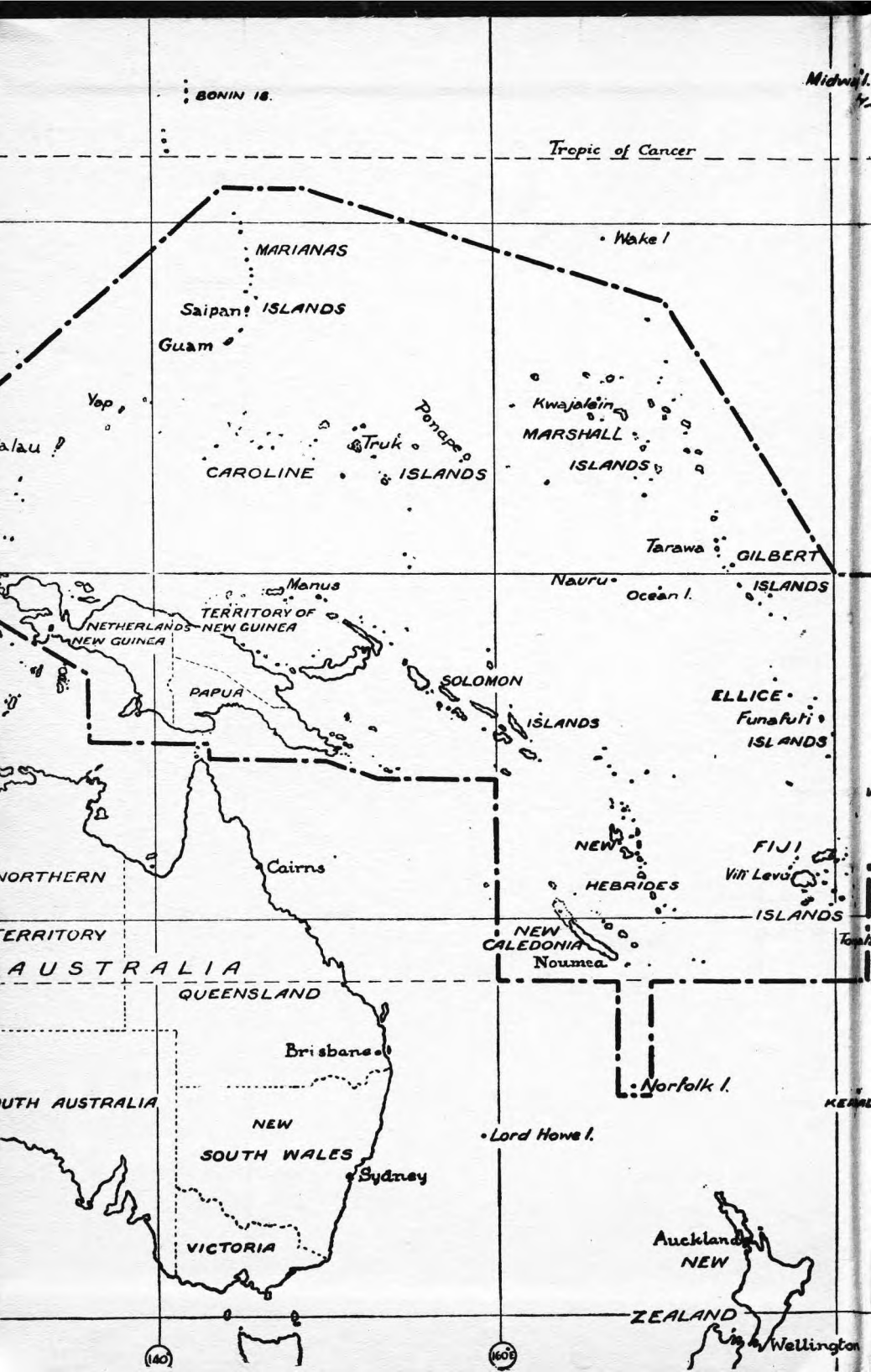


HB881 M23

1627414



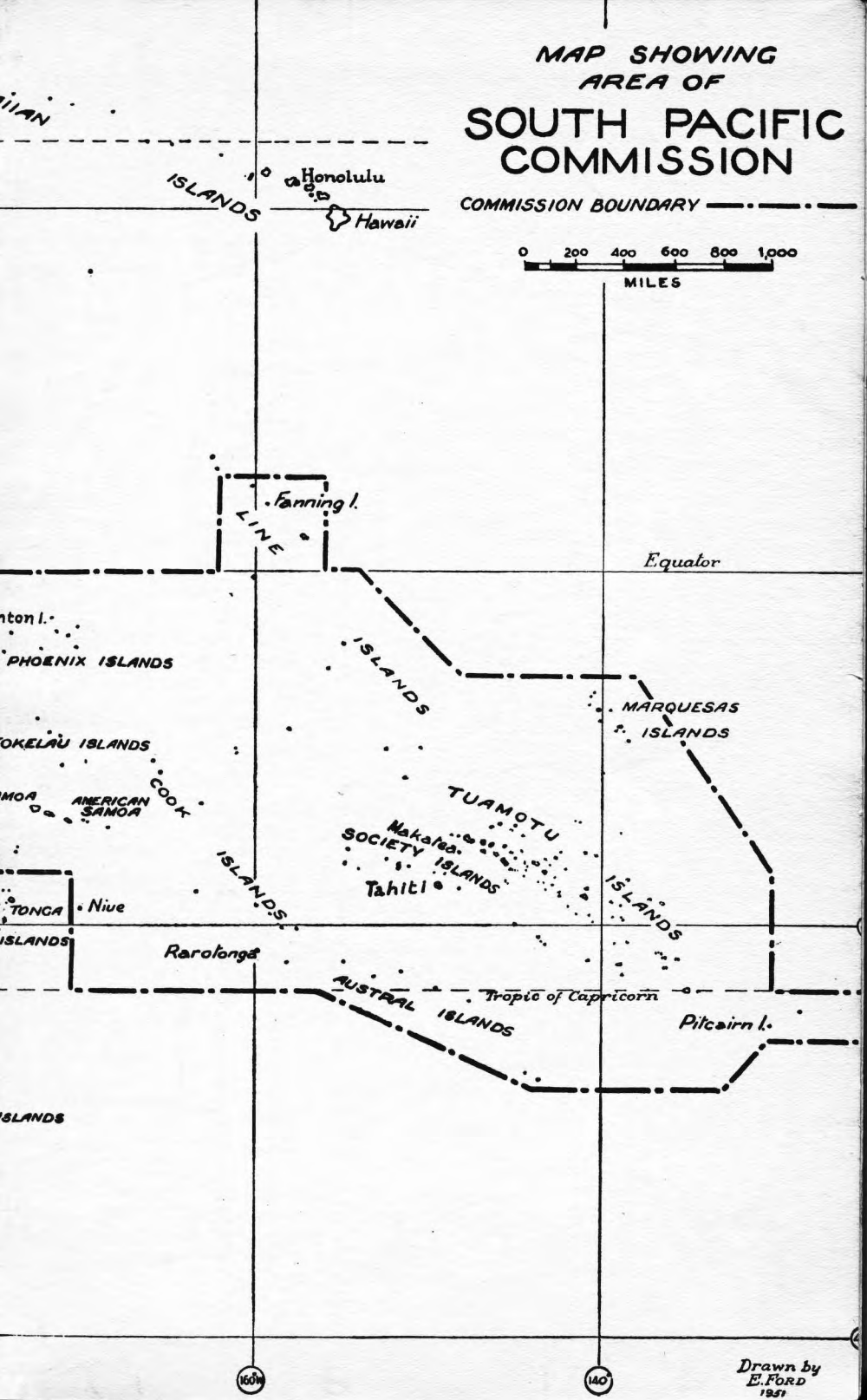
A.N.U. LIBRARY



MAP SHOWING AREA OF SOUTH PACIFIC COMMISSION

COMMISSION BOUNDARY

0 200 400 600 800 1,000
MILES



THE
AUSTRALIAN
NATIONAL
UNIVERSITY
LIBRARY

Presented by

The Publication
.....
Committee.

INTRODUCING
POPULATION STATISTICS

INTRODUCING POPULATION STATISTICS

NORMA McARTHUR

*Published under the auspices of the
South Pacific Commission and
The Australian National University, Canberra*

MELBOURNE
OXFORD UNIVERSITY PRESS
LONDON WELLINGTON NEW YORK

1961



15 JUN 1961
152133

Oxford University Press, Amen House, London, E.C.4

GLASGOW NEW YORK TORONTO MELBOURNE WELLINGTON
BOMBAY CALCUTTA MADRAS KARACHI KUALA LUMPUR
CAPE TOWN IBADAN NAIROBI ACCRA



Registered in Australia for transmission by post as a book
PRINTED IN AUSTRALIA BY HALSTEAD PRESS, SYDNEY

The South Pacific Commission is an advisory and consultative body set up in 1947 by the six Governments responsible for the administration of island territories in the South Pacific region (Australia, France, the Netherlands, New Zealand, the United Kingdom and the United States of America).

The Commission's purpose is to advise the participating Governments on ways of improving the well-being of the people of the Pacific island territories. It is concerned with health, and economic and social matters. Its headquarters are at Nouméa, New Caledonia.

The Commission consists of not more than twelve Commissioners, two from each Government. It normally holds one Session each year. There are two auxiliary bodies, the Research Council and the South Pacific Conference.

There is a Research Council meeting once a year. This may be either a meeting of the full Council, or of one or other of its three main sections, specializing in the fields of health, economic development and social development. Members of the Research Council are appointed by the Commission. They are selected for their special knowledge of the questions with which the Commission is concerned, and the problems of the territories in these fields. The chief function of the Research Council is to advise the Commission on what investigations are necessary. Arrangements to carry out those that are approved are the responsibility of the Secretary-General and other principal officers.

The South Pacific Conference, which meets at intervals not exceeding three years, consists of delegates from the local inhabitants of the territories, who may be accompanied by advisers. The first Conference was held in Fiji in April 1950, and was attended by delegates from fifteen territories and from the Kingdom of Tonga. The second Conference was held at Commission headquarters in April 1953. The third Conference was held in Fiji in April-May 1956, and the fourth Conference in New Britain in April-May 1959.

FOREWORD

A population being what it is, it can be effectively described only in numerical terms. Diagrammatic representations based on the numbers may serve to illustrate specific points about a population, or to highlight comparisons between populations but, in the long run, one cannot escape from tables of numbers. Unhappily and not infrequently, such tables inspire a feeling of ineptitude in many otherwise reasonable people. Should any of my readers share this prejudice, I would ask them to be patient for it may be that, by understanding the implications of the numbers, their array in a table will appear less formidable. More difficult to placate are those people who dislike population statistics, holding that these are the dullest, the most lifeless of all statistics, the prerogative of aged and desiccated eccentrics. Admittedly, the tedious arithmetic processes indulged in by some of us are dull and so, regrettably, are our reports. It is, however, easier to perform arithmetic manipulations than to translate their results into a lively text, and I would plead indulgence for the sacrifice of humour and light-heartedness to precision in the ensuing pages.

Written at the request of the South Pacific Commission, this book is not a fundamental treatise on the collection and interpretation of population statistics. Its aim is expressed in its title, and it was destined for the use and guidance of administrators and social scientists working in the territories of the South Pacific. All of the illustrative material presented was drawn from these territories, and while the apparent preoccupation with Fiji's population is regretted, the report on its 1956 census was the only one which had been published when the manuscript was being prepared, and it was the only territory for which the appropriate statistics of births and deaths were available. The very short list of references given at the end of the book excludes many books and papers which have contributed, indirectly and over a period of years, to this present work: the authors and titles cited are limited to those considered immediately relevant, but should any reader wish to know more about demographic measures and techniques (and there are some which have not even been mentioned here and some which deserve more attention than has been given them), he can trace them to their source through the bibliographies contained in some of these selected works.

Finally, I should like to thank Professors W. D. Borrie and P. A. P. Moran of the Australian National University for their helpful criticisms and still

more helpful corrections of the manuscript, at the same time absolving them from any responsibility for its form or content.

Norma McArthur

Canberra
January 1959

CONTENTS

	<i>Page</i>
<i>Foreword</i>	vii
<i>Illustrations</i>	
List of Figures	xi
List of Tables	xiii
 <i>Chapter I. A Population and its Structure</i>	
1. The Definition of a Population	1
2. Populations and Samples	2
3. The Fundamental Characteristics of a Population	3
4. The Composition of a Population with respect to Age	4
5. The Composition of a Population with respect to Sex	8
6. The Structure of a Population	11
7. Replacement and Dependency Ratios	12
8. Median Age	13
9. Factors Operating to Change Population Structure	14
10. The Effects of Exceptional Mortality on Population Structure	17
11. The Effect of Migration on Population Structure	20
12. The Interaction of Sources of Change and the Interdependence of these with Population Structure	21
 <i>Chapter II. The Statistics of Populations and Measures of Change</i>	
1. Sources of Population Statistics	23
2. Census Enumerations	23
3. Continuous Records	26
4. Assessment of Accuracy	27
5. Census <i>v.</i> Registration Data	28
6. Cohort Comparisons	30
7. The Risk of Mortality	34
8. Measures of Mortality: the Crude Death Rate	37
9. Age-Specific Mortality Rates	38
10. Comparisons of Mortality Rates	40
11. Summary Measures of Mortality	43
12. Standardized Mortality Rates	43
13. 'Indirect' Standardizing	46
14. Life Expectancies	47
15. Fertility Analyses	49
16. Measures of Fertility: Crude Birth and Fertility Rates	50
17. Age-Specific Fertility Rates	51
18. Reproduction Rates	52
19. Changes in Fertility Patterns	54

	<i>Page</i>
20. Cohort Fertility	55
21. Fertility Data from Censuses	56
22. The Average Number of Children per Woman	57
23. The Attainment of Motherhood	60
 <i>Chapter III. Censuses and Surveys</i>	
1. Planning a Survey	64
2. The Definition of Objectives	65
3. The Coverage of the Survey	67
4. Sampling and Sample Designs	67
5. The Size of the Sample	72
6. The Design of a Questionnaire	74
7. The Allocation of Interviewers	79
8. The Instruction of Interviewers	80
9. The Pilot Survey	81
10. Editing and Coding	84
11. Tabulation: Manual Methods	87
12. Tabulation: Sorter Cards	91
13. Tabulating Procedures	96
14. Card Preparation and Balancing	97
15. Sorting and Summarizing	99
16. Tabulation: Mechanical Methods	104
 <i>Chapter IV. The Uses and Abuses of Population Statistics</i>	
1. Review of Background Data	107
2. Urban/Rural Comparisons	109
3. Age at Marriage	109
4. Economic Background	111
5. The Interval Between Marriage and First Birth	114
6. Summary of Preliminary Investigations	118
7. The Objectives of the Survey	119
8. The Coverage of the Survey and the Design of the Sample	121
9. The Size of the Sample	122
10. The Design of the Questionnaire	123
11. Field Operations	124
12. Estimates of Mortality in Infancy and Childhood	125
13. Cautionary Comments	130
 <i>Appendices</i>	
A. A table of the limits of variation acceptable at the 5% and 1% levels of probability in samples of varying sizes	133
B. References and suggestions for further reading	134
 <i>Index</i>	 135

LIST OF FIGURES

	<i>Page</i>
1. Population pyramids representing the indigenous population of French Oceania at the Censuses of 1946 and 1951	7
2. Patterns of mortality: a graph showing the average annual age-specific mortality rates amongst males in American Samoa 1949-51, French Oceania 1950-52, and Western Samoa 1950-52	34
3. A graph showing the average annual age-specific mortality rates amongst Fijian males in 1946 and 1956	42
4. Design of a census schedule (actual size, 17" $\times$ 13½")	77
5. A specimen skeleton table for summarizing the classification of a population sample according to one characteristic, viz. marital status	89
6. A specimen skeleton table for summarizing the classification of a population sample according to two characteristics simultaneously, viz. marital status and religion	90
7. A specimen skeleton table for summarizing the classification of registered births according to the rank and sex of the child born and the age of the mother at the birth	92
8. A diagram of a 'sorter card' used in the manual sorting and tabulation of data	93
9. A specimen skeleton table for summarizing a population classified by age	98
10. A specimen skeleton table for summarizing the marital status of a population according to age	100
11. A specimen skeleton table for summarizing the total number of children born to women of each age group in the population enumerated in Provinces A, B, and C	100
12. A specimen skeleton table for summarizing the age of women at the birth of their first child in conjunction with their present age	102
13. A specimen skeleton table for summarizing the classification of first births registered in the years 1935-37 according to the calendar year of marriage of parents	115

LIST OF TABLES

	<i>Page</i>
1. The actual age structure of the indigenous population of French Oceania in 1951 classified according to age and proportion of population per 1,000 in each age group, shown side by side with its probable structure had there been no influenza epidemic in 1918 to diminish the size of the 1917-21 cohort	12
2. An abridged classification showing the composition of the two major components of Fiji's population with respect to age and sex in 1956	13
3. The number of Fijians enumerated in the Census of Fiji, 1956, at each year of age from 0 to 9 years and the numbers expected from the births and deaths registered between 1 October 1946 and 30 September 1956	29
4. The number of Indians in each birth cohort at the Censuses of 1921, 1936, 1946 and 1956, irrespective of sex	31
5. The numbers of deaths expected in a year in a population for which the age composition and risk of mortality at each age are known	36
6. The estimated annual age-specific fertility rates of females in selected territories of the Pacific: the average numbers of births per 1,000 females in each age group for the calendar years shown	51
7. The estimated annual age-specific fertility rates per 1,000 females in the Fijian and Indian components of Fiji's population in 1946 and 1951, and in the Polynesian and part-Polynesian population of American Samoa in 1940	54
8. The age-specific fertility rates experienced by cohorts of Indian females in Fiji since 1946: estimated average annual numbers of births per 1,000 women at each age	55
9. The average number of children born to women in each age group of the Fijian and Indian component populations: Census of Fiji, 1956	58
10. The proportion per 1,000 women in each age group who had borne their first child before the age of 20 years in the Fijian and Indian component populations: Census of Fiji, 1956	61
11. The numbers of females of the Indian component of Fiji's population classified according to their age at the time of the census and their age at the birth of their first child: Census of Fiji, 1956	61
12. The numbers of males and females in selected age groups of the Indian component of Fiji's population who were married or available for marriage (i.e. single, widowed or divorced) in September 1956	110
13. The numbers per 1,000 Indian males within each age range who were engaged in each major group of industries in September 1956	113
14. The fictitious distributions of first births according to the number of years separating the calendar year of marriage of mother and the year of first birth in each of three periods	115

	<i>Page</i>
15. The computational procedure for calculating the mean and standard deviation of the distribution of first births registered in 1935-37 according to the number of years separating the calendar year of marriage and the year of first birth	116
16. The means and variances of the distributions of first births shown in Table 14	117
17. The incidence of deaths in each year of age amongst cohorts of births in a population sample	127
18. Illustration of the computation of mortality rates and the standard errors of these rates	129

I

A POPULATION AND ITS STRUCTURE

1. *The Definition of a Population*

We have become so accustomed to the routine inclusion of population figures in contexts ranging from travel brochures to reports of international crises, that it is easy to regard a 'population' as no more than an inanimate, impersonal number. Sometimes the number has so many noughts at one end of it that its magnitude is beyond our comprehension; at best we may consciously register in our minds the fact that there are very many people in some particular region of the world, but we have no visual image of what this number of people, physically present in that region, really means in terms of where and how they live, nor of what each individual, the unit of that population, might do or look like. On the other hand, some of the islands in the Pacific are small enough for one person to come to know each one of their inhabitants intimately, and while this becomes less and less possible as the numbers increase, on the larger islands one person may yet have a nodding acquaintance with many of the people in a certain circumscribed area, even if he cannot distinguish each one by name.

Demographically, however, the populations of large countries and small islands are not so very different: no matter how many or how few persons there may be in some geographic area, each one of those persons is there because they were born, either in that area or in some other; if they were born in some other geographic area, then they have, at some time, moved or migrated from one place to another. Once in a particular population, they can be removed from it only by death or by emigration to some other region. The aggregate of persons present within a specified geographic area at some particular time is, therefore, the simplest definition of a population, and the size of this aggregate, the number of persons present, is the first characteristic of the population.

When describing a population, it is essential to specify both the geographic confines of the area inhabited by that population, and the time to which one refers. No population of human beings is static, and the itinerant ones of Pacific islands are perhaps more changeable than most. The total number of persons in the aggregate may change hourly for it can be altered by people coming into the population or by persons leaving. Births and immigration



will supply the accessions or gains to the population total; deaths and emigration will deplete it. The balance between the numbers of births and deaths which occur in some given period of time is known as 'natural increase' or 'natural decrease' depending on which is the larger of the two numbers, the gains through births or the losses through deaths; the balance between the numbers of persons coming into an area from outside and the numbers leaving for some other area is known as 'net migration'. These are the only sources of change possible in population numbers and the population will either increase or decrease according as the total gains to the population exceed or do not exceed the total losses.

2. Populations and Samples

By dividing up the total area which it inhabits, a population can be split into sub-populations, each of which is a 'sample' of the total population. Such samples are not necessarily 'random' in the statistical sense, nor are they necessarily representative of the total population. Depending on the form of the division of the area, in themselves the samples may constitute separate and distinct populations. The population of one island in a group, for example, is at one and the same time a geographically distinct population and a sample from the population of the whole group. Similarly the population of an urban area may be both a population in its own right and a sample of the larger population, again not necessarily representative nor statistically random. The dissection of a population according to their place of residence is not the only sort of division which can be made, and populations can be split into sub-populations or samples on the basis of criteria such as ethnic origin or language spoken. In the 1956 census of Fiji, the population was divided into eight component populations, each component being distinguished broadly on the basis of their acknowledged ethnic descent. In New Guinea, the population inhabiting one administrative district or division could similarly be broken down into sub-populations, each of which forms a distinct linguistic group, with perhaps a further dissection into samples according to certain variants in language within each of these broader groups. It may be that the definition of such sub-populations or samples is less precise than a distinction based on geographic dissection; but any population can be broken down and classified according to any one or more of the qualities possessed by the individuals who make up that population, grouping together into one class all of those who share certain attributes or characteristics.

This grouping together of people who possess attributes in common is termed the 'classification' of a population with respect to these attributes,

and with such a wide variety of personal attributes to choose from a population could be classified in an equally wide variety of ways. For a classification to be efficient, however, it must be based on criteria which will distinguish one group from another clearly; and if this rule is followed, the number of possible classifications of the population is immediately reduced to the number of personal attributes or characteristics which can be defined or distinguished precisely. What are the characteristics which distinguish, most clearly, one person from another within the same population? Further, are there any characteristics which are held in common by the members of different populations which might be used as the basis for the uniform classification of all populations? The characteristics which fulfil both of these requirements are clearly the most fundamental of personal characteristics, and therefore will provide the most essential classifications of any population.

3. The Fundamental Characteristics of a Population

The only two characteristics which fulfil these requirements simultaneously and which can be defined or measured precisely are the sex and age of a person. These are universal personal characteristics which have considerable bearing on the sort of life which each person lives, the contribution which each makes to the community of which he is a member, and the nature and extent of his demands upon that community. Of the two, age is perhaps the more important in determining the capacities, activities and needs of an individual. The resources and needs of the community to which an individual belongs is the sum of the resources and needs of all its members, and this sum is inevitably limited by the totality of the age and sex of the members and the relative importance of each in the total. The three fundamental characteristics of any population, then, are its size, its composition with respect to sex, and within this, its structure with respect to age. These represent the minimum requirements for an adequate description of a population, because it is the way in which the elements comprising the population are distributed between the two sexes over the range of possible ages which determines its potential. The demographic importance of the age and sex composition of populations will become clearer in the ensuing pages, but a few examples now may add emphasis to the importance and significance of these characteristics, not only demographically but in other fields as well.

A population will increase numerically when there are more people coming into that population over a given period of time than there are leaving it. In most populations of the Pacific area, the prime source of increase in population numbers is natural increase: this demands more accessions each year from

births than there are losses through deaths. Births can occur only if there are females in the population who are capable of producing children, and if there are males of appropriate ages to mate with these females. If there are no females of reproductive age in a population, there will be no births. Similarly, if there are females of reproductive ages but no males who are capable of reproduction, there will be no births. In a monogamous society, full reproduction requires more or less equal numbers of males and females of corresponding ages, and although polygamous societies may make do with fewer males than females, ultimately it is the number of females who are capable of reproduction which determines the number of births likely to occur in any population.

In the field of government and administration, the importance of the age composition of the population becomes immediately apparent in all aspects of planning and development. There would be scant need, for example, for the provision of day schools in areas where there were no children to attend them. The numbers of children of appropriate ages in an area, and the numbers likely to attain the age at which school attendance is considered necessary, are essential pre-requisites for the wise planning of the size of new school buildings and for the disposition of teaching staff throughout a territory. The planning of economic development, either by the initiation of new industries or by the expansion of the old, requires a different kind of person altogether. The desirable group here are the workers who are eligible for employment, and the most numerous of these will be the young adults who are just entering the work force. Although these might be enticed to move from one area to another, it will usually be more convenient, all other things being equal, to initiate a new industrial project or a new agricultural project in an area which has a reservoir of young adult males, who are already provided with accommodation, rather than to start the project in some other area, lacking such reserves of labour and in which the necessary housing would have to be provided. The expansion of an existing industry makes similar demands for workers, and neither initiation nor expansion is practicable unless there are sufficient numbers of workers available to carry out the plans. Long-term planning, either economic or otherwise, must inevitably be tied in with the resources or potential of the population, and thus, equally inevitably, related to the composition of the population with respect to age and sex.

4. The Composition of a Population with respect to Age

The numbers of persons at any age x years in a population represent the survivors of births which occurred x years previously. The expression of age as the number of years which have elapsed since birth immediately introduces

the concept of *birth cohorts*, each cohort consisting of all those persons born at the same time. Under certain circumstances, this is a more useful way of describing a population than by age, because the identity of a person by his birth cohort is invariable with time, whereas his age changes from year to year. A child born into a population during the calendar year 1950 belongs to the 1950 birth cohort. Once the year 1950 had passed, this cohort was closed, and in a population not subjected to migration, its size can never be increased. The only numerical change possible in this cohort is decrease, and with each year of life passed through, more and more of the original members will fail to survive. If migration occurs, the numbers in a cohort may increase or decrease depending on the direction of the migration, but for the time being it will be less confusing to consider only 'closed' populations in which no migration occurs. The population is then described by a series of birth cohorts, with the survivors of each cohort moving on year by year together, passing out of one year of age and into the next. The births which occur each year constitute the new cohorts, and in the interval between any two censuses, the only newcomers to a closed population are the survivors of the births which occurred during the intercensal interval.

Since some people will survive to the age of 100 years or so, the distribution of the population by single cohorts or single years of age becomes cumbersome. It is customary, therefore, to group birth cohorts or ages into five-year groups, the age groups running from 0-4 years, 5-9 years, 10-14 years, and so on. The age at the upper end of each group is always 1 year less than the age at the lower end of the next group so that there is no doubt as to where one group ends and the next begins. The age group 0-4 years includes persons at all single years of age up to, but not including, 5 years; persons aged 5 years or over but less than 10 years are included in the age group 5-9 years. For some purposes, and in some populations, five-year groups of age may be too fine a gradation, and ten- or even fifteen-year age groups may be preferable or adequate. Since the fifteen-year age groups correspond roughly to certain physiological states or changes which occur in humans (and more especially in females), these can be used with advantage in populations where exact chronological age is unknown and a person's age must be estimated. These fifteen-year groups are the broadest which should ever be used, for by increasing the range further, the structure of the population loses definition and the description becomes inadequate for virtually any purpose, be it analytical or comparative.

These physiological age groups must, of necessity, be less rigidly defined than are the groups of chronological ages, for the age at which individuals

of either sex reach puberty, for example, varies from person to person. The whole process of change from onset to full maturity in any individual may spread over two years, and there may well be a range of two or three years between the youngest age at which a child attains adolescence and the oldest. As with its age at onset, the duration of the female reproductive period is also variable. Despite the early beliefs or allegations of radical differences between populations or 'races' in the onset and span of the female reproductive period, the bulk of evidence now suggests that such differences as do exist stem from social rather than physiological causes, and that, on the average, females are physiologically capable of reproduction between the ages of 15 and 45 or 50 years. While reproduction at ages outside this range does occur, it happens but rarely, and reproduction after the age of 45 years is comparatively infrequent.

Even within this span of thirty or thirty-five years of potential reproduction, fecundity or the likelihood of conception varies with the age of the female. Starting with a relatively low probability at the lower end of the scale, fecundity increases with age during the first half of the reproductive period and then declines. Hence the first three divisors suggested in the physiological age scale are puberty, the middle of the female reproductive period and the end of this period. In males, the span of the reproductive period is less clearly defined than it is in females, but there may well be other criteria to guide an investigator in his classification of males into age groups corresponding with those of the two halves of the female reproductive period. The next oldest group can be regarded as those past reproduction, but not yet old enough to be considered too old to take an active part in the communal life of the community in which they live; this would be roughly equivalent to the age range 45 to 59 years, and the final group, 60 years and over, will consist of those who are regarded as 'aged'.

Reverting to five-year age groups, a population can be represented diagrammatically or schematically by a 'population pyramid', examples of which are shown in Figure 1. Age is scaled vertically on the line dividing the male and female sections of the population, and it is usual to depict the males to the left of centre and females to the right. The length of the rectangle built out from this central line for each age group indicates the numbers of males or females in this age group in the population. Each of these rectangles represents a cohort, and in a closed population, the size of any particular cohort in any given year depends on its size initially and on its age, or the number of years through which its members have lived. The original members of the cohort were all children born in the same year or five-year period, and once

a child is born, it is continually exposed to the risk of mortality. This risk is much higher at some ages than at others, and although the levels of mortality may vary considerably from one population to another, there is remarkable uniformity in the general pattern of risk with age.

There is usually a high mortality risk associated with the first year of life, but once this is past, the risk decreases with each succeeding year of life until adolescence is reached. From then on, the risk of dying at each year of age rises, slowly at first and then more steeply as age progresses. If the initial strength of a cohort is 1,000, the number likely to survive the first year of life may be as low as 850. In the next four years, a further 40 of these may fail to survive, so that only 810 members of the cohort will be alive at the age of 5 years. Some more of these will die with each year that passes, so that one would normally expect the numbers of persons in any cohort in a population to be smaller than the next youngest cohort, and larger than the next oldest. The number alive at any given age is, in fact, the complement of the cumulative mortality experience of that cohort.

The population pyramids shown in Figure 1 represent the indigenous population of French Oceania at the Censuses of 1946 and 1951. It will be seen that, in each case, the cohorts do not show the progressive decline in numbers which one would expect, and that the numbers of males and females

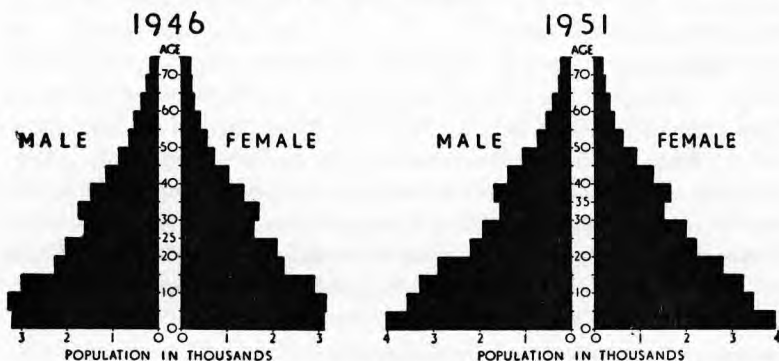


FIG. 1. Population pyramids representing the indigenous population of French Oceania at the Censuses of 1946 and 1951.

aged 25-29 years in 1946 are smaller than the numbers aged 30-34 years, while in the 1951 pyramid the indentation occurs at ages 30-34 years. This example was deliberately chosen to illustrate the ageing of cohorts between censuses: the persons aged 30-34 years in 1951 are, in fact, the survivors of

those aged 25-29 years in 1946, the cohort having aged five years in the intercensal period, 1946-51. It is somewhat unusual to find an irregularity such as this in a population which has not suffered from any large-scale or prolonged war, but there is probably no surer indication of the accuracy with which ages are reported in a census than to find such a pattern repeated from one census to the next, with the depleted cohort, in this case the 1917-21 cohort, advancing progressively in age as it should do. It is anticipating later sections of this chapter to suggest that the influenza epidemic of 1918 was responsible for the smallness of this cohort but, whatever its source, any such depletion of a cohort in a closed population is irremediable, and as the cohort moves through its life cycle, its smallness will be mirrored in all of the demographic changes which occur in the population.

