data on the population not only in accordance with the present place of residence, but also with residence at the time of the previous census. In that event, the earlier population within each administrative division can be compared with essentially the same population, at a more recent date, as modified by deaths in the intervening period and by migration beyond the country's national boundaries. Census tabulations of the type mentioned are rare, however. More commonly, censuses provide tabulations of the population according to place of birth, or according to place of residence at some specified earlier time (but not the date of the last previous census). In that event, the exact balance of the population between the two census enumerations cannot be established, but the census data furnish a certain amount of evidence regarding the direction and volume of migratory movements within the country between the two census dates.

The complications relating to internal migration are at a minimum when the balancing equation is used to test the completeness of birth registration for subdivisions of the country, in the same manner as explained above for the whole country. When births in the year preceding the census enumeration are compared with the number of children under one year of age enumerated in the census, the conclusion is not necessarily invalidated because the relatively small migration of infants, during a single year, can be disregarded.<sup>11</sup> However, this procedure suffers from the disadvantage that the completeness of infant death registration in the various subdivisions cannot be determined by the balancing-equation methods; unless some independent information on this score can be obtained, the estimates of variations in completeness of birth registration from area to area may be subject to considerable error.

The direct checks to be discussed in section D of this chapter are the best means of determining the completeness of vital statistics registration. If the application of direct checks has been limited, because of the cost, to certain parts of the country, it is still possible to estimate the completeness of birth and infant death reporting in those areas which were not directly checked by a procedure analogous to that of the balancing equation. The calculations are shown in table 4 below, with a hypothetical example. It is assumed that direct checks were made in subdivision A but not in subdivision B; therefore, the problem is to ascertain the probable com-

<sup>11</sup> If it can be assumed that death reporting is nearly complete in all the subdivisions, it is possible to estimate and make allowance for internal migration. In this way it is possible to compute the completeness of birth reporting by subdivisions using the numbers of children under 5 years of age enumerated in the census, and allowing for internal migration. See, A. J. Jaffe, *Handbook of Statistical Methods for Demographers*, Washington, United States, Bureau of the Census, 1951, pp. 145-148.

Table 4

PROCEDURE FOR ESTIMATING LIMITS OF COMPLETENESS OF BIRTH REPORTING FOR A COUNTRY'S  
SUBDIVISIONS (hypothetical data)

	Subdivision A	Subdivision B
1. Reported number of births, in year preceding census.....	110	350
2. Correct number of births as revealed by a direct check.....	120	(no check)
3. Reported number of infant deaths.....	14 <sup>a</sup>	35
4. Expected number of survivors at census (uncorrected figures, item 1 less item 3).....	96	315
5. True number of survivors at census date in subdivision A (item 2 less item 3).....	106	—
6. Children under one year of age enumerated at census.....	98	313
7. Ratio of true number to number enumerated, subdivision A...	1.082	—
8. Expected number of children under one year of age at census date in subdivision B, assuming equal completeness of enu- meration in both subdivisions (that is to say, 313 times 1.082)	—	339
9. Estimated number of births, subdivision B, assuming infant death reporting was complete (item 8 plus item 3).....	—	374
10. Estimated completeness of birth registration in subdivision B (item 1 as percentage of item 9) (per cent).....	—	94

<sup>a</sup> It is assumed that the direct check revealed this figure to be correct.

pleteness of birth and infant death reporting in subdivision B. It must be emphasized that this procedure only provides approximate answers, and that a direct name-for-name check affords a more accurate appraisal of the vital statistics.

The simple assumption used in table 4 can be varied. For example, on the basis of the calculations for several areas where the direct checks of birth statistics were carried out, it may be possible to estimate a range within which the correction factor for census under-enumeration of children (item 7) in subdivision B is almost certainly contained. If this range is, say, between 1.06 and 1.10, the expected number of children (item 8) falls in the range between 332 and 344. Adding 35 (item 3) we obtain the range from 367 to 379, and the completeness of birth registration (item 10) is then estimated in the range of 92 to 95 per cent.

Furthermore, the results of the checks of infant death registration in various areas may provide a basis for estimating its completeness in subdivision B. If it is estimated in this manner, for example, that between 80 and 90 per cent of infant deaths in subdivision B were registered, there would be, instead of 35, some 39 to 44 actual infant deaths, to be added, in item 9, to the range of 332 to 344 estimated for item 8. Thus the estimated number of births is placed in the range from 371 to 388, indicating that birth registration was 90 to 94 per cent complete.

#### D. Direct checks on completeness of vital statistics registration

The method of direct checks consists of securing an independent list of children born, or persons who died, during a particular period and in a particular area, and then checking name for name with the registrar's files to determine whether these events were registered. The simplest, but not necessarily the most accurate estimate of under-reporting is that proportion of the births or deaths on the independent list, for which no certificate of registration can be found. The procedures for a good estimate of the completeness of registration may be more complicated in some circumstances.

Lists of names of births and deaths can be secured from a wide variety of sources including church records, names reported in newspapers and other publications, and house-to-house inquiry. These names can be collected from the entire population or from a sample; in the latter event it is important that the sample be representative of the entire population living in the area. Any effort to cover the entire population of a country in order to determine the number of births and deaths which have occurred during a specified year, for example, is very expensive; indeed, its cost may be as great as that of a population census. The minimum sample which will provide reliable results is most economical and practical.

So far as births are concerned, the names of infants listed on the census schedules constitute a valuable basis for a check on registration.

In many respects these procedures are very similar to those described in chapter I for a direct check of the accuracy of a census enumeration, but a check on vital statistics can be carried on at any time whereas direct checks of the census enumeration are best conducted immediately after the enumeration is finished.

It is evidently desirable that the independent list be as nearly complete as possible; however, as will be explained below, it is not essential that the list be perfectly complete in order to obtain good estimates of the completeness of registration.

In matching the names on the independent list with the names of registered births or deaths, the events are tabulated in three categories:

(A) Matched events, that is to say, those recorded on both lists;

(B) Events found on the independent list but not registered;

(C) Events registered but not found on the independent check list.

In addition, unless the independent check list is known to be perfectly complete, a fourth category (which we shall designate as category D) must be considered, namely, events not included in either list, which actually occurred and should have been registered.

The sum of  $A + B$  represents all events found on the check list; the sum of  $A + C$  all events found in the register; and the sum of  $A + B + C + D$  all events which in fact occurred. A perfect measure of the completeness of registration would be the ratio of  $A + C$  to  $A + B + C + D$ .

If the independent list is very nearly complete, very few (if any) events should be found in category  $C$  and it may be presumed that category  $D$  is very small. The categories  $C$  and  $D$  can then be neglected, and the ratio of  $A$  to  $A + B$  is a good estimate of the completeness of registration.

If a sizeable number of events of both categories  $B$  and  $C$  is found, it may be presumed that the number of category  $D$  is also substantial. This is the situation which is usually to be expected. One estimate of the magnitude of  $D$  can be made on the assumption that, in the check list, the probability of inclusion of an event is the same for registered events as for those which escaped registration. On this assumption, the ratio of  $C$  to  $C + D$  would be the same as the ratio of  $A$  to  $A + B$ , or of  $A + B$  to  $A + B + C + D$ . The ratio of  $A$  to  $A + B$  then still constitutes a good estimate of the completeness of vital registration.

As a rule, however, the probability of omission from the check list will be greater for unregistered than for registered events. The reasons why certain events do not reach the registers are likely to apply also in some degree to the omissions from the coverage of the house-to-house inquiry or other source of the check list. The ratio of  $C$  to  $C + D$  will therefore usually be smaller than the ratio of  $A$  to  $A + B$ , so that the latter is a maximum estimate of the completeness of registration. The error of the estimate is likely to be the greater, the larger the numbers of cases in categories  $B$  and  $C$ . If the probable size of category  $D$  is large enough to have a substantial effect on the estimate of the completeness of registration, refined methods are necessary in order to arrive at a realistic estimate of completeness.<sup>12</sup>

<sup>12</sup> For a description of refined procedures, as applied in a test of the birth and death registers in an area of Bengal, where vital registration is known to be greatly deficient, the reader may refer to C. Chandra Sekar and W. E. Deming, "On a Method of Estimating Births and Death Rates and the Extent of Registration," *Journal of the American Statistical Association*, Vol. 44, No. 245, March 1949, pp. 101 ff.

#### BIRTH REGISTRATION TEST IN PUERTO RICO, 1950

The above procedures can be demonstrated with data obtained in a test of birth registration which was carried out in Puerto Rico. At the time of the 1950 population census special cards were filled out for all infants enumerated in the census, who had been born during the preceding three months. These cards contained the name of the infant, the names of the parents, and such other information as would be useful for identification purposes. The cards were then matched with the records of births registered during the same period.<sup>13</sup>

The figures for the municipality of Adjuntas, for example, are as follows:

Infants' cards matched with births registered.....	200
Infants' cards for which there were no births registered..	15
TOTAL	215

The completeness of birth registration was thus estimated at 93 per cent (200 divided by 215).

The results of this test for the entire island of Puerto Rico revealed that in the three months preceding the census of 1 April 1950, birth registration was about 95 per cent complete. The presumption is that birth registration had been approximately as complete for some time in the past, and would probably continue to be at least as complete in the future. In comparing this finding with that obtained by using the balancing equation (as described earlier in this chapter), it is seen that the tentative conclusion arrived at by means of the equation are substantiated by means of the direct check. The direct name-for-name check not only provides a more precise appraisal of completeness for the island as a whole, but also permits estimation of completeness of registration in each of the 77 municipalities, which is not at all feasible with the balancing equation.

It should be noted that this type of check can be carried out any time after a census has been taken, so long as the census schedules and the registration certificates are available.

<sup>13</sup> For further information on the steps involved in matching the records see Sam Shapiro and Joseph Schachter, "Methodology and Summary Results of the 1950 Birth Registration Test in the United States", *Estadística* (Journal of the Inter-American Statistical Institute), December 1952, Vol. 10, No. 37.

### CHAPTER III. THE ACCURACY OF AGE AND SEX STATISTICS

This chapter is concerned primarily with methods of testing the accuracy of population statistics classified by age groups. Tests of the accuracy of age statistics are necessary not only because such data are of major importance for population estimates and demographic analysis, but also because errors in these statistics are often indicative of deficiencies in the head count of the population, or in the records of the numbers of vital events. Statistics classified by age groups may be affected both by errors in the reporting of ages and by variations in the completeness of enumeration, or of recording of vital events, for the different age groups. Both types of error will be considered here.

The classification by sex is also of fundamental importance in demographic statistics. Sex, unlike age, is seldom reported incorrectly in census enumerations or records of vital events. It is therefore not necessary here to develop tests for the accuracy of sex reporting. However, statistics classified by sex may be in error because the enumeration, or the recording of events, is more nearly complete for one sex than for the other. Analysis of statistics by sex, like the tests of age statistics, may therefore be useful in evaluating the reliability of head counts and the completeness of vital statistics registration. In this chapter, certain tests applicable to population data classified by sex are presented together with the tests of age statistics. In addition, the methods explained in the preceding chapters for testing census and vital statistics for both sexes combined, can be applied to data for each sex separately.<sup>1</sup>

#### A. Evaluation of census results classified by sex only, or by sex and broad age groups

In some under-developed countries where a great majority of the population is illiterate, it is very difficult to obtain even approximately correct statements, in a census, of the ages of the people. In such situations, the census returns are sometimes classified by sex only, or for each sex by such indefinite age groupings as "infants", "children", and "adults". Although it is impossible to apply to data of this kind such rigorous tests as are possible where the exact ages in years are recorded, it is useful to examine the consistency of returns for the two sexes and for whatever broad age categories are presented. Such an examination may provide useful indications of apparent deficiencies in the head count.

Testing procedures of the types described below are especially appropriate where there is any *a priori* reason to suspect—in view of the prevailing customs and attitude of the people or the methods of enumeration employed—that there may have been a tendency toward an error in the count affecting especially persons of one sex or age category. If this is the case, the results of the tests should be interpreted with full regard to such

information concerning factors tending to produce an error. However, the application of these tests is advisable even where no *a priori* suspicion of error exists, since they may bring to light errors that have not been suspected.

A first test consists in the computation of sex ratios on the basis of the returns for the whole country and each of its parts. The sex ratio can be expressed as the number of males per 100 females, or females per 100 males. If broad age categories are also given, the sex ratios can usefully be computed separately for each of these categories. Computation of the ratio of children to women is another useful test.

Such ratios, computed for each area, should be compared and evaluated in the light of what is known about conditions in various parts of the country. In general, a population not much affected by migration has approximately (although hardly ever exactly) equal numbers of males and females. This is to be expected because, as a rule, slightly more males than females are born,<sup>2</sup> but female mortality is less than that of males; consequently, a slight excess of males among children is usually more or less counter-balanced by an excess of females among adults. Depending on the precise sex ratio of births and mortality conditions, the total number of females may or may not slightly exceed the total number of males.

This situation, however, is often modified in various ways. If the population is very small, the ratio of males to females is subject to chance variations. Following a war, with many military casualties, the ratio of males to females in the adult population may be abnormally low. Large-scale migration can have a great effect on sex ratios. Very often a majority of the migrants are young men in search of gainful employment, so that the relative numbers of males are reduced in areas of emigration, and increased in areas of immigration. Predominantly female migratory movements also occur, as in the case where many young rural women move to towns where they seek employment as domestic servants. Some knowledge of the types of movements and employment conditions in the country will enable the statistician to appraise the sex-ratios computed from the census data.

The following example, based on Canadian statistics, is given as an illustration of the variations in sex ratios which may in fact occur in different parts of a country. In this example, the ratios computed from the census figures give no basis for suspecting important errors in enumeration.

At the 1951 census of Canada, 7,089,000 males and 6,921,000 females were enumerated. For every 100 females, there were 102.4 males. Since Canada is an immigration country, the excess of males is not surprising. Among the rural population, the ratio was 112.3 males per 100 females; among the urban population, 95.8 males per 100 females. This difference

<sup>1</sup> See also the discussion of tests of vital statistics registration by reference to sex ratios and age patterns in chapter II.

<sup>2</sup> See chapter II, section B.



is related to the differences in employment conditions for the two sexes in the countryside and in towns.

For the rural and urban populations of the various provinces, the ratios are as shown in table 5 below.

Table 5

MALES PER 100 FEMALES IN RURAL AND URBAN AREAS OF THE CANADIAN PROVINCES IN 1951

Province	Sex ratio (males per 100 females)	
	Rural	Urban
Newfoundland .....	111	98
Prince Edward Island.....	111	86
Nova Scotia .....	110	96
New Brunswick .....	109	91
Quebec .....	109	95
Ontario .....	115	96
Manitoba .....	117	94
Saskatchewan .....	117	94
Alberta .....	121	99
British Columbia .....	121	98
Yukon Territory .....	164	120
North West Territories.....	122	184

Conditions in the Yukon Territory and the North West Territories are special; these are located in high northern latitudes and are very sparsely inhabited, by a largely migrant population. In the remaining provinces the sex ratios show a male excess in rural areas and a male deficit in the cities. In Alberta and British Columbia, where the rural excess of males is particularly great, the urban deficit of males is almost negligible. The provinces are listed in their geographic order, from East to West, showing that the rural excess of males diminishes slightly in the progression from Newfoundland to Quebec, and then increases continuously towards the West. A similar, but less regular pattern of variation appears also in the urban sex ratios. This pattern is undoubtedly related to the economic features and demographic history of the country.

A comparison of sex ratios can also be made for communities of various size classes. For the entire

country, the following sex ratios are obtained (males per 100 females) :

- 94 in urban places of 100,000 inhabitants or more;
- 96 in urban places of 30,000 to 99,999 inhabitants;
- 96 in urban places of 10,000 to 29,999 inhabitants;
- 99 in urban places of less than 10,000 inhabitants;
- 110 in rural non-farm areas;
- 117 in rural farm areas.

The sex ratio rises regularly in progression from the most urban to the most intensely rural categories. Analogous ratios for each of the provinces (not presented here) show the same systematic sequence, with only a few minor deviations, as, for example, in the case of Nova Scotia, where the sex ratio for urban places with 10,000 to 29,999 inhabitants was found to be 97, but that for urban places with less than 10,000 inhabitants was only 94. Perhaps some special reason exists to account for this deviant observation.

In other countries, conditions may differ greatly from those noted in the case of Canada. Where there is a predominantly male migration from rural to urban areas, the pattern observed in Canada may be altogether reversed. The presence of mining enterprises, or of plantations employing seasonal labor, for example, may result in selective migrations reflected in the sex ratios. Whatever the special conditions of the case may be, if sex ratios disclose a systematic picture, then there is reason to suppose that the two sexes have been enumerated with comparable accuracy. Deviations from the pattern, which cannot readily be explained, should be singled out for further investigation of the accuracy of the returns. If the entire pattern of sex ratios is erratic there are strong grounds for distrusting the completeness of the enumeration.

If the census also provides figures on "adults" and "children", various other ratios can be computed, such as ratios of children to adults, children to women, women to men, and girls to boys. An example of an analysis of this kind, based on the 1931 census in the Northern Provinces of Nigeria, is reproduced below:<sup>3</sup>

According to the 1931 "census" returns there were 3,497,807 men, 3,898,114 women, and 4,037,178 children. The proportion of children was 37.1 per cent. It oscillated between 29.2 per cent in Niger Province and 41.5 per cent in Zaria Province. The ratio of women to 100 men was 111.4, varying between 92.7 in Plateau Province and 129.5 in Bornu Province. The ratio of children to 100 women was 103.6, varying between 78.4 in Adamawa and 130.2 in Zaria. Large as these differences between Provinces may seem, the differences between the returns for the various Divisions of one and the same Province are in many cases not less startling. I shall quote a few striking examples:<sup>a</sup>

Province	Division	Men	Women	Children	Children per cent	Women to 100 men	Children to 100 women
Kabba	Igala .....	73,096	80,184	72,232	32.0	109.6	90.8
	Igbira .....	30,217	42,263	76,196	51.2	139.8	180.3
	Kabba .....	15,219	22,440	19,557	34.2	147.4	87.2
	Koton Karifi.....	12,339	13,664	5,319	17.0	110.7	37.5
	Abuja .....	23,634	26,912	23,870	32.1	113.8	88.7
Niger	Agai-Lapai .....	20,148	23,915	12,077	21.5	118.6	50.5
	Bida .....	57,759	66,546	47,653	27.7	115.2	71.6
	Kontagora .....	22,163	23,046	23,970	34.7	103.9	104.0
	Kuta .....	23,548	21,300	19,836	30.7	90.4	93.1
	Zungeru .....	12,958	13,176	10,556	28.8	101.6	80.1
Plateau	Jemaa .....	22,436	24,103	29,780	39.0	107.4	123.6
	Jos .....	57,282	41,330	33,685	25.5	72.1	81.5
	Pankshin .....	58,149	54,375	48,228	30.0	93.5	88.7
	Shendam .....	35,859	39,632	42,811	36.2	110.5	108.0
Zaria	Southern .....	28,969	28,459	23,640	29.2	98.2	83.1
	Katsina .....	219,037	284,457	422,354	45.6	129.8	148.5
	Zaria .....	142,279	148,266	141,016	32.7	104.2	95.1

<sup>a</sup> See *Census of Nigeria, 1931*, vol. ii, pp. 33, 52-3. Figures comprise total population.

<sup>3</sup> From R. R. Kuczynski, *Demographic Survey of the British Colonial Empire*, Vol. I, West Africa (1948), p. 608.



The Census Officer made the following comment:

One knows of no such conditions in Kabba Province as would account for the enormous disproportions of non-adults shown in the Division of Koton Karifi . . . and Igbira . . . or for the low figure . . . in Agaie-Lapai Division of Niger Province. In Jos Division, however, where there is a large immigrant adult male population to the mining areas, one would expect to find a lower proportion of adult females and non-adults.<sup>a</sup>

As a matter of fact all ratios of sex and age derived from the General Censuses are open to grave doubts, since these "censuses", also in the Northern Provinces, were based largely on the tag registers and, as regards women and children, on rough estimates or guesses. In some Divisions, such as Katsina, where a large excess of females (30 per cent) was recorded for adults and a large deficiency of females among children (18 per cent), this anomaly may be due to the custom of counting even the youngest wives as adults.<sup>b</sup>

<sup>a</sup> See *Census of Nigeria*, 1931, vol. ii, p. 33.

<sup>b</sup> It should be noted, however, that the ratio of children to women was extraordinarily high in Katsina Division, and that if the number of women was swelled by counting girls as adults, the ratio of children to women would have been higher still than shown in the census returns.

It should be noted that in those censuses in which "adults" and "children" are distinguished without any precise definition of the age limits of these categories (as has to be done in some areas of Africa where people are often quite unaware of their ages, in years), it very frequently happens that females are reported as "adults" at an earlier age than males. Consequently, among the children, there would often be a heavy preponderance of males because adolescent boys aged, say, 15 or 16 would be regarded as "children", while adolescent girls of, say, 13 or 14 years would be regarded as "adults". This results also in an undue preponderance of females among the population considered as "adults".

## B. Examination of detailed age classifications of the population at a single census date

For a proper interpretation of the results of tests applied to more detailed census statistics on ages of the population, it is important to bear in mind that errors in these statistics may be produced both by differences in the relative completeness of enumeration of persons in different age groups, and by mis-statement of the ages of those who are enumerated. The factors which may be responsible for both these types of error, under the conditions which exist in the given country, should also be taken into account.

The causes of mis-statement of ages include:

Ignorance of age, negligence in reckoning the precise age, deliberate mis-statement, and misunderstanding of the question.<sup>4</sup> If ages have been mis-stated mainly be-

<sup>4</sup> The standard definition of age, used in most censuses and vital records, is the number of completed years, that is to say, the exact age, in years, at the last birthday. Local customs in reckoning age sometimes diverge from this definition. In some cases, the question refers to the date of birth instead of the age attained. It is believed by the authorities in many countries that more precise answers are obtained if the question is put in this form, provided, of course, that date of birth is well known to a majority of the population; nevertheless, mis-statements of date of birth analogous to mis-statement of age also occur. The censuses of Brazil require statement of date of birth if known and statement of age if birth date is not known.

cause of ignorance, the returns may nevertheless represent a fairly close approximation to the true distribution of ages. If ages have been deliberately mis-stated, there is more likelihood of a bias in the distribution towards either overstatement or understatement of age, affecting certain broad age ranges.

