DIVISION OF HEALTH STATISTICS
DISSEMINATION OF STATISTICAL INFORMATION

MANUAL OF
MORTALITY ANALYSIS

A MANUAL ON METHODS OF ANALYSIS
OF NATIONAL MORTALITY STATISTICS
FOR PUBLIC HEALTH PURPOSES

WORLD HEALTH ORGANIZATION
GENEVA, 1977
Reprint 1980
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FOREWORD</strong></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td><strong>CHAPTER I</strong></td>
<td>SOURCES AND COLLECTION OF DATA</td>
<td>5</td>
</tr>
<tr>
<td>I.1</td>
<td>Concerning Populations</td>
<td>5</td>
</tr>
<tr>
<td>I.2</td>
<td>Concerning Registration</td>
<td>12</td>
</tr>
<tr>
<td>I.3</td>
<td>Concerning Migration</td>
<td>18</td>
</tr>
<tr>
<td><strong>CHAPTER II</strong></td>
<td>ANALYSIS OF POPULATION DATA</td>
<td>20</td>
</tr>
<tr>
<td>II.1</td>
<td>Population Composition</td>
<td>20</td>
</tr>
<tr>
<td>II.2</td>
<td>Types and Sources of Error in Population Data</td>
<td>26</td>
</tr>
<tr>
<td>II.3</td>
<td>Detection and Correction of Errors</td>
<td>28</td>
</tr>
<tr>
<td>II.4</td>
<td>Estimates of Populations</td>
<td>31</td>
</tr>
<tr>
<td><strong>CHAPTER III</strong></td>
<td>MEASURES OF MORTALITY</td>
<td>34</td>
</tr>
<tr>
<td>III.1</td>
<td>Introduction</td>
<td>34</td>
</tr>
<tr>
<td>III.2</td>
<td>General Statistical Measures</td>
<td>35</td>
</tr>
<tr>
<td>III.3</td>
<td>Primary Measures of Mortality</td>
<td>37</td>
</tr>
<tr>
<td>III.4</td>
<td>Adjusted Measures of Mortality</td>
<td>41</td>
</tr>
<tr>
<td>III.5</td>
<td>Standard Errors of Mortality Rates</td>
<td>49</td>
</tr>
<tr>
<td>III.6</td>
<td>Other Measures Related to Mortality</td>
<td>56</td>
</tr>
<tr>
<td><strong>CHAPTER IV</strong></td>
<td>THE LIFE TABLE AND ITS CONSTRUCTION</td>
<td>58</td>
</tr>
<tr>
<td>IV.1</td>
<td>Introduction</td>
<td>58</td>
</tr>
<tr>
<td>IV.2</td>
<td>Description of the Life Table</td>
<td>59</td>
</tr>
<tr>
<td>IV.3</td>
<td>Construction of the Abridged Current Life Table</td>
<td>65</td>
</tr>
<tr>
<td>IV.4</td>
<td>Other Methods of Constructing an Abridged Current Life Table</td>
<td>67</td>
</tr>
<tr>
<td>IV.5</td>
<td>The Variance of the Life Table</td>
<td>71</td>
</tr>
<tr>
<td>IV.6</td>
<td>Value of Life Tables and Difficulties in Their Use</td>
<td>74</td>
</tr>
<tr>
<td>IV.7</td>
<td>Generation or Cohort Life Table</td>
<td>76</td>
</tr>
<tr>
<td>IV.8</td>
<td>Life Tables from Incomplete Data</td>
<td>77</td>
</tr>
<tr>
<td><strong>CHAPTER V</strong></td>
<td>FOETAL, INFANT AND CHILDHOOD MORTALITY</td>
<td>80</td>
</tr>
<tr>
<td>V.1</td>
<td>Introduction</td>
<td>80</td>
</tr>
<tr>
<td>V.2</td>
<td>Definitions</td>
<td>81</td>
</tr>
<tr>
<td>V.3</td>
<td>Foetal Mortality</td>
<td>83</td>
</tr>
<tr>
<td>V.4</td>
<td>Induced Abortion</td>
<td>83</td>
</tr>
<tr>
<td>V.5</td>
<td>Infant Mortality</td>
<td>85</td>
</tr>
<tr>
<td>V.6</td>
<td>Perinatal Mortality</td>
<td>93</td>
</tr>
<tr>
<td>V.7</td>
<td>Early Childhood Mortality</td>
<td>94</td>
</tr>
<tr>
<td>V.8</td>
<td>Levels and Trends of Foetal, Infant and Childhood Mortality</td>
<td>95</td>
</tr>
<tr>
<td><strong>CHAPTER VI</strong></td>
<td>MORTALITY ANALYSIS IN TIME</td>
<td>100</td>
</tr>
<tr>
<td>VI.1</td>
<td>Trend Analysis</td>