5. The Composition of a Population with respect to Sex

To date, little has been said of the composition of a population with respect to sex, partly because one assumes roughly equivalent numbers of males and females at any age, partly because the sex composition is of secondary importance in the population structure. It is only when the numbers of males and females at each age are not equal, so that there is some distortion of the sex ratios, that this assumption need be modified. The reason why near equivalence is assumed in the numbers of males and females is because of the way in which sex is determined in humans. The sex of any child born depends on which of a particular pair of chromosomes is carried by the spermatozoon which fertilizes the egg released by the female parent. Genetically, females are designated XX, indicating that all female cells carry a pair of X chromosomes. When cell division occurs, each of the 24 chromosome pairs in that cell splits up, and all the eggs (or female gametes) which are produced will contain the X chromosome. Males have an unmatched pair of sex chromosomes, their genetic composition being one X and one Y chromosome. When cell division occurs in males, therefore, one-half of the gametes will carry the X chromosome and one-half the Y chromosome. At the time of fertilization, the chromosomes of the female gamete pair with the chromosomes of the male gamete and, theoretically at least, each resultant zygote has an equal chance of being male, with an XY chromosome pair, or female, with an XX chromosome pair.

The sex of each child born is determined independently, and if only a small number of births are involved, these may result in a comparatively large difference between the numbers of males and females born in any year. It is only when the numbers of births become large that the number of males

born during some specific period of time will equal, or nearly equal, the number of females born. The probability of a birth resulting in a male infant is the same as the probability of a tossed coin falling heads: in each case there is only one alternative in outcome, for if an infant is not a male it must be a female; if the coin does not come down heads, it must come down tails. If you toss a penny ten times, you are unlikely to find heads and tails alternately in successive tosses: you are more likely to find a sequence such as H T H H H T H H T H, in which the result of each toss is denoted by H or T. The frequency of heads in this sample of ten tosses is $7/10$ or 0.70 , whereas the frequency you would expect to find is $\frac{1}{2}$ or 0.50 . If the penny is tossed again and the next ten tosses produce a sequence such as T T H H H H T H T T, the frequency of heads in the sample of twenty tosses becomes $12/20 = 0.60$. Had you stopped after the fifteenth toss, the frequency of heads in the sample of 15 would have been $10/15 = 0.67$. As the sample size increased from 10 to 15 to 20, the frequency with which heads turned up decreased from 0.70 to 0.67 to 0.60 . Increasing the size of the sample still further to 25 and then 30, with a sequence such as T H T T H T H H T T, the frequency of heads becomes $14/25 = 0.56$ and $16/30 = 0.53$.

If you have enough patience to continue tossing, you will eventually arrive at the point where heads have come up no more frequently than tails, and if you continue beyond this, the frequency of heads will fluctuate up and down around the mean value of 0.50 . Because the result of each toss contributes relatively less to the total frequency of heads in a large sample of tosses than it does in a small sample, the range of fluctuation around the mean frequency of heads decreases as the size of the sample increases. In the sequences given above, the difference between the observed frequency of heads and the expected or mean frequency decreased from 0.20 to 0.03 as the number of tosses (that is, sample size) increased from 10 to 30. These sequences were not deliberately concocted to demonstrate this decrease towards the mean value, and the sceptics would be well advised to conduct their own experiment in coin tossing.

For an event which has a known probability or frequency of occurrence, it is possible to work out the probable range of frequencies which can be expected in a series of samples of various sizes. In coin tossing samples of size 10, for example, one would expect a sequence of 10 heads and 0 tails or 0 heads and 10 tails each to occur only once in 1,000 samples; sequences containing either 9 heads and 1 tails or 1 heads and 9 tails would be expected to occur not more than once in 100 samples. If any such sequences occurred more often than this, you could rightly suspect that the coin was biased. Any combinations

other than these—8 heads/2 tails, 7 heads/3 tails, etc. etc. through to 2 heads/8 tails—are within the normal range of variation to be expected with samples of size 10, when there is an equal chance in each toss for the coin to fall either heads or tails. For samples of size 20, sequences containing either more than 16 heads or more than 16 tails would occur less than once in 100 samples, and those containing either more than 14 heads or more than 14 tails would occur 6 times in 100 samples. All other varieties of sequences, 14 heads/6 tails through 10 heads/10 tails to 6 heads/14 tails, would be compatible with an expected frequency of heads of 0.50 in samples of size 20. It will be noted that, although the sample size has doubled, the upper limit to the frequency of heads which would be likely with an unbiased coin has decreased.

Statisticians have established a convention of significance at 5% and 1% levels of probability, and this means no more than the adoption of certain criteria which are not likely to be exceeded, *by chance*, more than 5 times in 100 (5% level) or more than once in 100 times (1% level). If our observations deviate from their expected value more frequently than this, we should re-examine the derivation of this expected value.

Reverting to the topic of male births, it was assumed that these occurred no more frequently in an infinite population of births than did female births. We have, therefore, an expected frequency of 0.50 male births amongst all births, and, theoretically, we would expect 100 male births and 100 female births in a total of 200 births. Since the sex of each birth is determined independently of what has gone before in the sequence there will be a certain amount of random variation, and with a sample size of 200, the deviation permissible by chance, at the 1% level of significance is 18.2. In other words, the number of male births in a sample of 200 births might be any number between 82 and 118 without upsetting the hypothesis that male births occur neither more frequently nor less frequently than do female births. Only if the number of male births was less than 82 or more than 118 would there be any doubt as to whether the frequency of male births in all births was, in fact, 0.50.

As an appendix to this volume, there is a table showing the extent of variation acceptable on the hypothesis of equal frequencies of two alternatives in samples of varying sizes. The upper and lower limits in samples of size N are shown as n_1 and n_2 respectively, and their values are given for the 5% and 1% levels of statistical significance. Provided a sample of N births contains no more than n_1 male births or no less than n_2 male births, this sample provides no evidence that male births occur either more or less frequently than female births. Many populations do show some excess of males amongst all births,

whether the excess attains levels of statistical significance or not, and for many years this was held to be a feature of declining populations. Subsequently it has been shown beyond all reasonable doubt that populations such as the Samoan and Fijian are increasing rapidly despite a continued, and statistically significant, excess of males amongst all children born.

6. *The Structure of a Population*

Knowing the numbers of males and females at each age, or in each five-year age group, in a population there are various ways of examining the structure and assessing the population potential. Probably the most convenient summary of the age structure is given by the proportion of the total population which is contributed by the males and females in each age group. If the total population is reduced to 1,000 persons, it is a simple matter to reduce the numbers of persons of each sex in each age group proportionately to this. Sometimes it may be preferable to examine the male and female populations separately, and to compute the proportion of the male total which is contributed by males in each age group, and the proportion of the female total which is contributed by the females at each age.

A comparative analysis based on such proportions may, however, need cautious interpretation, for if one age group or one cohort is numerically depleted, as was the 1917-21 cohort in French Oceania, each other cohort must contribute relatively more to the total population than it would do had there been no such irregularity in cohort size. This is demonstrated in Table 1, in which the actual age structure of the indigenous population of French Oceania in 1951 is shown side by side with its probable structure had there been no influenza epidemic in 1918 to diminish the size of the 1917-21 cohort. The contribution made to the total population by each age group is expressed as the proportion of each 1,000 persons in the total. An increment of 565 persons at ages 30-34 years (and therefore in the total population as well) increases the contribution of this cohort to each 1,000 persons by 10, and reduces the proportion per 1,000 population contained in each of the other age groups up to 45 years by 1. This change is comparatively small, but had age been recorded in a census of this population at the close of 1921, when the 1917-21 cohort was at its maximum size, the effect of its diminished size on the relative contributions of the other cohorts would have been much greater. In general, when assessing the structure of a population from the proportions of the total numbers contained in each age group, one must take care to isolate the most important feature: the sum of the proportions at each age must add to 1,000, and if there are few persons at ages beyond 60 years, the proportions

at ages less than 15 years, for example, must be higher than they would be in a population with many people aged 60 years or over.

Age Group	No. of persons enumerated	Proportion of total in each age group	
		Census Population	With no Epidemic
0 — 4	8,055	153	152
5 — 9	7,163	136	135
10 — 14	6,651	127	125
15 — 19	5,804	110	109
20 — 24	4,495	86	85
25 — 29	4,006	76	75
30 — 34	3,145	60	70
35 — 39	3,409	65	64
40 — 44	2,732	52	51
45 — 49	2,012	38	38
50 — 54	1,382	26	26
55 — 59	1,031	20	19
60 — 64	758	14	14
65 & over	951	18	18
Not stated	989	19	19
Total	52,583	1,000	1,000

Table 1. The actual age structure of the indigenous population of French Oceania in 1951 classified according to age and proportion of population per 1,000 in each age group, shown side by side with its probable structure had there been no influenza epidemic in 1918 to diminish the size of the 1917-21 cohort.

7. Replacement and Dependency Ratios

For comparative purposes, it is better to employ some measure which is relatively independent of biases of this kind, a measure such as the ratio between the numbers in two related sections of the population. The ratio between the number of children aged 0-4 years and the women who were responsible for their birth, that is women aged 15-44 years, or the ratio between the number of children aged less than 15 years and the number of males on whom such children are economically dependent, are two such likely indices. The number of children aged 0-4 years per 1,000 women aged 15-44 years is sometimes known as the replacement ratio; and the number of children aged 0-14 years per 1,000 males aged 20-64 years as the dependency ratio.

The indigenous population of French Oceania in 1951, for example, contained 8,055 children aged 0-4 years and 11,506 women aged 15-44 years, and this is equivalent to 691 children aged 0-4 years per 1,000 women aged 15-44 years. Children aged less than 15 years numbered 21,869 and as there

were 11,494 males aged 20-64 years, this leads to a dependency ratio of 1,903 per 1,000 adult males. Similar ratios can be constructed for the two major components of Fiji's population, whose composition with respect to age and sex in 1956 is given in Table 2. The replacement ratio for the Fijian component in 1956 was 771, and for the Indian component its value was 1,024; the dependency ratios were respectively 1,912 for the Fijian component and 2,633 for the Indian. The Indian population of Fiji in 1956, therefore, was unquestionably 'younger', with relatively more young people in its total, than either the Fijian component population or the indigenous population of French Oceania in 1951.

Age Group	Fijian		Indian	
	Males	Females	Males	Females
0 — 4	12,593	12,137	16,723	16,574
5 — 9	10,460	9,854	14,329	14,126
10 — 14	8,908	8,227	11,983	11,933
15 — 19	7,798	7,726	9,263	9,215
20 — 24	6,502	6,784	7,340	6,891
25 — 29	5,332	5,552	6,122	5,491
30 — 34	4,450	4,465	4,673	4,424
35 — 39	4,226	4,144	4,211	3,804
40 — 44	3,786	3,419	3,160	2,692
45 — 49	2,966	2,655	2,340	1,742
50 — 54	2,123	2,131	1,532	1,215
55 — 59	1,755	1,785	1,260	781
60 — 64	1,388	1,455	1,904	942
65 — 69	1,057	1,051	1,343	461
70 & over	1,496	1,602	2,128	722
Not stated	149	158	48	31
All ages	74,989	73,145	88,359	81,044

Table 2. An abridged classification showing the composition of the two major components of Fiji's population with respect to age and sex in 1956.

8. Median Age

A more concise summary than any of these of the age structure of a population is provided by the median age, this being the age which divides the population into two halves, one half being younger than the median age, the other half being older. If the population is classified by single years of age, the estimation of the median age is simple. Excluding the population for whom age was not stated, the population of known age is divided by 2: starting

with age 0 years (that is, less than 1 year), the numbers at each age are summed consecutively until the point is reached where the addition of persons at the next highest age exceeds half the total population. In the Fijian component population in 1956, the total number of males for whom age was specified was 74,840, and half of this number is 37,420. The number of males aged less than 18 years was 36,830 and the number aged less than 19 years was 38,354. This latter number exceeds half the population total, and the median age for male Fijians therefore lies somewhere between 18 and 19 years. Subtracting 36,830 from 37,420 and dividing this difference by the number enumerated at age 18 years gives the fraction of a year which is needed to attain the median age, in this case $590/1,524 = 0.4$ years. This assumes that persons in each year of age are evenly distributed over the year, and the median age for Fijian males is therefore 18.4 years.

If the population enumerated in a census has been classified, not by single years of age but by grouped ages, it is necessary to make either a similar assumption about the uniform distribution of population throughout the whole of each five-year age group (which is almost certainly not true), or to interpolate the number of persons at each year of age in the critical five-year age group which contains the median age. The first method will lead to an estimate of the median age which is almost certainly too high, and the second to an estimate whose accuracy is difficult to evaluate. Whichever one uses, it is deceptively easy to be mistaken in a median age, and to avoid adding the fraction of a year to the wrong base age, it is essential to think in terms, not of stated and tabulated ages, but of the upper limit to each year of age tabulated. One of the advantages of the median age is that inaccurate specification of age matters relatively little except at the ages around the median, because the value of the median age will be altered only by the transfer of persons from one side of the median to the other. The transfer of several hundreds of Fijian males from 15 to 16 years of age makes no difference to the estimate of the median age, but the transfer of the same number from age 19 years to age 17 years would lower it slightly.

9. Factors Operating to Change Population Structure

Since any change in the size of a population, either increase or decrease in numbers, inevitably entails some change in the structure as well, the factors which will change population numbers—births, deaths and migrations—will also be responsible for changes in population structure. The maintenance of a given structure demands exact replacement each year of the numbers of males and females at each year of age: those passing out of one year of age

must be replaced by their number at the beginning of the year, and their number at the end of the year should equal the number which started off in the next highest age the previous year. This imposes conditions about births and deaths which are not likely to be met with in any living population; in a population not subjected to migration, it implies that the births and deaths which occur each year should be numerically equal and be spaced evenly throughout each year, that the number of births in each year has been the same over a long period, and that the number of deaths at each year of age each year has been likewise invariable over a long period of time. In a population which experiences migration, the same sort of balance must be achieved with respect to the age and sex of the incoming population vis-à-vis the outgoing, be it by natural causes or by migration. It is patently impossible to achieve such a delicate balance with populations of human beings whose conditions of life and death are changing continually, and it therefore follows that the structure of the populations which they comprise must also be changing.

The two natural sources of change, gains or accessions through births and losses by deaths, are generated by the population itself. The amount of change possible, and the direction of this change as well as its extent, are determined by the fundamental characteristics of the population. For their occurrence, events such as births and deaths require the presence in the population of certain people who can be regarded as 'susceptible' or 'at risk'. If such people are not present, then the events cannot occur; if only a certain number are present, their number will coincide with the maximum number of events which can occur. In general, a large population will contain more 'susceptibles' than will a small population, and although a particular event may occur equally frequently in the two populations, the number of occurrences will be greater in the large population than in the small because of the greater number of persons at risk there. This is of course assuming that both populations are under observation for the same period of time, because if one population is observed for only half the time for which another population, identical with the first, is under observation, the number of events likely to occur in each will be directly proportional to the length of the period of observation as well as to the size of the population at risk.

Only part of the population is at risk for births, and they are the females of reproductive age. Taking into account the length of the human gestation period, the maximum number of births likely to occur in a given population in a year is not likely to exceed the number of females of reproductive age, even allowing for the occurrence of some multiple births. More generally,

the amount of increase possible from this source alone will be strictly proportional to the number of females who are capable of producing children, and any event which reduces the reproductive capacity of a population will reduce the number of births likely to occur within that population. It is almost redundant to add that these gains to the population numbers all enter the population at the age of 0 years.

Losses from a population through death will also be proportional to the size of the population at risk. While death can occur at any age, and the whole of the population is therefore at risk, there are certain ages when the likelihood, or risk, of death is greater than at others. The usual pattern of mortality risk with age was mentioned briefly earlier—the high risk associated with the first year of life, which decreases progressively with each succeeding year of age until adolescence, when it begins to rise again, gradually at first and then more steeply as age advances. Because of this, one would expect more deaths to occur in a population with concentrations of people at the ages which are associated with a high risk of mortality than in a population of the same size whose members were concentrated in the age range which has a low mortality risk.

This pattern of risk is the composite one of deaths from all causes, and there are occasions when one particular cause of death may leap into prominence and for a short time dominate the age incidence of mortality. Under such circumstances, all members of a population may be equally likely to die, or perhaps only persons within a certain age range will be at risk or will be more susceptible than others to this particular cause of death. Such exceptional mortality is usually associated with outbreaks of infectious diseases, some of which were peculiarly lethal amongst the indigenous populations of the Pacific in the late 19th century. Probably through the crews of trading vessels, diseases which were comparatively harmless in other communities were introduced into the Pacific islands, and these new diseases with which the Pacific peoples were unfamiliar tended to produce outbreaks of epidemic proportions, often entailing high mortality. One of the classic examples of such an outbreak was the epidemic of measles in Fiji in 1875, the description of which illustrates not only the immediate effects of such exceptional mortality on the structure of the population, but the consequences of this for the subsequent changes in population numbers. It was followed by a second epidemic of measles in 1903, when the age incidence of mortality was very different from that of the first outbreak; and the third example to be given here, when the pattern of risk was different again, offers the most plausible explanation for the indentation of the population pyramid for French Oceania.

10. *The Effects of Exceptional Mortality on Population Structure*

It is well known that the infective agent of measles is a virus, and that one dose of infection by this virus usually confers life-long protection from, or immunity to, any further attack. Hence, if measles is introduced into a population, the only individuals likely to succumb are those who have not previously been infected. In 1875, the whole population of Fiji was susceptible to measles and, when the infection was introduced, the conditions for its spread, even to the more remote islands of the group, were peculiarly favourable. Thakombau's son, Ratu Timothe, returned on the H.M.S. *Dido* from a visit to Australia. To celebrate his safe return, a great meeting was held at Levuka (then the capital of Fiji) a few days after his arrival, and this was attended by large numbers of Fijians who had come from all parts of the group. There was measles on board the *Dido*, and two weeks after its arrival, two other ships from Sydney called at Levuka, both carrying passengers who had developed measles during the voyage. Timothe and his servant became ill, and when the feasting and dancing were over and the gathering dispersed, many Fijians returned to their villages in a highly infective state. The disease spread rapidly through the village communities and, with so many people sick at the one time, often there was no one able to collect and prepare food. Many people died, either from the sequelae of measles or from starvation, and after four months, the epidemic petered out. Exact mortality figures are known only for some areas, but it was estimated that the total number of deaths in Fiji as a whole was 40,000. However accurate or inaccurate this figure may be, it is reasonable to assume that the epidemic mortality was spread evenly over the whole population so that although the population was reduced in size, its composition with respect to age and sex remained virtually unaltered.

From such descriptions of this epidemic as exist, it can be assumed that, after it had passed, all of the survivors were immune to further attacks of measles. Had the disease been re-introduced within a few years of this first outbreak in 1875, the only persons likely to become infected were those born since that date. In the absence of further doses of infection, the number of persons susceptible to infection by the measles virus would increase with each year that passed. In 1903, when there was once again a sizeable population of susceptibles, consisting of all those born since 1875, a second outbreak was recorded with measles responsible, directly or indirectly, for roughly 2,000 deaths. The brunt of this outbreak must have fallen on persons aged less than 27 or 28 years, and its effect on the subsequent growth of the population would be very different from that of the first outbreak. While both epidemics reduced the reproductive capacity of the population, and therefore reduced the number

outbreaks of a particularly virulent form of influenza were recorded in most parts of the world, and the notable feature of this particular virus was its attack on persons aged 15 to 44 years. These were the persons amongst whom the incidence of influenza was highest and the mortality most severe. It was reported from Fiji that 30% of the Tahitian population died as a result of this epidemic, and although this figure is probably exaggerated, a mortality rate of 30% amongst males and females of reproductive age would disrupt more than one-half of the existing matings. This disruption would effectively reduce the number of births likely to occur the following year by one-half, and even though new marriages or matings would be formed, the reproductive capacity of the population would still be much lower than it had been in the pre-epidemic period. This would be reflected in the number of births in each year of the post-epidemic period for some years. It would probably be at least five years before the number of births each year would begin to increase perceptibly, not merely as a consequence of the new marriages, but also because there would be more younger females moving into the reproductive ages than there were older females moving out.

The years with the fewest births would be those immediately following the epidemic, which would cover more than half of the five-year period of birth for the 1917-21 cohort. The initial size of this cohort would therefore be much smaller than the size of either the 1912-16 cohort or the 1922-26 cohort, for towards the end of this latter period there would have been some recovery in the number of births each year. Were the numbers of births which occurred during this period in Tahiti readily available, this sequence of events could be checked directly. In the absence of such evidence, however, the effects of the influenza epidemic provides the most likely explanation for the indentations previously noted in the population pyramids for French Oceania in 1946 and 1951. The Australian population pyramid shows an indentation similar to this because of the small number of births which occurred in Australia during the economic depression of the early 1930's.

Although population statistics for French Oceania before 1946 are rather inadequate, one can be reasonably certain that in or about the year 1935 the number of births occurring each year in that population decreased, and remained at a comparatively low figure for at least five years. These were the years when the females born in the immediate post-epidemic period attained reproductive age, and their entry into this section of the population would be marked by an almost immediate decline in the number of births. Thus the influenza epidemic of 1918-19 initiated a whole cycle of changes in that population. The initial impact resulted in a large number of deaths which

diminished the size of the population and distorted its structure so that the subsequent gains through births were reduced for some years. This distorted the population structure still further, and the consequences of this would be noticed only some sixteen or seventeen years later when the depleted cohorts themselves reached the ages when reproduction was likely to occur. At about the same time the number of deaths each year would decrease a little as the survivors of the cohorts affected by the epidemic attained the ages where the risk of dying was greatest. Such a cycle is less likely to be noticed now than it would have been a century ago because, during the past forty years, there has been a progressive lessening of mortality in many territories of the Pacific, and this would tend to mask these recurrent effects of the exceptional mortality caused initially by epidemic outbreaks of disease. There can be little doubt, however, that the numerical course of many island populations during the late 19th and early 20th centuries was determined by such outbreaks and their aftermath.

11. The Effect of Migration on Population Structure

Except under extreme conditions such as epidemic mortality, natural sources of change do not cause radical changes in population structure. Where levels of mortality and fertility are relatively stable over a period of years, structural changes in the population will be gradual. Migration, on the other hand, can bring about radical and rapid changes in population structure. The outstanding example of this in the Pacific is probably the introduction into Fiji of relatively large numbers of indentured labourers from India when the European settlers were building up first the coconut plantations and later the sugar industry. The ages of these labourers are not known, but it can be assumed that the majority would have been aged between 15 and 29 years, and with only one female to every four or five males, their impact on the structure of Fiji's population, both with respect to age and sex is obvious.

The sudden imposition of, say, 5,000 males and 1,000 females aged 15-29 years on the Fijian population of roughly 90,000 persons would alter the sex ratio of the population as a whole, and grossly distort the sex ratio of the population aged 15-29 years. With their arrival, the reproductive capacity of the population would have increased, and although the effect of this on births would be less than it would have been had equal numbers of males and females been brought in, the number of births each year would rise. The number of deaths in the population would be changed very little, because the mortality risk at ages 15-29 years is relatively small. Whatever had been happening in the population of Fiji at that time, the chances of favourable natural increase

would have been enhanced by the introduction of these 6,000 persons and such enhancement of the natural increase could be expected to continue for some years, until the introduced population passed out of the reproductive period into the ages where the risk of mortality is high.

Had these hypothetical immigrant people been thirty years older, their introduction into the population would have brought about almost no change in the number of births likely to occur each year, but would have increased the number of deaths quite considerably. With each year that passed, there would be fewer of the initial 6,000 surviving, and with no replacements from births, the immigrant group would gradually fade out of the population, and the distortion of the sex ratio and age composition with them. This parallels the demographic effects of emigration, where the emigrants are usually preponderantly young adult males. Their departure depletes the reproductive capacity of the population and tends to reduce the number of births without affecting the number of deaths, and thereby depresses the amount of natural increase of the population, the size of which has already been reduced by the loss through emigration.

12. The Interaction of Sources of Change and the Interdependence of these with Population Structure

These last examples illustrate, perhaps better than the earlier ones of epidemic mortality, the interaction which there must be between the various sources of change in population numbers and structure. Seldom will changes in the mortality pattern not be associated, either immediately or subsequently, with some change in the number of births, and the changes due to migration will usually be reflected in either or both the gains to the population from births and the losses through death. Any or all of these sources of change in population numbers may generate still further changes, because all operate to change the structure of the population in some way, and changes in structure, of themselves, will entail still further changes, either immediately or at some later date. The composition of a population at any particular time is the product of the mortality, fertility and migration experienced in the past, and it is this same composition which provides the potential for future changes. This inevitability about the course of change in populations is perhaps not generally appreciated. While the structure of a population at some given time provides some of the clues to its past history, at the same time it reveals much of what is likely to happen in the future.

In the absence of migration, the only gains to a population come from the births which occur in that population, and the only losses are through death.

Both events are dependent for their occurrence on the existence of persons of suitable age and sex in the population, and the more persons there are at such ages, the greater is the likelihood of the occurrence of these events. It requires no complex technique of projection to see, for example, from the numbers of females in each age group in the two major components of Fiji's population, that five years after the census date the number of births each year is likely to be larger than it was around 1956. An abridged classification of these populations with respect to age and sex was given in Table 2, and from this it can be shown that there were 32,090 Fijian females and 32,517 Indian females between the ages of 15 and 44 years in 1956. By 1961, the cohorts aged 10-14 years in 1956 will have become 15-19 years of age, and those aged 40-44 years in 1956 will have become 45-49 years. The persons in the reproductive ages in 1961 will therefore be the survivors of the females aged 10-39 years in 1956, and since far more women will have entered the age range 15-44 years than have left it, the reproductive capacity of both populations will have increased by 1961, the Fijian by roughly $(8,227 - 3,419) \times 100/32,090 = 15\%$, and the Indian by $(11,933 - 2,692) \times 100/32,517 = 28\%$. Under normal circumstances, one would expect roughly parallel increases in the numbers of births to each population each year.

Inferences concerning the number of deaths which are likely to occur in future years require more caution, but in the case of the Indian component in Fiji, one could be practically certain that there will be more males dying within five years of the census than there will be females dying, simply because there are more males than females at the ages where the risk of death is greatest. Because of the numbers of males at ages 60 years and over in 1956, one might go even further and predict some increase in the number of deaths each year in this component: should this happen, it will tend to retard the rate of increase in population numbers by offsetting some of the expected increment from the additional births. Further illustrations of the way in which the age structure of a population can be exploited by such simple, common-sense reasoning will be given later. For the present it will be sufficient if an impersonal series of numbers, such as those of either table in this chapter, begins to emerge as an animate population whose behaviour, both present and future, is regulated by the behaviour of each of its elements.

II

THE STATISTICS OF POPULATIONS AND MEASURES OF CHANGE

1. Sources of Population Statistics

The statistics needed for the measurement of population are of two kinds. There are those derived from the periodic censuses of a population, which depict the state of the population at a given moment of time, and there are the continuous records of the changes which are occurring in that population. These two varieties of population statistics are complementary, for the census data provide a base-line against which the current changes can be assessed, and the state of the population at times other than at a census is given by the grafting of the continuous records to the census population. The most effective combination of these two kinds of records is what is known as a 'population register', a system which might be adopted to advantage in some Pacific territories. Basically, this is a card-index of the population in which each individual is represented by a card on which all of the relevant demographic data are recorded. When a child is born, a new card, suitably cross-referenced to the mother and father, is inserted in the register: when a person dies, his or her card is removed from the register. Persons migrating into or away from the population are treated in the same way as these arrivals and departures from natural causes. Conscientiously maintained, such a register will provide knowledge of the state of the population at times other than at a census, and it is only by having such registers that one can achieve absolute consistency between the continuous records for any one year and the population as constituted at the beginning of that year. The accuracy of the register must be checked from time to time, and this requires either complete or sample censuses at regular intervals. The system is open to abuse but, under impartial direction and strictly limited access, misuse of the register should be no more likely than is the improper use of information divulged in a census.

2. Census Enumerations

In its simplest form, a census is a count or an enumeration of the population without regard to the differences which exist between individual members of that population. In its most elaborate form, it can be a complete inventory

of the population in respect of a large number of the characteristics possessed by its members. Most censuses nowadays fall somewhere in between these two extremes, with practicality as the key to the amount of differentiation between individuals which should be attempted. It is essential to achieve some compromise between the amount of information one would like to have and the amount one is likely to obtain reliably under the given conditions of the enumeration. There are certain fundamental details about the population which must be collected, and the elaboration beyond these will be determined by the capacity of the population themselves, the method of enumeration and various other factors which will be detailed in the next chapter. In general, so long as the census provides an adequate description of the population and of such of its activities as may be relevant at the time of the enquiry, nothing is gained by increasing the number of details known about each individual. In full-scale enumerations it is advisable to limit the scope of the enquiry to ensure accurate replies to relatively few basic questions, in preference to answers of less reliability to many, and if necessary, follow these with more detailed enquiries on selected samples from the population.

In many Pacific territories, the practice of census-taking was inaugurated in the initial stages of European administration. The scope of these early enumerations was necessarily limited to little more than a count of heads. Male and female heads were almost always distinguished, but since the calendar year is a peculiarly European measure of time, it is only in recent years that any attempt has been made to classify the populations according to chronological age. Important though this, or some approach to it, is for the analysis of population structure and change, the indigenous populations of the Pacific have shown an understandable reluctance to adopt the concept of chronological age. Even now, relatively few are conscious of age as the number of years which have elapsed since their birth; if asked their age, many will reply with a number, but often the answer is quite meaningless. The nearest approach the early census-takers could make to age was the differentiation of children from adults. Greater precision was attempted later, when populations such as the Samoan and Tongan, for example, were recorded according to some social or sociological status group, each category being roughly associated with some range of age. Gradually the classification of populations according to chronological age has been adopted, but it is only after prolonged contact with Europeans and their insistence on chronological age that this concept will be accepted meaningfully, and until then, the ages of people recorded in Pacific censuses can only be estimates of their real ages.

Other demographic statistics which it is desirable to obtain from a census

concern the reproductive habits of the population, the average size of completed families, and the pattern of family building. Since the number of children born to a female is partly dependent on her age at the time of asking, and her age when the first child was born, there is no point in obtaining this information without having some reasonably good estimate of her age. A woman of 30 years for example is more likely to have borne 3 children than is a woman of 18 years. Similarly, a woman of 30 years who had her first child at the age of 18 years is more likely to have 3 children than is a woman of the same age who had her first child at the age of 27 years. The likelihood of a woman having any child at all is sometimes related to whether she is or has ever been married, and consequently a question concerning marital status is usually included in a census. This is verging on the sociological questions which, with those related to the strictly economic aspects of the population, usually complete the remainder of the census questionnaire.

If reliable censuses are taken at regular intervals, with the same information recorded in each, much can be deduced about the population concerned from comparisons between the data from successive censuses. Provided that age is specified accurately, each cohort of persons can be followed through from one census to the next. The exception to this, of course, is the cohort or cohorts containing those persons born in the interval between any two censuses: once they have been recorded in a census, however, they too can be traced through succeeding censuses and their progress noted. If there is no migration either into or away from the population, the number of persons in any cohort can only decrease with the passing of time, and the difference in its size from one census to the next is a measure of the mortality experienced by that cohort in the intercensal period. Should a cohort increase in size between two successive censuses, then it is clear that the population is not 'closed', for immigration is the only possible source of increase in cohort size between censuses which are accurate both in their accounting and in the recording of ages. Emigration, on the other hand, is a supplementary source of the depletion of cohorts, and the decrease in size of any cohort in a population subject to migration is, therefore, the compounded loss through mortality and emigration.

An example may illustrate these points more clearly. Persons aged 10-14 years in 1956 were the survivors of those aged 5-9 years five years previously, who in turn were the survivors of the cohort aged 0-4 years in 1946 and of the births which occurred in the period 1942-46. Once the year 1946 had passed, the size of this cohort could not be increased by births, although it could be added to at any stage by persons of appropriate ages coming into the population from outside. There might be some compensating loss of people through

emigration, and over any period of time all members of this cohort are continually exposed to the risk of mortality. Hence, the change in its size between 1946 and 1956, say, will represent the net effect of the three factors, gains from immigration, losses from emigration, and losses through deaths. The contribution each of these makes to the total change cannot be deduced from census data. If one wishes to separate them out, the data needed are those which relate directly to each source of change. These are the complementary, continuous records of births, deaths and migrations occurring each year.