Balancing equations, such as those already described with reference to the total population, can be applied separately to each sex-age cohort.<sup>5</sup> Failure of these equations to balance may be accounted for by mis-statement of ages, by varying completeness of enumeration or by both. It is often impossible to distinguish clearly the effects of these two kinds of error if one test only has been used. Hence the need to apply several different tests. Examination of the data from different points of view increases the likelihood of a correct interpretation.

The subject of the present chapter cannot be entirely detached from the topics dealt with in chapters I and II. Some tests of completeness of enumeration or vital registration presuppose that age statements are fairly accurate. Some tests of age-accuracy, on the other hand, presuppose reasonably complete enumeration or registration. Direct checks of completeness of a census or vital registers can be conveniently combined with direct checks on the accuracy of age statements, as well as statements relating to various other population attributes.

Some effects of age errors are most readily apparent if statistics by single years of age are examined; others can best be observed in grouped data, preferably the conventional groups of 5 years. The age distribution should therefore be tested in both forms. Essentially the same principles can be applied to single-year and grouped data. Complete examination, in either case, involves four steps, namely: (1) inspection of the data, (2) comparison with an expected configuration, (3) analysis of ratios computed from the data, and (4) measurement of age-accuracy by means of an index.

The use of these methods will be illustrated by a detailed analysis of the age distribution of the 1945 census of Turkey. These data have been selected not only because errors in the age statements were evidently considerable, but also because the frequency and magnitude of errors differ greatly for the two sexes. The types of errors present in the Turkish statistics are common to most statistics relating to age, though they are often of much smaller magnitude.

### INSPECTION OF THE DATA

Table 6 reproduces the results of the 1945 census of Turkey, showing the population by sex and single years of age. The same data are presented in figure 1. It can be seen at a glance that unusually large numbers of individuals were reported at ages which are multiples of 5 and 10, and relatively small numbers at other ages.

Closer inspection of the figures reveals that deficiencies in reported numbers are greatest at ages with final digits 1 or 9; this is not surprising in view of the strong attraction exerted by the figures terminating in 0.

<sup>5</sup> An age-cohort is the number of persons surviving from the same birth-period. Thus, persons aged 35-39 years in 1940 and 45-49 years in 1950 belong to the same "cohort", born in the 1900-1905 period.

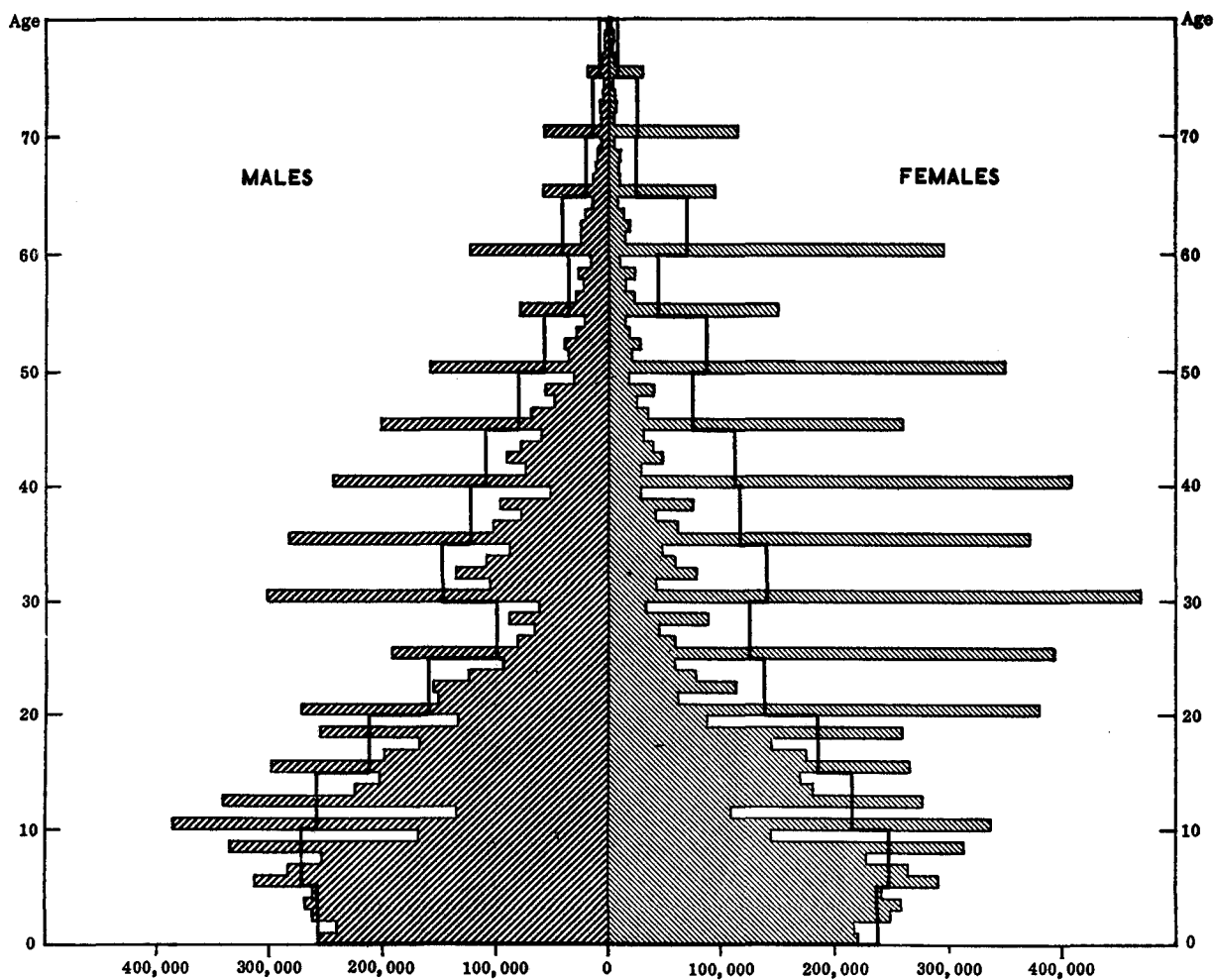


Figure 1. POPULATION OF TURKEY, 1945, BY SEX, BY SINGLE YEARS OF AGE AND 5-YEAR AGE GROUPS, ACCORDING TO CENSUS

Marked deficiencies at ages ending in 4 or 6 are likewise explained by the considerable attraction of the figures with terminal digit 5. As for ages ending in 2, 3, 7, and 8, it can be noted that the "even" figures are somewhat more attractive (or less repugnant) than the "odd" ones; numbers at ages ending in 2 are somewhat larger than those at ages ending in 3; numbers at ages ending in 8 are somewhat larger than those at ages ending in 7.

Finally, the forces of attraction of preferred final digits (or repulsion of disliked digits) are more marked in the case of females than of males, as can be seen from the wider fluctuations in the figures for females. Evidently, the age statements of females are less accurate.

These findings are common to inaccurately reported ages in almost any country. The intensity of attraction and repulsion of certain figures is not always as pronounced as in the present example, nor is the difference between male and female age-accuracy always as great. The details of the pattern of age mis-statements vary somewhat from country to country, but the characteristic features are usually the same.<sup>6</sup>

<sup>6</sup> For a study of the relative attractions and repulsions exerted by various figures of age in different countries, the reader is referred to Bachi's article in the *Bulletin of the International Statistical Institute*, Vol. XXXIII, Part IV.

Whereas these effects of age errors are readily apparent, it is usually difficult to interpret accurately their causes. Probably most age mis-statements result in statements of round figures (for example, persons aged either 39 or 41 may report their ages as 40), though this is not necessarily always the case (a person aged exactly 40 years may report his age as 41 or 39, or some other figure). Where ignorance is the main cause, the great majority of mis-stated ages are likely to be reported in round figures. If ages are deliberately mis-stated, they may often be given at other than round figures in order to create an impression of accurate reporting.

The extent of mis-statements cannot be directly inferred from the data. While mis-statements by one year only are probably most frequent, it is possible that in some cases the error is considerably larger. At advanced ages, mis-statements by several years are probably very frequent.

The data show no immediate evidence as to whether there are any tendencies, in certain age ranges, for overstatements of age to be more frequent than understatements, or *vice versa*. This question is of considerable importance. If errors in both directions are of nearly the same frequency and magnitude, combination of the returns into five-year groups, or the application

Table 6

POPULATION OF TURKEY, BY SEX AND SINGLE YEARS OF AGE, ACCORDING TO THE CENSUS OF 1945 (In thousands)\*

Age	Males	Females	Age	Males	Females	Age	Males	Females
TOTAL 9,447 9,344								
0-4.....	1,287	1,185	35-39.....	607	578	70-74.....	80	133
0.....	256	219	35.....	283	373	70.....	58	116
1.....	240	217	36.....	101	61	71.....	6	5
2.....	262	249	37.....	77	42	72.....	7	6
3.....	268	258	38.....	96	74	73.....	5	4
4.....	261	242	39.....	50	28	74.....	3	2
5-9.....	1,348	1,242	40-44.....	542	558	75-79.....	30	39
5.....	312	292	40.....	244	409	75.....	19	30
6.....	282	265	41.....	73	30	76.....	4	3
7.....	253	228	42.....	90	48	77.....	3	2
8.....	334	313	43.....	76	39	78.....	3	3
9.....	167	144	44.....	59	31	79.....	1	1
10-14.....	1,285	1,074	45-49.....	401	378	80-84.....	29	46
10.....	386	337	45.....	201	260	80.....	24	42
11.....	133	109	46.....	68	36	81.....	2	2
12.....	341	277	47.....	47	25	82.....	2	1
13.....	223	181	48.....	55	40	83.....	1	1
14.....	201	170	49.....	30	17	84.....	1	1
15-19.....	1,049	931	50-54.....	283	434	85-89.....	7.5	9.4
15.....	298	267	50.....	158	350	85.....	5.4	7.2
16.....	198	174	51.....	36	21	86.....	0.8	0.9
17.....	166	144	52.....	38	28	87.....	0.6	0.6
18.....	255	259	53.....	28	19	88.....	0.5	0.5
19.....	132	88	54.....	22	15	89.....	0.3	0.3
20-24.....	789	692	55-59.....	171	219	90-94.....	7.2	11.5
20.....	271	381	55.....	78	150	90.....	6.2	10.4
21.....	150	62	56.....	29	23	91.....	0.3	0.3
22.....	154	113	57.....	22	15	92.....	0.3	0.3
23.....	123	77	58.....	27	23	93.....	0.2	0.2
24.....	92	59	59.....	16	10	94.....	0.2	0.1
25-29.....	484	619	60-64.....	200	349	95-99.....	2.5	3.6
25.....	191	393	60.....	121	297	95.....	1.1	1.8
26.....	79	59	61.....	23	14	96.....	0.3	0.3
27.....	65	46	62.....	23	17	97.....	0.2	0.2
28.....	89	88	63.....	20	13	98.....	0.3	0.3
29.....	61	33	64.....	13	8	99.....	0.7	1.0
30-34.....	731	700	65-69.....	99	125	100 and over...	1.5	2.8
30.....	301	471	65.....	58	94	Unknown .....	11.3	13.1
31.....	103	42	66.....	13	9			
32.....	133	78	67.....	11	8			
33.....	107	60	68.....	10	9			
34.....	87	48	69.....	6	4			

\* Source: Turkey, Central Office of Statistics, *Recensement général de la population du 21 octobre 1945*, Publication no. 286, Vol. no. 65, Ankara, 1950.

of a rather simple smoothing formula, will suffice to remedy most of their defects. If the mis-statements are highly tendentious, it is doubtful whether even a complicated graduation formula can eliminate the resulting distortions in the age distribution.

Figure 1 also shows the 1945 population of Turkey in five-year groups of ages. The corresponding numbers appear in table 7. Since Turkey has a high birth rate, we should expect its age structure to resemble a pyramid, with each successive layer somewhat smaller than the preceding one. Deviations from this pattern in a high birth-rate country are possible if there have been marked previous changes in the birth rate, large immigration or emigration, or special mortality losses affecting particular age groups (for example, military casualties). Turkey has indeed experienced large popu-

lation transfers affecting certain ethnic groups but the transferred populations probably represented a fair cross-section of age groups; hence, no great irregularity in the age distribution is to be expected on this ground. Military casualties during the First World War and subsequent years must have resulted in some reduction in the number of male survivors aged between 40 and 60 years in 1945. During both the First World War and the Second World War, the male population of military ages was mobilized; the absence of many young men from their homes could have resulted in some reduction of the birth rate during both periods and, hence, in reduced numbers of survivors of either sex aged 0-4 and 25-29 by 1945.

Figure 1 and table 7 show the following departures from a regular age pyramid.

1. Numbers aged 0-4 are indeed smaller than those aged 5-9. As it is common, in the censuses of many countries, that small children are enumerated incompletely, one may at first suspect this to be the result of a great deficiency in child enumeration. However, as already pointed out, the differences may have been due partly to mobilization during the Second World War. Possibly both incomplete enumeration and a birth deficit combined to produce this effect.

2. Fewer persons are reported at ages 25-29 than either at ages 20-24 or 30-34. A possible explanation is a marked birth deficit (or increased infant and child mortality) during 1915-1920. If this was the case, it is probable that the relatively small number of children aged 0-4 should also be largely accounted for in the same terms. Closer inspection of the deficit at ages 25-29, however, reveals additional peculiarities. Females aged 30-34 outnumber not only those aged 25-29, but also those aged 20-24; in the case of males, this is not true. This observation cannot be explained in terms of a birth deficit, which would affect both sexes alike; it indicates either that the ages of females were misstated differently from those of males, or that many women aged 20-29 were not enumerated. It has been observed in many censuses that young women tend to understate their ages. In the present instance, it is possible that many women in their early 20's reported their ages in the 'teens, and also that some women in the early 30's reported themselves in the upper 20's. This conclusion is not certain since males may also have mis-stated their ages in some unknown fashion. Moreover, there might have been a tendency to omit

young women in the census enumerations. The First World War birth deficits, if any, could also have continued for one or two years after 1920, owing to internal warfare in Turkey affecting the birth cohort 20-24 years in 1945. A similar analysis of the age returns from an earlier or a later census of Turkey, using some of the methods to be discussed later in this chapter, would be very helpful in evaluating the possibility of explaining the irregularities in terms of a birth deficit. Even without reference to another census, a statistician with intimate knowledge of the country and its people could find the most plausible explanation for the apparent inconsistencies in the returns.

3. Females aged 50-54, 60-64, 70-74, 80-84, 90-94, and males aged 60-64, appear in each case more numerous than those in the preceding 5-year group. It is altogether improbable that birth rates could have fluctuated during the entire half-century from 1850 to 1900 so as to produce these effects. The alternating excesses and deficiencies in these five-year groups must be attributed to the very powerful attraction exerted by figures which are multiples of 10; they imply that many of the people over 50 were quite ignorant of their real ages, that the reporting was very negligent, or that there was a definite reluctance to report the ages exactly.

#### COMPARISON OF DATA WITH AN EXPECTED CONFIGURATION

Nothing is to be gained by a further examination of the Turkish data by single years of age because the effects of attractions of particular figures of age are immediately evident from inspection of the data. However, the statistician examining census age statistics which, though possibly inaccurate, are not so erratic as those of Turkey in 1945, may wish to use additional tests. For some populations, ages are reported fairly well with the exception of the advanced ages. In such instances, it is advantageous to compare the reported numbers at individual years of age with those obtained by some method of graduation or smoothing. For this purpose, comparison with a 5-year moving average may be entirely sufficient. A brief example is given in table 8.

In the case of the Turkish census, more can be learned from a comparison of the returns with an expected pattern of age composition in five-year groups. As has already been mentioned, the Turkish age structure should resemble a pyramid, with the exception of any effects of war-time birth deficits or military casualties. Such a pyramid can be constructed by computing a stable population<sup>7</sup> from the  $L_x$  values<sup>8</sup> of a suitable life table by applying an appropriate geometric rate of increase. Since great precision is not required, any stable population which fits the data tolerably well

<sup>7</sup> A stable population is a population of such age structure as would result if the same rates of mortality and fertility prevailed over a period of several generations. It differs from a stationary population in that the birth rate and the death rate may differ, whereas in a stationary population the birth rate equals the death rate. The actual population of a country with high birth rates often resembles a stable population fairly closely even though mortality has changed in the past.

<sup>8</sup> The  $L_x$  values in a life table indicate the age composition of a stationary population that would result from the mortality rates of the life table and a constant number of births (usually 100,000) every year.

Table 7

POPULATION OF TURKEY, 1945, BY SEX AND 5-YEAR GROUPS OF AGE, ACCORDING TO THE CENSUS (In thousands)\*

Age-group	Males	Females
TOTAL	9,447	9,344
0-4.....	1,287	1,185
5-9.....	1,348	1,242
10-14.....	1,285	1,074
15-19.....	1,049	931
20-24.....	789	692
25-29.....	484	619
30-34.....	731	700
35-39.....	607	578
40-44.....	542	558
45-49.....	401	378
50-54.....	283	434
55-59.....	171	219
60-64.....	200	349
65-69.....	99	125
70-74.....	80	133
75-79.....	30	39
80-84.....	29	46
85-89.....	8	9
90-94.....	7	11
95 and over.....	4	6
Unknown .....	11	13

\* Source: Turkey, Central Office of Statistics, *Recensement général de la population du 21 octobre 1945*, Publication no. 286, Vol. no. 65, Ankara, 1950.

Table 8

POPULATION OF THE UNITED STATES, 1940, REPORTED IN THE CENSUS AT AGES 80 TO 90, COMPARED WITH 5-YEAR MOVING AVERAGES

Age	Population reported at census	5-year moving average	Excess (+) or deficit (-) of reported numbers	Excess or deficit as percentage of moving average
78.....	65,045			
79.....	58,066			
80.....	53,512	50,189	+ 3,323	+ 6.6
81.....	38,307	43,245	- 4,938	- 8.9
82.....	36,016	36,831	- 715	- 2.2
83.....	30,324	30,318	+ 6	+ 0.0
84.....	25,997	25,844	+ 153	+ 0.6
85.....	20,946	21,232	- 286	- 1.3
86.....	15,938	17,134	- 1,196	- 7.0
87.....	12,957	13,481	- 524	- 3.9
88.....	9,834	10,610	- 776	- 7.3
89.....	7,728	8,166	- 438	- 5.4
90.....	6,596	6,162	+ 434	+ 7.0
91.....	3,716			
92.....	2,932			

can be used. In the present instance, a good fit was obtained by applying a 2 per cent annual rate of increase to the stationary population of the 1936-1938 life table for Egypt.<sup>9</sup> This was done by applying successive powers of 1.02<sup>5</sup> to the  $L_x$  values (that is to say, the stationary population in five-year age groups).

Ordinarily, a comparison of a stable with an actual population is made by multiplying the figures of the stable population by a constant, namely, the ratio between the totals of the two populations. In the present instance, in order to allow for possible birth deficits resulting in reduced numbers of persons aged 0-4 and 25-29, the stable population was multiplied, separately for each sex, by the ratio between the total stable and actual populations for all ages other than 0-4 and 25-29. The results are shown in table 9 and figure 2.

Since the comparison is rough, no importance should be attached to the exact size of the differences between the census data and the stable population figures for any particular age-sex category. Nevertheless, significant observations can be made. There is a considerable deficit at ages 0-4, which may represent in part the effect of a decline in the birth rate during 1940-1945. This deficit, however, appears to be larger for males than for females. On the other hand, the reported numbers of males appear markedly in excess at ages 10-14. A possible, but not necessarily correct, explanation would be a tendency to overstate the ages of boys under 10 and to understate the ages of boys just past 15 years of age. At ages 20-24 and 25-29, both sexes show deficient numbers, but this deficit is differently

<sup>9</sup> These assumptions do not necessarily imply that Turkish mortality has at any time been identical with that of Egypt in 1936-1938, nor that the Turkish population has always been increasing at 2 per cent per annum. Provided that the birth rate has been fairly constant, the stable population selected for this illustration may still provide a good fit if death rates in the past have been higher (and the rate of population increase consequently lower) or *vice versa*.

Table 9

POPULATION OF TURKEY, 1945, BY SEX AND 5-YEAR AGE GROUPS. COMPARISON OF CENSUS FIGURES WITH STABLE POPULATION

(Numbers in thousands)

Age-group	Males			Females		
	Census (1)	Stable population (2)	Ratio (1):(2)	Census (1)	Stable population (2)	Ratio (1):(2)
TOTAL	9,435*	10,148	0.93	9,330*	9,772	0.95
0-4.....	1,287	1,716	0.75	1,185	1,516	0.78
5-9.....	1,348	1,334	1.01	1,242	1,180	1.05
10-14.....	1,285	1,160	1.11	1,074	1,052	1.02
15-19.....	1,049	1,013	1.03	931	937	0.99
20-24.....	789	886	0.89	692	831	0.83
25-29.....	484	768	0.63	619	734	0.84
30-34.....	731	662	1.10	700	644	1.09
35-39.....	607	569	1.25	578	561	1.03
40-44.....	542	485	1.12	558	487	1.15
45-49.....	401	408	0.98	378	421	0.90
50-54.....	283	337	0.84	434	360	1.21
55-59.....	171	271	0.63	219	303	0.72
60-64.....	200	209	0.96	349	250	1.40
65-69.....	99	152	0.65	125	199	0.63
70-74.....	80	98	0.81	133	144	0.92
75-79.....	30	53	0.58	39	90	0.43
80-84.....	29	21	1.36	46	43	1.07
85 and over.....	19	7	2.60	27	20	1.34

\* Excluding persons of unknown age.

