Studies in Population

The Methods and Materials of Demography

Henry S. Shryock
Jacob S. Siegel
and Associates

Condensed Edition by
Edward G. Stockwell

The Methods and Materials of Demography

Henry S. Shryock
Jacob S. Siegel
and Associates

Condensed Edition by
Edward G. Stockwell
Bowling Green University
Bowling Green, Ohio

ACADEMIC PRESS New York San Francisco London

A Subsidiary of Harcourt Brace Jovanovich, Publishers

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1

Library of Congress Cataloging in Publication Data

(Studies in population)
Includes bibliographies.
1. Demography. I. Siegel, Jacob S.
II. Stockwell, Edward G. III. Title. IV. Series.
HB881.S526' 1976 301.32'01'82 76-18312
ISBN 0-12-641150-6

PRINTED IN THE UNITED STATES OF AMERICA

The Methods and Materials of Demography

Condensed Edition

STUDIES IN POPULATION

Under the Editorship of: H. H. WINSBOROUGH

Department of Sociology
University of Wisconsin
Madison, Wisconsin

Samuel H. Preston, Nathan Keyfitz, and Robert Schoen. **Causes of Death: *Life Tables for National Populations.***

Otis Dudley Duncan, David L. Featherman, and Beverly Duncan. **Socioeconomic Background and Achievement.**

James A. Sweet. **Women in the Labor Force.**

Tertius Chandler and Gerald Fox. **3000 Years of Urban Growth.**

William H. Sewell and Robert M. Hauser. **Education, Occupation, and Earnings: *Achievement in the Early Career.***

Otis Dudley Duncan. **Introduction to Structural Equation Models.**

William H. Sewell, Robert M. Hauser, and David L. Featherman (Eds.). **Schooling and Achievement in American Society.**

Henry Shryock, Jacob S. Siegel, and Associates. **The Methods and Materials of Demography. *Condensed Edition by Edward Stockwell.***

In preparation

Samuel H. Preston. **Mortality Patterns in National Populations: *With Special Reference to Recorded Causes of Death.***

Robert M. Hauser and David L. Featherman (Eds.). **The Process of Stratification: *Trends and Analyses.***

Preface

The present work is a condensed version of the two-volume work *The Methods and Materials of Demography* first published by the U.S. Bureau of the Census in 1971. The original work has enjoyed widespread use both in the United States and abroad.

The idea for a condensed edition of *The Methods and Materials of Demography* was suggested by Professor Halliman H. Winsborough, the editor for Academic Press's Series on Studies in Demography. This condensed edition is intended to serve several purposes. Among these are to reach a wider audience, particularly in academic circles, and to serve the needs of many users and prospective users of the previous work more effectively. It is believed that a one-volume edition of about 555 pages can much more easily serve as a text in a one-semester course in the techniques of demographic analysis and could be covered more thoroughly in a two-semester course, than the original two-volume work of 900 pages.

The authors are pleased that Professor Edward G. Stockwell was willing to undertake the time-consuming and demanding task of abridging the two-volume original. He worked from the latest edition then in print, namely the second printing of 1973; but he had available the record of the numerous but relatively minor corrections that were to be made in the third printing, which was completed in September 1975. In our opinion he has done a consistent and judicious job of selecting parts to be omitted and of providing the necessary editorial modifications.

Like the original two-volume work, this work attempts to present a systematic and comprehensive exposition, with illustrations, of the methods used by technicians and research workers in dealing with demographic data. The book is concerned with how data on population are gathered, classified, and treated to produce tabulations and various summarizing measures that reveal the significant aspects of the composition and dynamics of populations. It sets forth the sources, limitations, underlying definitions, and bases of classification, as well as the techniques and methods that have been developed for summarizing and analyzing the data.

This book is intended to serve both as a classroom text for courses on demographic methods, aimed at instructing students in how to use population data for analytic studies, and as a reference for professional workers who have occasion to use population data; but it is designed primarily as a classroom text. For courses that are focused on the subject matter of population, this condensed edition might well be used for supplementary reading. As a reference work we hope that the book will be helpful to statistical analysts and planners in private industry, city and regional planners, technicians in State and regional development agencies, and also statisticians and analysts in national agencies concerned with population matters.

The book could be employed as a text in a single semester, particularly with the omission of a few topics. On the other hand, with a more complete coverage of the book, and with supplementary readings from other appropriate publications, a full year could be devoted to the book. In addition to copious citations of relevant publications in the footnotes, there are lists of suggested readings at the end of each chapter that will identify for the interested student the more important works on demographic methods.

In the hope of making the book as widely useful as possible, we have presented the methodological material in the simplest mathematical form. For the most part, only high school algebra and elementary statistics are needed for a proper understanding of the book. However, a knowledge of elementary mathematical analysis is required for understanding some parts, particularly sections of Chapters 18 and 24, and Appendix C.

Considerable attention is paid to the data and measurement problems of the less developed countries, the kinds and quality of data that are available for these areas, and the special methods that have been worked out for handling incomplete and defective data. Although their demographic publications tend to be less adequate and less available, numerous illustrations for the less developed countries are given. The materials of the United States are also covered in relatively great detail, especially with respect to sources, definitions, and historical developments. Although national demographic materials differ considerably with respect to availability, definitions, classifications, etc., certain demographic principles and methods are essentially "culture free," and measures worked out for the United States could serve as well as for any other country.

The original work was intended to be comprehensive in its treatment of demographic methods and materials; yet, inevitably, choices had to be made as to what to include and what to exclude. These choices reflected the judgment of the authors as to what is most important and useful to demographers. This condensed version has tried to preserve the broad scope of the original work. As may be seen from the table of contents, the present book includes material on both formal demography and social and economic demography (i.e., the study of many social and economic characteristics of individuals as well as of such social groups as families and households). Several important but peripheral or applied fields are barely touched on, however. These include worker commuting, morbidity, the evaluation of family planning programs, optimum population, population quality, and various types of applied projections (e.g., needs for housing).