3. Continuous Records

Registration of births and deaths is generally required by law and such records are usually available. Although it has not yet proved practicable to institute the required legislation in some Pacific territories, registration systems are in operation in French Oceania, Cook Islands, Samoa, Tonga and Fiji. Commonly, local registrars of births, deaths and marriages are appointed, and these are responsible for the registration of all births and deaths which may occur throughout the area which is under their jurisdiction. The responsibility for reporting the occurrence of any such event to the local registrar lies in the first instance with the close relatives or persons who may have been present at either the birth or the death. The period of grace allowed for the reporting of the event varies from one to two months but once the event is registered, the completed forms of registration are forwarded to some central authority, where they are stored for reference. Usually a count is made of the number of births and the number of deaths registered in each calendar year, and in some territories more detailed statistical summaries are compiled, showing the age and sex of persons who have died, the sex of children born and perhaps the age of the mother at the birth of the child. The most complete statistical summaries published are those for Fiji, where the births are classified simultaneously according to the age of the mother at the birth, and the number of children she has already borne; and deaths are tabulated by sex and single years of age up to five years and thereafter in five-yearly age groups.

The institution of such a system of registration depends primarily on the availability of persons to act as local registrars, and on their accessibility. The efficiency of its operation depends on the policing of a system which offers no apparent gain to the people concerned. Without some inducement few people are going to travel several miles to register a birth or a death, especially when they cannot be sure that the local registrar will be at home. Should he

be absent the first time they wish to register a birth, their attitude towards the registration of the next child born will be less favourable. More likely than not, the registration will be delayed until there is some additional reason for visiting the town or the village where the local registrar lives, or it may be altogether forgotten. By demanding the birth certificates of children starting school or wishing to enter for certain school examinations, the Education Department of Fiji has provided some inducement to parents to register the births of their children, even though the registration may not be performed until six years after the event, when the child is about to start school. Incentives towards the registration of deaths are unnecessary if permission must be sought before a body may be interred, but any such system of burial permits is open to the same objections as the whole registration system and there would be many difficulties in policing it.

Whatever the difficulties of implementing an efficient system of registration, it is much more difficult to maintain satisfactory records of migration into or away from a territory. For island territories, people can come or go either by sea or by air, but provided they arrive or depart from one or two main ports, by some scheduled vessel or aircraft, reasonable records can be kept. Again one is dependent on the co-operation of the people concerned, and of the shipping and airways companies; if, in the flurry of departure, some records are incomplete, they may well remain so for few junior officials are likely to oppose an impatient captain anxious to be off however much they may delay irate passengers impatient to disembark. Not all movement to and from an island territory is restricted to main ports of call; small vessels may use any number of smaller ports and harbours from which arrivals and departures may be undetected. Although such small vessels may carry few passengers, multiplication over many small vessels several times a year may give rise to quite large discrepancies in the net migration figures.

4. Assessment of Accuracy

The accuracy with which any of these records are kept can be judged only from census data, and since censuses are periodic affairs, usually entailing an all-out effort, there is a much greater chance of having them accurate than there is of maintaining continuous records accurately. The simplest and most elementary check uses the total population enumerated in two successive censuses. Assuming that each represents an accurate count of the population present in the territory at each date, any change which has occurred over the intercensal period should be numerically equal to the sum of the natural increase and the net migration, the natural increase being the difference

between the numbers of births and the numbers of deaths, and the net migration being the difference between the numbers of persons coming into the territory and the numbers leaving. More specific tests can be made for certain groups within the population, checking the census records against the numbers expected if the records of births, deaths and migrations for the intercensal interval are complete and accurate.

This presupposes that the sex and age of deceased persons and migrants are recorded in sufficient detail to permit one to make such comparisons, and in many territories at present they are not. While this may be unavoidable for the time being, consistency between the various forms of population records is desirable and because the periodic censuses and the continuous records are complementary, exactly the same details should be recorded in each if the most is to be obtained from any of them. Whichever way one looks at it, the loss of 20 adult males aged 25-29 years, say, from an island community of 200 people would have very different repercussions on the life of that community than would the loss of 20 males aged 70-74 years. Moreover, such a loss would not be immediately offset by the birth of 20 male infants, nor by the immigration of 20 boys aged 10-14 years, although numerically both the total population and the number of males would be unchanged.

5. Census v. Registration Data

The first of the specific checks is directed at the completeness of the birth registrations. If the registration system functions efficiently, the numbers of children born in any intercensal interval should equal the number of births registered during that time; and the numbers of those who survive to be recorded in a census should be the difference between the number of births registered and the number of deaths, relating to those children, which were registered in the same interval. One would expect, therefore, to find reasonably good agreement between the number of children aged less than 1 year in a census and the number which would result from subtracting a proportion of the deaths amongst children aged less than 1 year during the year preceding the census from the number of births registered during that same year. One would not subtract all infant deaths because some of the deaths registered in the twelve months preceding a census would be the deaths of children born more than twelve months before census date. Roughly one-half of all infant deaths occur in the first month of life, and about 70% (or even 80% in some communities) before the age of 6 months, and since this is the average age of the population aged less than 1 year, the number of deaths at ages less than 6 months would approximate to the deaths amongst the year's births. This

balancing of the numbers enumerated at each year of age with the numbers expected from the births registered in each of the relevant years, adjusted for any deaths which may have occurred in each year of life, can be extended, quite simply, to higher ages. If a census was taken at the end of 1956, for example, the numbers expected at age 1 year in that population would be given by subtracting from the number of births registered in 1955 the deaths at ages less than 1 year in 1955 and half of those at age 1 year in 1956: the numbers expected at age 2 years would be the number of births registered in 1954 minus the sum of the deaths at age 0 in 1954, the deaths at age 1 in 1955 and half of the deaths at age 2 in 1956.

Table 3 reproduces some of the data relating to the Fijian population at the Census of 1956, showing the comparison between the population enumerated

Age in Years	Numbers enumerated (1)	Numbers expected (2)	Difference (1) — (2)
0	5,798	5,268	+ 530
1	4,506	4,504	+ 2
2	4,894	4,667	+ 227
3	4,900	5,043	— 143
4	4,632	4,087	+ 545
0 — 4	24,730	23,569	+ 1,161
5	4,265	3,947	+ 318
6	4,235	4,229	+ 6
7	3,793	3,909	— 116
8	4,139	3,887	+ 252
9	3,882	3,991	— 109
5 — 9	20,314	19,963	+ 351
0 — 9	45,044	43,532	+ 1,512

Table 3. The number of Fijians enumerated in the Census of Fiji, 1956, at each year of age from 0 to 9 years and the numbers expected from the births and deaths registered between 1 October 1946 and 30 September 1956.

at each year of age up to 9 years and the estimates derived from the registration data of the ten years preceding the census. There are discrepancies of several hundreds between the numbers enumerated and those expected at some ages, and although the initial difference of 530 at age 0 might conceivably be due to the failure to register immediately some of the births which occurred in the two months preceding census date, delay in registration cannot explain the differences observable at ages 4 and 5 years, for example. The inconsistencies

at other ages (2 and 3, and 7, 8 and 9 years) are smaller in magnitude, and while the negative deviations cancel out some of the excess in the total for each age group, the residual difference is still quite large, particularly for the age group 0-4 years. Since this amounts to roughly $4\frac{1}{2}\%$ of the recorded population, one must, justifiably, suspect the completeness of the birth registration records. In addition one might conclude that, if the ages of Fijian children were specified accurately on the census schedules, the births registered in any one year are not necessarily the same as the births which occurred in that year. While this may be straining faith in the census too far, both this and the fact that not all births are registered may prejudice inferences drawn from more detailed analyses of mortality and fertility in the population.

Had the reverse occurred in the Fijian data, with the numbers expected at each age exceeding the numbers enumerated, the situation would then be somewhat equivocal. The completeness of the enumeration would clearly be suspect and, at the same time, any assessment of the accuracy of the birth registration data would be impossible. Even reasonably good agreement between the two sets of figures, the numbers enumerated and the numbers expected at each age, is not wholly indicative of accuracy and reliability, because there might be defects of an equal order of magnitude in both sets of data. The omission of 5% of the population aged 0-4 years from the census records could be balanced by the non-registration of 5% of the births, so that no discrepancies would appear in the comparison between the numbers counted in the population and the numbers expected. The same argument might also be applied to the case in which the numbers enumerated at each age exceed the numbers expected: if the difference between the two numbers is used to estimate the proportion of births which are not registered, this proportion will be the minimum proportion because some of the children aged 0-4 years may have been missed in the enumeration. Had they been included in the census, the difference between the enumerated and expected numbers would have been increased, so that the proportion of unregistered births would be increased accordingly.

6. Cohort Comparisons

Hence, this check between the census and registration data is a two-edged weapon, for while it can be used to test the completeness of the birth registrations, it might also test the coverage of the census enumeration. As a group, the children aged 0 and 1 year are the ones most likely to be forgotten in a census, and it is reassuring for the census taker to find more of them recorded than would be expected from the number of births registered. In populations

which are not subjected to much migration, however, any under-enumeration in one census is likely to be revealed in the following one, by which time the children at ages 0-4 years in the first census will have moved into a higher age group, if they have not died in the meantime. In Table 4 the Indian population of Fiji as recorded in the various censuses since 1921 has been arrayed in birth cohorts, with the persons in each five-year age group in each census designated by their period of birth. This table refers to the total population, without distinction of sex, although it is more usual to make these cohort comparisons for the male and female populations separately.

Birth Cohort	1921	1936	1946	1956	Age in years 1956
1952-56				33,297	0 — 4
1947-51				28,455	5 — 9
1942-46			23,156	23,916	10 — 14
1937-41			20,635	18,478	15 — 19
1932-36		13,936	14,777	14,231	20 — 24
1927-31		13,452	12,349	11,613	25 — 29
1922-26		10,156	9,433	9,097	30 — 34
1917-21	10,291	8,868	8,521	8,015	35 — 39
1912-16	8,723	6,878	6,103	5,852	40 — 44
1907-11	4,393	4,643	4,378	4,082	45 — 49
1902-06	3,170	3,643	3,238	2,747	50 — 54
1897-01	4,318	3,403	2,784	2,041	55 — 59
1892-96	7,517	4,883	3,341	2,846	60 — 64
1887-91	7,809	3,699	3,294	1,804	65 — 69
1882-86	4,708	4,163	3,933	1,376	70 — 74
Before 1882	8,977	7,278	4,142	1,474	75 and over

Table 4. The number of Indians in each birth cohort at the Censuses of 1921, 1936, 1946 and 1956, irrespective of sex.

Although the recruitment of Indian labourers for Fiji ceased in 1916, and all unexpired contracts in Fiji were cancelled in 1920, the Indian population recorded in the 1921 census of Fiji contained some persons who were repatriated before the next census, and it was added to by new arrivals during the intercensal period. The number of Indians who returned to India during the years 1921-36 was approximately the same as the number of Indian immigrants to Fiji in that period, so that the total net migration between 1921 and 1936 was virtually nil. The immigrant group, however, would not be exact replacements of those repatriated, and their impact on the age and sex composition of Indian population in Fiji would not be balanced by the departure of the others. Despite suspected inaccuracies in the statements of age at both censuses, the comparison between the population structure in



1921 and 1936 amply demonstrates this point. The concentration of persons in the age range 25-34 years in 1921 (period of birth 1887-1896) is compatible with the continued presence of workers who had come out to Fiji under indenture; and the depletion in the size of these two cohorts between 1921 and 1936 would be due more to repatriation than to deaths. Some of them were accompanied by their wives and it is probable that some of the children aged less than 10 years in the 1921 census population subsequently returned to India with their parents. This would account for the inordinately large decrease in the size of the 1912-16 and 1917-21 cohorts between the two censuses. Without more detailed data than are available on either mortality or migration at this period, it is impossible to differentiate the relative contributions of these factors to the changes which occurred in the Indian population between 1921 and 1936. Over the next two intercensal periods there was little movement of population between Fiji and India, and cohort comparisons therefore become more meaningful, both as regards to the accuracy with which age is specified and with respect to the likelihood and the extent of under-enumeration.

In 1936 the two youngest cohorts were approximately equal in size. Of itself, this would arouse suspicions of under-enumeration in any population except one recently subjected to the disturbances of migration or of some outbreak of epidemic disease which caused high mortality, especially amongst persons of reproductive age. The Indian population of 1921 had experienced both—an outbreak of influenza in 1918 and continued migration before and after the census—and the combined effects of the two factors could perhaps have produced this equivalence of cohort size. In 1946, however, the census showed an increase of roughly 800 in the size of the 1932-36 cohort, and in the absence of migration for much of this intercensal period (at least from 1939 onwards), it appears extremely likely that not all of the children aged less than 5 years were enumerated in 1936. Confirmation of this is given by the fact that 15,960 Indian births were registered in the five years 1932-36, and if the census figures were correct, only 13,936 or 87% survived to be recorded in the census. This proportion would be still smaller if not all births had been registered.

Between 1946 and 1956 the only cohort which increased in size was again the group aged 0-4 years in the earlier census. The 1946 population distribution, however, shows a curious irregularity: the 1937-41 cohort contains only 2,500 fewer people than the younger and youngest cohort but nearly 6,000 more people than the next oldest cohort of persons born 1932-36. In the absence of migration, this implies a much larger number of births over the

period 1937-41 than had occurred in the previous five years, and a subsequent slackening-off over the next five-year period. From the age composition of the female population at this time, one would not expect this to have happened and, in fact, there were more children aged 5-9 years recorded in the 1946 census than there were births registered in the period 1937-41. Again some of the births which had occurred may not have been registered, but the proportion of unregistered births would need to be very high to make this excess of survivors plausible. Hence, while there may have been under-enumeration of children aged 0-4 years in 1946, it is more likely that some of these children recorded as being 5-9 years of age should have been included in one of the age groups on either side of this one: that some of the 4-year-olds were recorded as being already 5 years, and perhaps some of the children aged 10 years or over were recorded as being younger. There is no reason to suppose that a greater proportion than usual of Indian births were not registered during this period, and the numbers of births which were registered during each of the relevant five-year periods show no such irregularity in sequence.

Except for these youngest cohorts, all other cohorts decrease in size progressively from one census to the next. The rate of decrease is not particularly regular, however, and this indicates some unreliability in the personal statements of age on the census schedules. While it is not proposed to explain the methods here, it is quite a simple matter technically to examine the distribution of population by single years of age, and calculate the preferences shown by individuals for ages ending in each number between 0 and 9. Preferences for ages with certain unit digits, or the avoidance of ages with other unit digits, which are repeated from one census to the next may compromise the cohort comparisons. In the Census of Fiji, 1956, both the Fijian and Indian component populations tended to avoid ages with either 1 or 2 as the unit digit (that is, ages 21, 22, 31, 32, and so on), and this avoidance was more strongly marked than was their preference for any other unit digit. In consequence, the numbers of persons in the first half of each decade of age were rather less than the numbers which should have been included in each of these five-year age groups, and there were rather more people in the alternate age groups, 25-29, 35-39 years and so on, than there should have been. If a census is taken every ten years, and the population shows similar preferences for certain unit digits in each, this weighting of numbers in favour of one or other half of each decade of age does not disturb the cohort comparisons. If the intercensal interval is only five years, however, the survival pattern of cohorts over a period of years may be ragged and saw-toothed instead of being smooth and regular. The cohort aged 30-34 years for example may appear to have a

higher proportion of survivors five years later (at ages 35-39 years) than does the younger cohort aged 25-29 years, which is contrary to expectation because, in general, persons aged 25-29 years have a better chance of surviving to ages 30-34 years than they have of surviving the next five years, to ages 35-39 years.

7. *The Risk of Mortality*

The survival pattern of each cohort is complementary to the mortality risk associated with their particular stage in the life cycle, and the average pattern of risk with age was outlined briefly in Chapter I. Irrespective of the sex of individuals, there is high risk of mortality at and soon after birth which decreases continuously throughout the first year of life and becomes progressively

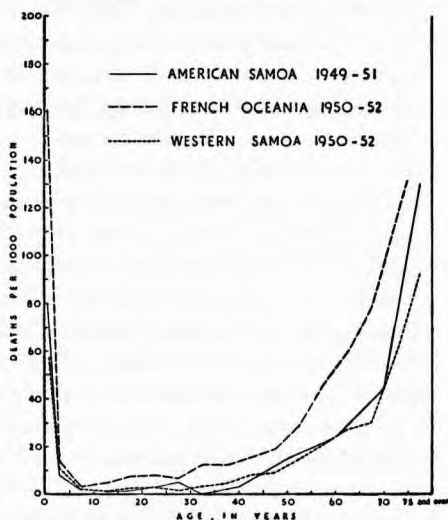


FIG. 2. Patterns of mortality: a graph showing the average annual age-specific mortality rates amongst males in American Samoa 1949-51, French Oceania 1950-52, and Western Samoa 1950-52.

less with each additional year of life until somewhere between the ages of 10 and 14 years. Thereafter the risk of dying increases gradually with each succeeding year of age, slowly at first and then more sharply as age increases. The magnitude of the risk may be different for males and females at specific ages, and it may vary quite considerably from one population to another, but the general pattern of risk with age remains uniformly a U-shaped curve.

It will not be a symmetrical curve, because whatever the magnitude of the risk at ages less than 1 year, represented by the height of the left arm of the U above zero, the right arm of the U must eventually rise to the point where death is certain, and the risk of dying is therefore 1. Some of the possible variations in the shape and span of this curve are illustrated in Figure 2, each of the curves shown having been derived from the mortality experienced by the male populations of one or other of the island groups in the South Pacific.

Since the mortality risk associated with any age is the chance that a person of that age will die within some specified time, the number of deaths likely to occur amongst a group of people of the same age is the sum of their individual chances of dying. This sum is obtained most conveniently by multiplying together the chance that one individual will die and the number of persons who are so exposed to the risk of dying. If the mortality risk at age n years is 1% per year, and 100 persons attain this age on some given day, one would expect only one of those persons to die in the course of the next year. If the mortality risk were 5% per year, one would expect 5 of them to die in the course of the year. If these 100 persons were divided into two groups, with 50 of them exposed to a mortality risk of 1% per year, and the remainder to a risk of 5% per year, the number of deaths likely to occur in the course of one year would be $50 \times 0.01 + 50 \times 0.05 = 3$. If each of these groups were divided again, and each subjected to different mortality risks, the number of deaths expected amongst the original 100 persons in the course of a year would be given by the sum of the number of deaths expected amongst each group. This is equivalent to dividing up a population into groups according to their age at some moment of time, and subjecting each of these groups to a mortality risk specific for their age. The number of deaths likely to occur in that population over the course of a year will be the sum of the number of deaths expected amongst persons at each age in the population.

Given the age distribution of the population, and the pattern of mortality risk with age which is experienced each year by that population, one can compute the number of deaths likely to occur by summing, over all ages, the products of the numbers of persons at each age, and the chance each has of dying within a year. Except at the extreme ends of the age scale, it is not usually necessary to work with populations and risks at each single year of age: for the middle range of ages, the mortality risk varies so little from one age to the next within a five-year age group that it is sufficient to apply a uniform risk to the population within an age group.

An example is set out in Table 5 to demonstrate this computation for a hypothetical population, the members of which are distributed over the range

of ages as shown in column (1). The only single year of age which has been distinguished is 0, that is, less than 1 year, and the risk of mortality run at this age and in each succeeding age group each year is given in column (2). Persons aged 0 have little less than a 1 in 4 chance of dying; those aged 75 years or more are more likely to die within the year than they are to live. The persons with the best chances of survival, those who are least likely to die, are the persons aged 10-14 years.

Age in years	(1) Population at each age	(2) Risk of mortality each year	(3) No. of deaths expected each year
0	2,970	0.236	701
1 — 4	9,620	0.045	433
0 — 4	12,590		1,134
5 — 9	10,705	0.009	96
10 — 14	10,275	0.007	72
15 — 19	9,904	0.008	79
20 — 24	9,430	0.012	113
25 — 29	8,802	0.016	141
30 — 34	8,102	0.018	146
35 — 39	7,350	0.022	162
40 — 44	6,536	0.026	170
45 — 49	5,592	0.037	207
50 — 54	4,464	0.054	241
55 — 59	3,204	0.084	269
60 — 64	1,926	0.133	256
65 — 69	864	0.220	190
70 — 74	233	0.370	86
75 and over	23	0.580	13
Total	100,000		3,375

Table 5. The numbers of deaths expected in a year in a population for which the age composition and risk of mortality at each age are known.

Each of the numbers in column (3) is the product of the number of persons in an age group and the risk of mortality associated with that age; and the number of deaths which would occur amongst the whole population, given these conditions of mortality, is the sum of the deaths expected at each age. A shorthand expression to describe this procedure, and one which will be used frequently in ensuing sections of this chapter, is the *application* of a frequency or a risk (or a series of either) to a population (or the relevant sections of a population). To apply a frequency or a risk to a population means no

more than the multiplication of the frequency or risk of an event with the population at risk: to apply a series of frequencies or risks, specific for certain ages or groups of ages, to a population similarly classified with respect to age means simply that each age-specific frequency or risk is multiplied by the number at the relevant ages in the population.

8. Measures of Mortality: The Crude Death Rate

Now let us suppose that the data of Table 5 related, not to a hypothetical population, but to the population recorded on the imaginary island of Tiara on 1 January 1960. At this date there were 100,000 people on the island, 3,375 of whom will die in the course of the calendar year 1960. Clearly, if the population had been only half this size initially, totalling 50,000 in all, with the numbers at each age diminished proportionately so that the population structure was not altered in any way, there would have been only half this number of deaths in the course of the year. Hence, the number of deaths which occur is not independent of the size of the population, and any measure or index we may wish to construct concerning the frequency of deaths or the level of mortality on Tiara during the year 1960 must contain some reference to the number of people who were living there at the beginning of that year. The simplest measure of the extent of mortality on the island, therefore, is one which relates the total number of deaths to the total population at risk. This is known as the crude death rate, and is usually expressed as the number of deaths per thousand population per year. In this case, Tiara has a crude death rate of 33.75 per thousand persons per year. Had the population been halved, so that there were only 1,687 deaths, or decimated so that only 337 deaths occurred, the value of the crude death rate would be unaltered.

Suppose now that instead of having 12,590 persons aged less than 5 years, and 9,904 persons aged 15-19 years in Tiara's population, these numbers had been reversed so that there were 9,904 persons aged less than 5 years and 12,590 persons aged 15-19 years. With these numbers in the population, the mortality risks associated with these ages would have produced 891 deaths at ages 0-4 years, and 101 deaths at ages 15-19 years. This would decrease the number of deaths to occur amongst the whole population in the course of a year from 3,375 to 3,154; and since the total population was not changed by this reversal of numbers in two age groups, the crude death rate for Tiara would fall from 33.75 per thousand persons to 31.54 per thousand per year. What then is the significance of this decrease of just over 2 per thousand in the crude death rate? Since the population was subjected to exactly the same risks of mortality each time, it does not reflect any change in the pattern of

risk in the population and it cannot, therefore, represent any real fall in the level of mortality on Tiara. It was brought about merely by interchanging the numbers of persons in two age groups and it is therefore merely the reflection of a change in the age distribution of the population. The age composition of a population is likely to vary much more than this from year to year, and if changes in the numbers at risk at each age can produce fluctuations such as this in the crude death rate, it is not a highly efficient measure of what it purports to be measuring, viz. the frequency of deaths in the population.

Should any further demonstration of its inefficiency be required, one can juggle with the age-specific mortality risks in such a way as to produce no change in the value of the crude death rate, while making highly significant changes in the pattern or level of mortality in the population. If, for example, in Tiara the mortality risk for infants aged less than 1 year was halved, the consequent decrease in the number of deaths could be offset by the doubling of the risk at all ages between 5 and 24 years. If these two changes occurred simultaneously in a population similar to Tiara's but for which one knew only the total number of deaths which occurred, changes such as these in the mortality pattern would be wholly masked by the constancy of the crude death rate. Yet halving the infant mortality rate in a population is of much greater significance than the doubling of a few rates which were initially small.

9. Age-Specific Mortality Rates

One of the principal objects in measuring the frequency of some event is to allow of comparisons either between populations of different geographic regions at the one time or between populations of the one region at different epochs in their history. A rate which is as susceptible as the crude death rate to changes and fluctuations which have no bearing on what is being measured is virtually useless, and because it may be gravely misleading, it is one to be avoided at all costs for comparative purposes. What then are the alternatives? The deaths which occur need to be related to the population from which they come, and with such a variable risk of mortality with age, the logical thing to do is to relate the number of deaths at each age to the population at each age. The rates so constructed are estimates of the risk of mortality at each age, and the pattern of risk in a population is usually derived from such estimates based on a census population, classified by age and sex, and the number of deaths which actually occurred, either in a single year or for a short period of years around the census year, similarly classified by sex and age. These data

would be in the form of the data in columns (1) and (3) of Table 5, and it is a simple matter to compute (average) annual mortality rates for males and females at each age (or age group).

The *age-specific mortality rates* which are thus obtained are not synonymous with the mortality risk, because the frequency of deaths amongst a group of people of some particular age is only an estimate of the probability that a person of that age would die in the course of the year or years in question. While such rates may be the best estimates available, statistical probability requires an infinitely large population for its precise estimation, and the reliability of any estimate of that probability decreases as the size of the sample used to compute that estimate diminishes. An illustration of this was given in Chapter I in the discussion of the sex composition of the population: with an expected frequency of 0.50 for a male birth, the observed frequency of male births amongst all births is likely to range from 0.40 to 0.60 in a sample of 100, while in the sample of 1,000 the observed frequency of male births will be within the range 0.47 to 0.53. With mortality rates which seldom have an expected value as high as 0.50, the contribution of each death to the total rate is relatively greater in a small population or sample than in a large one: D deaths in a sample of 10,000 represents a rate of $D/10$ per thousand and if this number is increased by 1, the increase in the rate is $1/10$ per thousand. If the rate in a sample of 1,000 is $D/10$ per thousand, there would be $D/10$ deaths per year and if this number is increased by 1, this is equivalent to adding 1 death per thousand to the rate.

In small populations the numbers of persons in each five-year age group are usually very small, and the smaller the number of persons at risk, the less reliable will the rates based on these numbers be as estimates of the risk of mortality run by such persons. It is therefore essential to have some index to the reliability of such rates, an index which will indicate the probable range of variation which can be expected in any future or alternative estimate of this rate. The index which meets this requirement is the standard error, and an estimate of the standard error of a rate is given by the formula $\sqrt{p(1-p)/n}$, where p is the rate (expressed as a ratio of unity, and not per cent or per thousand), and n is the number in the sample, or the population. The algebraic expression under the square root sign is known as the variance, and its square root might be described either as the standard deviation or the standard error. It is customary to use the term 'standard deviation' when referring to the dispersion of the elements of a frequency distribution about their mean value, and 'standard error' when referring to the dispersion of alternative estimates of a single statistic, such as the mean value, of that frequency distribution.

10. Comparisons of Mortality Rates

If we have a mortality rate p at age x in a population, and the standard error associated with p is s , the true risk of mortality at age x in the population will almost certainly lie within the range $p \pm 3s$, and within the narrower range of $p \pm 2s$ in roughly 95% of cases. If the risk of mortality at age x in another population falls somewhere within this range $p \pm 2s$, the chance that persons at age x in these two populations experience different mortality risks is less than 95 in 100. The evaluation of chances such as these is the basis of statistical tests of significance. It is customary in these instances to put up a 'null' hypothesis (such as there being no difference between any two rates), and ascertain by means of a test whether the null hypothesis is the most probable of all the alternative hypotheses. In comparing the magnitude of two mortality rates, the null hypothesis to be tested is that there is no difference between them, that they are merely alternative estimates of the one risk of mortality. If the difference between them is greater than one would expect to occur by chance 95 times in 100, then it is reasonable to deduce that the null hypothesis does not apply, and that there is a difference between them which is statistically significant at the 5% level of probability.

The test which is appropriate to comparisons between any two rates is based on the χ^2 -test, which is widely used to test the agreement between observation and hypothesis. In this case it is the square root of χ^2 which is calculated, and the test should be used with caution when the numbers of deaths on which the rates are based are small. When the number of deaths is less than 5 it is essential to introduce a correction for continuity into the computation of χ , and even with larger numbers it is preferable to use this correction. On the whole, it would be safer to leave the exact tests for the significance of the difference between mortality rates based on small numbers of deaths to a statistician.

With this warning, however, if p_1 and p_2 are the mortality rates experienced by two population samples of size n_1 and n_2 in a year, and the numbers of deaths d_1 and d_2 from which these rates were derived are not too small, then

$$\chi = (p_1 - p_2) / \sqrt{\left(\frac{d_1 + d_2}{n_1 + n_2} \right) \left(1 - \frac{d_1 + d_2}{n_1 + n_2} \right) \left(\frac{n_1 + n_2}{n_1 n_2} \right)}$$

If χ is greater than, or equal to, 2.58, the difference between p_1 and p_2 is such that only once in 100 times would you expect to find a greater difference than this between rates, based on samples of size n_1 and n_2 , which were alternative estimates of the one mortality risk, and it would be reasonable to assume that the mortality risk in the two populations differed significantly. If χ

is greater than, or equal to, 1.96, the chances of finding a difference of this magnitude or greater between two alternative estimates of the same frequency would be 5 in 100, and one would deduce from this that there probably were differences in risk between the two populations. The smaller the value of χ , the more confident one would be of the identity of the mortality risk in the two population samples.

For a thorough and reliable comparison of the mortality levels in two populations, distinct either geographically, ethnically or chronologically, the correct procedure is to calculate the age-specific mortality rates for the populations of each sex, and the standard errors associated with these rates; the parallel rates in the various series are then compared by means of the χ -test and the significance of the χ 's assessed one by one. With the males and females of each population classified into seventeen or eighteen age groups, such mortality comparisons become tedious, and often some of the laborious computation can be circumvented by inspection of the rates and their standard errors. If the mortality rates experienced by two population samples are p_1 and p_2 , and the standard errors associated with these are s_1 and s_2 respectively, there is a 95 % chance that the true risk of mortality in the first sample will lie within the range $p_1 \pm 2s_1$, and the risk in the second sample within the range $p_2 \pm 2s_2$. If $p_1 + 2s_1$ is less than $p_2 - 2s_2$, or if $p_1 - 2s_1$ is greater than $p_2 + 2s_2$, there is then no necessity to compute χ for it would be reasonable to assume that the risk of mortality in one sample differed from the risk in the other. The more rigorous comparison need be made only when there is overlapping between the ranges within which the risks of mortality lie, and with experience even some of these can be avoided.

From such comparisons between the parallel rates in two series of mortality rates one can pick out the ages at which there are significant differences between the mortality risks in two populations, or the ages at which significant changes have occurred in the one population over a period of time. Since the number of deaths likely to occur amongst the people of one age group in a calendar year is likely to fluctuate from year to year, it is advisable, when examining or comparing mortality risks in small populations, to use the statistics of deaths relating to a short period of 3 or 5 years rather than those of just a single year. The average of these annual statistics of deaths is then referred to the population alive at the middle of the period, and the resultant age-specific mortality rates have rather more veracity than would rates based on the deaths which occurred in one year only, in the course of which there may have been some outbreak of epidemic disease, or some other exceptional cause of increased mortality.

While these manoeuvres, averaging deaths at each age over a period of time, and the computation of the standard errors of age-specific mortality rates, overcome some of the interpretive difficulties of rates relating to small populations, few people can visualize just what is meant by such a series of rates in terms of the prevailing mortality in a population. Its meaning becomes clearer if the series is represented graphically, with age scaled horizontally along the x -axis of the graph, and the mortality rate at each age plotted, according to some chosen scale, vertically above the centre of the relevant age group. The points can then be joined, either by drawing a freehand curve between the points, or by straight lines joining one point to the next. When using mortality rates specific for five-year age groups, this latter is the more usual procedure, and this has been done in Figure 3, which shows the age-specific mortality rates for males in the Fijian component of Fiji's population

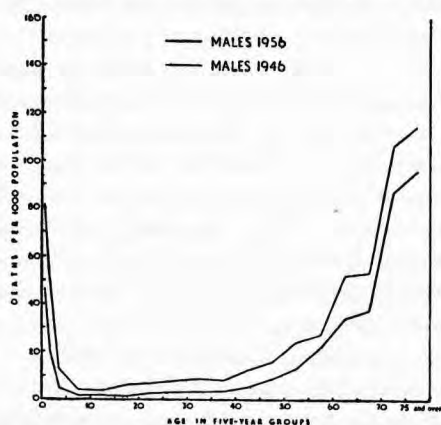


FIG. 3. A graph showing the average annual age-specific mortality rates amongst Fijian males in 1946 and 1956.

around the two censuses, 1946 and 1956. Both series are average annual rates based on the deaths which were registered during the relevant periods; the 1946 figures were derived from the deaths registered during the three calendar years 1945-47, and those for 1956 from the deaths registered during the two years from 1 October 1955 to 30 September 1957. The mortality at age 0 in both series is the infant mortality rate, in which the deaths of children at ages less than 1 year are referred, not to the census population at age 0, but to the number of live births which were registered during the same periods as the deaths.