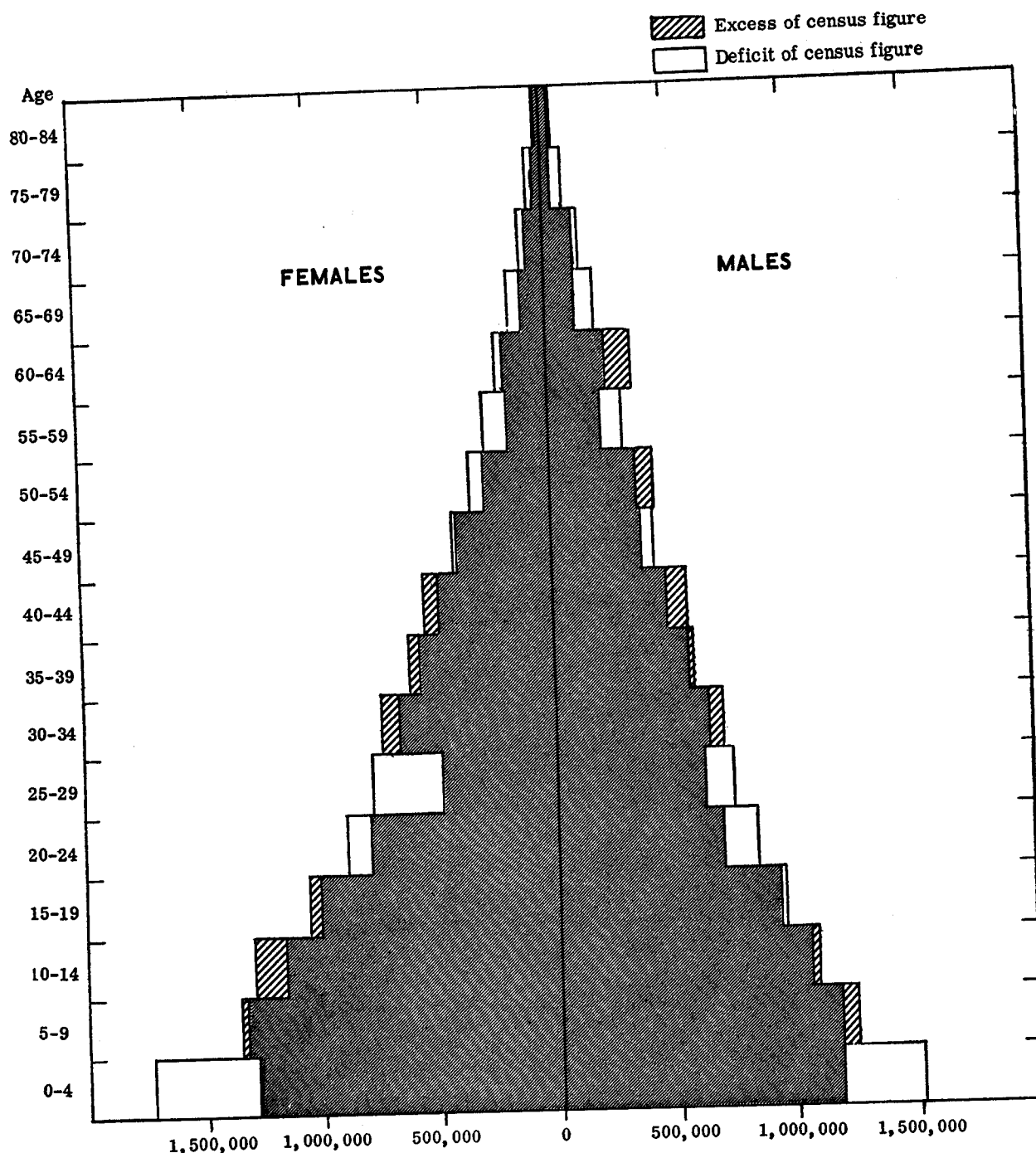


Figure 2. POPULATION OF TURKEY, 1945, BY SEX AND 5-YEAR AGE GROUPS. COMPARISON OF CENSUS FIGURES WITH A STABLE POPULATION

distributed in the two cases. The deficiency in the case of females is about the same at ages 20-24 as at 25-29. It could be explained as the result of a birth deficit only if the birth rate were as much depressed in 1920-1925 as in 1915-1920, which seems very unlikely. The relative deficits of males aged 20-24 and 25-29 could more plausibly be accounted for by the possible birth deficit. The observations tend to confirm that either the ages of many young women were mis-stated tentatively, or many women in the age-range 20-29 were omitted from the census count.

The comparison of the census data with the stable population becomes less and less accurate as we proceed to more advanced ages. It is unlikely that the hypotheses of 1936-1938 Egyptian mortality and an annual rate of natural increase of 2 per cent are realistic for the elder cohorts, who lived under the conditions of a more remote past. Nevertheless, the comparison does not lose all meaning. The consistent deficit of males above age 45 could be attributable to military casualties during the First World War and the years immediately following. The alternation of

excesses and deficits among females aged over 40, on the other hand, shows once more the attraction of ages ending in zero.

#### ANALYSIS OF RATIOS COMPUTED FROM THE DATA

For the examination of the accuracy of age statistics, it is useful to compute sex ratios for each age category, and ratios for each sex separately, between the numbers reported in adjacent age categories. These ratios can be defined in several ways. In the present context, we shall define the "sex-ratio" as the number of males per 100 females in the same age class; and the "age-ratio" as 100 times the number of persons in a given age class divided by the arithmetic average of numbers in the two adjoining age categories. Sex-ratios should ordinarily change only very gradually from one age to another, as they are determined mainly by the sex ratio of births and sex differences in mortality at various ages. Age-ratios should ordinarily deviate very little from 100, except at advanced ages or as a result of major fluctuations in past birth rates.<sup>10</sup> The variations in sex-ratios and age-ratios computed from grouped age data, which may occur as a result of age misstatement or for other reasons, have been studied by the Population Branch of the United Nations.<sup>11</sup>

<sup>10</sup> Migration and military casualties may disturb the normal patterns of sex-ratios and age-ratios.

Sex-ratios and age-ratios, computed by single years of age from the data of the Turkish census of 1945, are shown in table 10. It will be noted that the sex-ratios are invariably low at those round ages upon which the reports tend to be concentrated, and high at other ages, the reason for this being more frequent mis-statement of ages of females than of males. It would also appear that the enumeration of female infants was less complete than that of male infants, as witnessed by high sex-ratios at ages 0 and 1. Age-ratios reflect in a certain way the relative powers of attraction of certain figures of age; up to age 18, these ratios are very similar for males and females. At ages above 20, the attractive force of round figures is usually several times as great in the case of females as in that of males. At advanced ages, the reports for males are almost as erratic as for females.

Ratios of numbers reported at certain digits of age may or may not reveal any tendencies which exist towards overstatement or understatement of age. For example, one may compare, for each age which is a multiple of 10, the numbers of persons reporting the

<sup>11</sup> See "Accuracy tests for census age distributions tabulated in five-year and ten-year groups", United Nations, *Population Bulletin*, No. 2, October 1952, pp. 59-79. For a study of the effect of inaccurate age statements on sex ratios, see also Louis Henry, "La masculinité par âge dans les recensements", *Population*, 3<sup>e</sup> année, no. 1, 1948.

Table 10

SEX-RATIOS AND AGE-RATIOS COMPUTED FROM SINGLE-YEAR AGE DATA OF THE TURKISH CENSUS, 1945\*

Age	Sex-ratio	Age-ratios		Age	Sex-ratio	Age-ratios		Age	Sex-ratio	Age-ratios	
		Males	Females			Males	Females			Males	Females
0.....	117	—	—	33.....	178	98	96	66.....	149	39	18
1.....	111	93	93	34.....	182	45	22	67.....	137	90	86
2.....	105	103	105	35.....	76	301	688	68.....	111	127	154
3.....	104	102	105	36.....	166	56	29	69.....	130	16	7
4.....	108	90	88	37.....	182	73	63	70.....	50	985	2481
5.....	107	115	115	38.....	131	152	209	71.....	123	19	8
6.....	106	100	102	39.....	177	42	12	72.....	120	129	139
7.....	111	82	79	40.....	60	399	1399	73.....	139	101	90
8.....	107	159	168	41.....	239	43	13	74.....	134	25	14
9.....	116	46	44	42.....	187	122	140	75.....	66	563	1049
10.....	115	258	266	43.....	196	102	98	76.....	116	35	21
11.....	123	37	35	44.....	190	42	21	77.....	134	72	61
12.....	123	192	191	45.....	78	317	773	78.....	109	166	193
13.....	123	82	81	46.....	189	55	25	79.....	114	10	5
14.....	118	77	76	47.....	188	76	65	80.....	57	1616	3089
15.....	112	149	155	48.....	136	144	191	81.....	104	13	7
16.....	114	85	85	49.....	170	28	9	82.....	119	131	122
17.....	115	73	67	50.....	45	479	1806	83.....	123	86	87
18.....	99	171	223	51.....	170	37	11	84.....	120	25	17
19.....	150	50	27	52.....	138	118	137	85.....	75	671	937
20.....	71	193	508	53.....	147	94	89	86.....	93	27	22
21.....	241	70	25	54.....	142	41	18	87.....	97	88	86
22.....	137	113	162	55.....	52	310	781	88.....	102	114	109
23.....	160	100	90	56.....	125	57	28	89.....	96	9	6
24.....	156	59	25	57.....	148	79	65	90.....	59	2155	3269
25.....	49	224	667	58.....	118	141	186	91.....	84	9	6
26.....	133	62	27	59.....	169	22	6	92.....	99	134	128
27.....	140	77	63	60.....	41	613	2489	93.....	112	89	82
28.....	101	141	221	61.....	163	32	9	94.....	116	24	14
29.....	184	31	12	62.....	139	107	122	95.....	63	546	892
30.....	64	366	1248	63.....	153	109	104	96.....	95	37	27
31.....	244	48	15	64.....	159	33	15	97.....	113	86	65
32.....	170	126	152	65.....	62	442	1095	98.....	81	—	—

\* The sex-ratio is here defined as the number of males of a given age per 100 females in the same age group; the age-ratio is defined as the number of persons of a given age group per 100 of the mean of numbers of the two adjoining age groups, of the same sex.

preceding age (terminal digit 9) with the number reporting the next subsequent age (terminating in 1), by taking the ratio of those reporting age 9 to those reporting age 11; of those reporting age 19 to those reporting age 21 et cetera. If it can be presumed that a large part of age mis-statements are by one year only, then a tendency to overstate age might result in larger

numbers reported with terminal digit 1 than with terminal digit 9; conversely, a tendency toward understatement might be reflected by a low ratio of numbers at ages ending in 1 to numbers at ages ending in 9. It must be understood, however, that these ratios are also affected by various other forms of age errors. From the Turkish data, the following ratios are obtained.

Ages which are multiples of 10 (x)	Males				Females	
	Preceding age (x-1)		Following age (x+1)		Ratio of number at age (x-1) to number at age (x+1)	Ratio of number at age (x-1) to number at age (x+1) <sup>a</sup>
	Age	Number reported	Age	Number reported		
10.....	9	161,884	11	133,109	1.3	1.3
20.....	19	131,855	21	149,548	0.9	1.4
30.....	29	61,159	31	103,311	0.6	0.8
40.....	39	49,995	41	72,579	0.7	0.9
50.....	49	29,677	51	36,309	0.8	0.8
60.....	59	16,212	61	23,288	0.7	0.7

<sup>a</sup> Ratios computed in the same manner as in the case of males.

If high ratios in this case really indicate the prevalence of age understatement, and low ratios indicate overstatement, it appears that ages were understated around age 10 and overstated around ages 30, 40, 50 and 60. This interpretation is at best only partly correct. Relatively greater depletion of numbers at age 11 than at age 9, for instance, is probably in part a result of the strong attraction of age 12.

The differences between the ratios obtained for the two sexes are noteworthy. The ratio for females is much higher than that for males around age 20, an age at which women may readily be suspected of a desire to appear younger than they are. For ages 30 and 40, the female ratios are only slightly higher than the male, and for ages 50 and 60 the ratios for the two sexes are equal. Around ages 50 and 60, male and female sensitivities with respect to age may be nearly the same. The computation of sex-ratios and age-ratios for five-year age groups constitutes the basis of a method developed by the United Nations staff for examining the accuracy of age statistics.<sup>12</sup> It is useful to compute both ratios and examine them critically, particularly if there is some difficulty in constructing a stable population or some other expected age distribution with which the census data can be reasonably well compared. In the case of the 1945 Turkish census, analysis of the sex-ratios and age-ratios, for five-year groups, would yield essentially the same conclusions reached by the comparison with the stable population.

#### MEASUREMENT OF AGE-ACCURACY BY MEANS OF AN INDEX

It is of some interest to measure age-accuracy by an index for comparative purposes to establish, for example, whether the age statistics of one census are more accurate than those of another, whether there are great differences in age-accuracy between the urban and rural population et cetera. In the following, four methods of computing an index will be presented. Each of these methods has certain advantages and certain shortcomings. In particular, it is to be noted that three of these methods measure the extent of *digit-preference*

rather than age-accuracy in a wider sense. Since not all age mis-statements result in statements at preferred digits, and since digit-preference can be present whether or not age mis-statements have any directional tendencies, a measure of digit-preference should be interpreted with some reservations. The method mentioned above employed by the United Nations staff is sensitive to both tendentious mis-statement and digit-preference, but has other defects which will be discussed.

#### Whipple's index

This index is obtained "by summing the age returns between 23 and 62 years inclusive and finding what percentage is borne by the sum of the returns of years ending with 5 and 0 to one-fifth of the total sum. The result would vary between a minimum of 100, representing no concentration at all, and a maximum of 500, if no returns were recorded with any digits other than the two mentioned".<sup>13</sup> The choice of 23 and 62 as the limits of the age band to be examined is arbitrary but has been found most suitable for practical purposes.

For the male and female population of Turkey as of the 1945 census, Whipple's index can be computed as follows (numbers in thousands):

Age	Males	Females	Age	Males	Females
23.....	123	77	25.....	191	393
24.....	92	59	30.....	301	471
25-29.....	484	619	35.....	283	373
30-34.....	731	700	40.....	244	409
35-39.....	607	578	45.....	201	260
40-44.....	542	558	50.....	158	350
45-49.....	401	378	55.....	78	150
50-54.....	283	434	60.....	121	297
55-59.....	171	219	Sum,	1,577	2,703
60.....	121	297	ages		
61.....	23	14	multiple		
62.....	23	17	of 5		
Sum,	3,601	3,950			
ages 23-62					
Index for males			Index for females		
$\frac{5 \times 1,577}{3,601} = 219.0$			$\frac{5 \times 2,703}{3,950} = 342.2$		

<sup>13</sup> Census of India, 1921, Vol. I, Part I, Report by J. T. Marten, Calcutta, 1924, pp. 126-127.

<sup>12</sup> Loc. cit. See also pp. 42-43, *infra*.



Similar computations for the combined population of Turkish cities with more than 30,000 inhabitants yield an index of 146.3 for males and 223.3 for females. For the total of Turkish provincial and district capitals, the index is 165.8 for males and 292.8 for females. For the remainder of the country's population, which is mainly rural, the index amounts to 238.8 for males and 364.6 for females.

For purposes of comparison, the results of similar computations for other areas are shown below.

Bengal, 1901:		United States:	
Males .....	277.3	1880 .....	144.8
Females .....	292.3	1890 .....	131.3
Russia, 1897 .....	175.2	1900 .....	119.8
USSR, 1926 .....	159.1	1910 .....	120.9
Brazil, 1940:		1920 .....	115.7
Males .....	143.3	1930 .....	115.5
Females .....	153.1	1940 .....	109.7
Poland, 1921 .....	134.8	Sweden, 1920 .....	100.4

Whipple's index is a very effective measure of age-accuracy so far as digit preference is concerned, and has the advantage that it can be computed very easily. Its main drawback, apart from measuring digit-preference only, is that it measures the preferences for only two digits, 0 and 5. As the preference for particular digits may be influenced by linguistic and other habits, it is probable that the average degree of preference for 0 and 5 varies among different peoples though, in general, these are always the most preferred digits.

### Myers' index<sup>14</sup>

This index reflects preferences or dislikes for each of the ten digits, from 0 to 9. To determine such preferences, one might take successive sums of numbers recorded at ages ending in each of these digits; such a simple method, however, does not suffice since, with advancing terminal digits of age these sums will tend to decrease (for each successive digit, the population is one year older than for the preceding one). To avoid this inconvenience, the first step in Myers' method consists in the computation of a "blended" population in which ordinarily almost equal sums are to be expected for each digit. This being the case, the "blended" totals for each of the ten digits should be very nearly 10 per cent of their grand total. The deviations of each sum from 10 per cent of the grand total are added together, irrespective of whether they are positive or negative, and their sum is Myers' index.

The method of "blending" and the computation of the index are illustrated by means of the Turkish data in table 11. First, the sums of numbers at all ages terminating in each digit are computed for ages 10 and over, and for ages 20 and over. By multiplying the former with the successive coefficients 1, 2, 3 . . . 10 (for digits 0, 1, 2 . . . 9) and the latter with the successive coefficients 9, 8, 7 . . . 0 (for digits 0, 1, 2 . . . 9) and adding the results, the "blended" population

<sup>14</sup> This index is explained by R. J. Myers in "Errors and Bias in the Reporting of Ages in Census Data", *Actuarial Society of America. Transactions*, Vol. XLI, 1940, pp. 411-415.

Table 11

APPLICATION OF MYERS' METHOD TO AGE DATA FOR MALES, ACCORDING TO TURKISH CENSUS OF 1945  
(Numbers in thousands)

Terminal digit	Numbers at ages specified									Sum for ages 10-99	Sum for ages 20-99
	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89	90-99		
0.....	386	271	301	244	158	121	58	24	6	1,569	1,183
1.....	133	150	103	73	36	23	6	2	0	526	393
2.....	341	154	133	90	38	23	7	2	0	788	447
3.....	223	123	107	76	28	20	5	1	0	583	360
4.....	201	92	87	59	22	13	3	1	0	478	277
5.....	298	191	283	201	78	58	19	5	1	1,134	836
6.....	198	79	101	68	29	13	4	1	0	493	295
7.....	166	65	77	47	22	11	3	1	0	392	226
8.....	255	89	96	55	27	10	3	0	0	535	280
9.....	132	61	50	30	16	6	1	0	1	297	165
Sum....	2,333	1,275	1,338	943	454	298	109	37	8	6,795	4,462

Terminal digit	Ages 10-99			Ages 20-99			"Blended" Sum	Percentage distribution	Deviation from 10 per cent
	Sum	Coefficient	Product	Sum	Coefficient	Product			
0.....	1,569	1	1,569	1,183	9	10,647	12,216	21.7	+ 11.7
1.....	526	2	1,052	393	8	3,144	4,196	7.5	- 2.5
2.....	788	3	2,364	447	7	3,129	5,493	9.8	- 0.2
3.....	583	4	2,332	360	6	2,160	4,492	8.0	- 2.0
4.....	478	5	2,390	277	5	1,385	3,775	6.7	- 3.3
5.....	1,134	6	6,804	836	4	3,344	10,148	18.0	+ 8.0
6.....	493	7	3,451	295	3	885	4,336	7.7	- 2.3
7.....	392	8	3,136	226	2	452	3,588	6.4	- 3.6
8.....	535	9	4,815	280	1	280	5,095	9.0	- 1.0
9.....	297	10	2,970	165	0	0	2,970	5.3	- 4.7
Sum .....							56,309	100.1	39.3

for each digit is obtained. As was shown by Myers, all these sums should be very nearly equal if ages are reported accurately. In the present example, they are markedly unequal and, by adding together, irrespective of sign, the deviations from 10 per cent of the grand total, an index of 39.3 is obtained.

In the case of the Turkish data, positive deviations appear only at terminal digits 0 and 5; hence, in this instance, the advantage of Myers' over Whipple's method is not obvious. There are, however, populations for which positive deviations are also found at other age digits, such as 2 or 8, which would be overlooked if Whipple's index only were computed.

Theoretically, Myers' index can vary from 0 to 180. If ages are reported accurately, all "blended" sums are very nearly equal and deviations from 10 per cent are negligible, resulting in an index of almost zero. If all ages were reported with the same terminal digit (for example digit 0), then 100 per cent of the "blended" total would be at this digit; the absolute sum of deviations from 10 per cent would then amount to 180. The following indices, computed for male populations, provide a few examples of the results which may be obtained from actual observations:

Bengal, 1901 .....	62.6
Russia, 1897 .....	20.5
Brazil, 1940 .....	16.3
Australia, 1933 .....	4.0
Sweden, 1939 .....	1.2

#### *Bachi's index<sup>15</sup>*

Despite their demonstrated practical usefulness, both Myers' and Whipple's indices have some minor theoretical defects: it is not possible to formulate the precise theoretical conditions under which Whipple's index would be exactly 100 and Myers' index exactly zero. An ingenious method, developed by Bachi, avoids this defect. Because the computation of Bachi's index is somewhat more laborious than Myers' method, no detailed example is given in this manual.

Bachi's method consists, in principle, in a repetitive application of Whipple's method, to determine the relative preference given to each of the ten terminal digits. If the age range is suitably selected, very nearly 10 per cent of all persons within that age range should be expected at ages with each terminal digit. For theoretical reasons, as well as to satisfy practical requirements, the average of total numbers in two age ranges must be related to the number of persons reporting, within these ranges, the ages with a particular terminal digit. The following relations are then established:

1. The per cent relationship of age reports at ages 30, 40, 50, 60, and 70 to the average of all age reports in the age ranges 25-74 and 26-75;
2. The per cent relationship of age reports at ages 31, 41, 51, 61, and 71 to the average of all age reports in the age ranges 26-75 and 27-76;
3. The per cent relationship of age reports at ages 32, 42, 52, 62, and 72 to the average of all age reports in the age ranges 27-76 and 28-77;
4. The percentage formed by age reports at ages 33, 43, 53, and 63, as well as half the numbers at ages

<sup>15</sup> Roberto Bachi, "Measurement of the tendency to round off age returns", *Proceedings of the International Statistical Congress*, Rome, 1953.