The derivation of most demographic measures described is illustrated by step-by-step examples using actual or, occasionally, hypothetical statistics. The official statistics are usually accepted as published for use in these calculations, and the handling of a problem is sometimes simplified, in order to focus attention on the method being illustrated. In actual practice, the official statistics for many countries would need to be evaluated and perhaps adjusted, and related procedures might be required. Accordingly, the results from the illustrative examples given here should not be regarded as necessarily valid for substantive purposes.

The original work was written in the years 1967 to 1970. Thus, it takes account of recommendations made by the United Nations and other international agencies for the 1970 round of population censuses and the plans for the enumeration, processing, tabulation, and publication of the 1970 Census of the United States. In this condensed version, the material on the 1970 Census of the United States has been updated so that the treatment of the census more nearly describes what was actually done.

The condensed edition retains the essential organization of the earlier work. In particular, all the main topics covered in that work are retained, although necessarily many details had to be omitted in the condensation. The condensation did not involve a mere mechanical reduction to achieve the brevity intended, however. Where sections were deleted, the new text was woven together again to achieve logical continuity.

The earlier work consisted of 25 chapters, grouped under five broad headings, and four appendixes. The major structural changes in the abridged edition are (1) the deletion of Chapter 22 on "Selected General Methods" and the consequent renumbering of subsequent chapters; (2) the incorporation of what was Appendix A in the larger two-volume edition into the new Chapter 23 on "Population Projections," and the renumbering of Appendixes B and D as A and B; and (3) an expansion of Appendix C to include some of the material originally covered in the now deleted Chapter 22. As compared with the unabridged edition, the present one contains less emphasis on historical development, the recommendations of international organizations, the practices of national statistics agencies with respect to demographic data, and the more technical and mathematical aspects of demographic analysis.

Either Shryock or Siegel as principal authors wrote, or reviewed and edited, each chapter in the

original work. Associate authors contributed drafts of many of the chapters or parts of chapters, either working under the immediate direction of one of the principal authors or independently. These associate authors were Maria Davidson, Paul C. Glick, Elizabeth A. Larmon, Wilson H. Grabill, and Charles R. Kindermann of the Bureau of the Census; Robert D. Grove and Robert A. Israel of the National Center for Health Statistics; Charles B. Nam of Florida State University; Abram J. Jaffe of Columbia University; Francisco Bayo of the Social Security Administration; and Paul Demeny of the University of Hawaii. Elizabeth Larmon was editorial coordinator of the unabridged edition and was generally in charge of the technical production aspects of that work. Adriana Weininger performed these tasks in the final stages. We are grateful to the Bureau of the Census and particularly to the Population Division and the International Statistical Programs Division for its generous support of the work on this book. Acknowledgement is made of the role of the U.S. Agency for International Development in providing major financial support for the original publication.

The bulk of the retyping that was necessary for the condensed edition was done by members of the secretarial staff in the Department of Sociology at Bowling Green State University, with particular acknowledgement being due to Lauretta Lahman and Lynn Schmid. Thanks are also due to Helen S. Curtis, Anne Donnelly, and Mary Hartman of the Center for Population Research, Kennedy Institute, Georgetown University, for helping with the voluminous correspondence in connection with the present edition, and to Georgetown University for other courtesies.

The first-named author would like to acknowledge his appreciation to his wife, Annie Frances King Shryock, for her patience during his work on the various drafts and editions, and for her help with some of the editorial work. The second-named author would also like to express his appreciation to Rose V. Siegel and Lorise V. Siegel for their forbearance and sympathetic support during the long period when the original work was being prepared.

Henry S. Shryock
Jacob S. Siegel

Contents

CHAPTER 1

Introduction

THE FIELD OF DEMOGRAPHY

Demography is the science of population. The word was coined by a Belgian, Achille Guillard, who published his *Elèments de statistique humaine, ou démographie comparée* (Elements of human statistics or comparative demography) in 1855.[1] He defined it as the natural and social history of the human species or the mathematical knowledge of populations, of their general changes, and of their physical, civil, intellectual, and moral condition.

Like most other sciences, demography may be defined narrowly or broadly. The narrowest sense is that of "formal demography." Formal demography is concerned with the size, distribution, structure, and change of populations. **Size** is simply the number of units (persons) in the population. **Distribution** refers to the arrangement of the population in space at a given time, that is, geographically or among various types of residential areas. **Structure,** in its narrowest sense, is the distribution of the population among its sex and age groupings. **Change** is the growth or decline of the total population or of one of its structural units. The components of change in total population are births, deaths, and migrations. In analyzing change in structure, however, we have to include the transition from one group to another. In the case of age, this is expressed as a simple function of time.

A broader sense includes additional characteristics of the units. These include ethnic characteristics, social characteristics, and economic characteristics. Ethnic characteristics like race, legal nationality, and mother tongue shade into social characteristics. Other examples of social characteristics are marital and family status, place of birth, literacy, and educational attainment. Economic characteristics include economic activity, employment status, occupation, industry, and income, among others. Other characteristics that might be encompassed are genetic inheritance, intelligence, and health; but the usual sources of demographic data, such as censuses, seldom deal with these directly. Furthermore, demography may look beyond the basic personal units to such customary social groupings as families and married couples.

The widest sense of demography extends to applications of its data and findings in a number of fields including the study of problems that are related to demographic processes. These include the pressure of populations upon resources, depopulation, family limitation, eugenics, the assimilation of immigrants, urban problems, legislative apportionment, manpower, and the maldistribution of income. This book covers very little of these fringe areas of demography.