There has been quite a considerable reduction in mortality at all ages for Fijians over the intercensal period, the most spectacular of which concerns infants and persons aged 15 to 24 years. At all other ages up to 54 years for males and to 59 years for females, the rate of mortality has been roughly halved. Such a significant reduction in mortality over a ten-year period is a remarkable achievement, and there can be no doubt of its reality, even though the specification of age at death on the registration forms and of age for many of the population in the 1946 Census may not be wholly reliable, and the completeness of the death registrations is slightly suspect. In a case such as this, where the information with respect to age, either in the population or the statistics of deaths, is not particularly reliable it might be preferable to condense the series into age groups covering a wider span of years, such as those described earlier as 'physiological' age groups.

11. Summary Measures of Mortality

On the whole, it would be more convenient to be able to summarize the series of individual rates in some way, even at the expense of some of the accuracy, to obtain a single index or measure for each population. There are several alternatives for this, the first of which is to sum the age-specific mortality rates over the whole range of ages. If we do this with the age-specific rates of Tiara (Table 5), we obtain the figure of 1.877. Now it will be noted that more than half of this total is contributed by the two rates at 70-74 years and 75 years and over, which are the least reliable of all the rates because of the smallness of the numbers in the population at these ages. The standard errors of these two rates are of the order of 0.033 and 0.103 respectively, and in the following year, the rates at these ages might fall anywhere within the range 0.271 to 0.469 for the age group 70-74 years, and the range 0.271 to 0.889 for the persons aged 75 years and over. (Range = rate $\pm$ 3 S.E.) If in this year, both of these rates have the minimum value likely, the summary index of 1.877 drops to 1.469; if both have their maximum value, the index increases in value to 2.285. Thus, the straight summation of the series of age-specific mortality rates gives too much weight to these rates with low reliability, and the value of this index for comparative purposes is therefore vitiated.

12. Standardized Mortality Rates

If we introduce some system of weighting and assign a weight to each of the rates before summation, this bias can be minimized by the wise selection of weights. Having chosen an appropriate series of weights, the computational procedure is no different from that already demonstrated for the calculation

of a crude death rate for the population of Tiara in 1960. This technique is sometimes given the name of 'standardization procedure', and the end result is known as a *standardized mortality rate*. It is a particularly useful technique for comparative purposes, because it can be used on populations standardized for characteristics other than age and sex, and it can also be employed for the comparison of factors other than mortality. In comparisons of the mortality experienced by several populations, it is customary to concoct a 'standard population' which bears some resemblance in its age structure and sex composition to the various populations which are being compared.

An appropriate standard for the comparison of levels of mortality in five island territories of the Pacific in 1956, for example, would be a conglomerate of the populations recorded in their 1956 censuses, each of them preferably reduced to a common total of 20,000 persons so that no territory over-weighs any other in its composition with respect to age and sex. If Fiji's total was reduced from 345,737 persons to 20,000, there would be 10,324 males and 9,676 females, instead of 178,475 and 167,262 respectively, and the numbers of males and females in each age group would then be reduced so that each age group contributed just as much to these totals as it did in the full-scale population. The 31,668 males at ages 0-4 years, for example, would be reduced proportionately to 1,832 ($31,668 \times 10,324/178,475$) and the 30,917 females in this age group to 1,788 ($30,917 \times 9,676/167,262$). Western Samoa's population of 49,863 males and 47,464 females (97,327 persons) would reduce to 10,246 males and 9,756 females, and the numbers of males and females in each age group would be proportional to these total male and female populations. If Tonga, American Samoa and the Cook Islands were the three remaining territories in the comparison, they too would contribute 20,000 persons each to the standard population and varying numbers of males and females at each age.

To compute the standardized mortality rate for Fiji, the age-specific mortality rates for the male population of Fiji would be applied to the male section of the standard population in exactly the same way as was demonstrated in Table 5. After calculating the number of deaths expected in each age group, these are then summed over all ages, and the procedure is repeated with the age-specific mortality rates for females and the female section of the standard population to obtain the number of deaths expected amongst them. The number of deaths expected in the whole of the standard population is the sum of the numbers expected in the male and female populations, and the standardized mortality rate is obtained by dividing the number of deaths by the total standard population of 100,000. The standardized mortality rate

for each of Western and American Samoa, Tonga and the Cook Islands would be obtained by applying their age-specific mortality rates in turn to the male and female sections of the standard population, thereby deriving the number of deaths which would be expected if each of these populations were of the same age and sex composition as the standard population. This is the essence of all standardization techniques—the removal of differences between populations with respect to characteristics which may prejudice the frequency of occurrence of the event being studied.

For less ambitious comparisons than the one explained above, it may not be necessary even to devise a standard population. If one wished to compare only the levels of mortality amongst Fijians and Indians in Fiji in 1956, it would be simpler and quicker to choose one of the two enumerated populations as the standard population. The age and sex specific mortality rates of the other component population would then be applied to whichever was chosen as standard, and since the age and sex specific mortality rates of the latter, when applied to itself would produce only its crude death rate, only two series of age-specific rates need be computed instead of four. In general, when making a comparison between only two populations, it is advisable to select the smaller population as the standard and, because of the greater reliability of the rates pertaining to the larger population, apply the rates from the larger population to this. If the Fijian population of 1956 is taken as the standard population, and the relevant age-specific mortality rates for males and females in the Indian component are applied to it, the standardized mortality rates for the two populations are 8.30 per thousand for the Fijian, and 8.51 per thousand for the Indian. The rate for the Fijian population is the crude death rate amongst Fijians, and the rate for the Indian population is what their crude death rate would have been had the Indian population been of the same age and sex composition as the Fijian.

Had the Indian population been chosen as the standard population, and the Fijian age and sex specific mortality rates applied to it, the standardized mortality rates for the two populations would have different values from these: the standardized rate for the Indian population would be its crude death rate, which was 7.25 per thousand, and the standardized rate for the Fijian population would be 8.01 per thousand. Not only are the magnitudes of the rates different, but it will be noted that, in the first pair, the Indian rate was a little higher than the Fijian while in the second pair, this position is reversed and the Fijian rate is higher than the Indian. Two factors contribute to this turnabout, one stemming from differences between the populations with respect to their age structure, and the second from differences in the sex

composition of the two populations. The Fijian death rates at ages 1-4 years are significantly higher than the corresponding Indian rates, and the high Fijian rates are given more weight in the second standardization than in the first because the numbers in the Indian population at ages 1-4 years are larger than the numbers in the Fijian population. In addition, mortality amongst Fijian females at ages beyond 45 years is much lower than the corresponding male mortality, but because of the small numbers of Indian females at these ages, the low rates for Fijian females are accorded less weight in the total when standardizing to the composition of the Indian population. This example demonstrates, perhaps more clearly than mere words could, the pitfalls and difficulties of interpretation which may be met with in comparisons between summary measures of mortality. Were the age and sex specific mortality rates of the Fijian and Indian populations applied to an independent population, yet another set of rates would be obtained, and what should one deduce then about the relative levels of mortality in the Fijian and Indian populations of Fiji?

13. *'Indirect' Standardizing*

Nevertheless this standardizing technique is particularly useful in cases where the information relating to the frequency of some event in a population is inadequate. Before 1947, the only information available on mortality in Fiji was the total number of deaths and the crude death rates for Europeans, Half-castes, Fijians, Indians and Rotumans. Being so susceptible to changes in the age structure of populations, the crude death rates give no real information about the level of mortality in Fiji in 1936 as compared with its level in 1946 or 1956, because the age compositions of the populations recorded in the censuses taken in each of those years were so very different.

Knowing both the composition of the population with respect to age and sex in these years, and the sex and age-specific mortality rates for at least one of the three years, one can, however, compute the number of deaths which would have occurred in each component had the mortality rates of the third year prevailed in each of the other years. Comparisons between the numbers of deaths expected in these populations, and the numbers which were recorded, provides a basis for judging whether or not the level of mortality in the population has changed, even though one cannot be specific about the sort of changes which may have occurred. If the 1955-57 mortality rates are applied to the Fijian component of the populations recorded in the censuses of 1936 and 1946, the number of deaths expected in each of these years would have been 928 in 1936 and 1,013 in 1946. Since there were 2,755 Fijian deaths recorded in 1936, and 2,018 deaths in 1946, the numbers expected on the basis

The probability that persons of each age will survive the subsequent year (or 5-year period) is then applied, in turn, to some arbitrary number at age 0 to build up the life table population. The number chosen for age 0 is called the 'radix' of the life table, and it is customary to start with 100,000 persons at age 0. The number in the life table population at age 1 year is given by the product of the radix with the probability that a person of age 0 will survive the first year of life (p_0). Multiplication of this latter number with the probability of persons aged 1 year surviving for a further year (p_1) gives the number of persons at age 2 years in the life table population, and by a series of such multiplications, the life table population is constructed. These are the figures shown under the heading of l_x in an official life table, l_x representing the number of persons in the life table population at age x years.

The next step in the construction is to calculate the number of years of life lived by the life table population, and this is the sum of the years of life lived by persons at each age. If l_x persons start out on their $(x + 1)$ th year, and l_{x+1} persons attain age $x + 1$, each of the l_{x+1} persons have lived one full year of life between age x and $x + 1$ years, and the persons who died in the course of that year ($l_x - l_{x+1}$) lived, on the average, half of one year each. The number of years of life lived between the ages x and $x + 1$, therefore, is the average of the numbers of persons alive at age x and those alive at age $x + 1$. The summation of the years of life lived by all persons in the life table population is usually done by starting at the highest age and working back to age 0, so that the numbers recorded are the years of life lived at ages x years and over.

The average number of years lived after age x by each of a group of persons aged x is known as the complete expectation of life at age x , and is usually designated by e_x^0 . It is obtained by forming the ratio between the number of years of life lived from any age x onwards in the life table population and the number of persons at age x in that population. The complete expectation of life at birth, e_0^0 , is the total number of years lived by the whole of the life table population divided by whatever number was chosen as the radix of the life table, and either this or its reciprocal, which represents the life table death rate, can be used as a single comparative index of mortality. It is not a sensitive index, however, and some idea of its insensitiveness may be gauged from the fact that it requires mortality rates five times as high as those of Fijians in 1955-57 to reduce their expectation of life at birth from 60 years to 30 years. In view of the differences between the age-specific risks of mortality for males and females, separate life tables for persons of each sex are essential.

15. *Fertility Analyses*

Although birth and death are at opposite ends of the life cycle, and fertility and mortality are demographically complementary, the measurement of fertility in a population does not parallel the measurement of mortality because the two processes are conceptually as well as factually different.

Death is an event which can happen only once to any one person, and at some time or other, each must die. When a death occurs, only one person is affected, and that person might be either a male or a female, and of any age between 0 and more than 100 years.

While each individual is born only once, there is not the same inevitability about a birth as about a death, and the person being born is not the only person actively concerned in the birth process. Each birth which occurs in a population requires the presence there of a female physiologically capable of reproduction, and she possesses this capacity for only a limited number of years. Even this limited period may be interrupted while she goes about the business of gestating and lactating, for each birth is the termination of a gestation period of approximately nine months, and while lactation may not preclude the risk of another pregnancy, it does reduce its likelihood. Once her reproductive period is finished, no woman can make any further contributions to the births which might occur in the population, and some women will pass through the reproductive period without any births at all. The cause for this may not be biological for often the decision as to whether or not there will be a birth is a voluntary one, or one that is regulated by social custom and conventions.

If it occurs in the human species at all, parthenogenesis is an exceedingly rare occurrence, and no matter how fecund a woman may be, she is unlikely to conceive alone. If it is the custom that marriage should precede conception and the modal (or most popular) age for females at marriage is x years, few women aged less than x years, and fewer still of those unmarried, will produce children even though they are physiologically capable of doing so. Of those aged more than x years, the expression of female fertility is in reality the resultant of two independent levels of fecundity, for a fecund female monogamously mated to a sterile male is no more capable of producing children than is a sterile female.

Hence, the births which occur in any population in a given year reflect a very complex interaction of personal desires, opportunities and capacities which may be totally unrelated to the potential fertility of the population, or to the extent of eventual replacement of the population. Potential fertility, or the biological capacity to reproduce, is usually referred to as 'fecundity',

with the word 'fertility' reserved for the expression of this biological capacity.

The extent to which a population replaces itself is the crux of the measurement of fertility, and it is much less easy to devise adequate measures for a continuing process such as replacement than it is for an instantaneous event such as death. The effective replacement of either a female or of a couple can be determined only when the female passes out of the reproductive period or when the mating is dissolved at an earlier age by death or divorce. Since the process can be voluntarily controlled, the replacement of one generation by another may be concentrated within a short period or it may extend over many years. The extent of replacement of a population each year by its births, dependent as this is on individual replacement, is quite different from the more fundamental individual reproductive performance although both are expressions of the level of fertility of the population. Two sorts of measures and two kinds of data are therefore required: the statistics of births in any one year indicate current replacement and act as interim measures of the prevailing fertility of the population but, in the absence of continuous matched records of population, they need to be supplemented by data which can be obtained only from censuses. Not all of the women of reproductive age in a population are equally likely to reproduce in any one year, so that the births of one calendar year refer only to a selected sample of the total population of females of reproductive age; and it is as important to have information about those excluded from the sample as for those comprising it.

16. Measures of Fertility: Crude Birth and Fertility Rates

Since all measures of mortality were based on annual occurrences, it is logical to deal first with the measures of fertility which can be derived from the annual statistics of births. Corresponding to the crude death rate, we can construct a crude birth rate by relating the total number of births which occur in a year to the total population alive during that year. It need hardly be said that the efficiency of the crude rate as a measure of fertility in a population is of an even lower standard than was the crude death rate's as a measure of mortality. At least the whole of the population is exposed to the risk of dying, but the only persons in the population who are exposed to the risk of childbirth are the females of reproductive age, and the remainder of the population is virtually irrelevant. The crudest sort of birth rate which should be constructed as a fertility index, therefore, should be the one which relates the number of births in a year to the number of females of reproductive age in the population. However, before writing off the crude annual rates entirely, it should be mentioned that the difference between the crude birth and death rates can be

used, in a population which is not subjected to migration, as an indication of the annual rate of growth of the population. A population which has an annual death rate of 25 per 1,000 population and a crude birth rate of 45 per 1,000 population per year is increasing at the rate of 20 per 1,000 population, or 2 per cent per year. If the death rate rises to 30 per 1,000 population in one year, and the birth rate falls to 40 per 1,000 population, the rate of increase for that year would be only 1%.

17. Age-Specific Fertility Rates

If the age of the mother at the birth of each child is not recorded in the birth statistics, it is not possible to compute any more precise estimates of the level of fertility in the population than the number of births which occur annually to each 1,000 females of reproductive age. Data concerning the age of the mother at the birth of each child, and the birth rank of each child born, are not recorded in all territories of the Pacific: as yet there is no provision for this in the registration of births in Western Samoa, the Cook Islands and Tonga, although the age of the deceased is recorded on the death registrations in these territories. If the age of the mother at the birth of each child is recorded, one can compute age-specific fertility rates which parallel the age-specific mortality rates, but however uniform with respect to age the pattern of fecundity or potential fertility may be, the pattern of fertility which appears varies quite markedly from one population to another.

Age Group	Fiji, 1956		French Oceania 1951	American Samoa† 1950
	Fijian	Indian		
15 — 19	53	161	112	64
20 — 24	274	350	282	310
25 — 29	285	325	259	341
30 — 34	206	240	221	275
35 — 39	142	176	142	200
40 — 44	62	78	60	56
45 — 49	19	33	9	17
All ages	1,041	1,363	1,085	1,263
Total Fertility	5,205	6,815	5,425	6,315

† Polynesian and part-Polynesian.

Table 6. The estimated annual age-specific fertility rates of females in selected territories of the Pacific: the average numbers of births per 1,000 females in each age group for the calendar years shown.

Variations in the expression of the fertility of females in some of the territories of the Pacific are shown in Table 6. These are the age-specific fertility rates derived from the births which were registered in each population in either two- or three-year periods centering on a census year, and they can be interpreted either as the number of births in a year amongst each 1,000 females in each age group in the various populations, or as the proportion of each 1,000 females at each age who produced a child in each of the years specified. Amongst the Fijian and Samoan populations, the highest rate occurs at ages 25-29 years, whereas the Indian component of Fiji's population and the population of French Oceania show a maximum frequency at the younger ages, 20-24 years. Relatively more females in these latter populations have children at ages 15-19 years than do the females of the Fijian and Samoan populations, and this combined with the earlier peak indicate a probably earlier age at marriage for females in the Indian and Oceanian populations. The standard errors associated with these rates are computed in exactly the same way as were those for the age-specific mortality rates, their approximate values being given by the use of the formula for the standard error of a variable which is distributed binomially.

18. *Reproduction Rates*

As with mortality rates, these age-specific fertility rates can be combined into a single index either by straight addition, or by assigning some weight to each and forming the weighted mean of the rates. If each age-specific fertility rate is interpreted to mean the number of births which occur to every 1,000 females in each year of the five-year age interval, the sum of the age-specific rates multiplied by 5 can be regarded as the total number of children born to each 1,000 women living through the child-bearing period. This number then represents the 'total fertility' of the population concerned, and if it, in turn, is multiplied by the proportion of females amongst all live-born children, the result will represent the total number of female children born to 1,000 women in the course of their reproductive life. The number of female children born to each of these 1,000 women is known as the *gross reproduction rate*, and although this, and its companion, the *net reproduction rate*, have been discredited in recent years as indices of the extent of the replacement of females in a population, the intermediate stage in its computation, the sum of the age-specific fertility rates which equals $\frac{1}{5}$ of the total fertility of a population, is a useful comparative index of current fertility. Failing any alternative measures of the level of fertility in each of the populations given in Table 6, there could be some justification for assuming that, on the average, Fijian

women would have 5.2 births throughout their reproductive period, Indian women 6.8 births, Oceanians 5.4 births, and the Polynesians and part-Polynesians of American Samoa 6.3 births.

In the case of the age-specific fertility rates, there is probably little to be gained from the use of the direct standardizing procedures. The application of each series of rates shown in Table 6 to a standard population of some given age structure would indicate the number of births expected in that population in a year under different patterns of fertility. Since the reliability of the rates pertaining to any one population is roughly equivalent, the amount of additional information gleaned by this procedure scarcely justifies the additional computations involved over the straight summation described in the preceding paragraph. A few years ago it was fashionable to apply the series of age-specific fertility rates for a particular population to the life-table population which had been constructed for the females of that population. Having calculated the number of births expected in the life-table population, this was then multiplied by the proportion of female children amongst all children born, and divided by 1,000 to give the net reproduction rate. Whereas the gross reproduction rate ignored the effect of mortality in the replacement of one generation of females by another, the net reproduction rate attempted to allow for the effect of mortality on the extent to which a female of one generation passing through the reproductive period, would replace herself by another female who would pass through this period in the succeeding generation.

It was the demonstration of the inequality of male and female net reproduction rates (the male net reproduction rate measuring the extent of replacement of the male population in consecutive generations), which opened the attack on the validity of the reproduction rates or of any such idealistic index of reproductivity based on current annual statistics. A more fundamental criticism is that each age-specific rate relates to a particular cohort of women, and the fertility experienced at any given age by a cohort is influenced by its previous marriage and fertility history. As marriage rates rise in a population, the proportion of marriages of short duration will increase, and since these marriages are, in general, more likely to produce births in any year than are those of long duration, the number of births each year and all rates based on them would increase. A customary concomitant of increase in marriage rates is a fall in both the average and modal ages of females at marriage, and a lower age at marriage is likely to entail a longer period of exposure to the risk of pregnancy. It would be unwise to infer that this would lead automatically to a larger number of births amongst such matings: up to a certain point,

the number of children born to a married couple can be controlled voluntarily by them, and successive births may be spaced more or less at will throughout the female reproductive period. In the long run, therefore, marriages of long duration may prove no more fertile than marriages of comparatively short duration, in which the interval between consecutive births may be correspondingly lessened.

19. *Changes in Fertility Patterns*

Although fertility is probably less closely linked to marriage in some of the Pacific communities than it is in many European communities, the age-specific fertility rates for a population may vary from year to year, or their pattern may change over a period of time. There appears to have been some change, for example, in the age-specific fertility rates of both the Fijian and Indian components of Fiji's population in the ten years between the censuses of 1946 and 1956, and in those of the Polynesian and part-Polynesian population of American Samoa between 1940 and 1950. Estimated average annual

Age of Women	Fijian		Indian		American Samoa 1940
	1946	1951	1946	1951	
10 — 14	—	—	3	—	—
15 — 19	64	65	243	200	57
20 — 24	255	269	313	335	254
25 — 29	251	257	274	302	312
30 — 34	202	174	223	220	251
35 — 39	142	141	158	155	188
40 — 44	61	60	73	77	57
45 — 49	23	20	32	44	28
15 — 49	998	986	1,319	1,333	1,147
Total Fertility	4,990	4,930	6,595	6,665	5,735

Table 7. The estimated annual age-specific fertility rates per 1,000 females in the Fijian and Indian components of Fiji's population in 1946 and 1951, and in the Polynesian and part-Polynesian population of American Samoa in 1940.

rates for Fiji in 1946 and 1951, and for American Samoa in 1940 are given in Table 7, and by comparison with the corresponding rates of Table 6, it would appear that there has been an increase in the fertility rates for Fijians at ages 20-24 and 25-29 years, and a decrease in the fertility rates for the Indian component at ages 15-19 years which is compensated by an increase at ages 20-24 and 25-29 years. In American Samoa, only one rate changed enough between

1940 and 1950 to be statistically significant, and that is the fertility rate at ages 20-24 years, which is higher in 1950 than it was in 1940.

20. Cohort Fertility

Were these data for Fiji more extensive, the average annual fertility rates could be rearranged to show the fertility experienced by each cohort of women in each year of each five-year age group. As yet, there is hardly enough data available for useful comparisons, but the sort of pattern which could be built up is demonstrated in Table 8. These are the age-specific fertility rates for the Indian component of Fiji's population, and the figures along the diagonal represent the average annual rates at five-year intervals since 1946. The spaces in the bottom right-hand corner will be filled in in time; those of the top left-hand corner would require the average annual fertility rates for this population in the calendar years 1921, 1926, 1931, 1936, and 1941, and these are not readily available. The 1927-31 cohort (which was aged 15-19 years in 1946) can be followed through from the beginning of the reproductive period, and over the first half of its reproductive life the births to this cohort, in both the Fijian and Indian components, were more numerous than would have been expected on the basis of the cross-sectional age-specific fertility rates of 1946. In 1946 the age-specific fertility rates for Fijians in each age group between 15 and 30 years were 64, 255, and 251 per 1,000 females, which is equivalent to $5 \times 570 = 2,850$ births per 1,000 females in fifteen years. Following the 1927-31 cohort through from ages 15-19 years in 1946, the fertility rates for them at ages 20-24 years (in 1951) and 25-29 years (in 1956) were 269 and 285

Birth Cohort	Age in Years						
	15-19	20-24	25-29	30-34	35-39	40-44	45-49
1897-01	†	†	†	†	†	†	32
1902-06	†	†	†	†	†	73	44
1907-11	†	†	†	†	158	77	33
1912-16	†	†	†	223	155	78	
1917-21	†	†	274	220	176		
1922-26	†	313	302	240			
1927-31	243	335	325				
1932-36	200	350					
1937-41	161						

† These numbers would relate to years preceding 1946 for which the rates are not readily available.

Table 8. The age-specific fertility rates experienced by cohorts of Indian females in Fiji since 1946: estimated average annual numbers of births per 1,000 women at each age.

births per 1,000 females, and these rates, when added to their rate at ages 15-19 years, would be equivalent to 3,090 births per 1,000 females between the ages of 15 and 30 years. In the Indian component, the sum of the age-specific fertility rates for each 1,000 females aged between 15 and 30 years in 1946 was equivalent to 4,150 births per 1,000 females, while the rates experienced by the 1927-31 cohort at ages between 15 and 30 years were equivalent to their producing 4,515 births per 1,000 females.

This population illustrates the point made earlier about changes in the timing of births. The frequency of births to Indian females in Fiji aged 15-19 years has declined progressively from 243 births per 1,000 females in 1946 to 200 in 1951 and to 161 in 1956. The initial fall from 243 to 200 births per 1,000 females at ages 15-19 years between 1946 and 1951 was followed up by a slightly higher fertility rate at ages 20-24 years in 1956 than has been shown at these ages previously. It is therefore reasonable to conclude that this decrease in the fertility rate at ages 15-19 years in 1951 probably represented some postponement of births by the 1932-36 cohort rather than an absolute reduction.

If the fertility rate of Indian females aged 15-19 years in 1941 was known, we could compare the fertility of the 1922-26 cohort up to age 30 years with that of the 1927-31 cohort. If this rate had been correspondingly higher in 1941 than it was in 1946, with a value of 288 births per 1,000 females aged 15-19 years, the sum of the fertility rates at ages 15 to 29 years would have been the same for the 1922-26 cohort as it was for the 1927-31 cohort.

It will be interesting to see whether the rate at ages 25-29 years in 1961 will be high enough to bring the estimated number of births to the 1932-36 cohort up to the age of 30 years up to the level attained by the 1927-31 cohort. If it should be as high as 353 births per 1,000 females, this would indicate that the 1932-36 cohort had merely postponed births for a few years, and by having more frequent births subsequently, its total fertility may not, in the long run, be any different from that of older women. Similarly, the low fertility rate of the 1937-41 cohort at ages 15-19 years (in 1956) should not yet be taken as indicative of any reduction in the level of fertility in this component, because it might be offset by rates at ages 20-24 years (in 1961) or 25-29 years (in 1966) higher than those experienced by the older cohorts, and these would compensate for the late start made by the 1937-41 cohort.

21. Fertility Data from Censuses

It is changes such as those described above which confuse the issue with regard to estimates of completed fertility based on current reproductive performance. Nor can this be circumvented by fertility data which might be collected in

censuses, for this too will image changes in the timing of births and their spacing throughout the female reproductive period. Since the number of children already born and the time interval since the birth of her previous child are two of the several factors which will influence the likelihood of a birth to some individual woman in any particular year, it is essential to have the most complete data possible about the population at risk, if not for each year, at least for some base year, and this can be obtained only from a census.

Ideally, one needs to know the present age of each woman, the total number of children she has borne up until that age, the number of these who are surviving and, if possible, her age at the birth of each child. With parallel data recorded on the birth registrations, one could then compute the probability that a woman of some given age and parity will produce a child within a specified period of time. This leads to a more sophisticated analysis than is possible for most Pacific territories now or probable for some years in the future, but it is still worthwhile to aim for more complete background data concerning the reproductive performance of populations from which one might construct the fertility pattern for women at each age: the age at which the first child was born, the average interval between births of each rank, the average number of children born before they attained their present age, and the number of these surviving. In a full enumeration it is scarcely practicable to collect this amount of detail, and the best which can be done is present age of mother, total number of children born, the number surviving, and perhaps the age of each woman at the birth of each of her children.

22. *The Average Number of Children Per Woman*

Details such as these were obtained, with varying degrees of accuracy, in the Census of Fiji and the other censuses for which this served as a model in 1956. The tabulations which were derived from these data showed the number of women in each age group who had borne each number of children, the number who had each number of children surviving, and the number who had borne their first child before some specified age. From the first of these it is a simple matter to compute the average numbers of children born to the women in each age group. In the Fijian component population, for example, there were 7,726 females aged 15-19 years, 6,878 of whom had borne no children, 503 of whom had borne 1 child, 73 had borne 2, 8 had borne 3, 2 had borne 4, and 262 for whom the number of children borne had not been stated. Thus the number of children borne to each of 7,464 women aged 15-19 years was known, and the total number of children born

to them was therefore $6,878 \times 0 + 503 \times 1 + 73 \times 2 + 8 \times 3 + 2 \times 4 = 681$, which gives an average of $681/7,464 = 0.09$ children per Fijian woman aged 15-19 years. The average numbers of children born to women in each of the other age groups can be calculated in exactly the same way, and the results for the Fijian and Indian components of Fiji's population are shown in Table 9. For both populations the average number of children per woman increases with increasing age up to the end of the reproductive period, and remains relatively constant at ages beyond 45 years. This is what one would expect, for however much difference there might be between cohorts in the expression of their fertility, the average numbers of children born will be related to the duration of their exposure to the risk of pregnancy. The average numbers of children surviving to women at each age are computed in exactly the same way, but since the proportion of survivors is likely to decrease as the age of the woman increases, the maximum numbers of surviving children might occur at some age during the reproductive period, and not necessarily beyond it.

Age Group	Fijian	Indian
15 — 19	0.09	0.35
20 — 24	1.10	2.23
25 — 29	2.58	4.25
30 — 34	3.80	5.73
35 — 39	4.79	6.60
40 — 44	5.47	6.73
45 — 49	5.88	6.94
50 — 54	5.73	6.81
55 — 59	5.95	6.45

Table 9. The average number of children born to women in each age group of the Fijian and Indian component populations: Census of Fiji, 1956.

If one was confident that all births were registered as they occur, and that the age of the mother was recorded correctly on the birth registration forms, there should be close resemblance between the average number of children born per female in each age group as recorded in the census, and the average number of births expected on the basis of the age-specific fertility rates for each cohort. Estimates of the average number of births per female in each cohort are computed by summing the fertility rates for the age groups already past, adding one-half of the fertility rate for their present age group (because on the average, these women are only half-way through this age group),

of children are born well within the first half of the female reproductive period.

Had comparable data been collected in preceding censuses, and particularly in the 1946 census, it would have been possible to estimate the average length of the interval between successive births to each cohort. Knowing the average number of children already born to females aged 15-19 years in 1946, and the average number already born to females aged 25-29 years in 1956, the difference between the two numbers would represent the average number of children born to each woman over the ten-year period. From this one can estimate the average interval which elapsed between each successive birth to this cohort. Let us suppose that Indian females aged 15-19 years in 1946 had had an average of 0.55 children per female: in 1956, this cohort had an average of 4.25 children each. In the ten years between the censuses, therefore, they had had an average of 3.70 children, and the average interval between births for this cohort (at ages between $17\frac{1}{2}$ and $27\frac{1}{2}$ years) was $10/3.70 = 2.70$ years. If females of the 1922-26 cohort (aged 20-24 years in 1946) had had 2.51 children in 1946 and 5.73 children in 1956, each of the 3.22 children born in the intercensal period was born at an interval of approximately 3.10 years. The average interval between the births of the 1917-21 cohort would be a little longer than either of these, probably about 3.80 years.

23. *The Attainment of Motherhood*

The number of children born before a certain age is only partially dependent on the interval between successive births, and differences between populations in respect of the number of children born to women of each age may be contributed by differences between the populations in respect of the age at which the women have their first child. If this age is recorded in a census one can derive, from the double or cross-classification of women according to their present age and their age at the birth of their first child, the proportions of each cohort who attained motherhood before certain selected ages. This may also expose differences in the behaviour of the various cohorts within one population, for if the fertility pattern is consistent over a long period, the proportion of each cohort to attain motherhood before some specified age should remain constant. Differences between populations in this respect will appear in both the average proportions amongst all cohorts taken together, and in the proportions relating to women of similar ages in the populations.

Using the Fijian data again, the proportions of women in each age group up to 59 years, in each of the major component populations, who had their first birth before the age of 20 years are shown in Table 10, and the raw data

from which the frequencies for the Indian component were derived in Table 11. There are marked differences between the two populations in the proportions of the female population who attained motherhood before the age of 20 years. While the proportion of each cohort of Fijian women is relatively

Age Group at Census date	Fijian	Indian
20 — 24	306	685
25 — 29	315	731
30 — 34	315	766
35 — 39	320	734
40 — 44	296	723
45 — 49	294	695
50 — 54	296	661
55 — 59	285	651

Table 10. The proportion per 1,000 women in each age group who had borne their first child before the age of 20 years in the Fijian and Indian component populations: Census of Fiji, 1956.