23 and 73, in the average of all age reports in the age ranges 23-72 and 24-73;

5. The percentage formed by age reports at ages 34, 44, 54, and 64, as well as half the numbers at ages 24 and 74, in the average of all age reports in the age ranges 24-73 and 25-74;

6. The percentage formed by the age reports at ages 35, 45, 55, and 65, as well as half the numbers at ages 25 and 75, in the average of all age reports in the age ranges 25-74 and 26-75;

7. The percentage formed by age reports at ages 36, 46, 56, and 66, as well as half the numbers at ages 26 and 76, in the average of all age reports in the age ranges 26-75 and 27-76;

8. The percentage formed by age reports at ages 37, 47, 57, and 67, as well as half the numbers at ages 27 and 77, in the average of all age reports in the age ranges 27-76 and 28-77;

9. The percentage of age reports at ages 28, 38, 48, 58, and 68 of the average of all age reports in the age ranges 23-72 and 24-73.

10. The percentage of age reports at ages 29, 39, 49, 59, and 69 of the average of all age reports in the age ranges 24-73 and 25-74.

Deviations of percentages from 10 per cent can be taken as indices of preference or dislike for each of the ten digits. Bachi's index is then obtained by summing only the positive deviations from 10 per cent (the sum of positive deviations is theoretically equal to the sum of negative deviations). It is, therefore, of the order of one-half of Myers' index, with a theoretical range from 0 to 90. In practice, almost the same results are obtained by both these methods.<sup>16</sup>

#### *The United Nations Secretariat method<sup>17</sup>*

This method consists essentially in the computation of sex-ratios and age-ratios for five-year groups of ages, up to age 70.<sup>18</sup> In the case of sex-ratios, successive differences between one age group and the next are noted, and their average is taken, irrespective of sign. In the case of age-ratios, for either sex, deviations from 100 are noted and averaged irrespective of sign. Three times the average of sex-ratio differences is then added to the two averages of deviations of age ratios from 100, to compute the index. This procedure is illustrated, with the Turkish data, in table 12.

This method, unlike the ones described above, is applicable where single-year age data are not available. The resulting index is not very exact and should be regarded as an "order of magnitude" rather than a precise measurement. Various limitations must be borne in mind when applying this method. Exceptions must be allowed for in the case of irregularities arising in certain age groups as a result of real disturbances of the population trend, due to such factors as war casual-

<sup>16</sup> For a description of the method and presentation of results obtained, the reader is referred to R. Bachi, "The Tendency to Round off Age Returns: Measurement and Correction", *Bulletin of the International Statistical Institute*, Vol. XXXIII, Part IV, pp. 195-222.

<sup>17</sup> For a full account of this method, the reader is referred to "Accuracy tests for census age distributions tabulated in five-year and ten-year groups", United Nations, *Population Bulletin* No. 2, October 1952, pp. 59-79.

<sup>18</sup> Sex-ratios and age-ratios as defined on p. 22 *supra*.

Table 12  
COMPUTATION OF AGE-ACCURACY INDEX BY THE UNITED NATIONS SECRETARIAT METHOD FROM STATISTICS OF THE  
1945 CENSUS OF TURKEY

Age-group	Reported number		Analysis of sex-ratios		Analysis of age-ratios (males)		Analysis of age-ratios (females)	
	Males	Females	Ratios	Successive difference	Ratios	Deviation from 100	Ratio	Deviation from 100
0-4.....	1,286,705	1,184,799	108.6	— 0.1	104.9	+ 4.9	110.0	+ 10.0
5-9.....	1,348,446	1,242,281	108.5	+ 11.1	107.2	+ 7.2	98.8	— 1.2
10-14.....	1,284,952	1,074,080	119.6	— 7.0	101.1	+ 1.1	105.5	+ 5.5
15-19.....	1,048,701	931,461	112.6	+ 1.5	103.0	+ 3.0	89.2	— 10.8
20-24.....	789,205	691,680	114.1	— 35.9	63.7	— 36.3	89.0	— 11.0
25-29.....	484,328	619,069	78.2	+ 26.3	134.0	+ 34.0	116.9	+ 16.9
30-34.....	731,283	699,657	104.5	+ 0.5	95.4	— 4.6	92.0	— 8.0
35-39.....	607,377	578,390	105.0	— 7.8	107.5	+ 7.5	116.6	+ 16.6
40-44.....	542,301	558,000	97.2	+ 8.8	97.3	— 2.7	76.3	— 23.7
45-49.....	401,379	378,499	106.0	— 40.8	98.8	— 1.2	145.2	+ 45.2
50-54.....	282,856	434,107	65.2	+ 12.8	70.9	— 29.1	56.0	— 44.0
55-59.....	171,162	219,398	78.0	— 20.8	148.2	+ 48.2	203.0	+ 103.0
60-64.....	199,908	349,207	57.2	+ 22.0	70.5	— 29.5	51.7	— 48.3
65-69.....	98,664	124,606	79.2	—	—	—	—	—
70-74.....	80,007	132,964	—	—	—	—	—	—
Total (irrespective of sign)				195.4		209.3		344.2
Mean (total divided by 13)				15.0		16.1		26.5
Index (3 times mean difference sex-ratios plus mean deviations of male and female age-ratios.)						87.6		

ties, temporary birth deficits, or migratory movements involving mainly certain sex-age groups. For small populations, the measurement is also affected by chance fluctuations. The present example, however, illustrates an advantageous use of the method. The population of Turkey is large enough to permit neglecting any effect of chance fluctuations. Moreover, errors in age statements are sufficiently pronounced to overshadow whatever real irregularities of age structure may have existed.

The United Nations Secretariat method has the advantage over the methods of Whipple, Myers and Bachi that the index which is obtained is affected by differential omission of persons in various age groups from the census count and by tendentious age misstatement as well as by digit-preference and is therefore more truly a reflection of the general accuracy of the age statistics. Also, it provides an indication of accuracy of the data in the form in which they are used for most purposes, that is, in age groups rather than single years. The methods applied to date by single years of age may in some cases show a fairly large amount of age misstatement which has little influence on the grouped data.

### C. Examination of detailed age statistics from two or more censuses

It is often difficult, as the examples presented above have shown, to determine whether irregularities revealed by tests of the age returns from a single census are due mainly to errors in the data or to real peculiarities of the population structure. When the results of two or more successive censuses are available, it is often possible to clear up these uncertainties even with-

out the use of any more elaborate techniques than were described in the preceding section. For example, if the age statistics from a 1955 census of Turkey were at hand, the possibility of explaining certain irregularities in the 1945 data as the results of birth deficits or war casualties in the periods 1915 to 1925 and 1940 to 1945 would be greatly clarified. If the 1955 figures should show the same peculiarities in the age groups ten years higher, but not in the same age groups in which they appeared in 1945, there would be a strong basis for concluding that these peculiarities reflected the actual facts, rather than errors.

Still more definite information regarding errors can be obtained where data from two or more censuses at intervals of a few years are available, by using balancing equations or analagous calculations with the data for particular cohorts—comparing, for example, the numbers reported at ages 10-14 in an earlier census with those reported at ages 20-24 in a census ten years later. Where data from a series of three or more censuses are available, the returns may be linked in this manner over the entire series, with very illuminating results. For the purpose of explaining the techniques, however, it is sufficient to consider examples of the use of data from two censuses.

For the sake of simplicity in calculations, examples will be taken from censuses taken ten years apart. The balancing equation can be applied to any period of years, but it is arithmetically easiest when the interval is either five or ten years.

Not only the accuracy of age reporting, but also the completeness of enumeration of particular age and sex groups can be tested by proper use of the balancing equation under certain conditions which will be described below. On the other hand, balancing equation

tests, unlike the procedures described in the preceding section, cannot often be applied to the age statistics for component areas of the country, because of the lack, in most cases, of adequate information on internal migration. For this reason, the methods to be illustrated here are useful, generally, only for an entire country, unless adequate data on internal migration are available.

In some countries it may be possible to apply these tests to the various ethnic groups separately, if age and sex data are tabulated for such groups and if data are available on immigration and emigration of these groups (or if the groups in question are not substantially affected by international migration.)

For the purpose of this analysis, any particular age group can be defined as a cohort: boys under 5 years of age, women 50 to 54 years, all persons 10 to 19 years of age et cetera, at a given census date. If a second census is taken exactly one decade later, the members of each cohort will be exactly ten years older at the time of the second census. In the meantime their numbers will be reduced by deaths and they may be increased or reduced by the balance of immigration and emigration. Ordinarily, mortality is the main factor; if the migration balance is negligible, the change in numbers can be used to compute a survival ratio analogous to that of a life table. For example, the Japanese census of 1940 reported 3,670,000 women aged 15 to 19 and the 1950 census reported 3,354,000 aged 25 to 29. The survival rate between 1940 and 1950 for this cohort was 3,354,000 divided by 3,670,000 or .913 (on the assumption that the data were accurate and that net migration was negligible).

Computed for one cohort only, such a survival rate often reveals little, if anything, about the accuracy of the statistics. However, a patently absurd result may be obtained, giving clear evidence of error. For example, an increase in the numbers of a cohort, from one census to another, is obviously impossible, unless there has been a substantial amount of immigration. On the other hand, even under conditions of very high mortality, it is unlikely that a cohort aged, say, anywhere between 5 and 60 years at the beginning, will be reduced by one-half within a decade.

More accurate judgement is possible if the survival rates are compared for cohorts of each sex at different ages. Survival rates are functions of age-specific death rates, and, like these, generally conform to more or less the same pattern of variation from age to age whether mortality is high or low. The rate of survival increases after the earliest years of childhood and usually attains its maximum around age 10; thereafter it declines, at first very gradually, but more and more rapidly as advanced ages are attained. Also, at most or all ages females have a somewhat higher rate of survival than do males of the same ages. If the hypothetical survival rates computed for different cohorts deviate significantly from this pattern, and if no explanation, such as migration, can be found, inaccuracy in the statistics must be suspected.

Table 13 presents ten-year survival rates for five-year cohorts, according to the life tables of three countries.<sup>19</sup> The purpose of this table is to indicate the approximate magnitudes and the general patterns of survival rates which should be expected. In passing, we may observe that the survival rate for women aged 15 to 19 according to the Japanese life table of 1947 was 0.935, slightly higher than the rate computed from census statistics for 1940 and 1950. Since the 1940-1950 period includes most of the Second World War, when mortality was probably higher than in 1947, the rate of 0.913 from the census material may be quite consistent with the rate of 0.935 derived from the life table.

Under what conditions can such comparisons of cohorts in successive censuses be made most meaningfully? One condition is either the absence of a substantial net immigration or emigration or full knowledge about the age and sex composition of the migrants. A second condition, analogous to the first, is that of constant boundaries. If the country's boundaries have changed between the two censuses so that considerable numbers of people have been added to or subtracted from the population, the age and sex composition of these people must be known, if the cohort analysis is to give an accurate indication of the accuracy of the

<sup>19</sup> Survival ratios computed from a life table are ratios of the approximate  $L_x$ -values in the stationary population.

Table 13

TEN-YEAR SURVIVAL RATES FOR FIVE-YEAR COHORTS, BY SEX, ACCORDING TO LIFE TABLES OF THREE COUNTRIES

Age at beginning of decade	Age at end of decade	Thailand 1947/1948		Japan 1947		Mexico 1930	
		Males	Females	Males	Females	Males	Females
0 to 4	10 to 14	—	—	0.952	0.954	0.851	0.851
5 to 9	15 to 19	0.952	0.957	0.971	0.972	0.927	0.929
10 to 14	20 to 24	0.950	0.956	0.952	0.955	0.926	0.930
15 to 19	25 to 29	0.936	0.941	0.925	0.935	0.906	0.913
20 to 24	30 to 34	0.921	0.926	0.915	0.929	0.891	0.899
25 to 29	35 to 39	0.907	0.915	0.916	0.931	0.878	0.887
30 to 34	40 to 44	0.890	0.904	0.914	0.929	0.861	0.877
35 to 39	45 to 49	0.867	0.895	0.903	0.923	0.838	0.865
40 to 44	50 to 54	0.839	0.883	0.878	0.908	0.811	0.842
45 to 49	55 to 59	0.806	0.861	0.834	0.880	0.778	0.805
50 to 54	60 to 64	0.767	0.823	0.762	0.834	0.730	0.747
55 to 59	65 to 69	0.710	0.768	0.660	0.760	0.665	0.658
60 to 64	70 to 74	0.623	0.687	0.538	0.655	0.571	0.543
65 to 69	75 to 79	—	—	0.405	0.519	0.451	0.415
70 to 74	80 to 84	—	—	0.274	0.361	0.329	0.299

statistics. A third condition is identical coverage of the population at the two censuses. For example, if the entire male population is enumerated in one census, but the military are excluded at the second, the age cohorts involving the military cannot be compared without a suitable adjustment, unless the number of the military is negligible. If nationals living abroad are included in one census, and excluded from another, and if the numbers involved are large, especially if they are concentrated in any particular age or sex groups, this type of analysis is invalidated.

In the case of a country of immigration, under certain circumstances a cohort can be compared at two censuses even if migration data are lacking. If the native population (that is to say, persons born in the country) are known not to have emigrated in important numbers, the age and sex data can be tabulated by nativity, and comparisons of the two censuses limited to the native population. This point will be illustrated subsequently with data for the United States.

It should be noted that such survival rates can be calculated over any age span and time interval. Those shown in table 13 were calculated for a ten-year span of ages; but they can be calculated for single years, five-year, fifteen-year, or twenty-year periods, or whatever length of time is desired. A ten-year period was selected for the above illustrative purposes because the

actual census survival rates, which will be discussed below, were obtained from censuses taken ten years apart.

To illustrate the application of this test, six countries were selected, in each of which two censuses had been taken a decade apart, and the data had been tabulated by five-year age groups and sex, so that it was arithmetically simple to calculate ten-year survival rates for each of the age and sex groups. If the censuses had been less than ten years apart, or if the data had been tabulated by age groups of varying size (for example, some groups of five years, some in single years, and some of ten years) the arithmetic would have been much more complicated.

#### *Example of Egypt, 1937 to 1947*

Table 14 and figure 3 show survival rates computed from the age distributions of the Egyptian censuses of 1937 and 1947. The first observation is that in four cases for the males, and in two for the females, the survival rate is higher than 1.00. The following possibilities must be considered: (a) net immigration into Egypt in those particular age and sex groups; (b) errors in the 1937 census, and (c) errors in the 1947 census. Actually, there is no evidence of immigration into Egypt such as would explain these ratios, and the conclusion follows that the 1937 or 1947 census data, or both, for the cohorts mentioned were considerably in error.

Table 14  
COMPARISON OF AGE DISTRIBUTIONS ACCORDING TO CENSUSES OF EGYPT, 1937 AND 1947  
(Population in thousands)

Age group (years)	Males			Females		
	1937	1947	Survival rate	1937	1947	Survival rate
<i>(a) Five-year groups</i>						
TOTAL	7,966.8	9,391.7	—	7,954.0	9,575.0	—
0 to 4.....	1,021.9	1,279.6	—	1,085.7	1,305.3	—
5 to 9.....	1,107.9	1,208.9	—	1,101.0	1,191.2	—
10 to 14.....	1,030.9	1,142.3	1.117	878.2	1,071.2	0.985
15 to 19.....	713.2	983.9	0.888	633.1	917.4	0.833
20 to 24.....	539.7	677.8	0.657	565.2	706.1	0.804
25 to 29.....	616.7	685.7	0.961	692.9	786.5	1.242
30 to 34.....	557.9	620.1	1.149	634.5	689.5	1.220
35 to 39.....	600.4	659.3	1.069	540.6	653.6	0.943
40 to 44.....	474.8	569.1	1.020	472.2	566.3	0.893
45 to 49.....	345.1	428.5	0.714	313.2	415.2	0.768
50 to 54.....	330.3	421.2	0.887	335.5	448.5	0.950
55 to 59.....	144.7	171.1	0.496	134.3	173.1	0.553
60 to 64.....	201.7	252.0	0.763	231.3	298.8	0.891
65 to 69.....	72.3	83.8	0.579	72.8	82.1	0.611
70 to 74.....	100.9	107.8	0.534	122.1	136.7	0.591
75 to 79.....	26.9	23.4	0.324	28.9	24.0	0.330
80 to 84.....	39.0	34.8	0.345	58.6	52.6	0.431
85 and over.....	24.3	17.3	0.192	34.8	23.5	0.192
Not reported ...	18.2	25.1	—	19.2	33.3	—
<i>(b) Ten-year groups</i>						
TOTAL, age 5 and over .....	6,944.9	8,112.1	—	6,868.3	8,269.6	—
5 to 14.....	2,138.8	2,351.2	—	1,979.2	2,262.4	—
15 to 24.....	1,252.9	1,661.7	0.777	1,198.3	1,623.5	0.820
25 to 34.....	1,174.6	1,305.8	1.042	1,327.4	1,476.0	1.232
35 to 44.....	1,075.2	1,228.4	1.046	1,012.8	1,219.9	0.919
45 to 54.....	675.4	849.7	0.790	648.7	863.7	0.853
55 to 64.....	346.4	423.1	0.626	365.6	471.9	0.727
65 to 74.....	173.2	191.6	0.553	194.9	218.8	0.598
75 to 84.....	65.9	58.2	0.336	87.5	76.6	0.393
85 and over.....	24.3	17.3	0.192	34.8	23.5	0.192

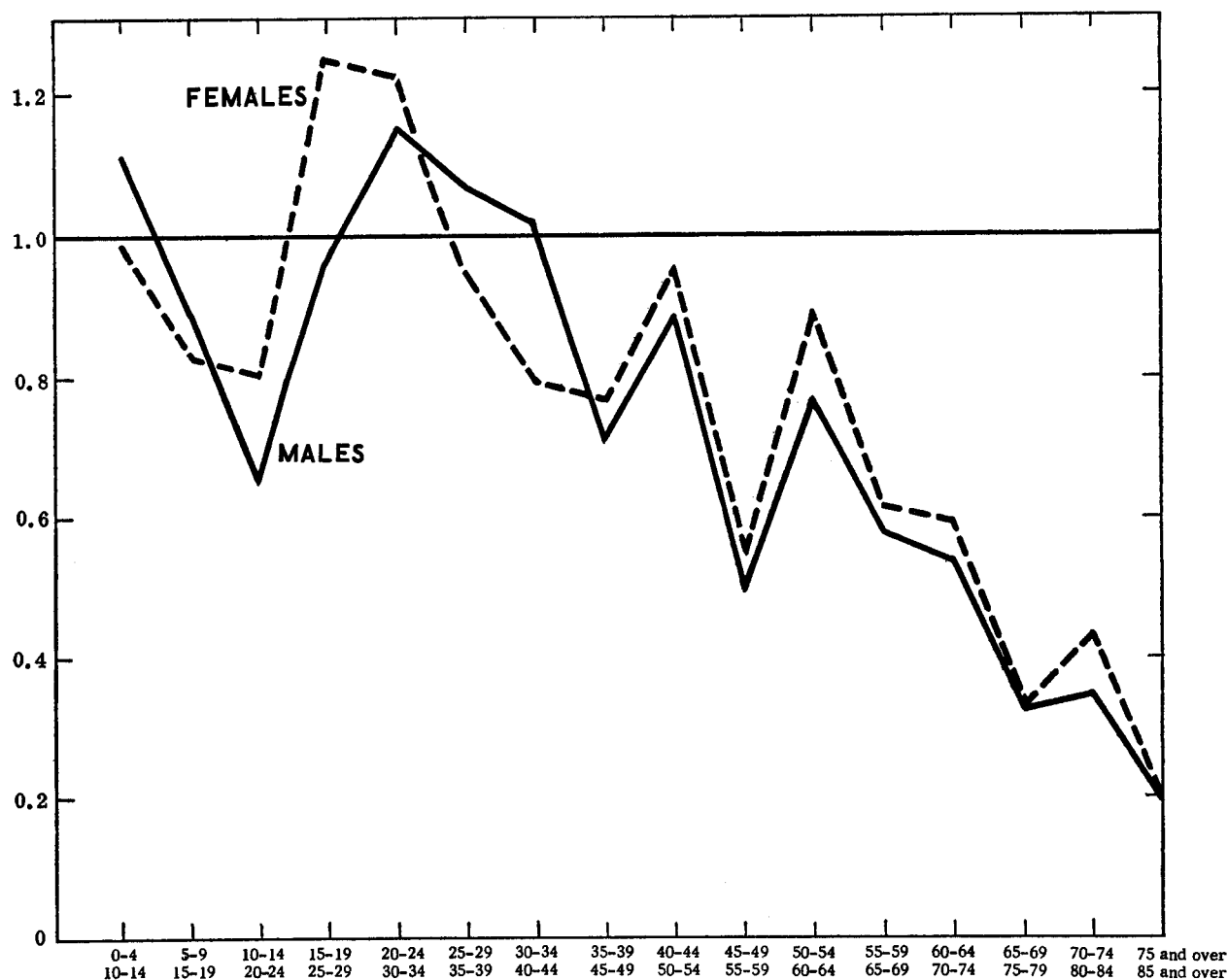


Figure 3. TEN-YEAR SURVIVAL RATES FOR MALE AND FEMALE COHORTS COMPUTED FROM EGYPTIAN CENSUS DATA FOR 1937 AND 1947

Examination of the entire pattern of survival rates reveals considerable irregularity. For example, among males 5 to 29 years old the rates were:

Between ages 5 to 9 and 15 to 19 .....	0.888
Between ages 10 to 14 and 20 to 24 .....	0.657
Between ages 15 to 19 and 25 to 29 .....	0.961

Furthermore, between ages 50 to 54 and 60 to 64, the survival rate (0.763) was higher than the rate between the ages of 10 to 14 and 20 to 24 (0.657) and higher than between the ages of 35 to 39 and 45 to 49 (0.714).