Hauser and Duncan regard the field of demography as consisting of a narrow scope—demographic analysis—and a wider scope—population studies. "Demographic analysis is confined to the study of components of population variation and change. Population studies are concerned not only with population variables but also with relationships between population changes and other variables—social, economic, political, biological, genetic, geographical, and the like. The field of population studies is at least as broad as interest in the 'determinants and consequences of population trends.' "[2] In the same work, Lorimer also discusses demography as a discipline.[3]

DEMOGRAPHIC DATA AND THEIR USES

The types of demographic data have already been suggested. Counts of persons are obtained from censuses and sample surveys and from the files of continuous population registers. Counts of events are obtained from registered vital events (births, deaths, marriages, divorces, etc.), and also from continuous population registers. Sometimes censuses and surveys inquire about events, e.g., the number of children women have borne in the preceding 12 months. Any of these counts may be shown in the form of multiple classifications, e.g., population by sex and age for urban areas or deaths by age and cause. Various demographic measures, such as percentages, ratios, and averages may be derived from them. Furthermore, numbers of registered vital events can be related to the corresponding population to produce vital rates, for example, the number of deaths per 1,000 of the population or the number of births to married women 20 to 24 years old. Such vital rates can also be derived from continuous population registers without the use of other sources of demographic data.

The resulting demographic statistics can then be used to describe the distribution of the population in space, its density and degree of concentration, the fluctuations in its rate of growth, its movements from one area to another, and the force of natality, nuptiality, and mortality within it. These demographic statistics have many and increasingly varied applications. The fields of application include public health; local planning for land use, school and hospital construction, public

---

[1] Adolphe Landry, *Traité de démographie* (Treatise on demography), Paris, Payot, 1945, p 7.

[2] Philip M Hauser and Otis Dudley Duncan, "Overview and Conclusions" in: *The Study of Population*, Hauser and Duncan, eds., Chicago, University of Chicago Press, 1959, pp. 2–3

[3] Frank Lorimer, "The Development of Demography" in Hauser and Duncan, op. cit., pp. 157–166.

utilities, etc.; marketing; manpower analysis; family planning programs; land settlement; immigration and emigration policy; and many others. An analysis of current demographic levels and past trends is the necessary first step in the construction of population forecasts that in turn form the underpinning of national plans for economic development and other programs, including explicit population policies in some cases.

COVERAGE OF TOPICS

As may be seen from the table of contents, the scope of this book corresponds pretty much to the first and second concentric circles of the field of demography that were described in the opening section, that is, it covers formal demography and many additional social and economic characteristics of the population as well as such social groups as families and households. Several important but peripheral or applied fields are barely touched on. These include worker commuting, morbidity, pregnancy wastage through miscarriages and induced abortions, the evaluation of the effectiveness of family limitation programs, and population quality.

ORGANIZATION OF THIS BOOK

Arrangement of Chapters

The book consists of 24 chapters covering the sources of demographic data, the major topics of population size, distribution, and composition, the basic components of demographic change, and population estimates and projections.

A major problem that had to be faced in the writing was the fact that the subject and methods of a given chapter cannot be adequately described without drawing on those of a later chapter. This problem would arise no matter what order of topics or chapters was followed. In the analysis of age composition in chapter 8, for example, it is essential to make use of survival rates, which are derived by methods explained in chapter 15, "The Life Table." To understand the life table, on the other hand, we have to know how age is defined in censuses and in vital statistics. More particularly, we need to be familiar with the sources of error in the underlying age statistics in order to have some appreciation of the accuracy of a given life table.

A related problem is that a given method may apply to a number of subject fields within demography. Standardization is used with almost all kinds of rates and averages—birth, death, and marriage rates; migration rates; enrollment rates; worker participation rates and unemployment rates; median years of school completed; mean income; etc.

We have tried to cope with these problems in several ways. Frequent use has been made of forward and backward references. Nonetheless, a certain amount of duplication in the exposition was inevitable.

Treatment of Subjects Within Chapters

In addition to the conventional treatment of standard methodological issues and measures, each chapter generally contains a discussion of their uses and limitations, the quality of the statistics, and some of the factors important in their analysis.

Uses and Limitations.—The uses to which the given statistics and measures derived from them are or can be put are sometimes illustrated by actual examples from the literature. These uses often represent contributions to the investigation of basic social and economic problems or of challenging scientific questions.

The limitations of the statistics have to do with their scope and pertinency. Because of the definitions used, the restricted number and form of the questions, the scope of the census, survey, or register, and the extent of cross-classifications in the tabulations, the statistics can at best give only approximate answers to the broad questions that they were designed to answer and may be even less adequate for the purposes of some imaginative analysts. Demographers often find new and previously unforeseen uses to which they put the statistics available to them. When they do so, they need to recognize that the statistics may not apply to just the universe that is of interest to them or may measure obliquely rather than directly the phenomenon they are investigating. In short, they should be aware of these limitations and the implications of them for their analysis and conclusions.

The demographer should also be aware of the assumptions that he has to make when he extends the interpretation of standard, readily available measures or models in certain ways. For example, a life table based on the mortality experience of a given year does not describe the mortality experience of any actual group of persons as they pass through life. Neither does a gross reproduction rate based on the fertility experience of a given year describe the fertility experience of any actual group of females who started life together. With due caution, however, these measures may be used in many important descriptive and analytical applications.

We may cite also the limitations of population projections when they are actually used to predict the future. The fact that, in some situations, this type of use is unavoidable does not relieve the forecaster of considering the realism and appropriateness of the assumptions underlying the projections. Often, too, he may realize that if he had additional data at his disposal, he could express his assumptions in terms of more meaningful metrics.

Quality of the Statistics.—By the "quality" of the statistics is meant the degree to which they measure what was intended to be measured when the questions and procedures were designed. Hence, the term quality as used here is measured within the limitations of the statistics and does not include those limitations. "Accuracy" could be used as a synonym for quality.

Factors Important in Analysis.—Only a limited understanding of demographic phenomena is obtained from statistics (percentage distributions, rates, etc.) presented for a topic in isolation. Such so-called "inventory" statistics have mainly a descriptive value. Inventory statistics for a large number of small geographic areas can be very useful in ecological analysis; but here an important additional factor, namely, geographic location has already been introduced. In some countries, racial or ethnic group may be associated with striking variations in the topic (mortality, literacy, family income, etc.) being investigated.