Age of Women	Total No. of Women§	Age at Birth of First Child						
		Under 15	15—19	20—24	25—29	30—34	35—39	40—44
15 — 19	8,806	71	†	—	—	—	—	—
20 — 24	6,718	148	4,452	†	—	—	—	—
25 — 29	5,386	268	3,670	956	†	—	—	—
30 — 34	4,347	218	3,112	625	161	†	—	—
35 — 39	3,708	176	2,545	629	127	48	†	—
40 — 44	2,609	162	1,723	463	86	35	11	†
45 — 49	1,691	96	1,079	337	67	26	7	2
50 — 54	1,157	79	686	251	49	19	6	3
55 — 59	750	32	456	132	27	13	2	4
60 — 64	864	26	494	190	42	17	5	2
65 — 69	415	7	203	103	36	11	6	2
70 — 74	326	8	121	106	29	20	7	1
75 and over	312	3	103	87	39	32	11	5

§ Excluding those for whom age at first birth was not stated.

† Not quoted, as the experience was incomplete for these women.

Table 11. The numbers of females of the Indian component of Fiji's population classified according to their age at the time of the census and their age at the birth of their first child: Census of Fiji, 1956.

stable at about 30%, the average proportion for the Indian component as a whole is more than twice this figure, indicating a very much younger age at first birth for them than for the Fijian women, and there are variations between their cohorts which suggest changes in the behaviour of younger Indian women in recent years. For those beyond the age of 40 years, the method of estimation of ages—both the current age and the age at the birth of their first child—and the differential effects of mortality on the older cohorts, tend to prejudice the reliability of the comparisons: the data concerning the cohorts recently attaining motherhood are probably more reliable, and these indicate a tendency amongst such Indian women to postpone their first conception to progressively higher ages. The underlying reason for this tendency, which is undoubtedly of great interest both socially and economically as well as demographically, lies beyond the scope of this chapter. The method of conducting an independent survey or enquiry to obtain a full explanation of this tendency is described in the next chapter.

III

CENSUSES AND SURVEYS

Since surveys of many types can act as new or supplementary sources of demographic data concerning the populations being surveyed, the opening sections of this chapter are designed to refer equally to all sorts of surveys, without restriction to those whose aim is solely the collection of population statistics. The word 'survey' will be used indiscriminately to mean either full-scale enumerations or sample surveys throughout the whole of this opening section, but where some remarks or comments are intended to refer specifically to one or other only, the appropriate distinction will be made. The scope of this book precludes anything more than a summary and even cursory treatment of the various aspects of planning, and this applies particularly to sample surveys, the detailed methodology and treatment of which are given in the books by Moser and Yates which are cited in the selected references. Full-scale enumerations of populations are rather simpler to describe, partly because their objectives are more standardized and of more limited scope than are those of most sample surveys, partly because they avoid the problems of sampling which complicate the discussion of other survey techniques.

Because of the differences which exist between the territories of the Pacific region—geographic as well as social and cultural—there can be no hard and fast rules for the practical conduct of a census or survey which will apply uniformly to all populations. There are certain broad principles which should be observed in the design and execution of any survey, but the details of both the design and the implementation of a particular survey can be evolved only within the context of what is already known about the population to be surveyed and what more it is desired to know. The very practical matter of costs must be taken into account and the final form of any survey should represent the most desirable compromise between all these factors—the best way of using the money available to obtain the maximum amount of the information required. There is no magic formula linking these factors, and each and every census or survey must therefore be unique. Some may think it presumptuous to attempt to condense this highly complex subject into a few pages with the minimum of statistical theory, but surveys appear to be becoming fashionable in Pacific territories, and those who sponsor them are often, and quite understandably, ignorant of the complexities of survey procedure, and unaware

of their own responsibilities in determining the course of surveys which may be undertaken. If this brief account does nothing more than point out some of the difficulties, and show what can and should be expected from a skilled and competent investigator, it will have been useful.

It will be assumed throughout that some form of questionnaire is to be used for recording the information to be collected, and the people who are supplying this information, the subjects of the survey, will be referred to generally as 'respondents'. Unless otherwise specified, it is tacitly assumed that the respondents are not completing the questionnaires themselves, but that this is being done by an interviewer or an enumerator. The concluding sections of the chapter deal with the summarizing and tabulating of data which are collected in surveys, and although demographic data are used for illustrative purposes, the methods outlined are not restricted to such data only.

1. Planning a Survey

The essence of all planning is anticipation, and efficient planning requires a certain skill in anticipating circumstances or predicting the outcome of a particular course of action. This skill may be acquired, somewhat expensively, by trial and error, or more economically by careful thought and the utilization of knowledge acquired elsewhere. When planning a party (a thing which most people have done at some time or other), the good host or hostess anticipates what the particular group of guests will eat and drink, and estimates how much of each kind of food will be required, or at least adequate, for the number of guests invited. Usually the party expenditure must be kept within certain limits, and the allocation of money between the various items of food and drink will be determined partly by the kind of party to be given, and partly by the object in giving it. No matter whether one is merely fulfilling one's social obligations or entertaining lavishly for some nefarious purpose, there must still be some anticipation of the demands of that particular group of guests, and they are the fortunate few who have not suffered the agonies of embarrassment because something runs out in the course of a party, or the physical discomforts of eating up the more perishable left-overs next day. With experience, anticipation of party requirements improves, and although no two parties are the same, a little thought beforehand can overcome the grosser errors of faulty anticipation.

The analogy between party-giving and collecting information, either about a population as a whole or about certain sections of a population, is not as superficial as it may seem, even though one is the dissipation of resources while the other purports to amass them. In planning and provisioning a party,

the end result is anticipated not merely in terms of the consumption of food and drink, but also in terms of social prestige, as measured by the reactions of the guests to the hospitality provided. Either consciously or subconsciously, the maintenance or enhancement of this social prestige is at the root of most party-giving, and each step in planning the party follows from this: the selection and number of guests, the kind of party to be given and the fare to be provided. If it is necessary to limit the scope for financial reasons, the plan may be modified, step by step, to keep the expenditure within the permissible limits, but the ultimate aim of the party will be preserved. In planning the collection of information by a census or a sample survey, the ultimate aim or purpose of the investigation must likewise be there, at the beginning; and since so many more people are involved in a survey and the financial outlay is so much greater, these aims and objectives need to be very much more precisely formulated than is the case with most parties. As with parties, it may be that only personal experience will prove that unless an investigator knows beforehand precisely what is required from the survey, what it is hoped to demonstrate in the final analysis of the information to be collected, he cannot even begin to plan its collection because he does not really know what questions need be asked of the participants in the survey.

2. *The Definition of Objectives*

This, then, should always be the starting point in any investigation—a clear and detailed exposition of the objectives of the whole project. If certain topics or problems are to be inquired into, an investigator must know precisely how it is proposed to inquire into them, what data are needed to show their existence or extent. Only when this has been done can he know what questions need be asked to ensure that the purpose of the investigation will be fulfilled, and that there is no waste of resources on futile and profitless questions. To the inexperienced, this always seems like starting at the wrong end, and the statistician who insists on this as a preliminary to the design of a questionnaire is apt to be accused of dogmatism.

Nevertheless, any survey, be it a full enumeration or a sample census or survey, must be envisaged as an operation complete in itself, the planning of which can begin only with the anticipation of its outcome. Purposeless collection of information is wasteful, and only those with unlimited time and resources can afford to be haphazard about the design of a survey. The demonstration of some interesting and possibly unsuspected phenomenon is no justification for a survey which is badly designed; such phenomena are just as likely to emerge, and with more point, from a well-designed

investigation. Surveys cost money, and those who make it available, be they governments or research institutions, are entitled to a return in the form of reliable information about the topic under investigation, and as much relevant information as can be obtained for the financial expenditure involved. It is the responsibility of the investigator to meet these conditions, and to plan his survey within the limitations set by the time and money available to him.

The budgeting of such resources demands the most careful planning of every stage of the survey, and this planning is not a job for amateurs. Under no circumstances should it be lightly undertaken, and all too often people who should know better waste both time and money because they fail to visualize the survey operation in its full sequence, fail to appreciate the all-important point that the form and contents of the questionnaire must be intimately related to the final presentation and interpretation of the data collected, and that the time needed to ask each question and to process the replies subsequently is a vital factor in the cost of the operation. Why ask a question at all unless it is clear at the outset why it is being asked, what use will be made of the answers, and how the various alternative replies should be interpreted? Respondents get bored by prolonged questioning, interviewers get tired of repeatedly ploughing through a lengthy questionnaire, the accuracy of the replies as given and as recorded declines, and at that point the survey might as well be abandoned. The whittling down of questions to the minimum that will provide the information required implies that the investigator knows exactly what information is required, and he has therefore arrived at the point when he can start to design and plan the survey.

This 'knowing what is required' might well be the most difficult part of the whole survey operation. It involves the clarification of the investigator's ideas and the sifting of those of others who may be interested in the subject or have worked in similar fields or on topics related to the subject matter of the survey. Almost always it seems that the background information which is so essential to the precise formulation of the objectives or hypotheses to be investigated is either lacking or inadequate, that it is impossible to predict what information will prove relevant, or the amount of detail which it will be feasible or practicable to collect. Nevertheless, unless an investigator is capable of overcoming these drawbacks to the extent of defining the objectives, of putting up a set of hypotheses and choosing the criteria by which they can be adequately tested, it is exceedingly doubtful if he is competent to conduct a survey at all.

Inevitably it takes much preparatory work to assemble and assess the merits of the various alternative hypotheses which have a bearing on the subject

matter of the survey, and to select the criteria for their testing. Some of these may subsequently prove redundant, others may need drastic revision: only careful examination of all imaginable alternatives, and continual review of them by the investigator and others in the early stages will show which are the most suitable for the purposes of the survey. When the pilot survey is taken, some further amendments in either hypotheses or criteria may appear desirable, but neither their formulation nor their pruning should be left until this comparatively late stage. The pilot survey should be thought of as the testing ground for the whole plan of operation, not merely for the opening phase of its planning.

3. The Coverage of the Survey

Implicit in the definition of objectives is the decision as to who are to be the subjects of the survey, what group of people is to be questioned. It is usually only in a population census that one is able to collect information equally about all members of a population, and most economic or social surveys cover only some group of people selected from the population. The first stage in the selection of participants for such surveys may be implicit in the topic to be investigated: if one wished, for example, to discover the incidence rates of various diseases amongst children aged less than 15 years in some specified area, the population to be studied has already been selected by the topic. If resources were not sufficient to permit the medical examination of all children aged less than 15 years, it would then be necessary to select from this population a group of children who could be regarded as being representative of this population. They would then constitute a sample in the statistical sense, and although any exposition of sampling theory is beyond the scope of this book, an outline of the elements of sample design and methods of sampling might be advantageous, if only to remove some of the more popular misconceptions about sample surveys.

4. Sampling and Sample Designs

If the results of a sample investigation are to form the basis for generalizations about the population from which the sample has been drawn, it is important that this sample be selected strictly according to the rules of statistical sampling theory. The various types of sample designs which are permissible in surveys all aim at ensuring that no bias enters into the selection of the sample, and that the sample chosen will afford the greatest precision in the results which can be expected for the financial outlay. The word 'bias' is to be interpreted as choice which is influenced by personal preferences, and the only way to avoid

bias is to make the selection according to some random process, 'random' being used in its statistical sense, to mean that each unit in a population has some known chance (other than zero) of being included in the sample. A *simple random process* is one in which each unit of a population has an equal chance of selection.

More complex processes may divide a population into clearly defined strata and allot each unit within a stratum an equal chance of selection; in such cases, the probabilities of selection associated with strata may differ one from another. The units of population in this context are not necessarily the individual members of a population: the heads of households or whole family groups may be chosen as the units of enquiry, and for the purposes of the survey, the population will then be composed of heads of households or family groups. Any of these populations may be divided into groups or strata according to characteristics which are common to all members of each group, and which are believed to be closely related to the subject of the investigation.

If one was conducting a survey into the level of consumption of some particular food in households, for example, the total consumption in any household will clearly be dependent on the number of persons within that household. Although this in turn may depend on the age of the head of the household, the grouping or 'stratifying' of the population of households on the basis of the number of persons each contains means that one source of variation between households is being controlled. Furthermore, it is probable that large households will provide a wider range of variation in their consumption than will the small households, and since households containing 9 or 10 persons will occur less frequently in the community being surveyed than will households of 3 or 4 members, it might be advisable to include a higher proportion of large households in the survey than of the smaller ones. The sampling fractions to be applied to each stratum might, therefore, be of the order of 1 in 5 for households of 9 or 10 members, and only 1 in 25 for households of 3 members and 1 in 20 for households of 4 members. These sampling fractions represent the probability or chance that each unit of population within a stratum will be included in the sample, and in a well-designed sample the various sampling fractions would be chosen so that the sample of households from any one stratum is large enough to permit valid comparisons between it and samples from the other strata.

If a uniform sampling fraction of 1 in 25 was applied to each stratum, the resultant sample would still be different from a simple random sample of population in which each household had a chance of 1 in 25 of being selected,

because one has ensured that households from each stratum would be represented in the sample. It is for this reason that stratification is used in virtually all sample designs nowadays; wisely applied, it does ensure a more representative sample of the population than one which is selected by simple random processes. The strata are delineated on the basis of known population characteristics considered pertinent to the topic of investigation, and the population within each stratum is then sampled randomly. This random sampling within strata is vital, for without it it is impossible to make any estimate of the precision of the results, and inferences drawn from the sample about the population are likely to be invalid.

Before any such random processes can be applied, however, it is necessary to have complete lists of whatever units of population are to be used, arrayed so that each unit can be identified by a number. This represents the 'sampling frame' and the units to be included in the sample are then selected from this frame either by the 'lottery method' or by using one of the sets of random numbers which have been compiled for such purposes, and can be found in books of statistical tables such as Fisher and Yates (1943) or Kendall and Babington-Smith (1939). If there are N units in the population from which n are to be selected by the lottery method, disks numbered from 1 to N are placed in a container and mixed well; the number of disks required for the sample are then withdrawn one by one, and these represent the units of population destined to be surveyed. If the selection is to be made from tables of random numbers, n numbers between 1 and N are taken in some systematic fashion from one of the several sets of random numbers available, and the units represented by these numbers will then constitute the sample.

When one is compiling a stratified sample, the N units of population are first arrayed in sub-populations, simultaneously classified according to the various stratification factors chosen. If the numbers of units in the various strata are $N_1, N_2, N_3, \dots, N_r$, where $N_1 + N_2 + N_3 + \dots + N_r = N$, the random selection of units from within each stratum proceeds in exactly the same way as before; if the sample is to include n_1 units from the first stratum, n_1 numbers are selected randomly from the series of numbers between 1 and N_1 ; n_2 units from the second stratum will then be selected at random from the series of numbers between 1 and N_2 , and this process is continued down to the r^{th} stratum, when n_r numbers are chosen randomly from all numbers in the range 1 to N_r . The difference between this and a simple random sample from the population is that instead of having just one sampling frame for the whole population, each stratum within the population has its own sampling frame. Usually it will be found that the identification



of units within the sampling frame for the population is inadequate for any elaborate stratification, and that it will not be possible to achieve the maximum homogeneity of sub-populations within strata which is desirable.

In addition to these wholly random processes, there are certain *quasi-random* procedures which are often used, where the starting point of the selection is determined randomly, and thereafter the selection is made according to some fixed rule. If one wished to take a 1 in 10 sample from a population, for example, some number between 0 and 9 would be chosen randomly, and this would be the starting point, the remaining units in the sample being identified by this number plus some multiple of 10. If the number chosen initially was 3, say, the 3rd unit listed in the sampling frame would be the first unit of the sample, and the remaining units to be included in this sample would be those numbered 13, 23, 33, and so on. The advantage of this sort of procedure is its convenience, for the sample can be compiled quickly and easily. It is not a random process, however, because except for the first unit, the selection of every unit of the sample is dependent on the selection of the previous one, whereas in simple random sampling (with replacement) each unit is selected independently of all others. Lists are usually compiled in some systematic fashion, and even though the reason for their ordering and arrangement may not be related to the topic of the survey, this will invalidate the estimate of error for the sample.

Of course if the list is not complete, no method of selection can achieve an unbiased, random sample of the population, and in the absence of complete lists of appropriate units of population, an alternative sampling frame will have to be found. One method which has been developed and used widely in the United States is *area sampling*, in which maps replace lists as sampling frames. This would probably be the most effective random process for sampling in Pacific territories, where the compilation of lists of household or personal units is often out of the question and there may be considerable practical difficulties in interviewing units scattered over a wide area. With this method of sampling, the area which is to be covered by the survey is divided into a number of smaller, contiguous areas, and a random selection is made of these areas to determine which are to be included in the sample. Every unit of population within each of these selected areas is then a member of the sample, or if this results in too large a sample, one could sample again within the areas to obtain the smaller numbers of units which would be subjected to a full enumeration. This is known as *multi-stage sampling*, with the sample selected at each stage being a strictly random sample of the relevant population—of areas in the first stage, and of units of enquiry within an area in the second stage.

This is quite different from *purposive sampling* in which areas are selected because of some particular characteristic which they possess, or because they might be regarded as being typical of all such areas. If one wished, for example, to study differences in the behaviour pattern of households in Fiji, one might elect to survey all households within certain areas which had been chosen on the basis of their predominant economic activity—the growing of sugar-cane, the cultivation of vegetables or coconuts, dairying, mining, and so on. Even if the person responsible for choosing these so-called typical areas is well-informed and has intimate knowledge of all areas in Fiji, the best one could hope for in the way of conclusions from such a survey would be a comparison of household behaviour patterns *between* these areas: a comparison between households dependent on sugar-cane and those engaged in dairying, or those growing coconuts as compared with those in the mining area. Because the initial selection of areas was based on somebody's personal judgment, there would be no real justification for generalizing from such a sample to the population of areas in which cane-growing, dairying, mining, etc. is carried out.

A better way of selecting the areas to be surveyed would be to stratify the total area of Fiji according to the various types of economic activity predominating, divide each of the resultant strata into areas of the required size and select, by one of the random processes previously described, which areas are to represent each of the various strata. Stratified sampling is just as applicable to area sampling as to any other type of sample design, and amongst the stratification factors likely to prove useful and relevant in surveys are the size of villages and their location, viz., coastal or inland, on a large or small island, with or without access to a vehicular road, within some radius of an urban centre or distant from it. Other pertinent factors may emerge from the survey topic, but once again, it will be essential to have complete lists of all villages, classified according to all of the stratification factors simultaneously, before any stratified random sample can be selected for survey purposes.

Before leaving this brief outline of types of sampling, one other type deserves mention, if somewhat unfavourably, and that is what is known as *quota sampling*. In its simplest form, an interviewer or enumerator is assigned a certain area in which to work, and is instructed to collect information from or about some number of units within that area, the choice of which units are interviewed being left entirely to the interviewer. It is in this respect, the identity of the basic sampling units, that quota sampling differs from random sampling. The protagonists for this survey technique claim that its time-saving qualities outweigh its non-randomness, and that the sampling errors of surveys are so small compared with all the other errors and biases which

enter into surveys and which cannot be estimated that no great harm is done by their confounding. It cannot be denied that much time may be saved in field work with quota sampling, because apart from the economies in travelling time which are possible, there is no necessity for 'call-backs' on respondents who were not at home when the interviewer called the first time. Nevertheless, it is exceedingly doubtful if the resultant sample can be regarded as being truly representative of the population being surveyed, even if the safeguards of stratification are introduced, and there is surely something to be said for the control of even one source of error.

Perhaps when there is no reasonably adequate sampling frame for either units of population or geographic areas, or when the nature of the questionnaire requires the fullest voluntary co-operation of participants (as in the surveys of sexual behaviour made by Kinsey and his collaborators) there may be no alternative to quota sampling, but such cases are rare. As a general rule, some sampling frame can be found or compiled to provide a basis for the random selection of samples, and skilful interviewing can overcome much of the reluctance of participants to respond to questioning. The refusal to answer questions will be encountered in all types of surveys, except perhaps in full-scale census enumerations when there is usually some legal provision for penalties for refusal, and it is fallacious to believe that the problems of non-response are circumvented by quota sampling: ignoring is not avoiding, and it is better to be on safe ground theoretically for some known percentage of a random sample than to have the dubious surety of full response from a co-operative but unrepresentative sample from the population.

5. The Size of the Sample

This description is by no means a full or exhaustive account of types of sample design; it is intended merely as an introduction (albeit salutary) and those who are concerned in the design of surveys would be well-advised to consult the books previously mentioned before proceeding further with their survey plans. The alternative would be to consult a statistician with some experience in this field as to what sample design is most appropriate to the objectives of the proposed survey. If, as in most modern surveys, the information to be collected is of a statistical nature or such as will require some sort of statistical analysis for its correct interpretation, the statistician will need to be brought in sooner or later, and since the planning of the survey must envisage the whole operation through to its final presentation, the sooner the statistician is introduced to it the better. The question of sample size, which is so closely linked with the sample design and the objectives of the survey, can then be handed over to

him, for although simple formulas can be given for the way in which sample size is estimated and the size of sample needed to yield results of the required precision, the general solution of this problem is not easy. Much depends on how the results of the survey are to be analysed, and what degree of precision is required at each stage of this analysis.

It will be remembered that the precision of an estimate is measured by the magnitude of its standard error, and it is from the solution of the appropriate algebraic equation for this that the optimum sample size, n , is obtained. Since this equation can be solved for n only by assigning some value to the standard error itself and anticipating the value (perforce approximate) of the estimate to be derived from the sample, the investigator must play his part at this stage and provide the statistician with all such information as may be required. For the novice this will be difficult enough, but the situation is further complicated if the survey aims to discover or estimate the value or frequency of more than a single attribute or characteristic of the population, because the attainment of the degree of precision desirable for each estimate may require samples of varying sizes. As a further part of the definition of the survey's objectives, therefore, the investigator needs to decide what priority is to be accorded each objective, what degree of precision is required in each estimate, and what use is to be made of such estimates. Given this and certain information about the population to be sampled, it is for the statistician then to achieve a compromise between the objectives and the design and size of sample needed to attain these objectives with the requisite degree of precision. There is no simple solution for what size of sample is needed for any survey, and it should be emphasized that the enlargement of a sample beyond a certain size will not necessarily increase the precision of the estimates derived from the sample. Should the optimal sample size imply costs in excess of the money available for the survey, there may have to be some sacrifice of precision, and such compromise should be the joint responsibility of the investigator and the statistician.

An intermediate stage which was not explained in the preceding paragraph is that leading to the decision as to what precision is desired in the various estimates to be made from the sample. Although this is primarily the responsibility of the investigator, the statistician can often give helpful advice and guidance, for the decision as to what estimates are to be derived from the sample entails the broad mapping of the types of analysis which are to be applied to the various items of information collected, and this in turn entails the plans for the summarization and eventual presentation of the data. To some extent the form of presentation and the type of statistical analysis which

will be necessary are mutually determined by the nature of the information collected, and there can, therefore, be no hard and fast rules concerning the arrangement of abstract data. Methods of summarizing and tabulating certain types of information will be dealt with in later sections of this chapter, and these will be found to have a wider application than may be immediately apparent. The intermediate stages in the processing of data from the questionnaires to the final summaries follow a logical sequence of simple steps, some of which can be omitted by those skilled in tabulating procedures. However, skeletons of the tables as they will appear in their final form mark the conclusion of this initial and complex phase of the planning, and with the most difficult part of the anticipation over, the investigator is ready to begin to design the questionnaire.

6. The Design of a Questionnaire

It is in this respect that there is the greatest divergence between censuses and surveys, and a questionnaire designed for a full enumeration of the population will usually be very different from one to be used with only a sample of people from that population. The form might well be different too if it is to be completed by the respondents themselves and not by trained interviewers or enumerators, for in the first case precise instructions need to be given with the questions, while in the latter, instructions on the form itself can be kept to a minimum. There are few general principles which can be formulated about the design of questionnaires because inevitably the format depends on the type of survey envisaged, the sort of information which is to be collected and the method of its collection. This is not particularly helpful for someone starting from scratch unless it is remembered that the objectives and coverage of the survey have already been precisely defined, and these mutually limit what questions can and need be asked. It is common sense not to ask questions which will not be fully and immediately understood by the majority of respondents, not to introduce (except by intention) concepts which are alien to their way of thought or subjects which are irrelevant to their way of life.

Hence one of the essential features in the design of the whole survey is that its objectives, and the questions needed to attain these objectives, should be pitched to the level of what can reasonably be expected of the population concerned at the time of the survey. In a full-scale enumeration, the questions which can be asked must be limited to those which can be answered readily by the majority of adults in the population. A difficult question or one requiring some thought might be included if it is absolutely essential, but it would be unwise to do so unless it were an enumerator census, and the enumerator

had been advised beforehand on how to detect and check an obviously incorrect answer. More ambitious questions might be included in a sample survey, when the interviewer has time and opportunity to prod for what seem like reasonable replies to such questions and cross checks can be inserted either by asking the same question in different ways or by asking questions which tie in with others asked elsewhere on the questionnaire. This is seldom possible in census enumerations which are designed to cover whole populations, and which need to be completed within some specified period of time.

Fortunately the basic information about the population which is needed from a census can be obtained from a few relatively simple questions, and the ordering of these on the census questionnaire follows a logical sequence. With any luck, someone else will have taken a census of the population previously, and if the schedule for that proved satisfactory, it might well be used again, either with or without modification. Should some additional information about the population be required at some particular time, it can usually be fitted in, either as a replacement of a question which had been asked previously and is no longer considered necessary, or as a new question altogether. It is inadvisable, however, to adopt a questionnaire from a previous census without a critical examination of its contents, and a realistic review of the value of the information obtained from it. Comparability between successive censuses is valuable only so long as the data being compared are of value.

On the other hand, the temptation to extend progressively the scope of the enquiry must be resisted unless one can be sure that the capacity of the population to supply the right sort of answers has also increased. Failing such reassurance, it would be preferable to restrict the scope of the full enumeration to what is practicable and feasible, and ask for such additional information as might be required only from a sample of the population. This combination of sample censuses with full enumerations of restricted scope has yet to be tried out in Pacific territories, but it could be used to advantage in many. The additional cost of an intensive census of certain sample areas, either at the same time as or immediately following the general enumeration, would be small compared with the cost of the full enumeration, and the quality of the information obtained would more than offset the extra cost which would be incurred in the collection and processing of the data. For the results to be generally applicable to the population as a whole, the sample areas would need to be selected randomly, and although the different levels of sophistication and educational attainment obtaining in the various parts of Melanesian territories in particular would create some difficulties in the framing of the questionnaire, these would not be insuperable.

If the unit of enquiry in a census is to be a household (however that may be defined), it is convenient to have as many members of a household as possible listed together on the one schedule. The design of the census schedule must then be such that one form will be sufficient for the majority of household units, a second form being required only for exceptionally large households. On the whole, it is preferable to have the members of a household listed vertically, with the particulars required about each individual ranged horizontally across the schedule so that each row contains all that it is necessary to know about one person. The schedule is ruled up into a network of rows and columns, with the rows evenly spaced down the page, and the columns of variable width to accommodate the answers which are required. The space needed to record a person's age, for example, is less than that needed for his name, so that the column for age on the schedule can be much narrower than that for name. Each cell of the schedule should be large enough to provide adequate space for what needs to be written there. If abbreviations are acceptable, and these should be allowed only so long as they are unambiguous, columns can be reduced in width beyond what would be possible if the replies were to be written out in full.

The general design of the census schedule which was tried out, with reasonable success, in several territories in 1956 is shown in Figure 4 and the measurements and column headings given there are those for the schedule which was used in Fiji. There were variations in size and content from territory to territory, and not all of the schedules were equally successful in operation because the convenience of the person completing the form was not always considered by those responsible for printing them. The loss of clarity engendered by this made the task of transferring the data from the schedules on to tabulating cards more difficult than it would otherwise have been, and the whole of the summarizing process was slowed down. Large though the household schedule for Fiji was, reduction in its size would have been practicable only if the scope of the enquiry had been reduced. As a footnote, it might be added that the particulars about dwellings, which occupied the top left hand corner of the schedule, was a topic better suited to a sample survey than to a population census: had this been excluded, and the geographic identification re-designed, provision for two additional persons could have been fitted on to the schedule.

The personal particulars which were recorded on these schedules probably represent the maximum which can be obtained with any degree of accuracy in a full enumeration of any Pacific territory at present. Even if the time allowed for the enumeration of a population is spread over two weeks, many enumerators will have to be used and the upper limit to the scope of the

enquiries possible is set by the calibre of the persons available as enumerators. Since it is fairly essential that an enumerator be reasonably familiar with the area for which he is responsible, inevitably there will be some range in the intellectual capacities of the various enumerators employed, even if the 'best' person in each area is cajoled into doing the job. However simple and straightforward a schedule may appear, whole batches of nonsensical or ambiguous

SCHEDULE NUMBER _____																			
DESCRIPTION OF DWELLING			CENSUS OF PACIFIC TERRITORIES, 19---				GEOGRAPHIC IDENTIFICATION OF VILLAGE, ENUMERATOR'S AREA, ETC.												
1 NAME	2 RELATIONSHIP TO HEAD OF HOUSEHOLD	3 SEX	4 AGE	5 FOR FEMALES YEARS & OVER	6 MARITAL STATUS	7 PLACE OF BIRTH	8 DATE OF BIRTH	9 ETHNIC ORIGIN	10 RELIGION	11 WHAT WORK DOES THIS PERSON DO ? FOR WHOM & WHERE DOES HE DO IT ?									
				a b c d															
1																			
2																			
3																			
4																			
5																			
6																			
7																			
8																			
9																			
10																			
a) TOTAL NUMBER OF CHILDREN b) BORN c) LIVING d) AGE AT 1ST BIRTH																			
<table border="1" style="float: right; border-collapse: collapse;"> <tr> <th colspan="3" style="padding: 2px;">NO. OF PERSONS</th> </tr> <tr> <td style="padding: 2px;">MALE</td> <td style="padding: 2px;">FEMALE</td> <td style="padding: 2px;">TOTAL</td> </tr> <tr> <td style="height: 20px;"></td> <td></td> <td></td> </tr> </table>											NO. OF PERSONS			MALE	FEMALE	TOTAL			
NO. OF PERSONS																			
MALE	FEMALE	TOTAL																	

FIG. 4. Design of a census schedule (actual size, 17" × 13½").

replies can be avoided only if enumerators are given explicit instructions as to how to proceed in their duties—how the schedule is to be filled in, what sort of information is to be recorded in each space and what are permissible alternatives. All of these instructions and directions must be remembered and followed at the time of the enumeration, and since there are likely to be many unexpected difficulties and hazards in the course of his unenviable task, the less an enumerator has to remember to do the better. Every additional question means more directions, more time per questionnaire, more exasperation when he strikes a difficult household, and probably more careless and cursory treatment of all the remaining households in his area thereafter. With more skilled interviewers or persons who were familiar with the use of questionnaires, more

ambitious schedules would be feasible, but so long as one has to rely on part-time and virtually untrained enumerators it is advisable to restrict the questions to be asked to those which are essential for an adequate description of the current state of the population.

In a sample survey one is not restricted to such a stylized questionnaire as a census schedule, and the form and ordering of its contents will depend on the objectives of the survey. To the general principles of clarity and convenience in handling, both in the field and in the processing later, should be added the further one that the questions be phrased as objectively as possible. Ambiguous or vague questions will produce ambiguous or vague answers; leading questions or questions indicating the sort of reply you expect are likely to produce either the answers indicated or their converse. Provided the survey is factual, and all questions concern matters of fact as distinct from opinions, the questions can usually be put to the respondents in a simple, straightforward fashion. Most questions can of course be phrased in several ways, and there is no general formula for what is the best way of asking a question. Experience in interviewing will help in deciding which of the several alternatives will prove most satisfactory, but if such experience is lacking, the various forms of a question might be tried out on friends and acquaintances.

If many questions are being asked, the interviewer's job can be simplified and speeded up by listing likely alternative replies to each question, and when filling up the questionnaire the interviewer need merely indicate (in some pre-arranged fashion) which of the alternative replies was applicable. In such a case, it is important to list all known alternatives, not merely the two or three first thought of, and these alternatives must be mutually exclusive. Some space must be left at the foot of the list for writing in answers which may not have been expected. This may add considerably to the bulk of the questionnaire, but if the alternative replies were precoded, the subsequent transfer of information from the questionnaire to a tabulating card would be accelerated. For example, a question on marital status could be written:

5. Are you married?	2
single?	1
widowed?	4
divorced?	3

If other than these,
please specify . .