Among females also the survival rates form an irregular pattern.

Comparison of the survival rates for males and females reveals that in four instances the rates are higher for the males. Although not impossible, this is improbable. As stated previously, higher mortality rates, and conversely lower survival rates, for females than for males may occur during the child-bearing ages in some countries having very high birth rates and high maternal mortality rates. Examination of the four instances in which the female survival rates were lower than those for males, reveals that only two of these were in the child-bearing ages; other explanations would have to be sought at least for the other two deviations; errors in the data are strongly suggested.

In table 14, the rates are also shown for ten-year cohorts. The results reveal some smoothing of the irregular pattern previously observed. Nevertheless, for two male cohorts, and one female cohort, the rate is over unity. For one cohort—between ages 25 to 34 and 35 to 44—the rate for females is lower than that for males; this is during the child-bearing period.

In summary, it must be concluded that the patterns of survival rates obtained by comparing the two censuses are so irregular, particularly in comparison with the patterns observed in other countries, as to imply a high degree of inaccuracy in the data. Further investigation would be required to arrive at definite estimates of the amounts of the errors and explanation of their causes (age mis-statements or under-enumeration in each census).

#### *Example of Honduras, 1940 to 1950*

Survival rates, calculated from census data by the same method as in the case of Egypt, are shown in table 15 for Honduras, the Philippines, Portugal, Thailand and Turkey. The rates for four of these countries are also presented in figure 4.

For Honduras, as for Egypt, the first observation is that for four cohorts for the males and six for the fe-

Table 15

SURVIVAL RATES OF 5-YEAR AND 10-YEAR COHORTS COMPUTED FROM CENSUS DATA FOR FIVE COUNTRIES

Five-year cohorts		Honduras 1940-1950		Philippines 1939-1948		Portugal 1940-1950		Thailand 1937-1947		Turkey 1935-1945	
Age at beginning of decade	Age at end of decade	Males	Females	Males	Females	Males	Females	Males	Females	Males	Females
0-4	10-14	0.966	0.918	0.988	0.979	0.952	0.971	0.928	0.927	0.898	0.828
5-9	15-19	1.016	1.034	0.783	0.875	0.945	0.991	0.940	0.980	0.866	0.826
10-14	20-24	0.973	1.078	0.935	1.048	0.928	0.963	0.883	0.965	0.930	0.927
15-19	25-29	0.913	0.958	0.894	0.908	0.894	0.920	0.871	0.905	0.867	1.274
20-24	30-34	0.912	0.885	0.718	0.753	0.833	0.877	0.876	0.871	0.976	1.093
25-29	35-39	1.050	0.984	0.860	0.821	0.916	0.941	0.912	0.877	0.977	0.802
30-34	40-44	1.000	1.020	0.824	0.822	0.926	0.952	0.833	0.826	0.928	0.869
35-39	45-49	0.977	0.977	0.815	0.780	0.914	0.934	0.813	0.858	0.825	0.743
40-44	50-54	0.933	1.046	0.689	0.742	0.901	0.940	0.751	0.814	0.874	0.916
45-49	55-59	0.985	1.031	0.588	0.618	0.853	0.905	0.753	0.788	0.705	0.697
50-54	60-64	1.115	1.147	0.847	0.766	0.812	0.875	0.682	0.666	0.885	0.908
55-59	65-69	0.870	0.910	0.639	0.612	0.754	0.842	0.565	0.572	0.609	0.638
60-64	70-74	0.660	0.755	0.456	0.509	0.600	0.672	0.357	0.472	0.456	0.447
65-69	75-79	—	—	0.415	0.441	0.496	0.588	0.290	0.410	0.303	0.370
70-74	80-84	0.718	0.797 <sup>a</sup>	0.538	0.565	0.323	0.415	0.256	0.337	0.266	0.303 <sup>b</sup>
75 and over	85 and over	—	—	0.385	0.351	0.149	0.211	0.230	0.257	—	—
Ten-year cohorts											
5-14	15-24	0.995	1.055	0.847	0.947	0.937	0.977	0.914	0.974	0.892	0.866
15-24	25-34	0.913	0.924	0.808	0.835	0.866	0.901	0.873	0.889	0.929	1.171
25-34	35-44	1.026	1.000	0.845	0.821	0.921	0.947	0.875	0.853	0.953	0.843
35-44	45-54	0.956	1.009	0.764	0.764	0.908	0.937	0.786	0.839	0.845	0.826
45-54	55-64	1.043	1.083	0.691	0.683	0.834	0.891	0.722	0.732	0.792	0.813
55-64	65-74	0.773	0.838	0.542	0.559	0.680	0.760	0.469	0.528	0.529	0.523
65-74	75-84	0.718	0.797 <sup>a</sup>	0.465	0.498	0.425	0.515	0.279	0.383	0.280	0.322 <sup>a</sup>
75 and over	85 and over	—	—	0.385	0.351	0.149	0.211	0.230	0.257	—	—

<sup>a</sup> Ratio of numbers aged 75 and over in 1950 to those aged 65 and over in 1940.

<sup>b</sup> Ratio of numbers aged 80 and over in 1945 to those aged 70 and over in 1935.

<sup>a</sup> Ratio of numbers aged 75 and over in 1950 to those aged 65 and over in 1940.

<sup>a</sup> Ratio of numbers aged 75 and over in 1945 to those aged 65 and over in 1935.

males, the survival rates are above unity. The question arises, whether immigration during this decade of persons in these particular age and sex groups could have produced this result.

The second observation is that the survival rates are all very high in comparison with those for the other countries shown. If there was considerable immigration at all ages or the death rate was unusually low in Honduras, these rates might be correct.

Comparison of the rates for the two sexes reveals three ages at which the males have higher survival rates. Two of these are in the child-bearing period.

The above short analysis is presented as an illustration of the procedure for detecting suspicious figures. Actually, the explanation for these unusual survival rates is that the 1950 census data were adjusted for under-enumeration whereas the 1940 figures were not so adjusted. Comparison of the unadjusted 1950 census data with those of 1940 reveals a much more plausible pattern of survival rates. More details are given in appendix C.

#### *Example of the Philippines, 1939 to 1948*

In only one instance is the survival rate above unity, namely, for females between the ages of 10 to 14 and 20 to 24. The pattern of decreasing survival rates with increasing age is not as regular or smooth as would be expected in a life-table population, but nevertheless, more regular than in the case of Egypt.

Comparison of the rates for males and females reveals seven cohorts in which the males had higher rates.

When the data are combined into ten-year groups, the patterns become much more regular; with increasing age the survival rates become progressively smaller. There is only one exception, among men between the ages of 15 to 24 and 25 to 34; this rate is lower than that of the preceding or following ages. One possible explanation for this lower rate is that it may have resulted from military casualties during the Second World War. There was considerable fighting in the Philippines and many men were killed; the group of men 15 to 24 years of age in 1939 probably would have been more involved in such fighting than other age groups.

Comparison of the rates for the 10-year cohorts of males and females reveals three cases in which the males have higher survival rates. This fact suggests that there may have been differences in the completeness of enumeration of males and females at either or both of the censuses. Such differences, of course, may have come from mis-statements of age. They could have resulted also from:

- (a) More under-enumeration of males than of females in 1939;
- (b) More under-enumeration of females than of males in 1948;
- (c) Over-enumeration of males in 1948 or of females in 1939;
- (d) Any combination of the above.

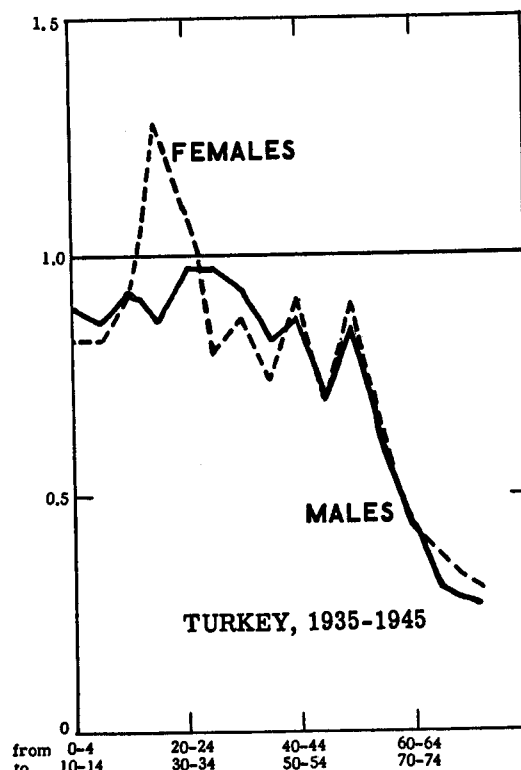
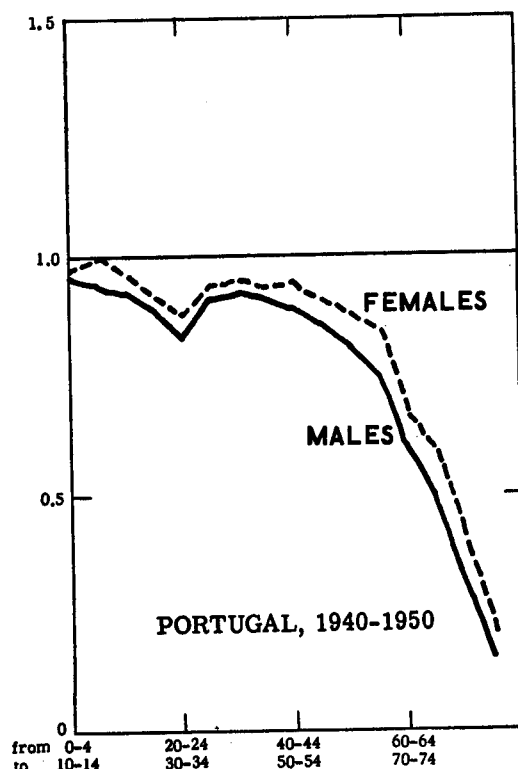
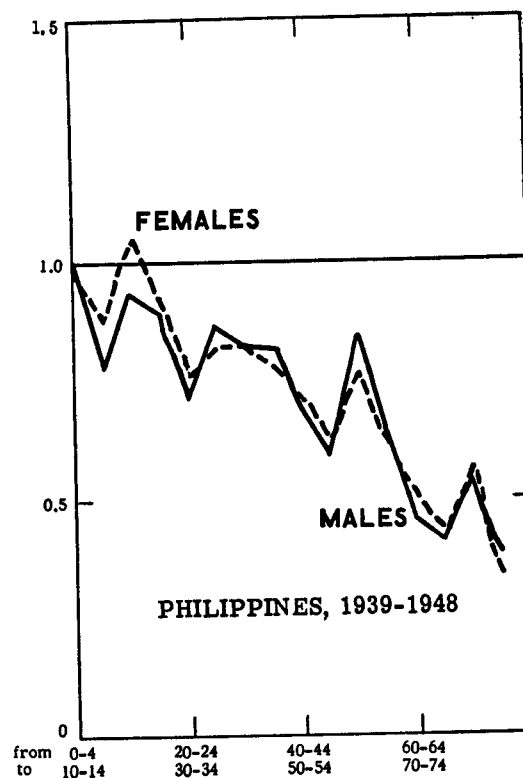
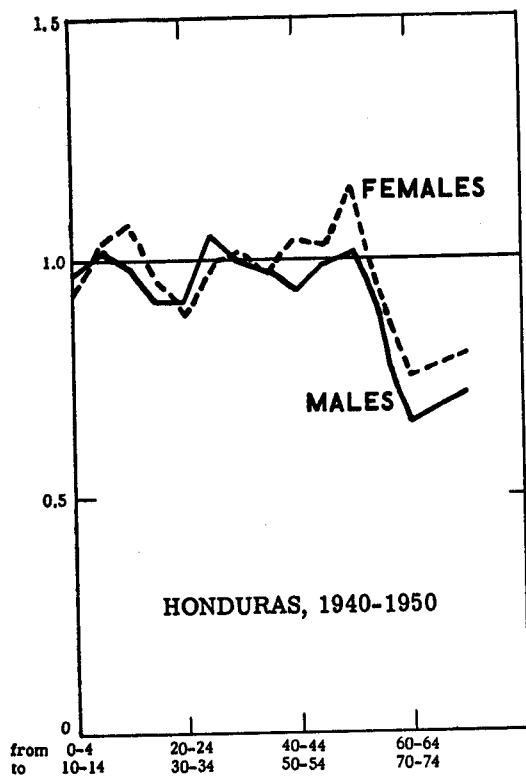


Figure 4. TEN-YEAR SURVIVAL RATES FOR MALES AND FEMALES COMPUTED FROM CENSUSES OF HONDURAS, THE PHILIPPINES, PORTUGAL AND TURKEY



### *Example of Portugal, 1940 to 1950*

There are no instances in which the survival rates are over unity, and no instances in which the rates for the males are higher than those for the females. In general, the survival rates for each sex decrease with advancing age, except for slight deviations in the middle ages.

Examination of the data for 10-year cohorts suggests that there is only one cohort which deviates sufficiently to attract attention; this is the cohort aged 15 to 24 in 1940 and 25 to 34 in 1950. The survival rates for this group, both for males and females, are lower than those for the preceding or following age groups. This could have resulted from emigration of members of these particular cohorts, or from the errors in the census. It is desirable to investigate the former possibility by reference to migration data.

In summary, Portugal is a good illustration of a country in which the census survival rates follow the expected life-table pattern reasonably well. Additional tests of the Portuguese age statistics can be made; one of these tests is the application of the balancing equation to be described subsequently.

### *Example of Thailand, 1937 to 1947*

Examination of the survival rates for Thailand by five-year age groups reveals approximately the same degree of regularity observed in the case of the Philippines. There are no instances in which the survival rates are above unity. On the other hand, in five instances the rates for the males are higher than those for the females.

Combining the ages into ten-year categories smooths out some of the irregularities. In only one cohort is the rate lower for females, the cohort which was 35 to 44 years of age—still in the child-bearing period—in 1947.

Pending further investigation it can be concluded tentatively that the age data for this country are perhaps not quite as accurate as those for Portugal, but perhaps somewhat better than those of the Philippines.

### *Example of Turkey, 1935 to 1945*

There are two instances in which the five-year survival rates are over unity, both among females. Also, there are eight instances in which the survival rates for males are higher than those for females.

The rates fluctuate from one age cohort to the next more than they do in the cases of Portugal and Thailand.

Combining the data into ten-year groups does not smooth the patterns of survival rates very much. There is still one cohort with a rate over unity, and four in which the rates for males are higher than those for females.

Comparisons of the Turkish patterns of census survival rates with those observed in some of the other countries, and with the expected patterns, indicates that there were large errors and that the quality of the data may be similar to that of the Egyptian material. Both mis-statements of age and differences in the completeness of census enumeration in 1945 and 1935 may be involved.

### COMPUTATION OF SURVIVAL RATES FROM AGE 5 AND OVER TO AGE 15 AND OVER

It is useful mainly as a check on the completeness of enumeration, to compare the population aged 5 and over at the earlier census with those aged 15 and over at the later census, in the six cases which have been chosen as examples. In the absence of migration the latter group should be the survivors of the former. These ratios, by sex, are presented below:

Country	Survival rate over decade	
	Males	Females
Egypt .....	0.830	0.874
Honduras .....	0.959	0.989
Philippines .....	0.785	0.819
Portugal .....	0.850	0.875
Thailand .....	0.822	0.852
Turkey .....	0.850	0.844

One observation is that the rates for Honduras are unusually high in comparison with the other countries. In fact, they are so high that they are impossible, unless there was a large volume of immigration.

Among the other countries the survival rates, for each sex separately, do not differ greatly. Among males the highest rate is 0.850 in Portugal and the lowest is 0.785 for the Philippines. Among women the range is from 0.875 to 0.819.

There are no hard and fast rules regarding the amount of variation that is possible for this survival rate given accurate data. Purely as a rule of thumb, the rates in most cases can be expected to fall within the range of about 0.800 to 0.900. If a rate is very much above or below these limits, it can be taken as an indication of probable errors in the census which should be investigated further.

### USE OF THE BALANCING EQUATION

Various forms of the balancing equations have been considered in chapters I and II. In those instances, the terms of the equation were so arranged as to bring most clearly into relief the effects of inaccurate census enumeration, incomplete birth registration, or errors in the registration of deaths, as the case might be. Another type of balancing equation can be used as a refinement of the comparisons of survivors in a given cohort which have been discussed in the preceding pages. Instead of computing an implied survival rate and attempting to evaluate it, in the case of each cohort, actual statistics of deaths are introduced and so a better basis for comparing numbers at the two censuses is obtained. For each cohort, the balancing equation can then be expressed by the formula:

$$P_1 = P_0 - D + I - E$$

where  $P_0$  is the number in the given cohort at the earlier census,  $P_1$  is the number in the same cohort at the more recent census,  $D$  is the number of deaths to members of the cohort during the census-interval,  $I$  is the number of immigrants of relevant sex and age, and  $E$  that of emigrants of such sex and age, during the interval.

Actually, few countries have the required migration data by age and sex. Accordingly, in most cases it is necessary to ignore the migration components and reduce the equation to:

$$P_1 = P_0 - D.$$

The remarks made previously regarding the limitations of cohort analyses where migration has taken place are applicable here.

In addition to the effects of migration not taken into account, failure of the equation to balance can be due to any of the following factors or combinations of them:

1. Differences in completeness of the two census enumerations, with respect to the cohort in question,
2. Age mis-statements at one or both of the censuses,
3. Incompleteness of death registration,
4. Mis-statement of ages at death.

Knowledge and good judgment are important in determining the relative weight of these factors in each instance. The results of other tests applied to the census enumerations and vital statistics are especially helpful. For example, if a direct check of the death statistics shows that registration is nearly complete and that the reporting of ages at death is not subject to such errors as would greatly affect the balancing equation, any sizeable errors can be attributed to the census figures alone. In the case of incomplete death registration or erratic reporting of ages at death, if estimates of the amounts of error from these sources can be made, on the basis of results of checks or otherwise, the figures for deaths can be corrected before they are used in the balancing equation. The same remarks apply, of course, to known errors in the census figures, where the equations are used to check death registration in the various age groups.

There are also some general considerations which are helpful in identifying the sources of errors. In the first place, errors due to both under-enumeration in the census and under-registration of deaths will usually conform to a predictable pattern of differences from age to age and between the sexes. The probability of omission from the census is likely to be particularly great in the case of very young children, and there may be some other sex-age groups for which the probability of omission is also high for reasons that will be known to persons familiar with the census operations and the reactions of respondents in the country concerned (young men subject to military service, young women of marriageable age, et cetera). With these exceptions, the amount of under-enumeration is likely to be similar for different sex-age groups, though it may well vary somewhat with sex and age, and the changes from one age group to the next will ordinarily not be very large. The same generalizations apply to under-registration of deaths. Consequently, if the discrepancies shown by the balancing equations change abruptly from one cohort to the next, mis-statement of ages, rather than under-enumeration, is likely to be the main cause of these variations. Age mis-statements are especially indicated if the discrepancies for successive cohorts are alternately positive and negative.

There is likely to be a relationship between mis-statements of ages in the census records and on death certificates. Although the frequency of mis-statements may not be the same in the two types of records, the pattern of their incidence in various age categories is likely to be similar, at least in the cases of mis-statements due to negligence or ignorance of the correct age. However, any errors in the reporting of ages on death certificates

will usually have much less effect on the balancing equations than errors in the census returns. Even within a period of ten years, only a minor fraction of each cohort, unless of advanced ages, is expected to die. If completeness or accuracy of age statements in death statistics is comparable to completeness and age-accuracy at the census count, the errors which the death statistics introduce into the balancing equations are fractional as compared to the census errors.

The effects of migration depend on the character of the migration. Migration for economic motives often affects particularly the cohorts of young men. Different types of migration may, however, vary as to their incidence by sex and age. The probable sex-age pattern of migration can usually be determined roughly, even without a complete statistical measure, by analysing available sources of information on the characteristics of the migrants and reasons for their migration.

### *Example of Portugal*

Data for Portugal will be used to illustrate the use of balancing equations in the test for the accuracy of age reporting, for selected five-year age cohorts. These data are the age and sex distributions from the 1940 and 1950 censuses, and the deaths during this decade, as reported by age and sex.

The discussion of the accuracy of death registration in Portugal<sup>20</sup> seemed to show some indication that death reporting was not entirely complete. For the purpose of this example, however, no attempt has been made to correct the death statistics. Also, it was assumed for this illustration that there was no migration, although it is known that there was some net out-migration during the decade. The method is illustrated with the following data for the cohort of males aged 5 to 9 years in 1940:

Number reported in 1940 census .....	428,000
Number 15 to 19 years reported in 1950 census ....	404,400
Apparent inter-censal decrease .....	23,600
Number of deaths during decade reported to cohort .....	10,600
<i>Residual</i> (excess of inter-censal decrease over number of reported deaths) .....	13,000

Table 16 presents the results of such calculations for selected age groups, for males and females separately.

There is considerable variation in the size of the residuals for the various age-sex cohorts shown. For example, the residuals are relatively small for the cohorts aged 50 to 54 years in 1940; among males the residual amounts to only 2.1 per cent of the apparent inter-censal decrease, and among women to 8.3 per cent. On the other hand, among females aged 5 to 9 in 1940, the residual is greater than the inter-censal decrease, and among males aged 5 to 9 in 1940, the residual amounts to 55.0 per cent of the reported inter-censal decrease.