The "factors" treated in these subsections can be regarded in several different ways. In ordinary statistical terms, they are the "independent variables" used to study variations in the "dependent variable" (or the topic under consideration). Hence, they may be introduced into the analysis in various ways as "controls" or as variables to be held "constant." The most important control factors in much demographic analysis are age and sex. Any of these factors may be introduced by means of cross-classification (and then by computation of factor-specific rates), by standardization, or by correlation.

To a lesser extent, the topic in question may itself be treated here as an independent variable. In that case, we study its effects on other demographic variables (or qualitative attributes).

More broadly, these subsections occasionally mention a few comprehensive studies in which the topic is an important element. These studies place the topic in the context of substantive research and give the reader a broader view of how the data and methods he has just examined in detail have been applied in some of the classical demographic works.

DESIDERATA IN DEMOGRAPHIC DATA AND STATISTICS

There are certain properties that demographers would like their data to possess. Taken literally these desiderata are of the nature of ideals in that even in the most advanced countries they are never fully attained. Nonetheless, they are goals that should be kept in view by agencies that produce demographic data. Moreover, whenever compromises are made with these goals because of practical restrictions, the nature of these deviations should be imparted to the public in the resulting statistical publications. Furthermore, when unintended errors of coverage, classification, etc., occur in the data gathering and processing, some resources should be devoted to evaluating them, on a quantitative basis if possible.

We are concerned with all types of demographic data—those from censuses, sample surveys, vital records, continuous population registers, and immigration and emigration controls. In general, the data should completely cover the units of the specified universe or a representative sample of them. The data should be comparable over time and space. They should be published promptly and in adequate detail. Some considerations apply to only censuses and others to vital statistics, and so on. Let us start with more specific consideration of censuses.

Censuses

A recent U.N. publication lists four essential features of a population census:

1. Each individual is enumerated separately and the characteristics of each person are recorded separately.
2. The census covers a precisely defined territory and includes every person present or residing within its scope.
3. The population is enumerated with respect to a well-defined point of time, and data are in terms of a well-defined reference period.
4. The censuses are taken at regular intervals.[4]

The Census as Part of an Integrated Program of Demographic Statistics

In planning a national population census, attention should be paid to its interrelationships with sample surveys, vital statistics, population registers, etc. The census may serve as a frame for sampling surveys to be taken shortly after the field work of the census is completed; such surveys may probe in greater depth topics included in the census, or they may investigate additional topics. In the case of periodic intercensal surveys, the census may serve as a benchmark in various ways. For example, it may serve as an occasion for updating the population "controls" to which the percentage distributions obtained in the survey are applied. As part of the census evaluation program, matching studies can be carried out between a sample of cases in the census and a survey taken at a nearby date. Such matching studies can also be carried out between the census and vital records, universal population registers, etc. Statistical totals can also be compared to measure the net differences. In the case of comparisons between the census, on the one hand, and a survey, population register, etc., on the other, both demographic programs will benefit from these matching studies and statistical comparisons unless one source is vastly superior to the other.

To achieve an integrated program of demographic statistics, it is essential that consistent concepts and definitions be employed throughout. When the respective programs are administered by different agencies, however, consistency may be difficult to achieve in some particulars since different program needs may seem to call for different treatments.

By the same token, international comparability is even more difficult to accomplish. Moreover, there may be such profound cultural, social, and economic differences among countries that, even if the phrasing and definition of questions are identical (and if identity can be achieved through translation), the resulting data from certain questions may not be comparable. Let us cite a few examples. If most of the people in a Far Eastern country are accustomed to figure their ages by the Oriental mode, simply asking for the age in completed years will not necessarily yield the Western-style age. More elaborate artifices may be required, and even these may yield only an approximation to the Western-style age distribution. If plural marriage or consensual unions are widespread, the conventional Western categories of marital status may give an unreal description of the population's distribution by marital status. Finally, if most of the working population are not part of a market economy and if there is considerable underemployment, the conventional labor force questions will yield meaningless statistics on unemployment. As Linder puts it, "A standard international definition for such characteristics can achieve comparable statistics only to the extent that the social and economic features of the country are similar. In two dissimilar countries, it may be that greater comparability of meaning for a census item can be obtained if the census question is asked in distinctly different ways."[5]

As source material, demography requires both regular censuses and the registration of vital events. In theory, a continuous, universal population register might suffice; but, in practice, even the best registers need to be checked periodically by a census. Moreover, a register cannot ordinarily accommodate as many items of social and economic characteristics of the person as a census can nor can new items be introduced so readily. Where census and vital records are not both adequate, as is true in many statistically underdeveloped countries, the demographer will work with what is available, using, for example, the techniques described in chapter 24.

Sample Surveys, Registration Systems, and Other Sources

Current sample surveys like censuses represent periodic stock-takings, whereas registration systems (universal population registers, vital records, etc.) and immigration and emigration control systems involve the continuous recording of data

[4] United Nations, *Principles and Recommendations for the 1970 Population Censuses*, Statistical Papers, Series M, No. 44, 1967, pp. 3–4.

[5] Forrest E. Linder, "World Demographic Data," in Hauser and Duncan, eds., op. cit., pp. 323–324.

even though the published statistics are extracted only at periodic intervals. Thus, some of the desiderata in census data are not applicable to registration data.

Sample Surveys.—Unlike censuses, many surveys are taken under private auspices. Although a periodic sample survey constitutes a continuing "micro-census" of a country or other area, a one-time survey can provide useful data for an *ad hoc* purpose. Surveys taken as census pretests or as experiments in survey methodology need not necessarily result in published tables based on the substantive demographic data. In other respects (individual enumeration, universality, and simultaneity), the scope of the survey and the methods of recording the data should be subject to the same principles as are censuses.

Registration Systems.—In the interests of national uniformity, it is best for a register to be operated by an agency of the national government. In some countries with a federal form of government or at least with relatively autonomous state governments, however, good vital statistics have been produced under an arrangement whereby the states handle the registration process and the national government exercises a coordinating function, setting standards and publishing national reports, for example.