. 5

The number to the right of each question would then be the code number to

represent marital status in the relevant section of the tabulating card. Whichever number was ringed by the interviewer to indicate the appropriate description of a respondent would subsequently be punched or marked directly on the card corresponding to each questionnaire, there being no necessity for reference to any code list first. The design and lay-out of a questionnaire can help or hinder the summarizing of the data, and any devices which will minimize the likelihood of errors in processing should be used to the full.

The essential elements in the design of a questionnaire are therefore that

- (a) it should be easy to handle in the field, easy for both the interviewer and the respondents;
- (b) the only questions to be asked should be those which are strictly relevant and to which reliable and pertinent answers can be expected; and
- (c) these questions be set out in some fashion convenient to the interviewer and to the subsequent processing.

Clearly these various conditions are not mutually exclusive. One of the factors which will make a questionnaire easy or difficult to handle in the field will be its length. If this is measured by the time needed to complete a questionnaire for each respondent, the nature of the questions must be taken into account as well as their number. It is both the nature of the questions and their number which will make the questionnaire acceptable or unacceptable to the respondents. While too long a questionnaire may induce carelessness in recording or non-cooperation in respondents, too short a questionnaire may defeat its purpose by not allowing the interviewer to establish rapport with his respondent. This is probably a rarity in practice, however, for the temptation which most people conducting a survey must firmly resist is that of covering too wide a field, of trying to collect every little piece of information which might have some remote bearing on the topic being studied, even if only to demonstrate eventually that it does not. There are always certain questions which can be considered to be on the border-line of what is relevant and what is not and a realistic appraisal of what can be hoped for from both interviewers and the survey population is necessary in deciding which of these questions can be included without endangering the success of the survey.

7. The Allocation of Interviewers

Fortunately for most investigators, the extraneous factors of time and money are likely to be the ultimate censors of the questionnaire. When its contents and form have been decided, it should then be possible to make tentative estimates of how long each interview will take and how many hours or days will be needed to cover the population to be included in the survey. Some

with a question on marital status for example, it would be necessary to define what is meant by 'marriage' for the purposes of the survey, and what constitutes 'divorce'. Although this may appear simple, and the investigator might feel that definitions given verbally would suffice, it is never safe to rely wholly on the memory of even a well-trained interviewer. It is much better to prepare written instructions which the interviewers can consult when there is any doubt as to how they should proceed. Since these are likely to need some amendment after the pilot survey has been run, the preliminary version of the instructions could be left in draft form until the questionnaire and all other phases of the survey plan have been tested in the field.

9. *The Pilot Survey*

Initially many aspects of the survey plan can be formulated only tentatively, and pre-tests of some phases may be necessary before the plans can be consolidated. These pre-tests are different from the pilot survey, and their aim should be to discover facts about a few isolated matters which will help the investigator in making decisions as to what is relevant or what is the best way to proceed in some set of circumstances. They are aids to the evolution of the plan, and are especially useful in cases where the background information about the population to be studied is inadequate or non-existent. They should not be thought of as substitutes for the pilot survey, the plan for which will incorporate all decisions which may result from the pre-tests.

The pilot survey itself should be conceived as the whole survey operation in miniature, its function being to test all phases of the organization for the collection and processing of data. Its scale will depend on the available resources of time, money and personnel, but it should be large enough to constitute an adequate test of the questionnaire under field conditions, and of the instruction and supervision of interviewers. Certain areas are selected for the trial and interviewers, armed with the questionnaires, are turned loose on the units of enquiry within these areas. Records need be kept of the time taken for each interview, of any difficulties which the interviewers encounter while administering their questionnaires, and of any suggestions they might have for the re-arrangement, re-phrasing or deletion of questions. Examination of the completed questionnaires will indicate how efficient were the instructions to the interviewers, and where such need alteration. It will also show which questions are worth asking and which can be eliminated from the questionnaire, either because they will not produce worthwhile answers, or because the nature of the replies received indicates that a particular question is on the border-line of relevance. If the questions are pre-coded, further alternative

replies might be suggested, or it might be found that those listed are not, after all, mutually exclusive.

If there is to be substantial revision of the questionnaire as a result of the pilot survey, this will entail some modification or amendment of all aspects of the plan, none more perhaps than the design of the tabulating cards and the plans for tabulation. Even if the questionnaire is not radically altered, experience with the transfer of data from the completed questionnaires to the tabulating cards might show that some re-arrangement of questions would facilitate their processing. Another important detail which should emerge from the pilot survey is the amount of non-response which can be expected. If the numbers unwilling or likely to refuse to participate in the survey are large, it would then be necessary to increase the size of the proposed sample to ensure that the number of units included will be sufficient to attain the survey's objectives. The pilot survey should also serve as the basis for rather more precise estimates of the probable cost of the survey than have been possible earlier. Finally it should act as a test of the efficiency of the field organization, to show how closely the interviewers need to be supervised, and how the essential liaison between field, supervisory and office staffs can best be maintained.

The pilot survey will not throw up all of the difficulties and inaccuracies which are likely to arise in the course of the survey, for the wider the population coverage, the more scope there is for break-downs in supplies or supervision. It can usually be relied upon to bring some incipient troubles to the surface, and their eradication can result in a general tightening-up of the operational plan and organization. Before making any changes, however, one needs to be sure that those which are proposed will be an improvement in the long run, and that some small gain from the roundabouts will not entail an enormous loss on the swings. This is why it is so important to see the operation as a whole, why so much preparatory work is essential for the right perspective on possible alternative procedures. The functions of the pilot survey are sufficiently manifold not to have the added complications of ill-conceived and vague ideas about the objectives of the survey. It can be used to resolve or clarify some problems of content, or some matters of organization, but it should not be strained too far in these directions if it is to fulfil its proper role. The execution of the pilot survey sets in train all of the final preparatory arrangements—the printing of the revised questionnaire and tabulating card, revision of the training plan and instructions for interviewers, such re-allocation of units or areas as may appear desirable, the revision of plans for the processing of the completed questionnaires and the tabulation of

the information collected. The time interval between the pilot and main surveys should be long enough to allow for all such preparations, for once the field work of collection is begun, the person in charge will have little time for dealing with anything other than the trivial and unsuspected crises which will be part of the daily routine.

The pilot survey marks the end of the calm before the storm of administrative matters which will, of course, include the recruiting and training of staff. Alterations in the questionnaire or in the design and size of the sample may change the former estimates of the number of interviewers required to cover the survey population in the time allowed. If the coverage of the survey is extensive, supervisory staff will be needed as well as interviewers. Even though practice runs with the questionnaire should be part of their training programme, it will still be advisable to have the first batch of questionnaires completed by each interviewer scrutinized very carefully. If a check is made after the completion of the first 10% of questionnaires, say, any errors or omissions which there might be can be rectified in good time. Such checks should be made continuously throughout the collection, and the training of adequate numbers of supervisory staff should be run simultaneously with the programme for interviewers.

Additional staff will be needed for processing the questionnaires, and this involves further checking and editing of the questionnaires before transferring the data on to tabulating cards and verifying that each transfer is correct. As for the interviewers, this procedure will require firm and fixed rules, in writing, which will cover the majority of contingencies likely to be encountered on the questionnaires, with decisions on dubious entries to be taken by the person in charge. The training of persons to handle this work should begin before the field work of collection is completed, and if the questionnaire is a long one, earlier still. This is probably the last vital stage in the planning, and from then on, the survey should proceed in a relatively orderly fashion.

In this outline of the planning and organization which is essential to a successful survey operation, the various stages have been listed serially, as though each stage leads on to the next. Little imagination is needed to see that few of the steps are entirely separate from the one before or the one after, and that many interlock with more than one other. It should also be apparent that all are essential, and while every survey will have its own particular planning problems and there may well be instances when this programme outline will need modification, none of the steps given here can be evaded if the survey is to produce tangible and meaningful results even from a very restricted

population coverage. There are no short cuts in designing a survey, and care in its planning yields handsome dividends. The plan of the whole operation may take several months to work out, and even if one is pressed for time, it is a grave mistake to embark on field work before all of the subsequent stages have been prepared for. None of these is as exciting as the planning and the collection, and all demand the most scrupulous care and patience.

10. Editing and Coding

The editing of the questionnaires when they are received from the interviewers is one of the more tedious jobs but also one of the more important because any errors and inconsistencies in the replies to questions which are overlooked at this stage will be carried through all subsequent stages of processing. Sometimes they may be detected in one of the final summaries of data, and it will involve considerable extra work to check through the tabulating cards to the questionnaire concerned, and thence through all the other summaries, amending each as required. Editing consists of checking each questionnaire for completeness and consistency. It is usual to expect some form of reply to each question, and even those questions which are designed for only some persons but not all can be given the provision of 'not applicable' as a reply. The job of editing is then to check that each question has been answered and that the replies are consistent one with the other. If not all questions have been answered, it might be possible to deduce from internal evidence what the answers should be.

Almost any questionnaire will contain some questions whose replies are complementary to or should be consistent with others on the questionnaire. Usually, for example, there are certain personal names which are used exclusively for persons of one sex, and relatively few which are given to either males or females. Hence, in a questionnaire which records name and sex, the personal name should be consistent with the sex stated. There may be other facts such as the relationship to the head of the household recorded, and provided that the terms for the male and female relationships are distinctive, this may help resolve some of the inconclusive instances of name/sex discrepancies. Other examples of replies which one would expect to find consistent would be those for age and date of birth for persons of either sex, age and marital status for persons aged less than 15 years, and if occupation is recorded, this would to some extent be consistent with a person's age and sex. When any two replies are inconsistent, there is always the difficult decision as to whether both are to be taken as stated despite the inconsistency, or one is to be altered to tally with the other, and if so, which should be altered?

Whatever decision is made, the same rule must be applied uniformly to all similar inconsistencies in the questionnaires.

If the replies to questions are to be coded on the questionnaire, this can be done at the same time as the editing. The joining of the two operations can be advantageous if the questionnaire is either very simple or very long because double-handling is thereby avoided. Some form of coding is essential if information is to be transferred from questionnaires on to tabulating cards, and the type of coding which is required depends on the method of tabulation planned. For manual sorting and summarizing and for some machine tabulating cards visual coding is often sufficient—this consists of having some abbreviation of each of the alternative replies anticipated printed on the tabulating card. As in all forms of coding, there should always be some provision for replies other than those specified, and also for instances when no reply was recorded. With visual coding, there is usually no necessity to write in the code on the questionnaire itself, and the information can be coded directly on to the card.

If numerical codes are used, one cannot avoid writing in on the questionnaire the number which is to be punched or marked on the card, and the construction of the coding frame relating to each question should properly be part of the design of the tabulating card. A simple example of a coding frame was given earlier in the pre-coded question concerning marital status: there were less than 10 alternative replies likely, and each of the alternatives could therefore be identified by a single number. If there were more than 10 alternative replies expected, it would be necessary to use a code involving more than one digit; a 2-digit code, comprising all numbers in the range 00 to 99, would be sufficient for 100 alternatives, and a 3-digit code, running from 000 to 999 would cover 1,000 alternatives. Whichever range is chosen, each alternative expected is assigned some number within that range, and this number thereafter identifies that particular reply.

It is not necessary to use all numbers in the range, and when constructing a coding frame it is customary to use one digit to designate some broad grouping of replies, and succeeding digits as particular classes within those broad groups. The coding frame for industry and occupation in Fiji, for example, used a 3-digit code. The first digit indicated broadly the nature of the industry in which a person engaged (*viz.* primary, secondary, construction and service, administration and government, etc.), the second digit particularized the type of industry within each of these broad groups, and the third the kind of work each person did within his industry. Occupations within each industry were grouped into five categories or classes of skill, and the final digit of the code indicated whether a person could be classed as proprietorial,

managerial or executive (=0), supervisory or clerical (=1), skilled tradesman (=2), semi-skilled worker (=3), or unskilled worker (=4). A first digit of 2 or 3 signified the broad group of secondary industries, and since sugar processing is the largest secondary industry in Fiji, all persons engaged in that were identified by the numbers 200 to 204. The numbers 210 to 214 covered all those engaged in the manufacture of food and drink exclusive of sugar processing; the numbers 220 to 224 those engaged in the manufacture of cigarettes and other tobacco products, and so on through all secondary industries.

When the coding frame was designed, it was by no means certain that it would be possible to classify the occupations stated on the census schedules in this amount of detail, even though the question concerning occupation was phrased in such a way as to indicate the nature of the work each person did and in which particular industry he was engaged. However, it is always easier to combine several classes in tabulations than it is to divide them, and when planning the coding frame it is therefore preferable to be optimistic about how detailed a classification of replies will eventually be possible. If the scrutiny of the completed questionnaires indicates that the retention of the detailed categories is not practicable, that distinctions between degrees of skill, for example, would not be justified by the nature of the replies recorded, then it is a simple matter to condense the coding frame, either during the coding operation or during the summarizing. In Fiji, it was found impossible to distinguish between the skilled, semi-skilled and unskilled workers in the agricultural and pastoral industries, and the sub-classes with final digits of 2, 3 or 4 were consequently all coded or tabulated as with a final digit of 4. Had it been possible to devise some scale to define degree of skill in these industries, there was provision for such distinctions within the coding frame, and this distinction between the degrees of skill would then have been carried through all of the tabulations automatically.

Finally, it is essential that each of the classes included in the coding frame be exclusive. With a coding frame such as that for Fiji in which each class of occupation within each industry was assigned a number, there was no provision for a person who combined two occupations. Although the proprietor of a store might also have been responsible for its accounting and clerical work, he could appear in only one place in the tabulation, either as a proprietor or as a clerk. In such a case, priority was given to his proprietorial status, and unless special provision is made in the coding frame for such a combination of occupations, an order of priority must be established. Had it been decided to allow for persons with dual occupations within one industry, a fourth digit

could have been added to the code to differentiate between those with one particular occupation only and those who combined this with one of several others. It would have been less easy to superimpose a scheme for the coding of a combination of occupations in different industries, for although there were many numbers unused in the coding frame, it is much easier and quicker to use a frame which is constructed according to some simple and logical system than one in which numbers are assigned at random to the various classes. Numerical coding offers much scope for error, particularly with inexperienced coding staff, and any devices which will reduce the likelihood or the extent of coding errors should be introduced into both the questionnaire and the coding frame. Some errors in coding are inevitable, but they can be kept at a minimum by scrupulous care and checking. As each batch of questionnaires is coded, they should be passed on to another person for verification of the codes marked.

11. *Tabulation: Manual Methods*

Tabulation means the ordering of the information collected in some systematic fashion so that the nature of the response to a particular question or combination of questions throughout the whole population can be assessed. It is the process of summarizing or serially grouping together into one category all those units for whom the same reply or combination of replies was recorded, thus leading to the classification of the population according to the various attributes or characteristics recorded for each individual or unit in the survey population. The process is perhaps easier to visualize if the information from a questionnaire is transferred on to a card so that there is one card corresponding to each unit in the survey population. When the transcription is complete, the cards can then be sorted into groups according to the items of information recorded on them. The procedure is no different from sorting a pack of playing cards into suits, each of which will contain 13 cards if the pack is complete. Each of these suits can then be sorted further into court cards and non-court cards, and there would then be 4×2 , that is, 8 piles of cards. If these were then counted, the result could be summarized in tabular form thus:

Card Suit	Court Cards	Non-Court Cards	Total
Clubs	3	10	13
Diamonds	3	10	13
Hearts	3	10	13
Spades	3	10	13
Total	12	40	52

The figures in the right-hand marginal column show the number of cards in each suit, those in the body of the table show the break-down of each suit into court and non-court cards, with the total number of court cards and non-court cards in the pack shown in the bottom row. Exactly the same result would have been achieved had the cards been sorted first into court cards and non-court cards: these totals could then be entered into the final row of the table, and the result of sorting each of these piles into suits would then be entered into the body of the table. Sorting according to one item only is termed a 'one-way classification'. If the cards in each of these suits are then sorted for another kind of marking, here court cards, the resultant classification is referred to as a 'simultaneous classification' with respect to the two items.

If it is proposed to compile only a series of one-way classifications of a survey population, transcription of information from the questionnaire on to a tabulating card might be avoided. Suppose that a survey population consisted of 1,000 adult males, who had supplied information about their marital status and their religion, and one wished to know only how many there were in each category of marital status, and how many belonged to each religious denomination, there being no interest in the simultaneous classification with respect to marital status and religion. To summarize this, the first step would be to draw up large skeleton tables with columns headed appropriately; one such table would list the categories of marital status, including a space for Not Stated, as in Figure 5, and the other would be similar but with the religious denominations listed instead of marital status. For each questionnaire which showed a single man, a vertical stroke would be made in the column headed 'single'; if a man was married, the stroke would be put in the column headed 'married', and this process would be continued until the replies recorded on all of the questionnaires had been scored. For convenience of counting, it is advisable to make every 5th stroke cross the four preceding ones, either diagonally  or horizontally . For obvious reasons this is known as the 'cricket-scoring' technique, and the number of strokes in each column represents the number of persons in each category specified. The sum for all columns taken together must add to 1,000 since there were a total of 1,000 questionnaires: if it does not, the whole operation must be repeated, this time more carefully. It is always advisable to repeat the operation to make sure that each questionnaire has been scored correctly. The summary for the religious denominations would then be compiled and checked in exactly the same way.

If it was thought that there might, after all, be some association between marital status and religion, this scheme could be modified to produce the

simultaneous classification of the 1,000 adult males with respect to marital status and religion. The skeleton of the summary table would then have the various categories of marital status heading the columns as in Figure 5, with the various religious denominations listed vertically, each one divided from the next by a line so that the body of the table consists of a series of cells as demonstrated in Figure 6. When scoring each questionnaire in the summary

Single	Married	Widowed	Divorced	Other	Not Stated
Number Single	Number Married	Number Widowed	Number Divorced	Number Other	Number Not Stated

FIG. 5. A specimen skeleton table for summarizing the classification of a population sample according to one characteristic, viz. marital status.

table, the stroke would be made in the cell which indicates both the marital status and religion which were recorded: for a man who was both married and a Roman Catholic, for example, the stroke would be put in the cell resulting from the intersection of the column headed 'married' with the row labelled 'Roman Catholic'. The marginal totals here should agree with the numbers obtained when marital status and religion were tabulated separately, and this could be used as a check on both of these tabulations.

This procedure is somewhat tedious, but it is useful if one wishes to tabulate only limited information about a population. This seldom happens in a census or a survey, but it often does happen with continuous or current records which are made for some purpose other than the accumulation of population statistics. In some territories the births, deaths and marriages are recorded in ledger-type registers, often with several entries on a page, which cannot be

and birth rank, if these are recorded in the register. The final table would be of the form shown in Figure 7 (in which dots and etc. are used to indicate the continuation of columns and rows), and the skeleton lay-out needed to summarize births of each sex from the birth register would be similar to this but on a larger scale to allow enough space in each cell for the scoring of all the births registered. A tabulation such as this requires painstaking care and patience: it is easy to put a stroke in the wrong place when scoring for only one variable and the likelihood of error increases with the number of variables being scored simultaneously. The maximum number of variables that should be attempted is 4, 2 across and 2 down.

The tabulation of the annual occurrences of births and deaths could be done much more simply and probably more accurately if each birth or death was recorded on a separate sheet or card which could be extracted from whatever container is used for their storage. In the previous example, the birth registration forms would first be sorted into two piles, one for males and one for females. These would then be checked and counted, and the totals written in on the skeleton summary in the cells marked A and B (Figure 7). Each of these piles would then be sorted according to the rank of the birth; taking the male births first, all first births would be placed in one pile, all second births in another, and so on. The number in each pile would be counted and entered in the cells A_1 , A_2 , A_3 , etc. of the summary, before going through each pile in turn to score them for maternal age. The last pile to be scored would be those for which the rank of birth had not been stated, and when all columns of the summary had been checked, the figures would be summed across each row in turn and their totals entered in the column headed 'All ranks of birth'. The sum of these figures should add to the total number of male births, just as the sum of all the subscripted A's in the final row should do. The female births would be treated in the same way: sorted first according to birth rank, and then scored for maternal age. The total number of female births in each age group of maternal age would be obtained by summing over all ranks of birth (including Not Stated) and these should add to the total number of female births. It will be noted that successive sortings of the forms reduced the tabulation to a series of one-way classifications according to the age of the mother, and thus reduced considerably the likelihood of mis-tabulation.

12. *Tabulation: Sorter Cards*

This is the essence of the procedure followed in summarizing data which have been recorded on some form of tabulating card. There are two sorts of cards

Age of Mother (5-year groups)	Rank of Birth and Sex													All Ranks of Birth			
	First		Second		Third		Fourth		Fifth		..	..	Not Stated				
	M	F	M	F	M	F	M	F	M	F	..	..	M	F	M	F	T
15 — 19																	
20 — 24																	
25 — 29																	
30 — 34																	
etc.																	
45 — 49																	
50 and over																	
Not Stated																	
All Ages	A ₁	B ₁	A ₂	B ₂	A ₃	B ₃	A ₄	B ₄	A ₅	B ₅			a	b	A	B	

FIG. 7. A specimen skeleton table for summarizing the classification of registered births according to the rank and sex of the child born and the age of the mother at the birth.

which can be used for manual tabulation: one is the usual sort of filing card on which each item of information required is written, and this is read from the cards when they are being sorted, and the other is a rudimentary punched card of a type known as Cope-Chat. An alternative name is 'sorter card', and the information is recorded on these, not by punching holes in the cards but by breaking holes which are already punched. One such card is illustrated in Figure 8 and the circles along the inner edges of the outline represent the holes which are punched along the edges of the card. To record some item of information, the piece of card between the outer edge of the appropriate hole and the edge of the card is removed by clipping with a ticket-collector's punch.

SEX M F	STATED AGE 0 1 2 3 4 5 6 7 8 9 NS	TOTAL CHILDREN BORN 0 1 2 3 4 5 6 7 8 9 NS	LIVING CHILDREN 0 1 2 3 4 5 6 7 8 9 NS
RECORD NUMBER Hundreds: 0 1 2 3 4 5 6 7 Tens: 0 1 2 3 4 5 6 7 Units: 0 1 2 3 4 5 6 7			
OCCUPATION & INDUSTRY Hundreds: 0 1 2 3 4 5 6 7 Tens: 0 1 2 3 4 5 6 7 Units: 0 1 2 3 4 5 6 7		AGE AT BIRTH OF FIRST CHILD 1 2 3 4 5 NS MARITAL STATUS M W D S NS	

FIG. 8. A diagram of a 'sorter card' used in the manual sorting and tabulation of data.

The sorting of a stack of these cards resolves itself into picking out those on which a particular hole is intact from those on which it has been clipped. This can be done very simply by running a skewer or knitting needle through the cards when they are stacked on the edge opposite the item required; those cards on which the hole is intact will be suspended on the skewer when it is raised vertically, and they can therefore be lifted out and away from those which are clipped. Since the cards have a tendency to stick together, it is advisable to riffle the stack slightly before lifting the skewer away from it. These cards come in various sizes, and can be punched and printed to any design required. The card illustrated in Figure 8 is of medium size, measuring

8" by 5", with 32 holes along the longer sides and 19 along each of the shorter sides, giving 102 spaces altogether for recording purposes. The corner holes cannot be used because one corner, usually the top right-hand corner, of each card is cut to ensure that the cards are stacked uniformly before sorting. The design shown here is deliberately incomplete, but it suffices to demonstrate the use of a second row of holes to accommodate a tens digit, such as might be required to record age or year of birth or number of children, and one of the methods for recording numerical information by using the '7 4 2 1 code'.

Some information has already been recorded on this card, and the clippings show that it relates to a married female aged 45 years, who has borne 12 children of whom 9 are still living, and who was 18 years of age when her first child was born. Her sex is indicated by the fact that the second hole along the top edge, the one marked 'M Sex', is intact; had the subject been a male, this hole would be clipped, and when sorting a stack of cards for sex, the cards belonging to females would be the ones withdrawn on the skewer. Her age is shown by the clipping of the hole representing a tens digit of 4 (which, it will be noted, eliminates the hole for the corresponding unit digit in the process), and the clipping of the hole for the unit digit of age of 5. One clipping is sufficient to record the age of persons for whom both digits are the same, viz. 00, 11, 22, 33, 44, etc. It will be noted that the first hole in the second row of holes in the section for stated age is labelled with a zero. In this system, all ages must be recorded as a 2-digit number, and those under 10 years will therefore be prefixed by a zero, 00 for persons aged less than 1 year, 01 for those aged 1 year, 02 for those aged 2 years, and so on up to 09 for the 9-year-olds. This means that every card representing a person aged from 1 to 9 will have both zeros clipped and one of the upper row of holes as well.

When sorting for age, the skewer is run through the cards in the position of the zero tens digit, and all cards representing people with some other tens digit of age will remain impaled on the skewer when it is raised, the only cards left behind being those for persons aged less than 10 years. The cards which have been lifted out are then sorted successively into 10-year age groups: those aged 10-19 years will be left behind when the skewer is inserted in the hole for a tens digit of 1 and raised from the stack, those aged 20-29 years will remain when the skewer is put through the hole representing the tens digit of 2 before lifting. Ten sortings will be needed to separate the cards into decades of age, and each of these piles or stacks then needs to be sorted on each unit digit of age. Whenever there is a double row of holes the cards must be sorted initially on the lower row, and subsequent sortings will distinguish the various sub-groups within these main groups. When sorting

cards for total numbers of children born, for example, the initial sorting would distinguish between women with 20 or more children and women with less than 20 children. The cards for this latter group would then be separated into those with less than 10 children and those with 10 to 19 children. With the cards then stacked in three groups, each of these would be sorted subsequently into 10 sub-groups corresponding to each number of children.

Instead of using a double row of holes for age and taking up ten spaces on the card, age could be recorded in eight spaces by using the 7 4 2 1 code. The magic of these four numbers is that by taking them singly or two at a time every number from 1 to 9 can be represented. The digits which do not appear are 3, 5, 6, 8, 9: 3 is the sum of 1 and 2, 5 is the sum of 1 and 4, 6 the sum of 2 and 4, 8 the sum of 1 and 7, and 9 the sum of 2 and 7. By taking all possible combinations of 7, 4, 2, and 1 numbers from 1 to 14 can be built up. The recording of age on this system requires 2 blocks of 4 holes each, one block representing the tens digit of age, and one the unit digit. On the card illustrated here, the 7 4 2 1 code has been used for 3-digit numbers such as Record Number and the numerical code for Occupation and Industry. Each has three blocks of 4 holes, one block representing the hundreds digit, one the tens digit and one the unit digit. The record number of the subject and her occupation have been marked on the card preparatory to clipping, and one can see at a glance that the record number is 549, and the code number of her occupation is 950, the unit digit of 0 being indicated by the lack of marking in the final block.

Sorting for occupation and industry on this system requires two skewers, with the cards stacked upside down and standing on the edge where age and numbers of children born and surviving are recorded. Starting with the hundreds digit, the skewers would be run through the holes marked 7 and 2. This separates out all occupations with code numbers in the range 900 to 999, which cards would be stacked and put aside, all of the other cards still being impaled on the skewers. Leaving one skewer through the 7 hole, the second skewer is withdrawn and then run through the hole labelled 1 to separate out all cards with numbers in the range 800 to 899. The sorting of cards with numbers between 700 and 799 needs only the one skewer already in position in the hole marked 7. Cards with numbers between 600 and 699 would next be extracted using the two skewers, one through the hole for 4 and one through the hole for 2, and the first skewer would stay in position 4 for the two further sortings, first of the numbers 500 to 599 and then for those in the range 400 to 499. Both skewers are needed again for the hundreds digit of 3, and only one for each of 2 and 1. The cards which will be left on

Age	Province or District of Enumeration							Total
	A	B	C	D	E	F	G	
0 — 9								
0								
1								
2								
3								
4								
0 — 4								
5								
6								
7								
8								
9								
5 — 9								
10 — 19								
10								
11								
12								
13								
14								
10 — 14								
15								
16								
17								
18								
19								
15 — 19								
20 — 29								
Not Stated								
Total								

FIG. 9. A specimen skeleton table for summarizing a population classified by age.

the province totted up from the counts for each area. The total number of cards used in all sortings of them subsequently must add to this total each time. The cards for females are treated similarly, combining the totals for each area into the total for the province to give the balancing total which must be obtained in each sorting.

15. Sorting and Summarizing

Since all of the tabulations retain the classification of the population by age, the male cards for each province would be sorted first for age. The initial sorting is performed on the zero digit of age, and starting with province A, the cards for males aged 0 to 9 years are extracted, counted and their number entered in the top left-hand corner of the summarizing lay-out, shown in Figure 9. These cards are left stacked, and those still impaled on the skewer are then sorted on the tens digit of 1. This extracts the cards for ages 10 to 19 years, which are then counted and their number entered in the appropriate cell of the summary table while they too are stacked and left. This procedure is continued until all of the cards for males in province A have been dealt with, and they are stacked in eleven (or less, depending on what is the highest age recorded) piles, ten (or less) relating to specific decades of stated age and one to those for which age was not stated. The sum of the numbers entered on the summary table should agree with the total number of males enumerated in the province, and this must be checked before proceeding further with the sorting. If the numbers tally, each stack representing a decade of age is then sorted on the unit digit of age. The cards extracted in each sorting operation are counted and their number entered in the appropriate cell of the summary, and the sum of the numbers at single years of age within the decade should agree with the total for this decade which has already been counted.

When the sorting of a decade into single years is completed, the cards can be combined immediately into two piles representing five-year age groups, ages with unit digits 0 to 4 in one and 5 to 9 in the other, because the remaining tabulations require age in no more detail than this. With the cards then in 5-year groups of ages, the male population of province A is then sorted into the six categories of marital status, one age group at a time, and the numbers of cards extracted at each sorting are entered in a summary such as that shown in Figure 10. The numbers entered in the cells relating to the marital status of males in each age group are then added across to check that the total number of persons contained in each age group coincides with the number shown in Figure 9. If this concludes all of the tabulations relating to the males enumerated in province A, the cards can be stacked together



[illegible]

FIG. 10. A specimen skeleton table for summarizing the marital status of a population according to age.

[illegible]

FIG. 11. A specimen skeleton table for summarizing the total number of children born to women of each age group in the population enumerated in Provinces A, B, and C.

again and put aside. If any further tabulations of these cards were required, they would be performed immediately, checking back each time to the numbers counted in each age group in Figure 9 for those in which the population is classified by age, or to the total for the province if the specification of age is not required. The programme for sorting and summarizing each of the remaining provinces is identical with this, and all of the tabulations which are required should be built up province by province, with all classifications of the male population within one province completed before embarking on the cards for the next province.

When the summaries for the male population have been completed, the cards for the female population of province A are then sorted into decades of age, and each of these stacks is then split up in turn into single years of age. The cards are counted, their numbers entered in the relevant summary table, and checked for completeness and consistency with the numbers counted in each decade of age and eventually with the total for the province. The cards for each decade are divided into two stacks corresponding to five-year age groups, and these piles are then separated out one by one into the various categories of marital status. When the cards have been counted and entered in the summary, the marginal total for the number in an age group should agree with the total for the same age group in the summary of the female population by single years of age. Distinctions of marital status are not observed in the tabulation concerning numbers of children born, and the cards are therefore re-combined into stacks according to five-year age groups at the conclusion of each step in the sorting by marital status.

The classification of the women in each age group according to the number of children they have borne proceeds as described earlier; starting on the highest tens digit (even for the younger age groups in case of mistaken clipping) cards showing 20 or more children are extracted first, then those with 10 to 19 children and finally those with less than 10 children are separated out from those on which the number of children was not stated. As each age group is sorted, the numbers of women with each number of children are entered on the summary table (abridged in Figure 11 to an upper limit of 12 and over), and when summed horizontally, these should add to the total number for that age group given in previous summaries. The cards relating to each five-year age group are put together again, and the sorting procedure for the number of surviving children follows the same pattern as that for the total number of children born, finishing again with the cards stacked in groups according to five-year age groups of ages.