How important are these residuals? Are they so small that they can be ignored for most analytical purposes? What are the sources of the errors? There is no simple and decisive answer to these questions. In arriving at any final judgements it is necessary first of all, to introduce any data available on the numbers and age and sex composition of migrants during this decade. It might be, for example, that the residual of 15,800 males

<sup>20</sup> See chapter II, pages 26-7.

Table 16

APPLICATION OF BALANCING EQUATION TO SELECTED AGE-SEX COHORTS, FOR PORTUGAL, 1940 TO 1950  
(Numbers in thousands)

(Numbers in thousands)

Age at 1940 census	Age at 1950 census	Number enumerated		Inter- censal decrease	Number of deaths reported	Residual		
		1940	1950			Number	Per cent of inter- censal decrease	Per cent of number enumerated in 1950
Males								
5 to 9	15 to 19	428.0	404.4	23.6	10.6	13.0	55.0	3.2
10 to 14	20 to 24	409.5	380.1	29.4	13.6	15.8	53.7	4.2
25 to 29	35 to 39	298.3	273.2	25.1	16.0	9.1	36.2	3.3
50 to 54	60 to 64	154.5	125.5	29.0	28.4	0.6	2.1	0.5
Females								
5 to 9	15 to 19	410.3	406.6	3.7	9.5	5.8	156.8	1.4
10 to 14	20 to 24	396.3	381.6	14.7	11.7	3.0	20.4	0.8
25 to 29	35 to 39	312.5	294.1	18.4	12.3	6.1	33.2	2.1
50 to 54	60 to 64	192.8	168.7	24.1	22.1	2.0	8.3	1.1

observed in the cohort aged 10 to 14 years in 1940 would disappear if emigration were taken into account; or it might become negative. On the other hand, it is also possible that the introduction of migration statistics into the equations would increase the residuals for cohorts aged 50 to 54 in 1940.

It should be noted that for some cohorts the residual is only one or two per cent of the number counted in the census. For most purposes to which the census data might be put, such differences are small enough to be ignored. Residuals which amount to four or five per cent or more, on the other hand, may have a significant effect upon the results of some types of analyses.

#### Example of Thailand

Several times in this manual reference has been made to the fact that children under 5 years of age are almost universally under-counted in censuses. For an illustration of the use of balancing equations to investigate under-enumeration in this age group, the following data on male children in *Thailand* will be used:

Number under 5 enumerated in the census of 1937....	1,230,500
Number 10 to 14 enumerated in the census of 1947..	1,142,000
Apparent inter-censal decrease .....	88,500
Reported number of deaths during decade to cohort under 5 years of age in 1937.....	118,400
Residual (excess of deaths over apparent inter-censal decrease) .....	29,900

The true excess of deaths over the inter-censal increase was probably greater than that shown above, since the previous analysis (in chapter II) indicated the probability of under-registration of deaths. One estimate<sup>21</sup> was that only 74 per cent of the deaths during the decade had been registered. Applying this proportion to 118,400 provides an estimate of about 160,000 for the true number of deaths during the decade to children who were under 5 years of age in 1937. If this were correct, the residual—the excess of deaths over inter-censal decrease—would be about 71,500.

What other errors could account for the residual? They may be listed as:

(1) Undercount in 1937 census of children under 5 years of age;

(2) Overcount of children 1947 census aged 10 to 14;

(3) Age mis-statements in 1937 so that some children under 5 were returned as 5 to 9 years old;

(4) Age mis-statements in 1947 so that some children under 10 or 15 and over were returned as aged 10 to 14.

More than one of these errors probably occurred to some extent, but what is the likelihood that the first error was the most serious and of the largest size? Further testing beyond the scope of this manual would be required to establish the size of each type of error. Such testing should take into consideration all available data, both published and unpublished, which the technicians in the country can obtain, together with any additional knowledge of whatever social and economic factors may be particularly important in the country concerned.

#### Example of the United States

Data are available for the United States for the decade 1930 to 1940 which are useful for illustrating apparent discrepancies in the two census counts. These data and the relevant calculations (shown in table 17) are for native white males. By limiting the analysis to the native group any possible distortions which might be introduced by migration are largely avoided since, by and large, very few natives emigrated. There were no boundary changes and no war to disrupt the comparison.

Instead of death statistics, which are known to have been somewhat incomplete, a life table constructed by the national statisticians was used; this life table includes corrections for under-reporting of deaths. The number of deaths in each cohort during the decade was estimated by calculating life-table survival rates on the basis of the  $L_x$  figures. These values (shown in column (b) of table 17) were obtained by dividing the sum of  $L_x$  for each five-year age group by the sum of  $L_x$  for the group 10 years younger.

The findings are as follows:

(1) There was an under-enumeration of children under 5 years of age in 1930. The number enumerated in 1930 was 5,036,900 whereas the number 10 to 14 years in 1940 was enumerated as 5,231,900. Altogether

<sup>21</sup> See chapter II, page 26.

Table 17

APPLICATION OF BALANCING EQUATION, BY MEANS OF LIFE-TABLE SURVIVAL RATIOS, TO CENSUS AGE DISTRIBUTIONS OF UNITED STATES  
NATIVE WHITE MALES, 1930 AND 1940.\*

(Numbers in thousands)

Age (years)	Enumerated in 1930 (a)	Survival rate 1930-1939 (b)	Expected number in 1940 (c)	Enumerated in 1940 (d)	Discrepancy	
					Absolute (e)	As per cent of enumerated (f)
0 to 4 .....	5,036.9	—	—	—	—	—
(0 years) .....	(965.2)	(0.96272)	—	—	—	—
(1 year) .....	(960.3)	(0.97707)	—	—	—	—
(2 years) .....	(1,023.0)	(0.98124)	—	—	—	—
(3 years) .....	(1,046.6)	(0.98332)	—	—	—	—
(4 years) .....	(1,041.9)	(0.98439)	—	—	—	—
5 to 9 .....	5,497.3	0.98432	—	—	—	—
10 to 14 .....	5,265.8	0.97838	4,926.0	5,231.9	+305.9	+ 5.8
15 to 19 .....	4,907.3	0.96131	5,411.1	5,433.5	+ 22.4	+ 0.4
20 to 24 .....	4,346.9	0.96805	5,152.0	5,014.7	-137.3	- 2.7
25 to 29 .....	3,731.8	0.96131	4,769.9	4,698.4	- 71.5	- 1.5
30 to 34 .....	3,408.6	0.94920	4,208.0	4,230.3	+ 22.3	+ 0.5
35 to 39 .....	3,278.8	0.93063	3,587.4	3,724.2	+136.8	+ 3.7
40 to 44 .....	2,771.5	0.90232	3,235.4	3,338.4	+103.0	+ 3.1
45 to 49 .....	2,411.9	0.86278	3,051.3	3,025.7	- 25.6	- 0.8
50 to 54 .....	2,092.8	0.80819	2,500.8	2,568.4	+ 67.6	+ 2.6
55 to 59 .....	1,670.6	0.73120	2,080.9	2,054.2	- 26.7	- 1.3
60 to 64 .....	1,305.3	0.62691	1,691.4	1,659.2	- 32.2	- 1.9
65 to 69 .....	944.8	0.49318	1,221.5	1,314.2	+ 92.7	+ 7.1
70 to 74 .....	690.0	0.33944	818.3	873.2	+ 54.9	+ 6.3
75 to 79 .....	376.6	0.19590	466.0	487.8	+ 21.8	+ 4.5
80 to 84 .....	165.8	0.09118	234.2	250.4	+ 16.2	+ 6.5
85 to 89 .....	57.0	0.03263	73.8	81.8	+ 8.0	+ 9.8
90 to 94 .....	11.7	0.00772	15.1	17.3	+ 2.2	+12.7
95 to 99 .....	1.9	—	1.9	2.9	+ 1.0	+34.5
100 and over .....	0.3	—	0.1	0.3	+ 0.2	+66.7
Unknown .....	36.5	—	—	—	—	—
<i>Totals</i>						
All ages .....	48,010.1	—	43,445.1	44,006.8	+561.7	+ 1.3
20 years and over .....	—	—	33,108.0	33,341.4	+233.4	+ 0.7

\* From A. Jaffe, *Handbook of Statistical Methods for Demographers*, U. S. Bureau of the Census, Washington 1951, p. 91.

an estimated 305,900 native white male children appear to have been omitted from the 1930 census.

(2) In 1940 there appears to be a deficit in the cohort 20 to 29 years old at that time and a surplus in the cohort 30 to 39 years (or 30 to 44 years). This suggests that young adult males may have been missed in both censuses, an observation which fits in with other evidence and which could be explained on the theory that many young men, at the age at which they leave their paternal home to seek employment and perhaps to migrate, and before they have established their families, do not have a well-established place of residence where they will be reported by other household members to the census enumerators. The reasoning is as follows. If, in the United States censuses there is a tendency to omit males about 20 to 29 years old, but not of men 30 to 39 years old, the 1930 count for ages 20-29 is presumably too low, but the 1940 count for ages 30-39 should be approximately correct, showing a surplus such as that observed over the expected survivors of this cohort. In the 1940 census on the other hand, there would be again a deficit in the age range of 20 to 29

years. On the other hand, young men may have tended to overstate their ages, with the result that unduly small numbers were reported in their twenties and unduly large numbers in their thirties. If this happened frequently on the occasion of both censuses, the same results would follow.

(3) At the older ages beginning about 65 there appears to be a tendency to exaggerate age in the census reports, a tendency which seems to increase with advancing age. Its appearance about 65 is probably related to the fact that at this age many persons become eligible for old-age benefits.

Comparison of the entire native white male population enumerated in 1930 with that 10 years of age and over in 1940 reveals on the whole but minor discrepancies, as follows:

Number enumerated in 1930.....	48,010,100
Number 10 and over enumerated in 1940.....	44,006,800
Apparent inter-censal decrease .....	4,003,300
Estimated number of deaths during decade to persons alive in 1930 .....	4,565,000
Residual (excess of deaths over apparent decrease).....	561,700

## E. Direct checks

Of the total residual, over one-half seems to be accounted for by the presumed under-enumeration of children in 1930. If the balancing equation is applied to the population 10 years of age and over in 1930 and 20 years of age and over in 1940, the residual is reduced to 233,400, as follows:

Number aged 10 and over enumerated in 1930.....	37,475,900
Number aged 20 and over enumerated in 1940.....	33,341,400
Apparent inter-censal decrease .....	4,134,500
Estimated number of deaths during decade to persons 10 years of age and over in 1930.....	4,367,900
Inter-censal decrease .....	4,134,500
Residual (excess of deaths) .....	233,400

The last residual is about 5.6 per cent of the inter-censal decrease, and only 0.7 per cent of the enumerated 1940 population—too small to invalidate the data for most purposes.

On the basis of a comparison of the 1940 and 1950 censuses the United States Census Bureau reported finding substantially the same types of errors as described above in the 1950 age and sex statistics.<sup>22</sup>

## D. Evaluation of the accuracy of statistics of deaths by sex and age groups

Errors in statistics of deaths classified by sex and age may be caused by the same two factors noted in the case of census returns, namely, mis-statements of age and omissions from the records that are more frequent in some sex-age categories than in others. In the case of infant deaths there is an additional factor, which may be of considerable importance in many areas, namely, failure to follow exactly the prescribed statistical definition of an infant death as distinguished from a still-birth. The latter error commonly takes the form of reporting as still-born, if not failing to report in any way, infants who are born alive but die very shortly after birth.

The balancing equations and analogous computations described in preceding sections of this chapter can be used to test the accuracy of death statistics, as well as population census returns, by sex and age groups. Valid conclusions with regard to the accuracy of the death statistics, however, will only be possible if the census data are very nearly accurate, or if the errors in them have been measured by means of direct checks. In the case of infant deaths it is also necessary to have accurate data on the number of births in order to establish an equation which will yield a useful test of death registration, and this requirement is not likely to be met where the registration of infant deaths is seriously deficient.

Various other methods of investigating the reliability of statistics of death classified by sex and age groups, and of data on infant deaths, have been described in chapter II.

Direct checks on the completeness of census enumeration and registration of vital statistics such as have been discussed in chapters I and II can be used also to test the accuracy of census and death statistics classified by sex and age groups, and to provide a basis for correction of any errors which may have resulted either from mis-statement of ages or from omissions varying in frequency within different sex-age categories. This can only be done, of course, if the independent records, against which the census returns or death registers are checked, contain reliable information on ages. When the independent record is established by means of a special field survey, it is well worth-while to include information on ages in this record and to take whatever steps are feasible to ensure accurate age returns. In a sample check enumeration of the population, specially selected and trained enumerators conducting more intensive interviews can obtain more accurate age reports than are obtained in the census; likewise, careful investigation in a house-to-house canvass of deaths may yield more accurate reports on the ages of decedents than are to be found on the death certificates. Not only the accuracy of age returns, but also the reliability of reports on other characteristics of the population and of decedents can be investigated in this manner.

On the other hand, no matter how carefully information on ages may be obtained in the check enumeration or survey, it is always subject in some degree to errors of the same kind which affect the data being tested. In particular, willful mis-statements of age and errors due to utter ignorance of the persons concerned may affect the results of even the most carefully conducted check.

A better method of checking age returns is to match the census records or death certificates with the birth certificates of the persons concerned. In a large country with a mobile population this procedure may be very laborious and expensive, even for a small sample of the records. Nevertheless, this is the most objective check on the accuracy of age statements which can be made. In the birth registers, the date of birth will rarely be mis-stated and, if so, probably only by a few days at the most. The age of an individual, as computed from the date of his birth according to these records, is for all practical purposes exact. This type of check was conducted on the occasion of the 1921 census of New Zealand. The following is an excerpt from the report:<sup>23</sup>

"...For the first time an attempt has been made to supply a more or less reliable measure of the extent and incidence generally of this assumed inaccuracy (i.e. of age statements). The *modus operandi* was of the simplest character, comprising little more than the checking of stated ages with birth registration records. A selection of names and persons of both sexes was chosen, following certain age-divisions. The method employed naturally restricted operations to native-born New Zealanders. Only those cases were chosen where the names and other particulars were such that certain identification was possible.

"Altogether 2,524 cases were investigated, resulting in the actual testing of 2,219; the remaining 305 cases either could

<sup>22</sup> United States, Bureau of the Census. 1950 Census of Population, *United States Summary*, Bulletin P-C 1, p. xiii, Washington, 1953.

<sup>23</sup> New Zealand, Census and Statistics Office, *Results of a Census of the Dominion of New Zealand taken for the night of 17th April 1921. General Report*. Wellington, 1925, pp. 93-96.

not be located, or, for one reason or another, could not be identified beyond doubt.

"The actual sample (2,219) is not a large one, consisting as it does, of only 0.24 per cent of the total native-born. That it is not larger is due to the considerable labour and inconvenience which larger numbers would have involved. Nevertheless the results show much consistency and are of no little value.

".....

"The following summary gives some of the facts in brief and shows that there exists a slight tendency to overstate the age of minors, whether male or female, and that misstatement is much more common with adults. Adult females are not greatly less accurate than males in the matter of stating ages; where, however, there is no defined tendency in the male figures, those for females show a decided preference for understatement."

	<i>Adults</i>		<i>Minors</i>	
	<i>Number</i>	<i>Per cent</i>	<i>Number</i>	<i>Per cent</i>
<b>Males:</b>				
Correctly stated .....	392	76.41	567	94.82
Understated .....	62	12.09	3	0.50
Overstated .....	59	11.50	28	4.68
<b>Females:</b>				
Correctly stated .....	352	68.75	568	95.30
Understated .....	115	22.46	7	1.18
Overstated .....	45	8.79	21	3.52

The ages reported in the census death records may also be compared with the ages or dates of birth stated in other records, for example, social insurance records. Likewise, for a sample of persons enumerated at one census, it is possible to examine ages returned at another census, either prior or subsequent. This was done in New Zealand for persons reporting advanced ages, and it was found that a large proportion of them had exaggerated their ages.

## CHAPTER IV. ADEQUACY OF MIGRATION STATISTICS

Previous chapters have presented tests for the accuracy of census data and of birth and death statistics. The remaining component needed for measuring population change is that of migration. Accordingly, in this chapter, tests will be presented for appraising the adequacy of migration statistics.

The term "adequacy," rather than accuracy or completeness, is here used because migration statistics often present peculiar problems other than mere accuracy of count, arising from the definition of categories of migrants to which they apply. The statistics sometimes refer only to certain types of migrants; in this case they may be accurate and yet inadequate as a measure of population change resulting from all movements into and out of the given area. Other statistics, relevant to migration, do not measure it directly. For example, census data on internal migration obtained by asking questions on place of birth or place of residence at a stated previous date do not measure the movements themselves but their residual effects on population distribution. Here the question of adequacy involves the sufficiency of the information given for a correct evaluation of the actual movements during a given time interval for which population estimates are desired.

It is beyond the scope of this manual to investigate the accuracy with which each of the various types of migratory movements is recorded. The tests considered here are designed to examine whether the available migration statistics furnish an adequate account of those gains or losses which accrue to a population as a result of migratory movements of all kinds.

Migration has relevance only in connexion with a specific geographic area; this can be either an entire country or territory, or a subdivision thereof. In either case, the essence of migration is the crossing of some boundary. By combining the migration statistics with the census and vital statistics for the population living within the given boundaries, the required population estimates are made for the specified geographic area. In case the census and vital statistics are recorded on a *de facto* basis, that is, with reference to the population actually present within the area at the census date and the vital events taking place in that area, it is desirable that the migration statistics refer to all persons who cross the boundaries, but in practice it is not always easy to record all movements. On the other hand, in the more usual case of *de jure* statistics referring to residents of a given area, it would seem more appropriate to have statistics of migrants who cross the boundaries in order to take up residence within the area, excluding visitors, et cetera. The problem of who should and who should not be counted is then not an easy one. For these reasons there are no easy tests for the completeness of migration counts. The problem can be seen more easily if two types of migration, international and internal, are separately examined.

### A. Adequacy of international migration statistics

In the migration statistics compiled in many countries, an attempt is made to differentiate between "permanent" and "temporary" immigrants and emigrants.<sup>1</sup> Where this is done, the category of "permanent" migrants seems in general to be appropriate, from a conceptual point of view for use in estimating changes in the *de jure* population. However, the proper classification of each migrant, or traveller, under either of these two categories, is rarely obvious and there are many problems in the definition of these, as well as other classes, of persons crossing a country's boundaries. The actual definitions of particular classes of migrants used for statistical purposes in most countries are the result of various forms of legislation concerning migratory movements, which are rarely alike in any two countries. It is not the purpose of this chapter to examine these problems of legislation and definition, but only to investigate whether the existing migration statistics, such as they are, furnish a reasonable measure of the amount of population gain or loss resulting from any form of migration.

The statisticians of each country, being familiar with the definitions of migration statistics currently in use and knowing generally whether the categories of migrants covered most of the important movements, will be able to judge whether the migration statistics—irrespective of their accuracy—are adequate for the purpose of population estimates. If definitions and coverage are not adequate, there is little need to investigate the accuracy of the data, so far as the present purpose is concerned.

### USE OF THE BALANCING EQUATION

The balancing equation can sometimes be adapted for testing international migration statistics. The particular adaptation to be made in a given situation is dependent upon the nature of the statistics which a country has, the main types of migration which occur, and the availability of the required census and vital statistics. However, unless the other components of the equation are highly accurate, the balancing equation as applied to the total population and its change during the interval between two censuses will not yield a very useful indication of the accuracy of the migration component. This is so, because migration is almost always the smallest component in the equation; hence errors which are small in proportion to the other components loom large in proportion to migration. For example, an error in the census count which amounts to one-tenth of one per cent is generally small enough to be ignored in so far

<sup>1</sup> United Nations, *Problems of Migration Statistics*, Population Studies No. 5, 1949. Recommendations for the collection and classification of migration statistics will be found in United Nations, *International Migration Statistics*, Statistical Papers, Series M, No. 20, 1953.

as the census count is concerned; this same error however may be larger than the recorded migration. In other words, a residual due to very minor errors in the other components, might give a false impression of a major error in the migration component.

For illustrative purposes, a balancing equation will be shown for the foreign-born population of the United States, in order to appraise the migration statistics for the decade 1940 to 1950. This form of the equation is justified by the fact that the United States is a country of immigration, and that very few of the native population or of the naturalized foreign born emigrate. What emigration does occur is mostly of persons who were not born in the country and are not citizens thereof, that is to say, previous immigrants who subsequently emigrate. The equation is:

$$P_1 = P_0 - D + I - E$$

where  $P_0$  = foreign-born population at the 1940 census,  $P_1$  = foreign-born population at time of 1950 census,  $D$  = number of deaths to foreign-born in inter-censal period,  $I$  = number of immigrants,  $E$  = number of emigrants.

Births are not taken into account because persons born in the United States are, of course, natives, even though of foreign-born parentage, and so do not add to the foreign-born population.

Actually the statistics are not available in exactly the desired form. The numbers of foreign-born persons living in the country in 1940 and 1950 are available from the census counts, by age and sex. The number of deaths to the foreign-born during the 1940's is not available; however, it is possible to estimate how many deaths occurred by assuming that the foreign-born had the same death rates in each age and sex cohort as the general population. The migration statistics show the numbers of aliens admitted and of aliens who departed, but no data on the movements of naturalized citizens. Many of the foreign born in the country are naturalized, but since very few of the naturalized emigrate, the omission of their movements does not seriously affect the balancing equation.

It should be noted that the statistics on migration and citizenship are highly complicated and it is not possible in this manual to present a complete account of the many technical problems connected with their use.