<td>100</td>
</tr>
<tr>
<td>VI.2</td>
<td>Long-term Trends by Age and Sex</td>
<td>101</td>
</tr>
<tr>
<td>VI.3</td>
<td>Short-term Trends</td>
<td>105</td>
</tr>
<tr>
<td>VI.4</td>
<td>Trends by Cause of Death</td>
<td>109</td>
</tr>
<tr>
<td>VI.5</td>
<td>Cohort Analysis</td>
<td>109</td>
</tr>
<tr>
<td>CHAPTER VII</td>
<td>DIFFERENTIAL MORTALITY</td>
<td>Page</td>
</tr>
<tr>
<td>------------</td>
<td>------------------------</td>
<td>-----</td>
</tr>
<tr>
<td>VII.1</td>
<td>Introduction</td>
<td>115</td>
</tr>
<tr>
<td>VII.2</td>
<td>Age</td>
<td>115</td>
</tr>
<tr>
<td>VII.3</td>
<td>Sex</td>
<td>117</td>
</tr>
<tr>
<td>VII.4</td>
<td>Marital Status</td>
<td>121</td>
</tr>
<tr>
<td>VII.5</td>
<td>Ethnic Group</td>
<td>123</td>
</tr>
<tr>
<td>VII.6</td>
<td>Area of Residence</td>
<td>127</td>
</tr>
<tr>
<td>VII.7</td>
<td>Occupation</td>
<td>128</td>
</tr>
<tr>
<td>VII.8</td>
<td>Social Economic Status</td>
<td>129</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER VIII</th>
<th>CAUSES OF DEATH</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIII.1</td>
<td>Introduction: Use of Statistics of Causes of Death</td>
<td>133</td>
</tr>
<tr>
<td>VIII.2</td>
<td>Certificate of Cause of Death</td>
<td>136</td>
</tr>
<tr>
<td>VIII.3</td>
<td>Classification of Diseases and Causes of Death</td>
<td>137</td>
</tr>
<tr>
<td>VIII.4</td>
<td>Specific Death Rates by Cause of Death, Sex and Age</td>
<td>141</td>
</tr>
<tr>
<td>VIII.5</td>
<td>Time Trends in Causes of Death</td>
<td>147</td>
</tr>
<tr>
<td>VIII.6</td>
<td>Proportionate Distribution of Deaths by Causes</td>
<td>161</td>
</tr>
<tr>
<td>VIII.7</td>
<td>Validity of Causes of Death</td>
<td>165</td>
</tr>
<tr>
<td>VIII.8</td>
<td>Multiple Causes of Death</td>
<td>168</td>
</tr>
<tr>
<td>VIII.9</td>
<td>Causes of Death as a Stage of Morbidity Statistics</td>
<td>171</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER IX</th>
<th>SPECIFIC PROBLEMS IN ANALYSIS BY CAUSE OF DEATH</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>IX.1</td>
<td>Infective and Parasitic Diseases</td>
<td>177</td>
</tr>
<tr>
<td>IX.2</td>
<td>Diabetes Mellitus</td>
<td>178</td>
</tr>
<tr>
<td>IX.3</td>
<td>Cancer</td>
<td>185</td>
</tr>
<tr>
<td>IX.4</td>
<td>Cardiovascular Diseases</td>
<td>189</td>
</tr>
<tr>
<td>IX.5</td>
<td>Respiratory Diseases</td>
<td>201</td>
</tr>
<tr>
<td>IX.6</td>
<td>Maternal Mortality</td>
<td>213</td>
</tr>
<tr>
<td>IX.7</td>
<td>Causes of Death During Gestation, in Infancy and During Childhood</td>
<td>220</td>
</tr>
<tr>
<td>IX.8</td>
<td>Deaths Attributed to Senility or Symptomatic Causes</td>
<td>226</td>
</tr>
<tr>
<td>IX.9</td>
<td>Death Due to Motor Vehicle Accidents</td>
<td>236</td>
</tr>
</tbody>
</table>
This Manual on the analysis of the mortality of populations deals with the basic methods of analysis which are commonly used in National Offices of Vital Statistics or in Departments of Health to demonstrate the changing pattern of health in a country from the certificates of death. It is intended to issue a second volume on advanced methods and possibly a third which will contain exercises with hints for their solution.