In separating out the women in each group according to their age when

Present Age Group	PROVINCE	AGE when FIRST CHILD BORN																																				TOTAL IN EACH PROVINCE												
		Under 15	15	16	17	18	19	15-19	20	21	22	23	24	20-24	25	26	27	28	29	25-29	30	31	32	33	34	30-34	35	36	37	38	39	35-39	40	41	42	43	44	40-44	45	46	47	48	49	45-49	50 & over	NOT APPLICABLE	NOT STATED			
15-19	A																																																	
	B																																																	
15-19 Total																																																		
20-24	A																																																	
	B																																																	
20-24 Total																																																		
25-29	A																																																	
	B																																																	
25-29 Total																																																		
NOT STATED	A																																																	
	B																																																	
Not Stated Total																																																		
ALL AGES	A																																																	
	B																																																	
TOTAL																																																		

FIG. 12. A specimen skeleton table for summarizing the age of women at the birth of their first child in conjunction with their present age.

their first child was born, there is a special category of 'not applicable' for all those women who have not borne a child. These will have all the holes in this section of the card intact and will be identified in the final sorting on the tens digit of maternal age at first birth by remaining on the skewer when the cards clipped for 'not stated' instances are extracted. The sorting of each group of present age according to age at the birth of the first child therefore involves sorting first into decades of age at first birth, then withdrawing the 'not applicable' cards from those remaining by skewering the 'not stated' hole, before separating each decade of age into single years of age. Each pile is then counted as usual, and the numbers entered in the summary (Figure 12), the number of cards for each group of present age again being checked against those of the initial summary for the province. Unless the provinces are very large, the numbers in each cell of these three tabulations will be small, and it may be more convenient to group the cards for several provinces together than to work with those of just one province at a time. If this is done, the card counts from each sorting must be checked back to the original combined total, in the same way as for separate provinces, and this combination of the cards relating to several provinces would not be done until all the other tabulations in which the distinction of provinces was to be retained had been completed.

This sorting and counting would not necessarily be done by one person only, and as staff and accommodation permit, it could be organized conveyor-belt fashion. When the classification of the population by single years of age is completed, the numbers of males and females in each five-year age group in a province are fixed and in all subsequent sortings the number of cards counted must tally with these totals. As one classification of the population of one sex within a five-year age group is completed, this batch of cards can be passed on to another assistant who will sort and count them for the next tabulation which is required, and pass them on down the line. The more detailed the classification the more time is needed to sort and count the cards, and care must be taken to ensure that the flow of cards is relatively smooth and that they do not pile up before any one assistant, causing confusion and muddling age groups or provinces. The continual checking back to the original totals for each age group is laborious, but essential to the accuracy of the tabulations. If on any sorting the total number of cards for an age group does not tally with the initial count for that age group, the sorting procedure must be repeated.

When the numbers of males and females in each age group are first derived, it is advisable to enter these numbers into the marginal total columns of all summary sheets on which they should appear so that the assistants, with these

summary sheets before them, know before they start sorting how many cards there are in any group, and the number which must be obtained at the end of each sorting process that is performed on each batch of cards. Systematic sorting of one batch at a time reduces the likelihood of errors in counting, and unless the process is tackled methodically, there is grave danger of having cards from one stack accidentally mixed with those of others. The inclusion of one card which is irrelevant or the loss of one which is not will throw out the whole of one sorting operation, and much time can be wasted in locating the fault. A desk computer or an adding machine is fairly essential for the prolific arithmetic which is involved in compiling these tabulations. This too needs careful manipulation, and no sum for a row or a column of figures should be considered correct unless the same answer is obtained when the figures are added from right to left for rows, or bottom to top for columns, as when the addition is performed from left to right or top to bottom as the case may be.

16. Tabulation: Mechanical Methods

This procedure with sorter cards has been explained in detail because they can be used for many different purposes, and their use requires no special equipment except ticket punches and skewers which are inexpensive, and patience which needs to be extensive if not inexhaustible. Sorter cards are tedious and laborious when there are large numbers of cards requiring cross-classifications of many of the items recorded on each. Neither the tedium nor the labour can be avoided, but if the cards are grouped into batches of manageable size, and the sorting arranged systematically so that it becomes a series of repetitive processes, the various classifications of the whole population which may be required can be built up gradually on the summary sheets. This is in fact no different from the summarizing procedures needed for many types of so-called mechanical tabulation, in which the conventional types of punched cards are used. Within the limits set by the type and capacity of the machines, the sorting and counting of cards is performed mechanically, and the number of cards in each of the categories specified will be printed out on to machine tabulation sheets; these may then need to be summarized further by hand, before being de-coded and put into shape for publication. This is necessary with the older types of collating machines, but work is going on continuously on the design of new models which will perform all summarizing operations mechanically, even to printing out each classification of the population in the form in which it is to be presented. The tabulating machinery which is generally available in the Pacific area at present is rather less elaborate than this, and the

final summaries of the data are generally obtained cumulatively, using the same sort of summarizing lay-outs as those illustrated in Figures 9 to 12.

Even with the most elaborate mechanical aids to tabulation, the items of information about each person or unit of enquiry must first be recorded on cards or tapes which are fed into the machines, and these must still be prepared by hand. The punching of both cards and tapes is done with a key punch not dissimilar to a typewriter, and the operator of the key punch reads an item of information from the questionnaire and then depresses the key or sequence of keys which will punch a hole or a series of holes in a card to represent this particular item. The punched cards contain 80 columns, each of which will accommodate 10 alternatives, ranged vertically in set positions on the cards. It is vital to ensure the correct sequence of items on the card by having the design of the questionnaire tied intimately to the design of the tabulating card, and the questionnaires must be fully edited and coded before card-punching is commenced.

All information on the questionnaire will need to be allocated a numerical or alphabetical code and unless the questions are pre-coded, these codes must be written on the questionnaire before it is passed to the key punch operator for the preparation of the cards. Each card punched must be verified from the original source data, and this requires another machine, similar to the key punch and similarly operated. Provided the amount of information recorded on the questionnaire is not large, coding on the questionnaire can sometimes be avoided by using another sort of card which allows visual coding. This card is 'marked' with a special pencil, and the marks are 'sensed' electrically in a machine which automatically punches holes in the card to correspond with the pencil markings. Since the cards must accommodate both marking and punching, preferably in separate areas, the 80 columns of the Hollerith or I.B.M. punched cards are reduced to a maximum of roughly 30 columns for mark-sensed cards, and this restricts their use more to full-scale enumerations than to sample surveys.

Although mechanical tabulation is not automatically preferable to tabulation by manual methods, the disadvantages in the processing of data remote from their source may be outweighed by the relative costs of the alternative methods. There are some facilities for mechanical tabulation in Australia and when planning censuses or surveys, it would be advisable to seek the advice of commercial operators in this field, in the preparatory stages of planning, as to what types of machines are available and what type of card will be the most efficient for the job to be done. Before being able to advise, however, they will need to know what questions are being asked, how it is

proposed to deal with the replies to such questions, what tabulations of these replies will be needed, and the approximate size of the population or sample to be covered. The number of questions being asked, and the number of alternative replies to these questions determines the number of columns which will be needed to record all possible answers on the card. A question which has more than 10 likely alternative replies will require at least 2 columns on the card: one which has less than 10 possible alternative replies will take up one full column, even if the number of alternatives is only 2. Hence, if the replies to all of the questions are to be tabulated, the number of questions which can be asked on the questionnaire is limited by the recording capacity of the tabulating card.

The tabulations which are required of the data recorded on the cards will be obtained by certain 'machine routines', and these routines can be facilitated, or unnecessarily complicated, by the ordering of the items of data on the tabulating card. This will reflect back to the ordering of the questions on the questionnaire since differences in ordering between questionnaire and tabulating card are a potent source of error. Finally, the time-consuming and expensive part of mechanical tabulation is the card preparation, and some estimate of the number of cards to be prepared will be necessary, not only to estimate the cost of the tabulation programme, but also to decide the relative merits of punched or marked cards for a particular census or survey. Although card preparation contributes so very largely to the cost of tabulation, the number of tabulations and the complexity of the machine routines necessary to derive these will add to their share of the total.

IV

THE USES AND ABUSES OF POPULATION STATISTICS

It is proposed now to choose a survey topic and indicate how data from both censuses and the continuous records of a population should be used to help build up the background for a survey operation. Quite apart from its tie-up to the hypothetical survey, the discussion of the census data in these terms affords further and more diverse examples of the interpretation of statistical summaries and of the methods which would be employed in their analysis. Earlier it was suggested that a topic suitable for a survey was the changing pattern of fertility amongst females of the Indian component of Fiji's population, and some justification for this choice of topic lies in the fact that this might not be an isolated instance of such change in the Pacific territories.

1. Review of Background Data

In the census of Fiji, 1956, the information sought about each individual in the population was

- (1) Name
- (2) Relationship to the Head of the Household
- (3) Sex
- (4) Age in completed years
- (5) For females aged 15 years and over
 - (a) Number of children born who were still living
 - (b) Number of children born who were dead
 - (c) Age of the woman at the birth of her first child
- (6) Marital Status
- (7) Place of birth
- (8) Date of birth
- (9) Ethnic origin (i.e. whether Fijian, Indian, European, etc.)
- (10) Religion
- (11) Occupation and industry.

The first two questions were asked to provide a check on the sex stated in column (3), although column (2) was also used to distinguish family units within households, from which the average size of family units and the number of family units per household were subsequently derived. A household was

defined primarily on the basis of communal eating, because this was the only definition which would apply equally to the two major components of the population. Needless to say the definition needed amendment under some circumstances, but the group of persons who commonly ate together and shared in the collection, cost and preparation of the food generally constituted a household for census purposes. The date of birth of each person served as a check on the age which was stated in column (4), and the check for consistency between the alternative replies for both age and sex constituted part of the routine editing of the schedules before tabulation. Similarly the stated age of females aged 15 years and over, the number of children they had borne and their age at the birth of the first child were examined for consistency during the editing and coding processes. The tabulations which finally emerged from these schedules showed the males and females of each component population classified in various ways, almost all of them in conjunction with age.

It was the analysis of the tabulation of women according to their age at the birth of their first child, in conjunction with their age at the time of the census, which suggested that some change might be occurring in the pattern of family building in the Indian component of the population, because the females in the younger cohorts appeared to be having their first child rather later than the women of the older cohorts had. If this is in fact happening, is there any difference in this respect between the populations in urban areas and those of rural areas? Is there any difference between those females who were born in urban areas and are still resident there, and those who were born in rural areas and migrated into the urban areas? More generally, and since the Indian population conventionally observes the precedence of marriage over first birth, is this change related to some change in the pattern of marriage? Are the females marrying at a later age than previously? If so, is this pattern consistent for the urban and rural populations, or are there differences between them? Is there any economic reason underlying the postponement of marriage? Is there any demographic reason for its postponement? No matter what has been happening to age at marriage, has the interval between marriage and first birth lengthened? If the first birth is delayed until a higher maternal age, does the second birth follow more quickly? Or is the interval between first and second births the same for all cohorts of women in all regions? These are the sort of questions which would need to be answered by a survey purporting to investigate the factors which might be involved in a changing pattern of fertility. Both the census data which were tabulated in 1946 and 1956 and the continuous records of registrations contain information relevant to many of these questions, and they should therefore be explored first.

2. *Urban/Rural Comparisons*

Few of the comparisons between urban and rural populations can be made from either source without additional tabulations of the data, but their compilation would present no insuperable difficulties. In the census data, the distinction between urban and rural populations would be reasonably clear from the definition of the enumerator's areas. There would be some areas on the borderline between the two, and it would have to be decided whether these should form a separate class or be excluded altogether. Whichever decision was made concerning these, the personal cards for the Indian population enumerated in urban areas and in rural areas bordering on urban areas would then be separated out from the remainder, and then the cards relating to females aged 15 years and over in each would be extracted. The subsidiary tabulations required from these cards all concern age and the first sorting of any one group would therefore be according to this characteristic. These would then be sorted in turn for the classifications by marital status, the total number of children born and the number still surviving, and the age at the birth of the first child. With each of the sub-populations classified in this way, the analysis would follow the same lines as were outlined in Chapter II for all Indian females aged 15 years and over, when their characteristics were being compared with those of the other components of population. The essential comparisons would be between the proportions married at ages 15-19 years, 20-24 years and 25-29 years in each sub-population, the average numbers of children born to women in each age group, and the proportions of each cohort who had borne their first child before each of certain specified ages.

3. *Age at Marriage*

Even without these additional tabulations, there is still much to be learned from the census data about the Indian component as a whole. The comparison between the proportions of persons married in each of the populations enumerated in the censuses of 1946 and 1956 indicates that this change in the age at which younger cohorts of Indian women have their first child has been accompanied by, and possibly associated with, a change in the age of Indian females at marriage. In 1946, 75 per cent of the female population enumerated at ages 15-19 years were married, whereas in 1956, this proportion had fallen to 46 per cent. At the same time, the proportions of females at ages 20-24 years who had not married had increased from 6 per cent in 1946 to nearly 9 per cent in 1956. These increases in the proportions of population at ages 15-19 years and 20-24 years who had not married were repeated for the male population. In 1946, 22 per cent of Indian males aged 15-19 years and 78

per cent of males aged 20-24 years either were or had been married; in 1956 the corresponding proportions were 6 per cent at ages 15-19 years, and 62 per cent at ages 20-24 years.

The parallel changes in the proportions of both males and females who had not married at the two censuses suggest that the reason for the smaller proportions married in 1956 as compared with 1946 is either economic or social rather than demographic. The demographic reason for a higher age at marriage for females would apply only if there was a shortage of eligible males in the population. On the contrary, there is positive evidence that such shortage of eligible parties as exists concerns the female population rather than the male section. An abridged classification of the Indian component by marital status in conjunction with age is shown in Table 12, and from this it is clear that there were more males available for marriage in 1956 than there were females. Between the ages of 15 and 29 years there were 12,281 eligible males and only 6,063 eligible females, so that there is a distinct shortage of females of marriageable age. Under normal circumstances, this would tend to depress the age of females at marriage and there must, therefore, be other factors operating to have raised it over the intercensal period 1946-56.

Age group	Married		Available for marriage		Total excluding not stated	
	Male	Female	Male	Female	Male	Female
0 — 4	—	—	16,723	16,574	16,723	16,574
5 — 9	—	—	14,329	14,126	14,329	14,126
10 — 14	—	—	11,983	11,933	11,983	11,933
15 — 19	593	4,155	8,666	5,054	9,259	9,209
20 — 24	4,434	6,168	2,899	718	7,333	6,886
25 — 29	5,398	5,195	716	291	6,114	5,486
30 — 34	4,392	4,193	277	227	4,669	4,420
15 — 29			12,281	6,063		

Table 12. The numbers of males and females in selected age groups of the Indian component of Fiji's population who were married or available for marriage (i.e. single, widowed or divorced) in September 1956.

More direct evidence on the ages of persons at marriage could of course be obtained from the marriage registers. Since one is trying to establish some trend in the age at marriage, either that there has been a change or that there has been no change, it would be advisable to examine the records of more

than just one year, and one could begin with the marriages recorded in the three-year period 1935-37, following these up with those registered in 1945-47 and 1955-57. All first marriages which occurred in each 3-year period would be tabulated according to the age of the bride and the age of the groom at marriage, and although a simultaneous classification of marriages according to these two ages would be desirable, it would not be essential. It would, however, be important to record the ages at marriage in single years because the grouping of ages into the customary 5-year age groups in this tabulation would mask any changes which might have occurred within an age group. From the age distribution of brides and grooms at marriage one would then compute the median age of the brides in each period, and the median age of the grooms. If there has been any change in the age at marriage, the direction of this change should be apparent in the median ages of either brides or grooms or both, because the median age is the mid-point in the distribution of ages such that half of the persons who married did so at a younger age than the median, and the other half married later than the median age.

4. *Economic Background*

Assuming that the median age at marriage for both males and females has increased in recent years, what information was collected in the census which might indicate some likely reason for this change? Having already decided that there is no purely demographic explanation, one could look at the economic conditions as revealed in the distribution of the population throughout the range of occupations and industries. There would be no point in examining these data for the female section of the population because even if the number of females engaging in industry had increased, one could not distinguish how much this was the cause of the postponement of marriage and how much the effect. The classification of occupation and industry used in the 1956 census tabulations was different from that used for the 1946 census, and the data are not therefore comparable. Comparisons must accordingly be based on indirect arguments.

If we define the work force as the number of males in the population aged between 15 and 64 years, one can use the age distribution of the population to make certain inferences about the difficulty or otherwise of finding employment. In 1956 there were 9,263 males at ages 15-19 years in the Indian component, and therefore the recruits to the work force over the five years preceding the census date numbered roughly 9,000. The number of males at ages 65-69 years in the census population were 1,343, and it is therefore probable that approximately 1,500 persons retired from the work force in the same

period. This means that six times as many persons entered the work force as retired from it, and the size of the work force increased by approximately 25 per cent in the course of that 5-year period. Unless there was considerable expansion of economic activity in Fiji during this period, leading to increased opportunities for employment, this excess of recruits would create some difficulties for young males seeking employment. This situation is likely to be repeated in the future because by 1961 the males aged 10-14 years in 1956 will be aged 15-19 years, and the ratio between the number of Indian males aged 10-14 years and the number aged 60-64 years in 1956 was 11,983:1,904. Once again there will be six times as many persons entering the work force over the period 1956-61 as there are retiring from it. It should be noted that these figures relate to the Indian component alone, and although the ratio is a little more favourable in other components, the search for remunerative employment was not and will not be confined to the Indian component alone.

Turning to the more specific tables of occupation and industry, more than half of the economically active males of the Indian component were engaged in agricultural and pastoral industries at the time of the 1956 census, and three-quarters of all so engaged were employed in the cultivation of sugar-cane. The proportions of each 1,000 males within each age range who were engaged in each major group of industries are shown in Table 13. Although the proportion of males aged 15-29 years who were engaged in agriculture was not much lower than the corresponding proportions at ages 30-44 years and 45-59 years, the picture is markedly different if the age range 15-29 years is split into its three component age groups. While 574 per 1,000 of males aged 15-19 years were engaged in agricultural and pastoral industries, and 433 of them in the cultivation of sugar-cane, only 490 per 1,000 of the males aged 20-24 and 25-29 years were engaged in agricultural and pastoral industries, with 377 of them cultivating sugar-cane. There may be some distortion in these figures because the census was taken in the cane-cutting and crushing season when the maximum number of persons was likely to be engaged in this industry, and although no distinction was made between cane-cutters and other workers on cane farms in the tabulations, at least some of the younger males would have been employed only temporarily as cutters. However, it is reasonable to conclude from the figures that males aged 20-29 years are tending to drift away from agriculture into secondary industries and construction.

Nevertheless, the cultivation of cane remains the predominant economic activity of the Indian component as a whole, and the census showed that roughly 9,000 of the 9,500 cane farms in the Colony were owned or leased by Indians. In addition to the cane farmers themselves there was a total of 8,900

Industry group	Age range (years)		
	15 — 29	30 — 44	45 — 59
Agricultural and Pastoral	521	537	536
Other Primary	17	20	15
Secondary	103	104	81
Construction	99	77	64
Commerce	72	78	96
Transport and Communications	39	49	31
Other Service	9	8	9
Entertainment and Personal Service	19	26	37
Administration and Government	13	11	7
Professional and Allied Workers	24	33	19
Miscellaneous†	42	28	25
Economically Active	959	971	921
Male Population (nos.)	21,046‡	12,044	5,132

† Chiefly not specified industries.

‡ Excluding persons attending school.

Table 13. The numbers per 1,000 Indian males within each age range who were engaged in each major group of industries in September 1956.

other workers on cane farms, from which it can be deduced that on the average, these cane farms are already small enough to be worked by less than two adult males, even when the cane is being cut. The best that the existing cane farms can offer as a future source of employment would appear to be a one to one replacement of the numbers retiring. The classification of the population by industry in conjunction with age indicates that within the five-year period 1956-61, less than 1,500 of the older Indian males will retire from the cultivation of cane, and between 1956 and 1966 the number retiring will be in the vicinity of 2,000. These represent one-eighth and one-twelfth respectively of the numbers of Indian males who will be entering the work force in the same periods.

It is unlikely, therefore, that the cultivation of sugar-cane on the land which is already available to them will continue to be the predominant economic activity of Indian males, and in seeking alternative employment they will be competing with males from the other component populations. The only conclusion possible from all these data is that, given present conditions, it might well have been difficult in recent years for males of marriageable age to find remunerative employment, and that if these conditions continue unchanged, it will become more and more difficult as time goes by. Theoretically at least, this should act as a deterrent to early marriage, and it might well

constitute a counterbalance to the effect that the lack of marriageable females in the population would have on the age at marriage.

5. *The Interval Between Marriage and First Birth*

The next point which could be checked in a general fashion is the interval between marriage and first birth. In Fiji both the date of marriage and the rank of birth are recorded on the birth registration form, although the only tabulation of births by rank relates to the ages of mothers. It would therefore be necessary to compile the classification of first births according to year of marriage directly from the birth registration forms. At best the information on the interval between marriage and first birth will be imprecise because some of the births which occur in any one year may result from marriages in the previous year, and these marriages might be of less than one year's duration at the time of the birth or they might precede the birth by nearly two years. However, one could assume that the distribution of marriages throughout a year would not vary much from year to year, so that one could make valid comparisons between the distributions of first births registered in each of two or three 3-year periods, the births being classified according to the number of years separating the calendar year in which the marriage occurred and the year in which the birth was registered.

Starting with the year 1935, for example, the summarizing lay-out would resemble that in Figure 13, and all first births which were registered in 1935 would be scored according to the calendar year of marriage recorded. The first births registered in 1936 and 1937 would be tabulated in the same way and these three summaries would be combined into one table covering the whole period. The figures in brackets after each of the years of marriage in Figure 13 indicate the difference between the calendar year of marriage and the calendar year of the first birth, and the summary for the period 1935-37 would show the total number of first births which occurred in the same calendar year as marriage, the number which occurred in the calendar year following marriage, the number which occurred in the second calendar year after marriage, and so on. Similar summaries would then be compiled for the first births registered in the periods 1945-47 and 1955-57, and three fictitious distributions, which probably resemble the sort of figures one would get, are shown in Table 14.

To compare the three distributions one with another, the arithmetic mean and standard deviation of each distribution, and the standard error of each mean would then be computed. The arithmetic mean here measures the average number of years separating the year of marriage and the year of first

Year of Marriage	First births of 1935	Total Number of births
1935 (0) 1934 (1) 1933 (2) 1932 (3) 1931 (4) 1930 (5)		
1936 (0) 1935 (1) 1934 (2) 1933 (3) 1932 (4) 1931 (5)	First births of 1936	
1937 (0) 1936 (1) 1935 (2) 1934 (3) 1933 (4) 1932 (5)	First births of 1937	

FIG. 13. A specimen skeleton table for summarizing the classification of first births registered in the years 1935-37 according to the calendar year of marriage of parents.

No. of years between year of marriage and year of 1st birth	First Births of		
	1935-37	1945-47	1955-57
0	275	373	426
1	1,115	1,711	2,438
2	304	512	732
3	56	134	193
4	31	30	57
5	14	5	34
6	—	—	2
7	1	—	—
8	—	—	1
Total first births	1,796	2,765	3,883

Table 14. The fictitious distributions of first births according to the number of years separating the calendar year of marriage of mother and the year of first birth in each of three periods.

birth, and the standard deviation measures the spread or dispersion of the distribution around this mean. The computational procedure is illustrated in Table 15 for the fictitious distribution of first births registered in the period 1935-37. Columns (1) and (2) are the same as the first two columns of Table 14, and the mean is obtained by multiplying the corresponding entries in these two columns together, summing them to the total (2,092) and then dividing by the total number of first births registered. This gives a mean value of 1.165 years separating the calendar years of marriage and first birth.

(1) No. of years	(2) No. of 1st births	(3) (1) × (2)	(4) Square of (1)	(5) (2) × (4)
0	275	—	0	—
1	1,115	1,115	1	1,115
2	304	608	4	1,216
3	56	168	9	504
4	31	124	16	496
5	14	70	25	350
6	—	—	36	—
7	1	7	49	49
8	—	—	64	—
Total	1,796	2,092		3,730
x	n	Σx	x ²	Σx ²

$$\text{Mean number of years} = 2,092/1,796 = 1.165$$

$$\text{Variance of distribution} = [3,730 - (2,092)^2/1,796]/1,795$$

$$= 0.720455$$

$$\text{Variance of mean} = 0.720455/1,796$$

$$= 0.000401$$

Table 15. The computational procedure for calculating the mean and standard deviation of the distribution of first births registered in 1935-37 according to the number of years separating the calendar year of marriage and the year of first birth.

The calculation of the standard deviation is a little more complicated, because it involves first squaring each number of years in Column (1) and forming the sum of the products of these squares with the corresponding numbers of first births as in Column (5). The square of the standard deviation, which is the variance of the distribution, is then obtained by subtracting from the total of Column (5) the product of the total number of first births (1,796) with the square of the mean (1.165)², and dividing by a number one less than the total number of first births (1,795). Algebraically this would be written

$$\text{Var.} = (\Sigma x^2 - n\bar{x}^2)/(n - 1)$$

where x is the general form for the number of years separating the calendar years of marriage and first birth, $\bar{x}$ is the mean number of these years, and n

is the total number of observations. Since $\bar{x} = \Sigma x/n$, the formula for the variance can be written

$$\text{Var.} = [\Sigma x^2 - (\Sigma x)^2/n]/(n - 1)$$

This latter leads to a slightly more accurate estimate of the variance, and the variance formula shown at the foot of Table 15 is in this form. The variance of the mean is equivalent to the variance of the distribution divided by the total number of first births, and since the distribution of first births of 1935-37 has a variance of 0.720455, the mean of the distribution has a variance of 0.000401. Taking the square root of each of these numbers, the standard deviation of the distribution is 0.849, and the standard error of the mean is 0.020. The means and the variances of the fictitious distributions of first births registered in 1945-47 and 1955-57 can be derived in the same way, and their values, together with those given above for the period 1935-37, are shown in Table 16.

Parameter	1935-37	1945-47	1955-57
Mean	1.165	1.187	1.262
Variance	0.720455	0.605764	0.726471
Standard deviation	0.849	0.778	0.852
Variance of mean	0.000401	0.000219	0.000187
Standard error of mean	0.020	0.015	0.014

Table 16. The means and variances of the distributions of first births shown in Table 14.

From this it appears that, were these figures real, the mean interval separating the calendar years of marriage and first birth increased between 1935-37 and 1945-47, and again between 1945-47 and 1955-57. The difference between the first two means is 0.022 years, which barely exceeds the standard error for the mean of the 1935-37 distribution, and this difference cannot therefore be regarded as indicating a statistically significant lengthening of the interval between the calendar years of marriage and first birth. The difference between the means for 1945-47 and 1955-57 is 0.075 years, and this is approximately six times the standard error of either mean. Such a difference is probably greater than would occur by chance, but to make sure that it is, it could be tested for significance by means of a *t*-test. The *t*-test was designed to test the significance of the difference between the means of small samples, and although the numbers of births in these distributions are not small, the principle of the *t*-test can still be applied.

The observed value of *t* is the ratio between the difference in magnitude of the two means and the standard error of this difference, this latter being the

square root of the sum of the variances of the two means being tested. The theoretical value of t varies according to the number of degrees of freedom in the two samples from which the means were calculated, and the number of degrees of freedom is two less than the total number of observations. Provided that the number of degrees of freedom is greater than 120, there is no need to refer to the statistical table of the t -distribution to assess the significance of the difference between the two means, and if t is equal to or greater than 2.58, the difference between the two means is such that only once in 100 times would a difference of this magnitude or greater occur by chance. The 5 per cent level of probability coincides with a value of t equal to 1.96, and a calculated value of t which lies within the range 1.96 to 2.58 indicates a chance of 5 or less in 100 that such a difference between the two means would occur by chance. In this example, $t = 0.075 / \sqrt{0.000219 + 0.000187} = 0.075 / 0.020 = 3.75$. With considerably more than 120 degrees of freedom, it is therefore unlikely that the difference of 0.075 years between the two means occurred by chance, and it is probable that the interval between the calendar years of marriage and first birth was longer in the period 1955-57 than it had been in the period 1945-47.

If one wished, one could tabulate all second births registered in these periods in the same way, and compute the mean interval at each period between the calendar year of marriage and the year of the second birth. On the average, these second births would relate to marriages of longer duration and of an earlier date than were the marriages which sponsored the first births, and one could therefore deduce nothing concerning the interval between first and second births from the comparison between the two sets of data. Nor would an increase in the average interval between the calendar years of marriage and second births between 1945-47 and 1955-57, for example, help much for this could mean either that the first birth had been delayed for these women as for the mothers whose first births had been registered in 1955-57, or that there was a longer interval between first and second births, or that both intervals, that between marriage and first birth and between first and second births, were longer in 1955-57 than in 1945-47. In view of such difficulties in interpretation, the labour of tabulating second births or births of any higher rank according to the calendar year of marriage could easily be conserved without prejudice to the quest for knowledge.

6. Summary of Preliminary Investigations

All in all, this probably exhausts the information which could be obtained from the census and registration data on the changing pattern of fertility in

the Indian component, and although several of the initial questions remain unanswered, the information we have got is by no means negligible. There is some indication that age at marriage has risen in recent years, although one might feel sceptical about this apparent change unless it has been confirmed directly from the tabulations of age at marriage from the marriage registers. Accepting the fiction of Table 14, there is also some indication that the interval between marriage and first birth has lengthened, although more precise information on this would be desirable. Nothing at all is known about the intervals between subsequent births, nor about the marriages which had produced no births during the three 3-year periods examined. Completely unknown too are the differences in several respects between the populations living in urban areas and those in the rural parts of Fiji. There is indirect evidence that there may have been some economic stringency in recent years which would encourage the sort of changes thought to have occurred, and on the face of it, this pressure is likely to increase in the future and may induce still further changes in the pattern of fertility in this component. If it was proposed to investigate this topic further by means of a sample survey, we begin to know the sort of questions worth asking and the sort of answers we could expect to get.

7. The Objectives of the Survey

The main objective of the survey would be to discover whether or not there has been any change in the pattern of female fertility in the Indian component, and if there has been a change, how and why this has occurred. The pattern of fertility can be assessed only by comparing the behaviour of cohorts of women born at different times, and the behaviour of each cohort will be revealed by the answers to questions concerning their present marital status, and for those who are or have been married, their age at marriage, the duration of the marriage, and the spacing of children after marriage, that is, the interval between marriage and first birth, and between each pair of consecutive births. If it were possible to elicit the information, it would be preferable to record full-term pregnancies rather than births, noting the outcome of each pregnancy and the subsequent fate of each live birth.

Only full-term pregnancies could be considered, because a pregnancy which terminates in a miscarriage after only three months' gestation may be followed by another pregnancy within the next six months, and this is a risk which would not be run had the first pregnancy gone to term. Further, if all pregnancies were recorded, the interval between any two consecutive pregnancies would have to be measured from the commencement of one pregnancy to the

beginning of the next, which would mean antedating their dates of termination by estimates of the probable gestation period. By considering only full-term pregnancies, the interval between conceptions is assumed to be the same as the interval between their dates of termination, and one is probably on safer ground with dates of termination of pregnancies than on dates of conception. While it is probable that such details would be less well-remembered by the older women whose reproductive life was past than they would be by the younger women who were still in the reproductive ages, it might be worth while to aim for the record of pregnancies as distinct from live-births at least until a pre-test or the pilot survey indicated that the time spent in collecting these details was not justified by the results.

The picture which is to emerge from the information to be collected is one which shows the number of pregnancies which each of a group of women who married at a particular age in a particular year have had at the end of each segment of their married life. Any pattern of fertility which is dependent on marriage is likely to be disrupted by the death of a husband or the termination of the marriage by divorce before the wife has passed out of the reproductive period, and even if she remarries, there is likely to be an interval of time during which the likelihood of her having any further births is less than it would otherwise be. Not all cohorts of women will be equally affected in this respect, and it would have to be decided at the outset whether to consider only first marriages or whether to include all marriages, making due allowance for interruptions in the period of exposure to the risk of pregnancy. This latter procedure will introduce difficulties into the coding and possibly also the tabulating of the data, and since the derivation of the main tabulation tracing each group of women throughout each year of marriage is likely to be somewhat complicated, careful consideration would need to be given to elaboration beyond this.

Apart from these tabulations concerning the pregnancies of women who are or have been married, there will also be the more general ones concerning the numbers in each birth cohort who married at specified ages, the numbers in each birth cohort who had births of each rank before some particular age irrespective of their marital status, and the same sort of classification as in the census of women according to their present age in conjunction with the numbers of children borne by them before the date of the survey. Although the survey aims to collect more detailed information from a sample of the population than was possible in the census, the representativeness of the sample needs to be checked, and this can be done only by comparing parallel data relating to both the sample and the population as a whole.