The equivalent of the desideratum "individual enumeration" in the census or survey is the individual vital record. As Hauser and Duncan put it, "To provide adequate data for demography, a vital registration system must include procedures to assure the filing of a uniform record for every vital event—for example, live birth, death, stillbirth, and marriage; to provide for complete and usable answers to the inquiries on the record form; and to enable the information in the record form to be processed for statistical purposes—that is, edited, coded, tabulated, and presented, preferably through some central office which provides vital statistics for the nation and its subdivisions on a comparable basis. Principles and procedures for achieving these objectives have been evolved over the years."[6] (See also United Nations *Handbook of Vital Statistics Methods,* Studies in Methods, Series F, No. 7, 1953.)

A registration system should likewise have "universality within a defined territory." The principle of "simultaneity" does not apply, of course; but it suggests another criterion that is pertinent to registration, namely, a specified maximum interval of time between the occurrence of the event and its recording. For most demographic purposes, moreover, the data should be tabulated as of the date of occurrence, not the date of recording. On the other hand, tabulation by the place of occurrence is less useful than tabulation by the place of residence of the person concerned.

Again, "defined periodicity" would not apply to the continuous recording of data, but it does suggest to us that compilations of the records for statistical purposes should be made periodically. (The date—day, month, year—of both occurrence and registration should be on the record form.) The usual interval for publication purposes is the year; but some series should be published quarterly, or even monthly. Finally, the compilers of statistics from registers, like the agencies that collect and tabulate census data, are obligated to publish, evaluate, and, to some extent, analyze them.

[6] Hauser and Duncan, "The Data and Methods," in Hauser and Duncan, eds., op. cit., p. 62.

SOME BASIC DEMOGRAPHIC METHODS

As was noted above, it is not possible adequately to discuss demographic methods following a simple linear path. At many points, we have to refer to what lies ahead, or to the left or right, as well as to what lies behind. Partly for this reason, an early overview of how demographers organize and analyze their data seems necessary.

A good summary account of the essence of demographic methods is given by Hauser and Duncan.[7] After first pointing out that demography shares many of its methods with science in general and especially with statistics, they list certain groups of techniques that tend to be peculiar to demography. These techniques include: (a) techniques of data collection; (b) techniques of data evaluation and adjustment and of statistical estimation; and (c) techniques of analysis including demographic projections or forecasts. We will not comment here on (a) but rather will focus on (c) with some attention to (b), since the respective techniques are interdependent to some extent.

> "For dealing with population 'statics' the demographer depends largely on general statistical descriptive techniques with some special rates and graphic devices which have become widely used. In this latter category are such things as the sex ratio, the dependency ratio, the index of displacement, and the population pyramid . . . To deal with population dynamics, demography has developed a rather comprehensive and elaborate set of 'rates' designed to measure vital events or components of population change, such as natality, mortality, morbidity, marriage, divorce, and migration. . . . On the whole, demographic rates are designed to measure change and are calculated as approximations to a posteriori probability statements."[8]

The Balancing Equation

The most basic method of demography is the decomposition of population change into its components, or, conversely, the synthesis of the components to estimate the total population change. Schematically, we may express this process in terms of the fundamental equation

$$P_t - P_o = B - D + I - O, \tag{1}$$

where P_t is the population at the end of the period, P_o that at the beginning of the period, B is births, D is deaths, I is in-migration, and O is out-migration.

This simple equation, which is called the "balancing equation" (or the "inflow-outflow relationship" or the "component equation") has many forms and many uses. To be exactly true (i.e., represent a necessary relationship), it must apply to a fixed territory and there must be no measurement errors. In fact, the equation may be used to estimate the net error in this system of demographic statistics. If we find that the right-hand side differs from the left-hand side by an amount e, then we can write,

$$P_t - P_o = B - D + I - O + e \tag{2}$$

Here e can be called the "residual error" or the "error of closure." On the basis of additional knowledge about the

[7] Op. cit., pp. 70–73.

[8] Ibid., p. 70.

accuracy of the various terms, one may be able to decide whether e can be attributed as a measurement error almost wholly to a particular term in the equation. For example, if there is evidence that the right-hand terms are all based on very accurate registration data and the population figures come from successive censuses, then e would represent the relative accuracy of coverage of the two censuses. If e is positive, P_t is more nearly complete than P_o; if e is negative, then the reverse would be true.

Let us consider some other illustrations of the uses of (1) or its variations. Suppose that a country has adequate vital statistics and statistics on immigration and emigration. Then t years after the last census but before the next census, it is desired to make a postcensal estimate of the current national population. We have

$$P_t = P_o + B - D + I - O \tag{3}$$

In this form, the equation may be thought of as the "basic estimating equation," which uses a straightforward bookkeeping procedure.

If we are interested in projecting the population to a future date, we can use equation (3) in principle by making assumptions about the future births, deaths, and migration. Especially in the case of births and deaths, however, these assumptions are ordinarily made in the form of fertility and mortality **rates**, not in the form of the absolute numbers of births and deaths. The nature of these rates is another fundamental part of demographic methodology and will be discussed presently.

For another application, suppose that we have two successive population counts for a subnational area (province, county, commune, etc.). We also have vital statistics on births and deaths but no statistics on internal migration (or on the extent to which external migration affects the individual subnational areas). Then we may write

$$M = I - O = (P_t - P_o) - (B - D) \tag{4}$$

where M is the net migration to or from the area. In other words, to estimate the intercensal net migration for the subnational area, we subtract the natural increase, $B-D$, from the total population change, $P_t - P_o$.