The balancing equation then, can be stated as follows:

Foreign-born population in 1950 .....	10,161,000
Foreign-born population in 1940 .....	11,420,000
Apparent inter-censal decrease .....	1,259,000
Deaths among foreign-born persons who were living in the United States in 1940, during decade 1940 to 1950 (est.) .....	2,320,000
Aliens admitted during decade 1940 to 1950 .....	3,496,000
Aliens departed during decade .....	2,263,000
Migratory balance (excess of admissions over departures) .....	1,233,000
Expected inter-censal decrease (net excess of deaths over migratory balance) .....	1,087,000
Residual (excess of apparent over expected inter-censal decrease) .....	172,000

The residual would actually be smaller than 172,000 if deaths among those admitted during the decade had been taken into account. Data are not available for estimating with any degree of precision the number of

persons admitted after 1940 who died before the 1950 census enumeration. Of the 1,233,000 excess of admissions over departures during the decade, perhaps as many as 125,000 may have died; this number could account for more than one-half of the residual.

In summary, this equation balances out fairly well and leads to the conclusion that the migration statistics were probably not affected by any very gross errors.<sup>2</sup>

It would be possible to compute a similar balancing equation for each sex and age cohort separately if the data were available. The census enumerations of the foreign-born by age and sex are available, but tabulations of the migration statistics in the required age and sex groups are not available.

The equation can also be computed for the immigrants from any given country, and the results may be useful in the interpretation of the residual for all immigrants.

An example is afforded by the United States census statistics for persons born in Puerto Rico who were living in the continental United States at the time of each census. Another measure of the migration from Puerto Rico to the continent is afforded by statistics on the numbers of passengers carried to and from the island by airplanes; there is almost no passenger flow by water transportation; virtually all emigrants from Puerto Rico go to the continental United States, and nearly all Puerto Rican immigrants to the United States come from that island. Under these circumstances it is possible to calculate the following balancing equation:

Persons born in Puerto Rico but living in continental United States according to census of 1950 .....	226,100
Persons born in Puerto Rico but living in continental United States according to census of 1940 .....	70,000
Apparent inter-censal change, that is to say, net immigration from Puerto Rico during decade .....	156,100
Number of air passengers carried out of Puerto Rico during decade .....	770,000
Number of air passengers carried into Puerto Rico during decade .....	610,000
Excess of passengers carried out .....	160,000
Residual (excess of passengers carried out over apparent inter-censal change) .....	3,900

It will be noted that this equation balances out quite well but not exactly. Part of the reason for the discrepancy is that no information is available on the number of deaths during the decade, among those persons who were born in Puerto Rico but were living in the continental United States in 1940, or migrated to the continent during the 1940's. An approximate estimate of the number of such deaths is 5,000 to 6,000. This test, then, suggests that the data on place of birth, so far as Puerto Rico is concerned, do not seem to have serious errors.

#### INTERNATIONAL COMPARISON

A useful check is obtained by comparing the immigration statistics of one country with the emigration statistics of another, and *vice versa*. Persons leaving country

<sup>2</sup> It should be noted, however, that considerable numbers of Mexicans who entered the country illegally and who would, of course, not have been included in the migration records, may have also been omitted from the 1950 census count or returned erroneously as natives of the United States. These errors would not be shown by the balancing equation in the form used here.



A to settle in country B should in most cases be identical with persons arriving in country B whose origin was country A; the exceptions would be births and deaths which occurred *en route*, and persons who either changed their intentions while travelling or were not admitted to country B. If the definitions used for migration statistics in the two countries are very nearly the same, this comparison can serve to check the accuracy of migration records in either country. Where the basic definitions are at variance—which is the more usual situation—the comparison may still serve as a check on the adequacy with which the net migratory balances are measured.

#### DIRECT CHECKS

Post-enumeration checks can be made in the same manner as described previously, in order to ascertain the accuracy of enumeration of the foreign-born population. Furthermore, during the course of such a check or at any other time, it is possible to obtain a list of names of persons who have entered the country during a specified period, and to compare this list with the names of the arrivals officially recorded. This type of test is exactly the same as that which can be used for a direct check of the completeness of birth reporting.

If the census shows immigrants classified according to the year in which they immigrated (as, for example, in the case of Canada), there may be some advantage in matching the census records with migration records for a particular year of immigration. Naturalization records may also serve as a basis for checking the immigration records. It may be noted that the population of foreign nationality is diminished as a result of naturalizations, a factor which must be taken into account in a balancing equation of the population of foreign nationality.

A direct check on statistics of emigrants is usually very difficult. An individual who has left cannot be interviewed; only if someone remains to report to a field worker that he left the country can his name be obtained for the purpose of a check.

#### B. Adequacy of statistics on internal migration

Very few countries collect any current data on internal migration. The few countries that have done so have collected these data in various ways, and there are no internationally "approved" standards for such statistics. Consequently, appraisals have to be devised on an *ad hoc* basis to fit the type of data which are available in the country.

#### CONTINUOUS POPULATION REGISTERS

In a few countries, notably the Scandinavian countries and the Netherlands, continuous registers of the population are maintained in each locality. An individual record is kept for each person, a record being added for every birth and withdrawn for every death, and the record is transferred from one locality to another when the person changes his place of residence. Internal migration statistics are provided by tabulations of such transfers.

The accuracy of these registers is checked periodically by a balancing equation in which the recorded changes due to births, deaths, and migration are compared with

the changes in the numbers of persons on the register for each locality. If the balance is not accurate, apparent errors are investigated. But even a perfect balance is not a sufficient guarantee that all the changes which actually took place have been recorded and that the register actually corresponds to the present population. One of the shortcomings of population registers is the lack of any sure check on persons who have left the country, whose departure may not have been reported to the registration authorities. Because of these and other errors which may accumulate in the course of time the registers must be checked from time to time against independent records. These are provided by censuses. Comparison of the census returns with registration records provides a check on the accuracy of both, and any errors discovered are eliminated by local investigation.

The details of the procedures are complicated and are not the same in all countries where continuous registers are kept.<sup>3</sup>

#### CENSUS STATISTICS RELATING TO INTERNAL MIGRATION

Three types of questions which relate to internal migration have been asked in various censuses or other house-to-house surveys: place of birth; place of residence at some specified time in the past; and duration of residence in the present place. The resulting statistics, though very useful for analysis of internal population movements, are subject to important weaknesses. Of the three questions, the one relating to birthplace is most commonly asked; for a person no longer living in the area where he was born it yields no information on the time of migration and therefore permits no precise calculation of internal migration during any specified time period. The question on place of residence at a specified time in the past is free of this weakness, but the results measure only the net effect of movements since that time. If a person has moved more than once during the interval, that fact is not revealed; if he has moved away from the place of previous residence and then returned, even the fact that he has migrated is not revealed; furthermore, there is no record of the migrants who have died prior to the time of the census. The question on duration of residence in the present place gives no information on the places of origin of the migrants, apart from such inferences as may be possible if the question of birthplace or residence at a specified past date is also asked.

In evaluating the adequacy of internal migration data obtained from censuses, these weaknesses as well as any inaccuracies in the data must be taken into account. Errors of reporting, even though quite frequent, may be less important than the other weaknesses. The tests which can be applied vary with the forms in which the relevant census questions are asked.

#### TESTS OF DATA ON PLACE OF BIRTH

The accuracy of data on place of birth from two successive censuses can be tested, if the required data are available, by means of a balancing equation for persons born in each part of the country. Such an equation takes the form:

<sup>3</sup> The reader interested in further detail may consult the census publications of the countries concerned.

$$P_1 = P_0 + B - D$$

where  $P_0$  and  $P_1$  represent the total population born in a given part of the country according to each of the two censuses,  $B$  the number of births in that part of the country during the census interval, and  $D$  the number of deaths, during the interval, to persons born in that part of the country. The latter quantity is ordinarily not given in the statistics, deaths being rarely classified according to place of birth, but it may be estimated, possibly on the assumption that the death rate for persons born in the given part of the country is similar to the death rate for the whole population in that part of the country. The discrepancy obtained may be due to errors in reporting of birthplace, incomplete enumeration at either census, incomplete birth registration, error in estimating the number of deaths, or failure to take international migration into account. In some circumstances equations of this type may be more useful for testing the accuracy of birth statistics than of census reports on place of birth.

More useful balancing equations can be made if statistics of the population by place of birth are tabulated by age groups. The data for persons of a given age group born in a certain part of the country at one census date can then be related to the records of births during a certain period of years in the past, corresponding to the age of the group. The equation takes the form:

$$P_1 = B - D$$

$P_0$  being zero by definition. This case will be illustrated with statistics of the United States, for the age group under 5 years.

According to the 1950 census there were 77,400 persons under 5 years of age living in the United States who were born in the state of North Dakota. If state of birth and age were reported correctly, this number should correspond to the number of births in North Dakota during the five years preceding the census—that is to say, 1 April 1945 to 1 April 1950—minus the deaths. The figures are as follows:

Number of persons enumerated .....	77,400
Number of births reported .....	78,900
Excess of births .....	1,500

The number of deaths which occurred among these children during the five-year period is not shown exactly by the records, since some of these deaths occurred outside North Dakota, while some of the deaths registered in North Dakota for this age group were to children born elsewhere. The figure for deaths could be estimated by assuming that the death rate was about the same for the children born in North Dakota as for those resident in North Dakota during the five-year period; on this assumption it would be found that the number of deaths was not very different from the excess of births noted above, and the residual representing errors would be small. To complete the test, several age groups should be tested in the same manner.

Still more illuminating results can be obtained by introducing into the equations for various age groups, data from the tabulations of birth place by age from an earlier census, if these data are available. For example, let us suppose that the population born in a given part of the country, aged 15 to 19, has been obtained from the 1950 census and the number born in the same area, aged 5 to 9, from the 1940 census. These figures can be compared with each other and with the recorded number of births in the given area during the period from 5 to 9 years prior to the 1940 census date, by means of the following simultaneous balancing equations:

$$P_2 = P_1 - D_2$$

$$P_1 = B - D_1$$

where  $P_2$  = number aged 15-19 in 1950,  $P_1$  = number aged 5-9 in 1940,  $D_2$  = deaths among this cohort, between ages 5-9 and 15-19, during the decade 1940-1950,  $D_1$  = deaths among the same cohort, between birth and ages 5-9, during the decade 1930-1940,  $B$  = number of births recorded during the period 5 to 9 years prior to the 1940 census date (*circa* 1930-1934.)

#### TESTS OF DATA ON PLACE OF RESIDENCE AT A PREVIOUS DATE

The accuracy of data on residence at an earlier date can be appraised by means of balancing equations if the specified date coincides with the date of an earlier census.

For example, if two censuses have been taken 10 years apart, and at the time of the second census all persons (10 years of age and over) are asked where they lived 10 years ago, it is possible to reconstruct the population at the earlier census date for each part of the country, with appropriate allowances for mortality. This reconstructed population can be compared with the count obtained in the preceding census, and if the two are found to agree fairly closely the data on place of earlier residence can be considered good. This test can be made separately for age and sex groups if the census data on place of earlier residence are tabulated by age and sex.

Actually, where questions on place of residence at a specified previous date are asked in censuses, the date chosen is likely not to be that of an earlier census, but some intervening date, so that information is obtained on internal migration during a particular portion of the inter-censal period.

#### DIRECT CHECKS

Direct checks on the accuracy of returns relating to either place of birth or place of residence at an earlier date can be obtained from check re-enumerations for samples of the population after the census.

Data on place of birth may also be verified by searching for the birth records of a sample of the population.

## Appendix A

### THE HAITIAN CENSUS OF 1918/19

A census, admittedly incomplete, was taken in this country between September 1918 and August 1919.<sup>1</sup> Apparently the census was limited to a head count only for the various political subdivisions of the country. The results were first published in a Haitian newspaper and subsequently reproduced by Victor. These are the census data with which this appendix is concerned.

Since 1900 (if not earlier), the Roman Catholic church has provided annual estimates of the Roman Catholic population of the country, by departments. Since the population is almost entirely Roman Catholic, these figures have been used as population estimates for Haiti.

The comparison of these two sets of data in chapter I showed such disagreement that no appraisal of the census could be made on this basis alone. It is necessary to apply more tests, some of which will be illustrated in this appendix.

The results of the tests presented here should not be considered as definitive; they are presented as a detailed illustration of methods of testing a census count with non-census data. There is probably no other country with identical conditions, where the tests illustrated here could be applied without modification. On the other hand, it is believed that the reasoning followed here could be adapted to the solution of a similar problem elsewhere.

This problem has several aspects each of which will be considered separately, namely:

- (1) Appraising the church estimates;
- (2) Comparing the church estimate and the 1918/1919 census count with the 1950 census count;
- (3) Comparing the rate of growth with rates for other countries;
- (4) Using the 1950 census data by age and sex to estimate the population 30 years earlier, that is to say, as of 1920.

If the original records from the 1918/19 census were available for examination, perhaps other appraisal procedures could be employed; also, detailed information on the social, economic and geographic factors in each of the parts of the country could be used to advantage in a definitive analysis, but that is beyond the scope of this manual.

#### Appraising the church estimates

The number of church members for each of the dioceses is available for each year since 1900 (if not

<sup>1</sup> See: M. Lubin, *Du Recensement en Haiti*, 1951. Also, René Victor, *Recensement et Demographie*, Port au Prince, 1947.

earlier).<sup>2</sup> Since the boundaries of the dioceses correspond fairly well to the boundaries of the departments, these figures are taken as the equivalent of population estimates for departments.

It is not clear exactly how the church estimates were made. It is clear, however, that the pattern of population growth which they imply is most improbable. Figure A1 shows the annual estimates for each department. It appears that the changes are in the nature of "steps." For several years the figure for a given department is carried unchanged; then it is abruptly raised to a new plateau. Occasionally, the figure is abruptly lowered. For example, in the Sud, the population for the years 1907 to 1913 is reported uniformly as 325,000. In 1914 the figure is suddenly boosted to 521,000, and it is kept at this size through 1917. In 1918 it is boosted to 541,000 and allowed to remain at this level through 1920. Between 1921 and 1924 the figure is decreased to 509,000. It is then reported without change until 1929, when it is raised to 628,000. It is reported at this level annually until 1943, when the round figure of 700,000 is substituted. It is not credible that the population of this department grew in this fashion.

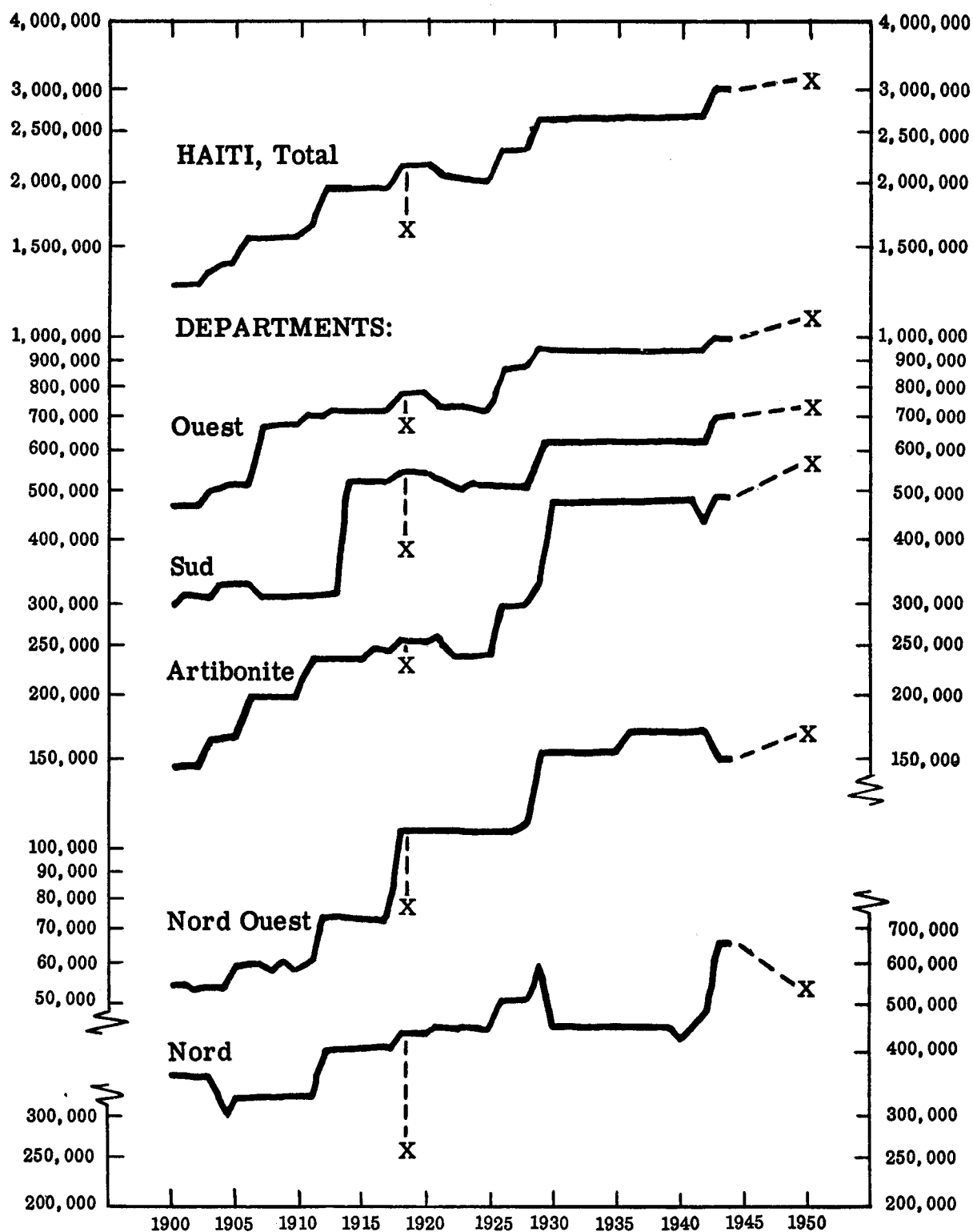
Victor noticed this type of behavior in the figures and concluded that these data must be used with caution and reservations. He added that some of the variations were explainable in terms of epidemics and other events; many of the variations, however, are not explainable in such terms and may have resulted simply from the desire to present a larger estimate of the number of church members.

The totals reported for Haiti often do not agree with the sums of the figures for the departments. In 1918, for example, the church figures are as follows:

Artibonite .....	255,000
Nord .....	437,000
Nord-Ouest .....	108,000
Ouest .....	780,000
Sud .....	541,000

The total for Haiti obtained by adding these figures is 2,121,000; the total reported for the country is 2,150,000. For this year the discrepancy is small, but for some years it is as high as 10 per cent of the total for the country.

<sup>2</sup> These data are taken from United States, Department of Commerce, *Haiti, Summary of Biostatistics*, Washington, June 1945. This report is a compilation of all data on Haiti which could be located in both official and unofficial Haitian sources. The data are reproduced without alteration, but are fully annotated; for further information regarding sources, the reader is referred to this publication, as well as others mentioned in this appendix.



The census results for Haiti and departments are indicated on the chart by X

Figure A1. ANNUAL POPULATION ESTIMATES, 1900-1944, FOR HAITI AND DEPARTMENTS, AND CENSUS RESULTS, 1918/19 AND 1950

This type of error is mainly important in so far as it suggests that the figures were not carefully compiled and that all of them may be subject to error. Each of the totals for a department presumably was obtained by adding its component parts, and the possibility of arithmetic errors in these sums has to be considered. Those who compiled the figures were apparently not much concerned with accuracy. Perhaps approximate estimates were sufficient for church purposes, but these figures cannot be accepted as accurate standards against which to evaluate the census count of 1918/19. The fact that the two sets of figures do not agree cannot be taken as proof that the census count was wrong.

#### Comparing the church estimate and the 1918/19 census count with the 1950 census count

This comparison can be made department by department, in an effort to determine whether one of the earlier sets of figures seems more reasonable than the other in light of the 1950 census count. Such a comparison is shown in table A1.

The two sets of earlier data show entirely different rates of change when compared with the 1950 census count, except for the Artibonite. In the Nord, it appears that the population increased by 111 per cent on the basis of the 1918/19 census, but only 23 per cent on the basis of the 1919 church data. Which is more nearly correct? The answer should be sought by evaluating the indicated changes in the light of known social, economic and geographic conditions in each of the departments. This analysis is beyond the scope of the present manual.

#### Comparison of rate of growth with that of other countries

The annual geometric rate of increase shown by the 1918/19 census and the 1950 census is 2.1 per cent; the rate indicated by the church estimate for 1919 and the 1950 census is 1.2 per cent. For comparison, the rates of growth in neighbouring Caribbean areas were:

Dominican Republic (1920 to 1950) .....	2.9
Jamaica (1921 to 1953) .....	1.7
Other British West Indies (1921 to 1946) .....	1.2
Cuba (1919 to 1943) .....	2.1
French West Indies (1921 to 1946) .....	0.5

Since both rates for Haiti fall within the range of rates found in the neighbouring islands, the results of this test are inconclusive.

#### Estimation of 1920 population from 1950 census data

Given the 1950 population by age and sex, and a set of mortality rates, it is possible to work backwards and estimate the population of Haiti in 1920, of which the population in 1950, aged 30 years and over, were the survivors.<sup>3</sup>

The age and sex distribution of the Haitian population in 1950 is available, but the statistics by age are markedly inaccurate. Numbers reported at ages under 5, for example, appear much too small and, in the absence of any other explanation, it must be presumed that the enumeration of children was very incomplete in 1950, and this suggests the possibility that there were important omissions also in the enumeration of adults. Inaccurate enumeration of children does not affect our computation, since only the persons aged 30 years and over are relevant to an estimate of the population in 1920; however, if the adults were under-enumerated, the 1920 estimate may be too low.

A set of mortality rates for the period 1920-1950 is not available from Haitian sources. Accordingly, it is necessary to select a set of rates from among the life tables available for various countries. Mexico as of 1930 was chosen as a country which had about the same age composition of population as Haiti in 1950, where the level of economic development and health facilities, though probably superior to those of Haiti, were not grossly different, and for which a life table was available. The life table of some other country might have been used equally well, but the Mexican life table will serve for illustrative purposes.

Thirty-year survival rates were obtained from the Mexican life table and applied to the 1950 Haitian age and sex composition. These procedures can be illustrated very briefly as follows. For males, according to the life table, the survival rate between ages 20 to 24 and 50 to 54, is 623 per 1,000. In 1950 in Haiti there were 51,300 men aged 50 to 54 years; this number divided by 0.623 provides an estimate of 82,300 males aged 20 to 24 in 1920.