8. *The Coverage of the Survey and the Design of the Sample*

The population to be covered by the survey has already been decided by the topic, and the respondents in the survey will be a sample of the population of females aged 15 years and over. Assuming that the subsidiary tabulations in regard to the number and timing of children born to such females before the census in 1956 had shown differences between the behaviour of the females enumerated in urban centres and those from the rural areas, one of the first things to do in planning the sample investigation would be to stratify the whole of Fiji into areas which could be described as rural, or urban, or on the borderline between the two. The enumerator's areas which were used for the census in 1956 covered the whole territory, and they could therefore form the basis for this stratification. By defining what was meant by each of the labels, the areas to be assigned to each category could be distinguished relatively easily, and these groups of enumerator's areas would then constitute the sampling frames for the various strata. Again some decision would have to be taken in regard to the rural areas bordering on the urban areas; they would be relatively fewer in number than either the purely rural or purely urban areas, and if they were to be included as a separate stratum, the sampling fraction which was to be applied to them would need to be large enough to provide sufficient females aged 15 years and over to permit of valid comparisons between their behaviour and the behaviour of females from the other two strata.

The use of the enumerator's areas as sampling frames would have the additional advantage that the number of households within each at the time of the census was known, and these would be an invaluable guide in estimating what would be the most desirable sampling fraction to apply to each stratum. The sampling fractions themselves will be decided by many factors but even at this early stage, the ratios between the sampling fractions for the various strata can be decided by rough calculations based on the number of enumerator's areas in each stratum, and the average number of Indian households in areas within strata. Suppose for the sake of argument that there were 100 urban areas, 200 rural areas and 45 borderline or intermediate areas, with an average of 250 households in the urban areas, 100 households in the rural areas, and 150 households in the intermediate areas. There were therefore roughly 25,000 households in urban areas, 20,000 households in rural areas and 6,750 households in intermediate areas. Assuming that there is no difference between the strata in respect of the average number of females aged 15 years and over per household, samples of equal size from the three strata would require sampling fractions in the ratio of 4:5:15 for urban, rural and intermediate areas respectively, because $4 \times 25,000 = 5 \times 20,000 = 15 \times 6,750$

approximately. A 4 in 50 sample of the urban stratum, for example, would yield 2,000 households, and the same number of households from each of the other strata would be given by sampling fractions of 5 in 50 rural areas and 15 in 50 intermediate areas ($15 \times 45 \times 150/50 = 2,025$).

9. *The Size of the Sample*

The census showed an average of 6.46 persons per household in the Indian component in 1956, and 5.11 persons per family unit. The basic family unit consisted of a married couple, either with or without offspring, and in the editing of the schedules, the general rule was to separate out the youngest family units within each household first, leaving the head of the household, his wife and their unmarried children, together with all unaccompanied servants, boarders and visitors as the one remaining unit. Provided they were accompanied by one or more of their offspring, widowed and divorced persons or those separated from their spouse on census night, who were present in another's household, constituted distinct family units within that household. On the other hand, aged parents of either the head or his wife, who were clearly dependent on the head, did not constitute a separate family unit unless they too were accompanied by other offspring. Of all the Indian households enumerated, 80 per cent consisted of one family unit only, 15 per cent contained two family units and the remaining 5 per cent contained three or more family units. One could therefore assume that 80 per cent of households contained at least one female aged 15 years and over, 15 per cent contained at least two such females, and 5 per cent contained three or more. A sample of 2,000 households could then be expected to yield at least $1,600 + 600 + 300 = 2,500$ females aged 15 years and over.

These 2,500 would be distributed over the range of age groups in much the same way as was the total population of Indian females aged 15 years and over, and starting with the age group 15-19 years, the numbers in each five-year age group up to 75 years and over would be 600, 450, 358, 288, 248, 175, 113, 80, 51, 61, 30, 23, 23. These numbers represent the numbers one would expect to get in each age group in a random sample of 2,500 females aged 15 years and over, and the question then to be decided is whether or not the numbers in each birth cohort will be large enough to fulfil the objectives of the survey. It could be, for example, that the numbers of women just past the reproductive period—113 at ages 45-49 years, 80 at ages 50-54 years and 51 at ages 55-59 years—are too small for the very detailed analysis proposed; on the other hand, to increase their numbers by roughly 50 per cent would mean enlarging the whole sample by 50 per cent, and the additional costs involved

in this would have to be weighed against the possible gain in precision in all the estimates concerning the behaviour of this very important group of women of completed fertility. This would be a job for the statistician who, with the background of the census data and skeletons of all tabulations proposed and their analysis mapped out, would then be able to decide on the optimum size of sample, reconciling this with the cost of administering the questionnaire.

10. The Design of the Questionnaire

By this time the contents of the questionnaire would be fairly clear, because the catalogue of questions needed to collect the fertility history of the women would be relatively straightforward. If the enquiries into economic aspects were to be pursued, certain questions would have to be asked of or about the husbands. Questions which would be relevant in this context would be their age and occupation at marriage, and any changes in occupation which they may have made in the course of their married life. If there is any association between economic status and the fertility pattern, it is the economic status at marriage and throughout the years when the children are being born which is important: present occupation or economic status is relevant only if the wife is still capable of reproduction. It would probably not be easy to devise a scale or a set of criteria by which to measure economic status so that there is comparability throughout the whole period of time between the earliest marriage recorded and the most recent. However, it might be that some simple criterion, combining continuity of employment or income with an initial status category, would suffice: certainly, any elaborate classification would require a very large sample if it was to be analysed fully in conjunction with the age of the wife and the duration of the marriage.

A questionnaire relating to fertility which could hardly be bettered for its arrangement and ordering of questions was that used in the Family Census of 1946 of Great Britain. This census was instigated by the Royal Commission on Population, and the questionnaire was completed by or for more than a million married women. Simple and short though the questionnaire was, it was sufficiently comprehensive to permit a very sophisticated analysis of the trend and pattern of fertility in Britain. The questions were arranged in three blocks headed 'Yourself', 'Your Children', 'Your Husband'. The first block contained questions concerning present marital status and dates of birth and marriage of the respondent. The second block provided spaces in which to record the month and year of birth of each child born alive in the order of their birth. The questions concerning the nature of the husband's occupation, which were the only questions asked about husbands, were used

to assess the social class of the respondent. The questions to be asked in the supposed survey of Indian women in Fiji could be fitted into this pattern. Being on a much smaller scale than the British survey and using enumerators to complete the questionnaires, the questions could be amplified while retaining the same order and general arrangement.

If it were decided to record pregnancies instead of live births, details concerning the outcome of each pregnancy and the subsequent fate of children born alive could be inserted alongside their date of birth. The most cogent reason for collecting such information would be to estimate the still-birth rate and the rates of mortality in infancy and childhood. Since this is of particular importance for populations in which there is no system of registration operating, the method of estimating such rates from the fertility histories of the females in the population will be explained later.

If the exact intervals between marriage and first birth and between any pair of consecutive births were to be tabulated, these would need to be computed and marked on the questionnaires during the coding process. This would lead to more precise estimates of the length of intervals than were possible in the British Family Census, when coding facilities were restricted and the coding of intervals between marriage and successive births was avoided by an ingenious tabulating programme which measured the intervals between the calendar years in which each pair of events occurred. The risk of error in manually computing the number of months separating any two events would be large, and the advantages of the more exact intervals computed by a coding staff as compared with the automatic estimation of intervals would have to be viewed in relation to the capacity of the coding staff available and the method of tabulation proposed. This latter would also determine the type and design of the tabulating card, and the programme of summarizing routines needed to compile the tabulations required from the cards.

If the estimation of mortality during infancy and childhood was to form part of the survey, it could be that one tabulating card per questionnaire would not be sufficient, and that subsidiary or 'trailer' cards would be needed to record some of the particulars about the fate of pregnancies and the children born alive. These would need to be cross-referenced to the original card, even though most of the tabulations to be compiled from them would be independent from those of the women.

11. Field Operations

The time needed to complete a questionnaire would depend largely on the number of pregnancies a woman had had, and on her remembrance of their

dates of termination. With a questionnaire of this nature, it would not be possible to reduce the average time needed for an interview by eliminating some of the questions, and in the initial planning, one would need to be fairly generous in the time allowed for each interview. Other factors which would need to be taken into account in assigning some number of households to an enumerator would be the pattern of settlement and the habits of the women in each type of area. In the rural areas dwellings are dispersed and many of the women would probably be working in the fields and away from their dwellings on most days. In the urban areas the houses are closer together and the women could probably be found in or around the houses when they were not out shopping or visiting. Since such absences may be quite erratic it is perhaps as well that 'call-backs' would be more feasible in urban areas than in rural areas; nevertheless, it would be necessary to stipulate the maximum number of call-backs required in any area.

Having decided how many enumerators were needed to cover all households in the sample, the selection of suitable persons and the programme of training would be the next problems to be faced. With such a simple questionnaire, instruction on how it should be completed would not be difficult, nor would its completion be in practice, provided the respondents knew the answers required of them: the real difficulty here would be in training enumerators to probe for the information, and to check for internal consistency while they were recording it. Some pre-tests or preliminary interviews might be revealing in regard to the ease with which the details of pregnancies could be obtained, and the devious ways and means of encouraging respondents to remember. It would be preferable to have all such intricacies in interviewing ironed out before the pilot survey is run, and this would be designed to cover households in at least one urban and one rural area, using more than one enumerator in each. From the experience gained and the results obtained from this, the survey plans would then be modified or amended, reconciling all the while the costs of the operation as planned with the money available to pay them.

12. Estimates of Mortality in Infancy and Childhood

The main object in recording all full-term pregnancies and their outcome in a survey of this nature in Fiji would be to compile estimates of the levels of mortality in rural areas as distinct from urban areas. These are not readily available from the registration data, and by appropriate weighting, such independent estimates could be used as a check on the completeness of the registration data. There would be even more point in the collection of this

information in populations for which there were no records of births and deaths, or in small populations for which the records did not extend for more than a few months. An anthropologist working in a village or a group of villages might record the numbers of births and deaths which occurred during his stay, but because of the small numbers of such events which are likely to occur in a small population in the course of a year or less, these data would have relatively little import compared with other data which he might collect. By collecting histories of pregnancies and their outcome from each of the women in the village, recording the approximate date of occurrence of those pregnancies which terminated in live births, and if a child did not survive, its age at death, then it is possible to form some estimate of the level of mortality amongst children at different ages in that village. With the year of birth known, all live births are then arrayed in birth cohorts, with the deaths which occurred amongst each cohort at each year of age ranged alongside as in Table 17. These are imaginary figures to depict the sort of data which could be collected from a village population which is under observation in the year 1960, and they show the total number of births, irrespective of sex, to women of all ages in that village in the calendar years 1945-59.

Estimates of the infant mortality rates in two 5-year periods and one period of 4 years can be derived from these. In the first period there were 143 births and 18 infants died before attaining one year of age, so that the infant mortality rate amongst them was $18/143$. With 18 dead, only 125 started on their second year of age, and in the course of this year, 5 of these died. The mortality rate at 1 year of age was therefore $5/125$, and there were 120 survivors who started on their 3rd year of age. Each successive rate is computed by referring the number of deaths at each age to the population at risk of dying during that year, and the population at risk of dying in any year of age is the population alive at the beginning of that year. Starting with the total number of children born, the numbers of persons born in the period 1945-49 who survived each year of age up to 9 years are 125, 120, 117, 116, 116, 115, 115, 114, 113, and these are the denominators for the mortality rates at ages 1, 2, 3, . . . 8, 9 years, the numerators of the rates being the number of deaths listed for each age.

When we come to the mortality rate at age 10 years, not all of the children born in 1949 will have run the full risk of mortality during this year of age; some will attain the age of 10 years towards the end of 1959 and must survive through to their 11th birthday before the risk of dying at the age of 10 years has been passed. The 23 survivors of the cohort of infants born in 1949 must therefore be excluded from the estimate, and the mortality rate at age 10 years

Year of Birth	Total no. of births	Numbers who died at ages —															
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
1945	25	3	1	1	—	—	—	—	—	—	1	—	—	—	—	§	
1946	30	4	2	—	—	—	—	—	1	—	—	—	—	—	§		
1947	32	2	1	1	1	—	1	—	—	—	—	1	—	§			
1948	27	5	—	1	—	—	—	—	—	1	—	—	§				
1949	29	4	1	—	—	—	—	—	—	—	1	§					
1945-49	143	18	5	3	1	—	1	—	1	1	2						
1950	30	3	2	—	1	—	—	—	1	—	§						
1951	35	5	1	1	—	1	—	—	—	§							
1952	25	4	1	—	—	—	—	—	§								
1953	27	2	2	—	1	—	—	§									
1954	28	3	1	1	—	—	§										
1950-54	145	17	7	2	2	1											
1955	30	3	1	1	—	§											
1956	31	4	2	—	§												
1957	26	2	2	§													
1958	27	3	§														
1959	30	§															
1955-59	144																
1945-59	432																

§ Mortality experience incomplete for these cohorts.

Table 17. The incidence of deaths in each year of age amongst cohorts of births in a population sample.

will be based on the 111 survivors to age 10 years minus the 23 whose mortality experience at this age is incomplete. One person died at this age, and in computing the mortality rate at age 11 years, this 1 and the 20 survivors of the 1948 cohort must be excluded from the denominator. There are no deaths shown at ages 12 and 13 years, but if there had been deaths at these ages, the denominators for the rates would have been respectively 41 and 19 or less if there had been any 12-year-old deaths amongst the 1945 cohort.

Amongst the births of the second period 1950-54, the mortality rates at ages 0, 1, 2, 3, and 4 years can be calculated as above for the 5 birth cohorts combined, and it is only from the age of 5 years onwards that the survivors of the youngest cohort need be excluded progressively from the population at risk. In computing the infant mortality throughout the most recent period, the births which occurred in 1959 must be excluded because some of the children born in 1959 would still be less than 1 year old in 1960. This rate would therefore be based on 114 births and not on the total 144 births shown, and the denominator for the rate at each other age for this group would need progressive adjustment similar to that used for the rates at ages 10-13 years in the first group and at ages 5-8 years in the second. The standard error of each of these rates would be computed concurrently with the rates, and the computing lay-out for the calculation of the rates, variances and standard errors is shown in Table 18. The population at risk and the deaths which occurred amongst them are derived from Table 17 and are repeated in Table 18 for convenience in computing the rates and variances. It will be remembered that the variance is the square of the standard error, and the formula for the derivation of the variance of a rate p is $p(1 - p)/n$, where n is the population at risk. The true risk of mortality at each age in this village population will lie somewhere within the range of $p \pm 3s$, p being the age-specific rate and s the standard error of this rate.

With numbers as small as these, the data should be handed over to a statistician for tests of significance because tests based on such small numbers would involve the computation of χ^2 either with the correction for continuity or by the exact method. If there was no difference between comparable rates for the three groups, the deaths and the populations at risk at each age could be combined into a single group which will provide more reliable estimates of the age-specific rates. Summing the births and infant deaths over the 14-year period, for example, the infant mortality rate becomes $47/402 = 0.117$, which has a variance of 0.000257 and a standard error of 0.016. This average rate has a standard error which is little more than half the magnitude of the standard errors for the rates in the smaller groups. As an exercise, enthusiasts

Age in Years	1945-49 Cohorts					1950-54 Cohorts					1955-59 Cohorts				
	No. of deaths	Pop'n at risk	Mort. rate	Variance of rate	S.E.	No. of deaths	Pop'n at risk	Mort. rate	Variance of rate	S.E.	No. of deaths	Pop'n at risk	Mort. rate	Variance of rate	S.E.
0	18	143	0.126	0.000769	0.028	17	145	0.117	0.000714	0.027	12	114	0.105	0.000826	0.029
1	5	125	0.040	0.000307	0.018	7	128	0.055	0.000404	0.020	5	78	0.064	0.000769	0.028
2	3	120	0.025	0.000203	0.014	2	121	0.017	0.000134	0.012	1	51	0.020	0.000377	0.019
3	1	117	0.009	0.000072	0.009	2	119	0.017	0.000139	0.012	—	25	—	—	—
4	—	116	—	—	—	1	117	0.009	0.000072	0.009					
5	1	116	0.009	0.000074	0.009	—	93	—	—	—					
6	—	115	—	—	—	—	71	—	—	—					
7	1	115	0.009	0.000075	0.009	1	51	0.020	0.000377	0.019					
8	1	114	0.009	0.000076	0.009	—									
9	2	113	0.018	0.000154	0.012										
10	1	88	0.011	0.000128	0.011										
11	1	67	0.015	0.000220	0.015										
12	—	41	—	—	—										
13	—														

Table 18. Illustration of the computation of mortality rates and the standard errors of these rates.

for arithmetic could work out the mortality rates for the combined populations at all other ages shown, there being no significant differences between the groups in respect of mortality at any age.

13. *Cautionary Comments*

In general, the analysis and interpretation of data collected from small samples is better left to a statistician who understands the limitations imposed on the data themselves and on the various tests of significance by the small numbers of observations. He may quite justifiably refuse to do the arithmetic involved, but in this case he will advise on the general procedure to be followed and assess the significance of the results. It is, however, advisable to seek his advice *before* collecting the data to ensure that the sort of deductions you hope to draw will be valid and that the details you propose to collect will be meaningful. Estimates of fertility, for example, which might be derived from the analysis of genealogies stretching back through several generations are apt to be biased because those women of older generations who had large numbers of children are more likely to have descendants amongst the present-day informants than are those who had few, and childless women in early generations are likely to be forgotten completely. Hence such data are weighted in favour of women with many children, and the average number of children per woman in the various generations is likely to be distorted. Usually one would be unable to guess the extent of this distortion, and it is therefore better to exploit what can be discovered about the behaviour of the people who are living, in this case by comparing the behaviour of different cohorts of women as they passed through their reproductive cycle.

The essential demographic data required about population samples are simple enough: the classification of persons by sex and age in 5-year age groups (or in the broader physiological age groups if this finer grouping is not practicable); for the women in the sample, the number of children borne by them up to their present age and their age at the birth of each child, the number of children surviving to each woman at the time of the survey and the age at death of those who failed to survive. Advisable though it would be to seek the advice of the statistician before trying to analyse such data, there are certain comparisons which can be made, if not tested, between the sample and the population from which the sample is drawn, or if nothing is known about that population, between the sample and some other population which might be considered comparable.

Let us suppose that the people in the sample are divided into categories according to some characteristic such as age, or the women in some age group

according to the number of children they have borne. Given the corresponding data about some population, one can compute the numbers expected in each category in the sample if there was no difference between the sample and the population in respect of this characteristic. This is no different from the procedure of calculating the number of deaths expected in one population if the mortality risks experienced by it were the same as the risks experienced by another, nor from the estimation of the numbers expected in each age group in a sample of size 2,500 from the Indian component of Fiji's population. The expected frequency of some characteristic or event is derived from the population data, and this is then applied to the relevant section of the sample to give the number one could expect to find possessing this characteristic, or the number of events which could be expected to occur, in a sample of that particular size. A χ^2 -test will pick out significant deviations between the observed and expected numbers in the sample, and indicate the statistical significance attaching to them.

Such comparisons between samples and populations are of more than academic interest because there are now several populations in the Pacific whose age structure and patterns of mortality and fertility are known reasonably accurately, and while the numbers in the sample may be too small to provide direct evidence on demographic trends, certain inferences about these may be drawn merely from the resemblance of the sample to any one of these 'known' populations. Still better evidence of trends will of course be given by comparisons between data relating to the same population sample at two different epochs, particularly if these epochs are separated by a period of five or ten years. With a register of the persons comprising the sample on the initial visit, and records of all arrivals and departures, either by natural causes or migration, which have occurred in the years separating the visits, one can construct a relatively complete picture of the patterns of mortality and fertility which prevailed in that population during the period. Those persons who survived the intervening years will be that number of years older; those who failed to survive will provide the basis for an estimate of the mortality risk experienced by the population at each age. The children born since the first visit will be easily identified and they, plus the live-born children who failed to survive, will reveal more about the pattern of fertility and its variations than it is possible to deduce from a single census. This is the essence of the study of populations—knowing the people who are living within some geographic area at some particular time, knowing them well enough to know their potential, and then to note its expression over the ensuing years. The statistics of populations are merely summary records of the potential of the

aggregate of individuals comprising each population and the changes which are being made in that aggregate by its members. While the study of population statistics may seem less engrossing to many people than other investigations which might be made of human populations, few of these can dispense with a demographic analysis of the population concerned.

APPENDIX A:

A TABLE OF THE LIMITS OF VARIATION ACCEPTABLE AT THE 5% AND 1% LEVELS OF PROBABILITY IN SAMPLES OF SIZE N WHEN IT IS EXPECTED THAT $n_1 = n_2 = \frac{1}{2}N$.

5% level of probability		Size of Sample	1% level of probability	
n_1	n_2	N	n_1	n_2
8	2	10	9	1
14	6	20	15	5
20	10	30	22	8
26	14	40	28	12
31	19	50	34	16
37	23	60	39	21
42	28	70	45	25
48	32	80	51	29
53	37	90	56	34
59	41	100	62	38
69	51	120	73	47
80	60	140	84	56
91	69	160	95	65
101	79	180	106	74
112	88	200	117	83
123	97	220	128	92
133	107	240	138	102
144	116	260	149	111
154	126	280	160	120
165	135	300	171	129
175	145	320	181	139
186	154	340	192	148
196	164	360	202	158
206	174	380	213	167
217	183	400	224	176
227	193	420	234	186
238	202	440	245	195
248	212	460	255	205
258	222	480	266	214
269	231	500	276	224
295	255	550	303	247
321	279	600	329	271
346	304	650	355	295
372	328	700	381	319
398	352	750	407	343
424	376	800	433	367
449	401	850	459	391
475	425	900	485	415
501	449	950	511	439
527	473	1,000	537	463
782	718	1,500	795	705
1,037	963	2,000	1,052	948
1,292	1,208	2,500	1,309	1,191
1,545	1,455	3,000	1,564	1,436
1,799	1,701	3,500	1,819	1,681
2,052	1,948	4,000	2,074	1,926
2,306	2,194	4,500	2,328	2,172
2,559	2,441	5,000	2,583	2,417

APPENDIX B:

REFERENCES AND SUGGESTIONS FOR FURTHER READING

Statistics of Populations and Measures of Change:

- Jaffé, A. J. *Handbook of statistical methods for demographers*
(U.S. Bureau of the Census, Washington, 1951).
- Myers, R. J. 'Errors and bias in the reporting of ages in census data',
Trans. Act. Soc. of America, 41 (1940), 395.
(Reprinted in Jaffé, above)
- 'Accuracy of age reporting in the 1950 U.S. census',
Jour. Am. Stat. Ass., 49 (1954), 826.
- Dublin, L. I., Lotka, A. J. *Length of life* (Ronald, New York, 1949).
and Spiegelman, M. (A study of mortality)
- Glass, D. V. *Population policies and movements in Europe*
(O.U.P., Oxford, 1940).
(See appendix for measurement of fertility)
- Kuczynski, R. R. *The measurement of population growth*
(Sidgwick & Jackson, London, 1935).
- United Nations *The determinants and consequences of population trends*,
Document No. ST/SOA/Ser.A/17, Population Studies
No. 17, (New York, 1953).

Censuses and Surveys:

- United Nations *Handbook of population census methods*,
Vol. I: General aspects of a population census,
Document No. ST/STAT/Ser.F/5/Rev.I (New York,
1958).
- Moser, C. A. *Survey methods in social investigation*
(Heinemann, London, 1958).
- Yates, F. *Sampling methods for censuses and surveys*
(Griffin, London, 2nd ed., 1953).
- Glass, D. V. and Grebenik, E. *The trend and pattern of fertility in Great Britain*,
2 pts. (H.M.S.O., London, 1954).
- Fisher, R. A. and Yates, F. *Statistical tables* (Oliver & Boyd, London, 2nd ed., 1943).
- Kendall, M. G. and Babington Smith, B. *Tables of random sampling numbers* (C.U.P., Cambridge,
1939).

Continuous Records:

- United Nations *Handbook of vital statistics methods*,
Document No. ST/STAT/Ser.F/7, (New York, 1954).
- *Principles for a vital statistics system*,
Document No. ST/STAT/Ser.M/19, (New York, 1953).

INDEX

- Age, accuracy in reporting, 8, 25, 31-3, 59; at birth of first child, 57, 60-2, 130; at marriage, 49, 90, 109-11, 119; composition of a population with respect to, 3-6, 44, 130; definition, 4, 6; groups, 5, 6, 33, 111, 122; importance of, 3, 4; in Pacific territories, 24; median, 13, 14, (at marriage), 111; mortality risk with, 7, 16, 19, 34-48; of birth cohorts, 5, 25; pattern of fertility with, 51-6; physiological age scale, 5, 6; preferences for unit digits of, 33; reproductive, 4, 6, 16, 19-22, 49-62; structure of population, 11-22; summarizing of data with respect to, 90-103
- Arithmetic mean, 57-9, 116
- Birth cohorts, 5, 6; ageing, 7, 8, 31; attainment of motherhood, 60-2, 108; change in size, 25, 31-3; deaths, 126; fertility, 53, 55-9, 120; in French Oceania, 11; of Indians in Fiji, 31-3
- Births, balance between births and deaths, 2, 3; crude birth rate, 50; expected increases in Fiji, 22; interval between marriage and first birth, 114-19, 124; interval between successive births, 60, 124; occurrence of, 4, 15, 16, 18, 20-2, 49, 126; probability of males and females in samples, 9, 10, 133; records and registration of, 26-33, 51, 126; recovery in numbers of, 18, 19; reduction in numbers of, 16, 18, 19; requirements for, 4, 15, 16, 22, 49; sex, 8-10; tabulation, 89-91, 114
- Census enumerations, accuracy, 27-33; conduct, 63-106; scope, 23, 24, 74, 75
- Census schedule, design of, 76-8; *see also* Questionnaire
- Characteristics, of population, 1, 3, 15; personal (as basis for classification of population), 2, 3
- Chi (χ), 40, 41
- Chi-squared (χ^2), 40, 128, 131
- Code numbers, 94-6; *see also* Tabulation
- Cohort comparisons, 31-3; *see also* Birth cohort
- Deaths, balance between births and deaths, 2, 3; crude death rate, 37, 38; infant, 28; life table death rate, 48; occurrence of, 15, 16-18, 20-2; probability of in epidemics, 17, 18; records and registration of, 26-33, 126; tabulation by age and sex, 90; *see also* Mortality
- Dependency ratio, 12, 13
- Emigration, 2, 25; demographic effects, 21; *see also* Migration
- Epidemics, effects on population structure, 8, 11, 16-20; in Pacific, 16; of influenza (1918), 8, 11, 18-20, 32; of measles in Fiji, 16, 17
- Expectation of life, 48
- Family, Census of Great Britain, 123, 124; size of completed families, 25, 59; unit, 107, (among Fiji Indians), 122
- Fecundity, 6, 49
- Fertility, age-specific fertility rates, 51-3; association with economic status, 123; average number of children per woman, 57-9; changes, 54, 119; cohort fertility, 119, 120, (of Indians in Fiji), 55, 56, 108; measurement of, 49-59; rate, 50; 'total fertility', 52; variations in patterns, 51, 52, 60, 107, 108, 119, 131; *see also* Births; Reproduction
- Fiji, age at birth of first child, 60-2; census, 2, 29, 33, 107, 122; coding frame for industry and occupation, 85, 86; dependency ratios, 13; economic activity, 112, 113, 119; fertility rates, 51-6; increases in births, 22; influenza epidemic, 32; measles epidemics, 16-18; median age of males, 14; migration to, 20, 31, 32; mortality comparisons, 45-7; mortality of males, 42, 43; opportunities for employment, 112, 113, 119; population, 13, 22, (Indian), 31; proportions married, 109, 110; replacement ratios, 13; statistical summaries of births and deaths, 26; work force, 111-13

- French Oceania, dependency ratio, 13; epidemic of influenza, 19; fertility rates, 51-3; mortality rates of males, 34; population of, 7, 8, 11, 12, 19; replacement ratio, 12
- Immigration, 2, 25; to Fiji, 20, 31; *see also* Migration
- Influenza, *see* Epidemics
- Interviewers, allocation of, 79, 125; instruction and training, 80-2, 125
- Life expectancies, 47, 48
- Marriage, age at, 49, 90, 109-11, 119; changes in pattern of, 108-10; fertility of marriages, 53, 54; interval between marriage and first birth, 114-19, 124; tabulation of ages of brides and bridegrooms, 90, 111
- Measles, *see* Epidemics
- Migration, effect on population structure, 20; Indian, to and from Fiji, 31, 32; net, 28; records, 26, 27
- Mortality, age incidence of, 7, 16, 19, 34, 42, 131; age-specific mortality rates, 38-47, 126-30; changes in, 20, 21; from epidemics, 16-19; in childhood, 125, 126-30; in Fiji, 42, 45-7; infant, 42, 125, 126, 128; life table, 47; measures of, 37-48; risk of, 7, 34-6; standardized mortality rates, 43-7; *see also* Deaths
- Natural decrease, 2
- Natural increase, 2, 3, 27
- Parthenogenesis, 49
- Pilot survey, 67, 81, 82, 120, 125
- Population, accessions to, 2, 3, 15, 21; annual rate of growth, 51; 'at risk', 15, 16, 37, 38, 50; changes in, 1, 14-21; characteristics, 1, 3, 15; classification, 2, 3, 130; 'closed', 5, 6, 8, 25; composition (or structure), 3-8, 11-21, 130; definition, 1, 3, 5; economic aspects of Fiji's, 111-13; increase in, 3, 14-16, 18-22; life table population, 48, 53; losses from, 2, 3, 15, 21, 25; of Fiji, 2, 11, 13, 16-18, 20, 22, 29, 31-3, 44; of French Oceania, 7, 11-13, 19; of New Guinea, 2; of Samoa, 11, (Western), 44; potential of, 3, 4, 11, 131; proportions of total at each age, 11, 12; pyramid, 6, 7, 19; register, 23; reproductive capacity of, 15-22; sample, 2, 120, 130, 131; size, 3, 15; sources of change in, 2, 14-22; 'standard population', 44, 45, 53; structure, 3-8, 11, 14-21; unit of, 1, 68, 69
- Pregnancies, 49, 119, 120, 124-6
- Probability, levels of, 10, 118; of deaths at each age, 39, 47; of disruption of matings through epidemic mortality, 18, 19; of males and females in a sample of births, 9, 133; of surviving to each age, 47, 48
- Questionnaire, coding of, 78, 84, 85; design of, 74-9, 123; editing of, 84, 108
- Registration, systems, 26; tests for completeness, 28-30
- Replacement, of population, 50; ratio, 12, 13
- Reproduction, in monogamous societies, 4; in polygamous societies, 4; physiological capacity for, 6, 49; rates, 52, 53
- Reproductive period, duration in humans, 5, 6
- Samoa, fertility rates, 51-4; mortality rates of males, 34; population of, 11
- Sample, designs, 67-72; of births, 9, 10; of population, 2, 120, 130, 131; random, 2, 68; representative, 2, 67, 72; size of, 9, 72, 73, 122; surveys, 63-83, 119-25
- Sampling, area, 70; fractions, 68, 121, 122; frame, 69, 70, 121; multi-stage, 70; processes, 68-72; purposive, 71; quasi-random, 70; quota, 71, 72; simple random, 68; stratified, 68, 69, 71, 121
- Sex, chromosomes, 8; composition of a population with respect to, 3, 4, 8, 44, 130; determination of, in humans, 8; ratio, 8, 20, 21
- Sorter cards, 93-6, 104
- Standard deviation, 39, 115-17

- Standard error, as index of precision, 73; of a rate, 39, 43, 128, 129; of an arithmetic mean, 116, 117
- Statistical significance, conventional levels, 10; of differences between means, 117, 118; of differences between rates, 40, 41
- Surveys, *see* Sample surveys
- Tabulation, cards for, 78, 79, 82, 83, 85, 91-6, 105, 106, 124; methods of, 87-106, 114; numerical codes for, 85-7, 105; of first births and calendar year of marriage, 115; plans for, 74, 82, 105, 106
- t*-test, 117, 118
- Variance, of a frequency distribution, 116, 117; of a rate, 39, 128, 129; of an arithmetic mean, 117
- Work force, 4, (in Fiji), 111-13



A
HB 881
M 23

CANCELLED 152133 ✓

1627414



A.N.U. LIBRARY

Presented by
Publication
Committee
9.6.6
See

THE AUSTRALIAN NATIONAL UNIVERSITY
THE LIBRARY

This book is due on:

- 8 NOV 1967 - 2 OCT 1979

- 8 AUG 1968

10 MAY 1982

31 OCT 1968

25 OCT 1985

30 JUN 1969

25 NOV 1985

18 JUL 1966

27 OCT 1969

23 MAR 1970

27 AUG 1986

28 DEC 1970

16 DEC 1986

22 MAR 1971

19 JUL 1971

CANCELLED

6 NOV 1972

30 APR 1973

25 JUN 1973

1 JUL 1974