First, however, we should mention another kind of elaboration of the balancing equation, namely, its use for a population subgroup, such as the male population, the female population of childbearing age, the native population, or university graduates. For some subgroups, the males in the native population, for example, we simply have to obtain the corresponding components, i.e., statistics on births, deaths, and migration for that subgroup. This restriction may be expressed by using the superscript i, to denote the subgroup, thus:

$$P_t^i - P_o^i = B^i - D^i + I^i - O^i \tag{5}$$

In the case of an age group, however, the very specification of the group, i.e., its age, changes over the period. For example, if t is 10 years, then we should compare age x at time 0 with age $x+10$ at time t. The identification of age is more complicated in the case of the components. For example, we shall need to have death statistics for ages x, $x+1$, $x+2$, . . ., depending on the time elapsed since the first census, and then we shall need to redistribute them into other age groupings. For persons aged x at that census, for example, we shall need to have part of the deaths at age x, part of those at age $x+1$, $x+2$, etc., depending on the time elapsed since the census.

In this situation although age changes, there is something about the population subgroup that remains the same. A group having such a common property is what demographers call a "cohort." Here the cohort is all the people who were born in a given year or their survivors. There are other types of demographic cohorts, such as marriage cohorts, or all the marriages that occurred in a given year. A great deal of demographic analysis is carried out in terms of cohorts.

Not all types of population subgroups are entered or left only by means of birth, death, migration, or growing older. The numbers in most social and economic subgroups are also affected by changes of status. For example, the number of citizens of a country is changed by naturalizations and losses of citizenship as well as by demographic factors. The number of married persons is increased by marriages and decreased by the number of "widowings" and divorces. In general, there are at least two gross components, a minimum of one positive and one negative one, that must be added to the right hand side of equation (3) to allow for these changes in status.

We analyze changes in population subgroups not only because we are interested in particular subgroups *per se* but also because this form of decomposition of population change gives us a better understanding of the nature of changes in the total population and of variations over time in these changes. The demographer's interest in the factors to be taken into account in this decomposition process is often limited only by the data that are available to him. The decomposition is more effectively carried out by multiple cross-classification rather than by simply distributing the population by one factor at a time. An example of a multiple cross-classification would be the distribution of the population by age, sex, and marital status, in which one subgroup or cell is the number of single females 15 to 19 years old. "Having accomplished what he regards as a suitable decomposition of population changes (or variations in rates of change from one period or place to another), the demographer may, of course, put the components back together again, e.g., in the form of a 'balance sheet,' a mathematical model, a statement of the relative importance of the several components, or the backward or forward projection of population changes, which may involve assumptions about the several components and the ways in which they are likely to change." [9]

The general topic dealing with deaths is called **mortality** and that with births, **natality** or **fertility**. These terms apply both to absolute numbers and to rates, and both to totals and to births and deaths specific with respect to various characteristics.

Rates and Ratios

As we have mentioned, demographic analysis also makes abundant use of rates. The term "rate" most appropriately applies to the number of demographic events in a given period of time divided by the population at risk during that period. Thus, we may speak of the number of deaths divided by the population as the death rate. The population at risk is usually only approximated. It may be the population at the middle of the period (which is roughly the average population during the period), the population at the beginning of

[9] Hauser and Duncan, "Overview and Conclusions," in Hauser and Duncan, eds., *op. cit.*, p. 4.

the period, or a more complex definition. The period is usually a year and the rate is often expressed per 100 or per 1,000 of the population. For example, the **crude** birth rate, which is based on the midperiod population, is

$$b=\frac{B}{P}\times 1{,}000 \quad (6)$$

where B is the number of births in the year and P is the midyear population. Similarly, the **crude** death rate is given by,

$$d=\frac{D}{P}\times 1{,}000 \quad (7)$$

The adjective "crude" is especially appropriate in the case of the crude birth rate because obviously men, children, and old people are not at risk of having a baby. (Although rarely used, there are "paternal fertility rates," which relate the number of children sired to the number of adult males.) In the case of the crude marriage rate, persons who are in the married state throughout the year are likewise not at risk. Various refinements are introduced in the population base in order to obtain a more meaningful rate. These may not only omit that part of the population for which the risk is zero, but they may also take account of the fact that the risk is much greater for some population subgroups than for others. For example, all persons are at risk of dying but the risk is much greater at age 90 than at age 10. Accordingly, we have age-specific rates, for example, the number of deaths during the year to persons aged 20 to 24. The general formula is

$${}_nr_x=\frac{{}_nE_x}{{}_nP_x} \quad (8)$$

where E is the number of events, x is the initial age, and n is the number of years in the age group. Note that here the numerator is restricted so as to correspond to the age restriction in the population. Rates may be specific for any other characteristic into which the demographic data are subdivided. Common characteristics are sex, race, nativity, and marital status. Moreover, the rates may be specific for two or more characteristics simultaneously.

In addition to the distinction between crude and refined rates, demography follows actuarial science in distinguishing central rates from probabilities. In central rates, the denominator is the population at the midpoint of the period (or the average population during the period). Probabilities, on the other hand, are based on the population at the beginning of the period, which is then viewed as the population at risk of experiencing the event during the period. The distinction is somewhat blurred for short periods when the population is not closed, e.g., when it is subject to immigration or emigration. In an "open" population, the midperiod population may be viewed as the average population at risk during the period, or the number of person-years that the population is at risk.

Suppose we are concerned with deaths at a given age, x, in a given year. The central death rate would be given by the number of deaths occurring at age x in the given year divided by the population of that age at the middle of the year. The probability of death, on the other hand, would be given by the population at exact age x at the beginning of the period divided into the deaths occurring at age x during the year; the deaths are in the same cohort as the population. When the population is not closed, the inclusion of any deaths of immigrants at age x and the exclusion of any deaths of emigrants make the resulting rate not strictly a probability because the deaths are no longer restricted to those occurring to the initial population.

The term rate is also loosely used to refer to the ratio between a population subgroup and the total population where the definition of the subgroup reflects a prior event. For example, the number of persons reported in the census as migrants to an area during t preceding years divided by the total population of the area is often called the in-migration rate. The underlying justification here is that the number of enumerated migrants roughly approximates the number of persons who migrated into the area during the specified t years—although it always falls short of that number. The population counted in the census is likewise only an approximation to the population at risk of migration during a specified period. A rate like an illiteracy rate is even more similar to a simple proportion or percentage. It is simply the percent of persons in a given population subgroup who are classified as illiterate. The acquirement of literacy may have occurred many years earlier.