These calculations were carried through for each five year age and sex cohort, and the results summed. With the addition of suitable estimates for the highest age groups in 1920, from which there were few or no

<sup>3</sup> The census as of 1918/19 was taken about 31 years prior to the 1950 census. The 1950 data could have been projected backwards 31 years, but the computations are much simpler to make for a period of 30 years and the results are good enough for illustrative purposes.

Table A1

COMPARISON OF THE 1950 CENSUS COUNT WITH ESTIMATES FOR 1919 FOR HAITI, BY DEPARTMENTS  
(Numbers in thousands)

Department	1950 census	1918/19 census	1919 church estimate	Per cent increase up to 1950 from	
				1918/19 census	1919 church estimate
Artibonite .....	569	241	255	136	123
Nord .....	540	256	437	111	23
Nord-Ouest .....	168	78	108	115	55
Ouest .....	1,095	671	780	63	40
Sud .....	740	385	541	92	37
TOTAL	3,112	1,631	2,121	91	47

survivors in 1950<sup>4</sup>, an estimated population in 1920 of 1,900,000 is obtained. This should be reduced, to allow for increase from 1918/19 to 1920, to perhaps 1,850,000, which is intermediate between the 1918/19 census count and the 1919 church estimate. It must be admitted that this estimate is subject to a rather wide margin of error. So far as it can be trusted, it suggests that the census figure is too low and the church estimate too high. A figure as low as the 1918/19 census count could

---

<sup>4</sup> These estimates are necessarily rather arbitrary, but the age groups concerned are too small to have an important effect on the result.

be accepted as roughly correct only on the assumption that Haitian mortality was much lower than that indicated by the Mexican life table. This life table gives an expectation of life at birth of about 33 years; if the expectation for Haiti had been as high as 45 years during the period in question—that is, about the same as for Trinidad and Tobago in 1930-1932—the 1950 and 1918/19 censuses would be reconciled; but this is highly improbable. Acceptance of the 1919 church estimate, on the other hand, would imply either extremely high Haitian mortality, at least equal to that of India for a comparable period, or a gross under-count of the adult population in the 1950 census.

## Appendix B

### THE RELIABILITY OF CENSUS DATA FOR LIBYA<sup>1</sup>

A check on the accuracy of the census figures for 1931 and 1936 is provided by a comparison of the annual rates of growth (geometric mean) of the native population during the interval for major geographical divisions of the country (table 1).

Table 1

INTERCENSAL INCREASE OF *de facto* NATIVE POPULATION  
BY REGIONS, 1931-6

Provinces	Population		Annual percentage rate of growth (geometric mean)
	1936	1931	
Libya .....	732,973	654,716	2.3
Tripoli .....	343,093	296,946	2.9
Misurata .....	203,922	182,953	2.2
Benghazi and Derna*..	137,582	136,215	0.2
Libyan Sahara .....	48,376	38,602	4.6

\* Data for Benghazi and Derna are combined because of large interprovincial migration during this period. Source: Vol. v, p. 21.

The rate of 0.2% for Benghazi and Derna stands in marked contrast to the rates for the other regions, reflecting the possibility of higher mortality, lower fertility, and much larger emigration resulting from the Italian military custody of one-third to one-half or more of the regional population during 1923-35. The large increase in the Libyan Sahara is possibly due to underenumeration there in 1931. The increase of 3% per year in the province of Tripoli seems hardly possible without heavy immigration. It was alleged that many exiles who left the country during 1915-30 returned between 1931 and 1936. Their return may have in part accounted for the rapid increase in the province of Tripoli.

A higher rate of increase for females than for males is shown by the census figures for natives in all provinces except Benghazi and Derna (see Appendix, table 1). For the territory as a whole the figures show a ratio of 1094 males per 1000 females among the *de facto* native population of 1931, 1062 in the *de facto* population of 1936, and 1075 in the resident native population of 1936. These figures suggest a relatively greater underenumeration of women than of men in the 1931 census. The departure of exiles prior to the 1931 census and their return during 1931-6 does not explain these figures. On the contrary, since the exiles must have been mostly men the trend of the sex ratios is contrary to that which would have been expected to result from these movements.

The hypothesis of relatively greater underenumeration of women than of men in the 1931 census is also sup-

<sup>1</sup> Reproduced by permission of the author, from Chia-lin Pan, "The Population of Libya", *Population Studies*, Vol. III, No. 1, June 1949, pp. 106-108.

ported by a comparison of the total native population by sex for 1931 with the corresponding population aged 5 years and over in 1936. The percentage reduction of females, which should reflect in the main intercensal mortality if the data were accurate, is much lower than that of males (table 2).

Table 2

TOTAL *de facto* NATIVE POPULATION OF LIBYA 1931 COMPARED  
WITH THOSE AGED 5 YEARS AND OVER IN 1936

	Population		Difference	
	1931	1936	Number	Percentage
TOTAL	654,716	623,848	30,868	4.7
Males .....	341,984	322,691	19,293	5.6
Females ...	312,732	301,157	11,575	3.7

Source: Italy, *VIII Censimento generale della popolazione* . . . , vol. v, pp. 24-5, 1939.

Differences in the numbers of persons enumerated in 1931 and 1936 as nomads and semi-nomads in various districts also suggest the possibility of errors in enumeration. . . .<sup>2</sup> The figures for certain districts show fairly large numbers of nomads or semi-nomads in 1936 where none were enumerated in 1931, and *vice versa*. To be sure, the differences may be due partly to changes in the application of the terms "nomad" and "semi-nomad," which were rather loosely defined, or perhaps to errors in the identification of geographical areas, but they suggest the possibility that sizeable groups of these people were omitted from the enumeration in 1931 or 1936, or in both censuses. Even the figures for whole provinces, as given by the two censuses, seem inconsistent; they are also contrary to the generally held views that the people of Benghazi and Derna are relatively mobile and that those of the Libyan Sahara are more nomadic than those of the northern provinces. The figures by provinces are summarized in table 3.

Table 3

Province	Percentage of native population enumerated as nomads or semi-nomads	
	1931	1936
Tripoli .....	23.6	18.1
Misurata .....	35.0	15.4
Benghazi .....	23.8	6.0
Derna .....	—	13.7
Libyan Sahara .....	4.5	10.8

The population figures by sex and age groups also show evidence of errors. They are available only for

<sup>2</sup> As appears in a table (Appendix, table 3) of the work cited.

the resident native population of 1936. The sex ratios computed from these figures are presented in table 4.

The ratios for the age groups under 5, 20-24, and 25-29 seem to be too low in relation to those of the neighbouring age groups. On the other hand, those for the group 10-14 and for all groups above 50 seem too high. To some extent these variations may be due to mis-statement of ages of men or women; and the continued absence of male exiles at ages 20-29 may help to explain the ratio for that age. However, the figures suggest the possibility of serious underenumeration of males of ages 20-29 and of females of ages 10-14 and 50 and over.

Table 4

Age (years)	Males per 1000 females
Under 5	1016
5-9	1143
10-14	1344
15-19	1111
20-24	885
25-29	968
30-39	1001
40-49	995
50-59	1173
60-69	1152
70-79	1093
80 and over	1083
All ages	1075

The proportionate distribution by age of the resident native population in 1936, which is shown in table 5, suggests underenumeration at ages under 10 and 20-29 for both sexes.

Table 5

Age (years)	Males	Females
0-4	14.5	15.3
5-9	15.0	14.1
10-19	17.6	15.1
20-29	11.6	13.4
30-39	13.7	14.7
40-49	10.5	11.4
50-59	8.1	7.4
60-69	5.1	4.7
70-79	2.7	2.7
80 and over	1.2	1.2
All ages	100.0	100.0

The census data for Italians appear to be much more accurate. The distributions by sex, age and marital status for the Italian population are characteristic of a population built up mainly by immigration. . . .<sup>3</sup>

<sup>3</sup> Data are given in the Appendix (Tables 14-16) of the work cited.

## Appendix

Table 1

PRESENT NATIVE POPULATION OF LIBYA, 1931 AND 1936, AND RESIDENT NATIVE POPULATION, 1936, BY SEX AND MAJOR ADMINISTRATIVE DIVISIONS

Administrative division and sex	Present population			Resident population 1936
	1931	1936	Percentage increase	
Libya, total	654,716	732,973	12.0	750,851
Males	341,984	377,416	10.4	388,948
Females	312,732	355,557	13.7	361,903
Province of Tripoli	296,946	343,093	15.5	351,774
Males	154,834	174,881	12.9	182,265
Females	142,112	168,212	18.4	169,509
Province of Misurata	182,953	203,922	11.5	213,486
Males	95,109	103,712	9.0	111,230
Females	87,844	100,210	14.1	102,256
Province of Benghazi and Derna*	136,215	137,582	1.0	137,426
Males	72,376	75,079	3.7	71,908
Females	63,839	62,503	2.1	65,518
Libyan Sahara	38,602	48,376	25.3	48,165
Males	19,665	23,744	20.7	23,545
Females	18,937	24,632	30.1	24,620

\* Data for Benghazi and Derna are combined because of large inter-provincial migration during this period.



## Appendix C

### THE CENSUSES OF HONDURAS

Because censuses have been taken at frequent intervals, the population statistics of Honduras provide excellent material for the application of various methods for their appraisal. In this example, it is intended chiefly to show how a series of several successive population counts can be utilized in a tentative evaluation of their relative degrees of accuracy. More definite appraisal becomes possible only after application of various balancing equations to specific census intervals, as described in the several chapters of this manual.

The series of census results for Honduras since 1881 shown in table C1 is very irregular. If all censuses had been accurate, the series would indicate moderate rates of population growth in the periods 1881-1887, 1910-1916, 1916-1926, and 1940-1945; rapid growth in the periods 1905-1910, 1930-1935, 1935-1940, and 1945-1950; extraordinarily rapid growth in the periods 1887-1901 and 1926-1930 and a decline in the period 1901-1905. Only very unusual events could have resulted in such fluctuations of the rate of growth. Actually, it is highly improbable that all censuses could have been correct; if some of them were fairly accurate, some others must have been very inaccurate.

Table C1

POPULATION OF HONDURAS ACCORDING TO CENSUSES TAKEN FROM 1881 TO 1950, AND APPARENT INTER-CENSAL RATES OF INCREASE

Date of census	Population	Apparent annual rate of increase <sup>a</sup>
1881 .....	307,289	—
1887 .....	331,917	1.30
1901 .....	543,841	3.59
31 XII 1905 .....	500,136	— 2.07
18 XII 1910 .....	553,446	2.06
18 XII 1916 .....	605,997	1.52
26 XII 1926 .....	700,811	1.46
29 VI 1930 .....	854,184	5.80
30 VI 1935 .....	962,000	2.41
30 VI 1940 .....	1,107,859	2.86
24 VI 1945 .....	1,200,542	1.63
18 VI 1950 .....	1,368,605 <sup>b</sup>	2.66
Apparent average annual rate of increase, 1881-1950 .....		2.19

<sup>a</sup> Average annual geometric rates.

<sup>b</sup> Unrevised result, as obtained from the enumeration.

Aware of these inconsistencies, the statisticians of Honduras have expressed their doubts as to the accuracy of the most recent census as well as of earlier censuses. Tosco and Mondragón have proceeded to construct a series of revised population estimates which they consider to be closer to reality than those which could be made on the basis of the census results, such

as they were.<sup>1</sup> It is not within the scope of this manual to discuss methods for the revision of past estimates. Although the revised figures to Tosco and Mondragón have been adopted for official purposes, we shall, in the present appendix, concern ourselves mainly with the raw census data, which were obtained directly from the admittedly inaccurate census enumerations.<sup>2</sup>

#### The consistency of the census totals

The inconsistency of the series of census results shown in table C1 is very obvious mainly because censuses have been taken in close succession. The effect of an error in a census enumeration on the apparent inter-censal rate of increase is very large if the census interval is short. For longer census intervals, the calculated average rates of increase are less severely affected by errors. For example, although both the 1881 and 1950 figures may have been in error, the apparent rate of increase for the entire period—2.19 per cent per annum—is plausible and is probably not much in error.<sup>3</sup> We may compute apparent rates of increases from the results of any two censuses in order to determine which figures could be consistent with one another, and which not. In this fashion, by accepting some censuses and eliminating others, we can obtain various series of census results which, taken by themselves, would not seem to be necessarily inconsistent.

One such possible series is obtained by accepting only the results of the censuses of 1887, 1905, 1910, 1916, 1926 and 1945, while rejecting the censuses of 1881, 1901, 1930, 1935, 1940 and 1950. Another possible series may include the censuses of 1881, 1910, 1930, 1935, 1940 and 1950, while rejecting those of 1887, 1901, 1905, 1916, 1926 and 1945. (See table C2.) Apparently consistent series may also be obtained by selecting other combinations of censuses. The census of 1901, however, cannot easily be fitted consistently into any series. Either the enumeration of 1901 was exaggerated (possibly intentionally) or else most, if not all, of the other enumerations were deficient.

<sup>1</sup> Tosco, M., and Mondragón, R., *Análisis dinámico y económico-social de la población de Honduras*, Servicio Informativo del Banco Central de Honduras y del Banco Nacional de Fomento, Tegucigalpa, 1952.

<sup>2</sup> The revised census age distribution has been used in the example in chapter III on page 46 for the computation of hypothetical survival rates. It was found that the revised data for 1950 were not consistent with the unrevised census age distribution of 1940.

<sup>3</sup> This rate may be compared with long-term inter-censal rates of increase for neighbouring areas. For Nicaragua, from 1906 to 1950, the apparent rate amounts to 1.70 per cent per annum; for Costa Rica, from 1864 to 1950, 2.23 per cent per annum; for Panama, from 1906 to 1950, 2.48 per cent per annum; for Cuba, from 1899 to 1950, 2.45 per cent per annum; and for Mexico, from 1921 to 1950, 2.04 per cent.

Table C2

APPARENTLY CONSISTENT SERIES OF SELECTED CENSUS RESULTS  
FOR HONDURAS, 1881-1950

Date of census	Population enumerated	Apparent annual rate of increase
<i>First alternative</i>		
1887 .....	331,917	—
31 XII 1905 .....	500,136	2.24
18 XII 1910 .....	553,446	2.06
18 XII 1916 .....	605,997	1.52
26 XII 1926 .....	700,811	1.46
24 VI 1945 .....	1,200,542	2.95
<i>Second alternative</i>		
1881 .....	307,289	—
18 XII 1910 .....	553,446	2.01
29 VI 1930 .....	854,184	2.25
30 VI 1935 .....	962,000	2.41
30 VI 1940 .....	1,107,859	2.86
18 VI 1950 .....	1,368,605	2.14

The first alternative suggests population growth at a declining rate until 1926, followed by very rapid growth.

A possible explanation for this mode of growth might be the political disturbances from which Honduras suffered during the first quarter of the present century. The second alternative, which seems more plausible, portrays a gradually accelerating trend, which might be accounted for by a gradual fall in mortality rates owing to improving knowledge in matters of health and personal hygiene; after 1940, however, there follows an unexplained drop in the rate of growth.

#### Consistency of census results with vital statistics

A series of the numbers of births and deaths registered in Honduras, beginning with the year 1910, is available.<sup>4</sup> There are minor gaps in this series, the figures not being available for the years 1917, 1922, 1924 and 1932; the missing figures can, however, be easily interpolated without the risk of any major error (the series being, on the whole, fairly smooth). Since migration was probably of small numerical importance in most years, these figures probably suffice for balancing equations of the total population in each of the census intervals beginning with 1910.

Table C3

APPARENT INTER-CENSAL INCREASES AND APPARENT EXCESS OF BIRTHS OVER DEATHS IN EACH  
CENSUS INTERVAL, HONDURAS, 1910-1950

Census interval	Apparent inter-censal increase	Apparent excess of births over deaths	Discrepancy		
			Amount	Per cent of census total at end of period	Per cent of apparent inter-censal increase
1910-1916 .....	52,551	60,126	— 7,575	— 1.25	—14.4
1916-1926 .....	94,814	111,435	— 16,621	— 2.37	—17.5
1926-1930 .....	153,373	52,474	100,899	11.81	65.8
1930-1935 .....	107,816	81,645	26,171	2.72	24.3
1935-1940 .....	145,859	100,880	44,979	4.06	30.8
1940-1945 .....	92,683	110,239	— 17,556	— 1.46	—18.9
1945-1950 .....	168,063	162,924	5,139	0.38	3.1
1910-1950 .....	815,159	679,723	135,436	9.90	19.9

The interpretation of these balancing equations depends mainly on the assumptions made with regard to the accuracy of the vital statistics. Three assumptions can be made, namely: (a) that the vital statistics were accurate; (b) that the vital statistics, though inaccurate, were of comparable accuracy throughout the period from 1910 to 1950; and (c) that the accuracy of vital statistics has improved in the course of time.

Assumption (a) would favour the interpretation that the censuses of 1945 and 1950 were fairly accurate or, if incomplete, were both incomplete to nearly the same extent. On the same assumption, the 1940 census would be somewhat more complete, but earlier censuses considerably less complete. The latter conclusions are reached by calculating backward the population according to the observed natural increase. For example, if the population in 1950 was 1,368,605, and the natural increase of the past 40 years was 679,723, then the population of 1910 should have been 688,882, whereas the census of 1910 reported 553,446. The 1910 census, then, would grossly be deficient. On the other hand, the census of 1901 would then appear more acceptable than it appeared at first.

Assumption (b) is more likely if we are inclined to believe that the census of 1910, and some of the subsequent ones, though perhaps not very accurate, were not as grossly in error as would follow from assumption (a). If, during the entire 40-year period, population actually increased by 815,159, as implied by the censuses of 1910 and 1950, but a natural increase of only 679,723 is accounted for by vital statistics, we might consider a hypothetical constant ratio of true increase to reported natural increase which is obtained by dividing 815,159 by 679,723. This ratio amounts to 1.20. Multiplying the figures for natural increase of each inter-censal period by this ratio, we obtain hypothetical estimates of true natural increase. Subtracting the latter from the census figure for 1950, we obtain alternative population estimates for each census year which may then be compared with the census figures actually obtained. This leads to the following results:

<sup>4</sup> United States, Bureau of the Census, *Honduras, Summary of Biostatistics*, Washington, D. C. 1944, and United Nations, *Demographic Yearbook*, 1952.

Year of census	Correction factor	Natural increase up to 1950		Population estimate for census year	Population reported at census
		As reported	Hypothetical estimate		
1950 .....	1.20	—	—	1,369,000	1,369,000
1945 .....	1.20	163,000	196,000	1,173,000	1,201,000
1940 .....	1.20	273,000	328,000	1,041,000	1,108,000
1935 .....	1.20	374,000	449,000	920,000	962,000
1930 .....	1.20	456,000	547,000	822,000	854,000
1926 .....	1.20	507,000	608,000	761,000	701,000
1916 .....	1.20	620,000	744,000	625,000	606,000
1910 .....	1.20	680,000	816,000	553,000	553,000

If the assumptions are correct, it appears that, relative to the censuses of 1910 and 1950, the censuses of 1916 and 1926 were deficient while those of 1930, 1935, 1940 and 1945 were in excess. Alternatively, the census of 1940 might have been nearly correct, in which case the censuses of 1930, 1935, 1945 and 1950 would have been slightly deficient and those of 1916 and 1926 greatly deficient.

In accordance with assumption (c), namely, the progressive improvement in the accuracy of vital statistics,

we might assume for the sake of illustration that the reported natural increase was nearly accurate by 1950, that it requires a correction factor of 1.20 for the entire 40-year period, but that the correction factor diminishes by 0.005 each year so that, by 1950, it equals unity. It must be admitted that the sizes of the correction factors assumed for different years are highly arbitrary, but the hypothesis of registration improving over a long period of time is in accordance with the experience of many countries. The results of this computation are presented below:

Year of census	Correction factor	Natural increase up to 1950		Population estimate for census year	Population reported at census
		As reported	Hypothetical estimate		
1950 .....	1.000	—	—	1,369,000	1,369,000
1945 .....	1.025	163,000	167,000	1,202,000	1,201,000
1940 .....	1.050	273,000	287,000	1,082,000	1,108,000
1935 .....	1.075	374,000	402,000	967,000	962,000
1930 .....	1.100	456,000	502,000	867,000	854,000
1926 .....	1.120	507,000	568,000	801,000	701,000
1916 .....	1.170	620,000	725,000	644,000	606,000
1910 .....	1.200	680,000	816,000	553,000	553,000

The population estimates made on this assumption show a surprisingly good agreement with the results for 1930, 1935, 1940, 1945. The average annual rates of natural increase implied by the estimates on this assumption have been computed for each time-interval. From 1916 onward, the rate of increase rises gradually, from 2.0 per cent per annum in 1916-1926 to 2.6 per cent in 1945-1950; this trend is quite similar to the experience of other countries in Central America. For the 1910-1916 period, however, the implied rate of natural increase is 2.5 per cent, a figure which is too high to fit the pattern. This implies that recorded natural increase in 1910-1916 was probably more nearly correct, before the adjustment, than that for later periods.

Because of their dependence on unproven assumptions relating to the vital statistics, these tests are in-

sufficient. Nevertheless, the reasoning underlying assumption (c), namely, that vital statistics improved in the course of time, though perhaps with some temporary relapse after 1916, is probably in the right direction. If so, it seems that the census series is fairly consistent except for the censuses of 1916 and 1926, which were probably incomplete. The evidence, however, requires strengthening by means of balancing equations designed to appraise the accuracy of the vital statistics series in the several inter-censal periods.

Statistics are available for additional tests. For several censuses, the sex-age distribution of the population, as well as its geographic distribution, by department and municipality, have been published. Statistics of deaths are given by sex and age. With these data, many of the tests described in this manual can be performed.