Public Health and Demography owe so much to the quantitative study of mortality. For centuries, the primary determinant of population trends has been mortality and it still remains so in many less developed countries; it was mortality that formed the primary challenge to the medical profession; it was the prevention of early death that was the primary objective of public health workers and of social legislation. Nowadays, this central role of the study of mortality has gradually yielded way to concern for other phenomena such as fertility and morbidity and the definition of positive health and the study of the provision and use of health services. Nonetheless the analysis of mortality data is still an indispensable part of informed decision-making and of the evaluation of policies in the health services. New problems have arisen even in the area of mortality analysis; the growing importance of chronic diseases have raised new issues and problems; demands for statistical analysis have become ever more sophisticated; the improved quality of certification of the causes of death has created a demand for a detailed study of the difficulties encountered in their interpretation; the use of computers has changed the problems of data processing and facilitated more complex methods of analysis. It was with these considerations in mind that the work on an up-to-date manual of mortality analysis was initiated.

This manual is intended for statisticians, physicians and health administrators who are charged with the analysis of national mortality. It is hoped it will be useful for postgraduate students at Schools of Public Health in courses of biometry, demography, epidemiology, or public health administration, and it is hoped that teachers of medical undergraduates will use it in the preparation of courses on social and community medicine. It is intended that the manual will also serve as a background for training activities and refresher courses in health statistics organised or sponsored by the WHO.

The Manual is intended primarily for use in countries which have already developed a system of accurate mortality registration and certification, but it is hoped that it will also be useful in those areas where such systems are being developed. Problems of data collection and estimation of basic vital and health statistics parameters in countries lacking an efficient vital registration scheme are discussed only marginally in the chapter on Sources and Collection. In these countries the basic problem is that of sampling and there already exists a wide range of relevant textbooks, some published by the U.N., such as "Methods of Estimating Basic Demographic Measures from Incomplete Data".

/continued
This volume is the result of an international collaborative effort between
Professor H. Campbell, Cardiff Wales, U.K., Professor S. Koller, Mainz, German
Federal Republic, with Dr. H. Hansluwka and Dr. S. Mandel, of World Health
Organization. Many other have taken part in the discussions on the Manual and
the comments of many staff members of W.H.O. on a preliminary draft have been most
valuable.

The individual contributors are listed below:

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Description</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Sources and Collection of Data</td>
<td>H. Campbell</td>
</tr>
<tr>
<td>II</td>
<td>Analysis of Data Concerning Populations</td>
<td>H. Campbell</td>
</tr>
<tr>
<td>III</td>
<td>Measures of Mortality</td>
<td>S. Mandel and H. Campbell</td>
</tr>
<tr>
<td>IV</td>
<td>Life Table Construction</td>
<td>H. Campbell</td>
</tr>
<tr>
<td>V</td>
<td>Foetal, Infant and Early Childhood Mortality</td>
<td>H. Hansluwka</td>
</tr>
<tr>
<td>VI</td>
<td>Mortality Analysis in Time</td>
<td>H. Campbell</td>
</tr>
<tr>
<td>VII</td>
<td>Differential Mortality</td>
<td>H. Campbell</td>
</tr>
<tr>
<td>VIII</td>
<td>Analysis of Causes of Death</td>
<td>S. Koller</td>
</tr>
<tr>
<td>IX</td>
<td>Specific Problems in Analysis of Cause of Death</td>
<td></td>
</tr>
<tr>
<td>1)</td>
<td>Infectious Diseases</td>
<td>S. Koller</td>
</tr>
<tr>
<td>2)</td>
<td>Diabetes</td>
<td>S. Koller</td>
</tr>
<tr>
<td>3)</td>
<td>Cancer</td>
<td>H. Campbell</td>
</tr>
<tr>
<td>4)</td>
<td>Cardiovascular Diseases</td>
<td>S. Koller</td>
</tr>
<tr>
<td>5)</td>
<td>Respiratory Diseases</td>
<td>S. Koller</td>
</tr>
<tr>
<td>6)</td>
<td>Maternal Mortality</td>
<td>H. Hansluwka</td>
</tr>
<tr>
<td>7)</td>
<td>Causes of Death in Foetal Life in Infancy and in Childhood</td>
<td>H. Hansluwka</td>
</tr>
<tr>
<td>8)</td>
<td>Deaths attributed to Senility, Symptomatic and Ill-defined Causes</td>
<td>S. Koller</td>
</tr>
<tr>
<td>9)</td>
<td>Deaths Due to Motor Vehicle Accidents</td>
<td>S. Koller</td>
</tr>
</tbody>
</table>

The chapter on Life Table Construction benefited greatly from the advice and
assistance of Professor Chin Lon Chiang, Berkeley, U.S.A.

The encouragement and advice of Dr. W.P.D. Logan and of Mr. K. Usura as Directors of
Health Statistics in W.H.O. at Geneva, are gratefully acknowledged.

Professor H. Campbell, Cardiff, Wales, U.K., acted as overall editor.

These studies were undertaken through grants from the United Nation's Fund for
Population Activities.

Footnote: It is not always possible to print a table in the correct place in the text. When
a table is first referred to in the text but it is not printed in the relevant position, it
will be indicated by the phrase "Table XY.j. Here". This means that the reader should turn
to Table XY.j. at this juncture.
**CHAPTER I**

**SOURCES AND COLLECTION OF DATA**

<table>
<thead>
<tr>
<th>I.1. Concerning Populations</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Enumeration and Registration</td>
<td>5</td>
</tr>
<tr>
<td>b. Census Methods</td>
<td>6</td>
</tr>
<tr>
<td>c. Sampling Methods</td>
<td>10</td>
</tr>
<tr>
<td>d. Register of Populations</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>I.2. Concerning Registration</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Registration Methods</td>
<td>12</td>
</tr>
<tr>
<td>b. Birth Registration</td>
<td>12</td>
</tr>
<tr>
<td>c. Fetal Death Registration</td>
<td>13</td>
</tr>
<tr>
<td>d. Legal Termination of Pregnancy</td>
<td>14</td>
</tr>
<tr>
<td>e. Death Registration</td>
<td>15</td>
</tr>
<tr>
<td>f. Sampling methods of vital events</td>
<td>16</td>
</tr>
</tbody>
</table>

| I.3. Concerning Migration    | 18   |
| a. Definition of Migration  | 18   |
| b. Registration of birth events amongst migrants | 19 |

I.1. **Concerning Populations**

It is the purpose of mortality analysis to describe the pattern of mortality in a population in such a way as to identify those groups of men, women or children, who are especially at risk from some particular hazard or who may be free of a type of death common in other groups. The intention is to learn from this pattern of mortality how the pattern of living may be adapted or modified, so that the people concerned may live a more healthy life free of disease.

In studying the pattern of death, the essential purpose is to understand the pattern of life of a population and what fatal consequences may arise from it.

Mortality analysis is, therefore, a bridge between two disciplines, the first being demography or the study of populations, and the second epidemiology or the study of diseases in populations and an analyst must be able to understand these two distinct modes of study.