Still other types of ratios are called "rates" in demography. For example, the infant (or infantile, in British usage) mortality rate,

$$\text{IMR}=\frac{\text{Deaths to children under 1 year of age during the year}}{\text{Births during the year}}\times 1{,}000 \quad (9)$$

Refinement of the rate takes account of the facts that some of the deaths in the numerator occurred to births of the preceding year and that some of the births in the denominator are at risk in the following year.

Survival rates are very frequently used in demographic analysis. In a closed population (i.e., a population with no external migration), a survival rate is the ratio of the number of persons in a cohort at one date to the number at an earlier date. Survival is from a given age to a subsequent age. There are 1-year, 5-year, 10-year, etc., survival rates. Survival rates are typically calculated from age distributions in two successive censuses or from a life table. Survival rates, too, may be specific for various characteristics or population subgroups, e.g., we may have age-specific survival rates for females, married men, or the country's aboriginal population.

By multiplying a population subgroup at one census by its appropriate survival rate, we obtain the **expected population** t years later. The expected population can be used in a number of ways—to make estimates or projections of the population, to estimate net migration during the period, to help in estimating the contribution of certain factors to the total change during the period, or to estimate the relative completeness of two successive census counts. By dividing the survival rate into the population, one obtains an estimate of the size of the cohort t years earlier, for example at the preceding census date.

In general, rates are of interest in their own right in describing the dynamics of population. They may also be applied, as just illustrated, by multiplication or division to demographic aggregates to obtain various types of estimates. For example, a death rate for a prior period or one from a life table may be multiplied by an appropriate population to estimate the number of deaths in a more current period. Conversely, on the assumption that the death rate has remained constant, a rate for an earlier period (say the census year) may be divided into the number of registered deaths to obtain a postcensal estimate of the population.

We have shown that the distinction between "rates" and "ratios" in demography, as elsewhere, is somewhat fuzzy.

Restriction of the term "rate" to probabilities or to fractions where the numerator is part of the denominator is not observed in actual practice. There is a tendency for what are called "ratios" (e.g., the sex ratio or the dependency ratio) to be used for descriptive purposes in population studies and for rates to be used in the analysis of change, i.e., in population dynamics: but this is only a tendency not a universal rule. Some writers even use the term "survival ratio."

Life Tables

Several references have already been made to the life table. A whole chapter, chapter 15, is devoted to this subject—the definition and interpretation of life table functions, the construction of life tables, and some of the applications of life tables in demography. Other applications are scattered through the book.

The primary purpose of the life table for the actuary is to measure the expectation of life at each age, $\mathring{e}_x$. The table is built up from age-specific death rates, conventionally those observed in a single year or period of years for all cohorts that were alive and hence subject to the risk of dying in that period. An example of a life table is given in table 15-1. In addition to the expectation of life at age x, it may be observed that the life table displays a number of other functions. These are:

- $_nq_x$ proportion of persons alive at beginning of age interval dying during interval
- l_x number living at beginning of age interval of 100,000 born alive
- $_nd_x$ number dying during age interval
- $_nL_x$ number living in the age interval
- T_x number living in this and all subsequent age intervals

The subscript n is the number of years of age in the age interval. For a complete life table, it is 1; for an abridged life table, it is usually 5. A more precise definition of $\mathring{e}_x$ is the average number of years of life remaining at the beginning of the age interval. $_nL_x$ may also be defined as the number of man-years lived in the age interval, and $_nT_x$ as the number of man-years lived in this age interval and all subsequent age intervals.

A life table ordinarily starts with 100,000 births, a benchmark which is called the radix. $_nL_x$ may also be viewed as describing the age distribution of a population called the **stationary population** because it is constant over time. This population is generated by the constant 100,000 annual births and the death rates of the life table. The survival rates previously referred to may be derived from the $_nL_x$ values. Thus, the probability of surviving k years for persons in the age group x to $x+n$ is given by

$$\frac{_nL_{x+k}}{_nL_x} \qquad (10)$$

The survival rate is a particularly useful measure in demography.

Unabridged life tables are usually constructed by actuaries, who are concerned with the careful graduation or smoothing of the basic data or the calculated values. There are, of course, many actuarial applications of life tables in the field of insurance. Demographers often construct their own life tables from deficient data, for special population subgroups, or by shortcut methods.

Both actuaries and demographers are interested in what are called **multiple-decrement tables**. In these tables, the population is subject to attrition not only from the force of mortality but also from that of some other factor or factors as well, for example, marriage, widowhood, or entry into the labor force. Thus, we might start with 100,000 single persons and apply to their "survivors" at successive ages not only mortality rates ($_nq_x$ values) for single persons but also a proportion that represents the probability of marrying during the interval.

Demographic Models

Increasing use is being made of model-building in demography. The life table is one such model. More complex models link together component models of fertility, mortality, and migration. Most models, however, represent closed populations. In addition to the life table, one of the best-known models is the "stable population."

Stable Population.—Demographers have long been interested in the question of what population structure would result if fixed age-specific schedules of mortality and fertility rates remained in effect indefinitely. More specifically, what would be the age-sex composition of such a population and what rate of growth would it have? The solution was found by Lotka and was presented in a classic paper.[10] He proved that the resulting population would eventually have a fixed age composition and a fixed rate of growth and provided the formulas for these. This population is called the **stable population** because its age composition is stable. Unlike the life table population, which is stationary as well as stable, it may increase or decrease in absolute numbers.

There are other parameters of (measures associated with) the stable population that are widely used in demographic analysis. These include the **true** (or **intrinsic**) **birth and death rates**, the **mean length of generation**, and the **gross and net reproduction rates**. The reproduction rates are particularly important and deserve a brief explanation at this point.

The gross reproduction rate represents the average number of daughters that would be borne by a cohort of females all of whom lived to the end of the childbearing period if the cohort bore children according to a given set of age-specific fertility rates. The net reproduction rate removes the assumption of no mortality before age 50, say, and represents the average number of daughters borne by a cohort of females starting life together, if there were no changes in the age-specific schedules of fertility and mortality.

The concept of stable population has proven itself a powerful tool in demographic analysis. For example, when an underdeveloped country has very inadequate demographic statistics but can be assumed to have had roughly constant fertility for many years, its age structure, life table, gross reproduction rate, etc., can be estimated fairly closely under some circumstances on the basis of a very few simple statistics using stable population theory.

On the other hand, when demographers tried to interpret or even apply such measures as the net reproduction rate in situations where fertility was changing very sharply, certain inconsistencies and limitations in these measures became apparent. Efforts were then made to refine the concepts and measures. Probably the chief modification was the shift from period to generation (cohort) measures, not only for making

[10] Louis I. Dublin and Alfred J. Lotka, "On the True Rate of Natural Increase," *Journal of the American Statistical Association*, 20(150) 305–339, September 1925

projections of reproduction but also for gauging its force more realistically Another refinement was to take account of the parity of the women and order of birth of the child when computing age-specific fertility rates. **Parity** is the number of children previously borne. Zero-parity (childless) women are at risk of having a first birth, one-parity women of having a second birth, and so on. Measures using the **interval** between marriage and a birth of a given order and between births of successive order have also been developed.

Cohort Analysis

In an earlier section, we defined and illustrated a cohort. Other examples are the persons immigrating in a given year and the persons completing a given year of school in a given year. A great deal of demographic analysis is carried out by arranging the data in cohort form, particularly when changes over time are being studied.

Cohort statistics are distinguished from **period statistics,** the latter applying to a combination of cohorts in a given year or other period. A life table or a conventional reproduction rate describes a period such as a year or group of years. There are corresponding cohort forms, which are called "generation life tables" and "generation reproduction rates," respectively.

The statistical ingredients are the same for cohort as for period analysis, the difference lies in how they are put together. The death rate of a given age group in a given year can be viewed as applying either to a period of time or to a cohort. When we compare the number of persons 40 years old at two different censuses, we are comparing two different cohorts at the same stage of life, whereas when we are comparing the number of persons 40 years old with the number 50 years old 10 years later we are analyzing change within a single cohort. Both comparisons are of value to the demographer.

Cohort analysis may be carried out either forward in time or in the reverse direction. For example, **reverse survival rates** are applied to estimate the size of a cohort at earlier dates.

Controlled and Uncontrolled Factors

Much of demography is concerned with the comparisons of populations and of demographic processes in space and time. In making such comparisons, one quickly encounters a common problem in scientific analysis—that of "uncontrolled" factors. For example, if the force of mortality is being compared for two populations, one must ask whether the difference in the crude death rates can be explained by differences in their age composition—since mortality is so highly correlated with age. The demographer applies a number of methods that are in general use in social statistics in order to cope with this problem. Some of these methods, however, have been adapted in specific ways in demography; indeed, some of them have had their greatest technical elaboration in demographic applications.

One general method, namely that of **cross-classification,** has already been mentioned. Reverting to the previous illustration, we may compare the mortality of two populations using specific rates, age for age. If there are no errors in reporting age in the census or in the death records, the finer the age detail, the better the control on that factor. We may also control or "hold constant" more than one factor at a time by multiple cross-classification. Thus, we can have death rates specific for age, sex, and race.

This method of decomposition enables us to make mortality comparisons without the disturbing influence of the specific factors used in the cross-classifications. We have lost, on the other hand, one advantage of the crude rates, namely an overall summary comparison. By the technique of **standardization,** we can derive from these factor-specific rates a summary measure that, to some extent, holds constant the influence of the factor. Thus, we can compare age-standardized mortality rates for two populations or for the same population at two different dates. What is required here is the age distribution (or distribution by the factor to be controlled) for a **standard population.** The standard population may be one of the two populations concerned, their averaged distributions, or a third population, for example, in the case of comparison of two or more provinces, the total population of their country. The population of England and Wales in 1901 was long used as such a standard, and its distribution by age and sex was referred to as the **standard million.** The stationary population of a life table may also be used as the standard for the corresponding observed population.

What we have just sketched is called the **direct** method of standardization. There is also an indirect method. Suppose we know, for the two areas to be compared, their crude rates and their age distributions and, for a standard population, its factor-specific rates (age-specific, etc.). We can then compute another kind of standardized rate. There are still more types of standardization, and this general technique is closely related to **Westergaard's Method of Expected Cases** used in conjunction with multiple classifications.[11]

The concept of an expected population is a very general one in demography and is very widely used. Suppose we have a population distributed by age and sex and, for each age-sex group, the percent having some other characteristic such as being married, enrolled in school, or in the labor force. Such percentages are sometimes called **participation** rates. Then we assume that the particular participation rate remains constant. By applying such rates to the observed age-sex structure at the later date, we can calculate the expected number of married persons, students, workers, etc., on the assumption of constant participation rates. Comparison of these expected numbers with those actually observed tells us what part of the total change in the number of married persons, etc., was attributable to the change in the participation rates as opposed to change in the age-sex structure of the populations.

Summary demographic factors that are used to control the influence of age include the expectation of life from the life table and gross and net reproduction rates. There are other general statistical methods that hold constant a given factor or that measure the relative effects of two or more factors upon a dependent variable. These include partial correlation, analysis of variance, analysis of covariance, and factor analysis.

In this section we have given a brief overview of the typical approaches in demographic analysis and some of the methods used. This discussion was by no means intended to explain those methods to the extent that the reader would feel equipped to apply them. Detailed definitions and explanations are given in the body of the book. As he reads the following chapters, the reader will sometimes encounter certain terms before he encounters their fullest explanation. We hope then that he will have received a general notion of their meaning and uses and will be able to look ahead, if necessary, by means of the forward references that are given, the table of contents, or the index.

[11] Robert M Woodbury, "Westergaard's Method of Expected Deaths as Applied to the Study of Infant Mortality," *Journal of the American Statistical Association*, 18(139) 366–376, September 1922