This first chapter will describe the sources of data available concerning populations and the difficulties and biases that occur in collecting this information so that the interpretation of mortality data may take these into account.

**I.1.a. Enumeration and Registration**

Two separate and distinct sources of information concerning population are required for all types of analysis of vital statistics of which mortality analysis forms a part. The first concerns the enumeration of the numbers and the characteristics of the population itself, and the second is the record of the vital events of birth, marriage, migration and death which occur within the population. From these two sources it is then possible to calculate ratios, proportions and rates which measure the frequency with which events occur and to compare patterns of events in different sections of the population.
The important characteristics of the population which should be enumerated are first sex, age and place of residence, but it is also valuable to know marital status and certain occupational, social, cultural, or racial characters which may affect mortality.

It is clearly expensive and difficult to have a system that would provide such detail on a continuous basis so this information is usually obtained by a single census made at infrequent intervals of time (usually 10 years). Annual adjustments are then made to the census figures from the known facts of births, deaths and migration to estimate the current population distribution.

In contrast to the census, the events of birth, marriage, migration and death, are occurring at all times and in all parts of the population as continuing processes and it is essential to record these as closely as possible to the place and time where they occur; to do this adequately requires a complete system of registration which operates permanently throughout the whole country and which must be, of its nature, a complex administrative operation.

These two information systems of vital registration and census are very different from each other in organisation and concept and the type of errors and biases which may occur in them can be quite different. These will be discussed separately in the remainder of this chapter and the problems that arise when they are related to each other in the analysis of mortality will be discussed later.

I.1.b. Census Methods

The most important method used to enumerate the characteristics of a population is a census. Although a health department responsible for mortality analysis will not usually be responsible also for organising a census, yet it is essential that they should fully understand the methods used and the limitations and difficulties concerned. They should be consulted by the census department at an early stage so that they can state their requirements and should be closely involved in the design of questions and the definition of terms as these will affect their own analyses.

The basic principles of census taking are adequately described in "Principles and Recommendations for the 1970 Population Census", UN Publication, No. 67, 13, 1, and here a summary is given to emphasise the difficulties which may be relevant to mortality analysis.

A census attempts to enumerate the total number of persons alive within a geographic area at a point of time and in addition to record certain demographic and socio-economic characters concerning the individuals within that population.

A National census will usually be based upon a complete national area and its completeness will be enforced by some form of legal compulsion, but the basic principles of the UN recommendations apply whether the census is national or local, compulsory or voluntary, complete or sample.

Human beings spend some time at rest and asleep, usually at night, and they almost all have some specific location in place to which they return to eat, to sleep and to live with a family or social unit. It is, therefore, possible to count human populations by dividing a geographic area up into small units, identifying within these small units all the possible places of habitation and then enumerating all persons living in these places during one short interval of time, so that omission and duplication as a result of movement during the time of the census is reduced to a minimum.

There is an important distinction to be made between a "de facto" population census which enumerates all those persons who are actually found to be resident in a specific area on a specific night and a "de jure" census which enumerates those persons who normally have a legal right to live in that area and who are usually resident there, but may be absent on that specific night. The use of a "de facto" census gives a better assurance of complete enumeration, but it requires considerable post census adjustment to allocate all persons to the area of their usual residence at which future births, marriages and deaths will be registered. The "de jure" method avoids these difficulties of allocation, but increases the risk of omission or duplication. Most countries use the "de facto" method for enumeration and re-allocate the population according to some "de jure" method.
The problems of how to enumerate nationals of the country who are abroad at the time of the census and the enumeration of foreigners within the country are complex and different countries use varying methods, some of which will be discussed in the section on migrants, I.3.

The geographical division of a country into small administrative units will be predetermined by the physical characteristics and by the provincial and administrative areas customarily in use, but these will be too large for census tracts and it is the responsibility of the census office to divide these basic administrative units into census tracts which are small enough for a team of workers to enumerate within a short period of time. Usually an enumerator is responsible for approximately 200 households.

An adequate number of census teams have to be recruited to cover all census tracts throughout the country who require a certain level of education before recruitment. These enumerators have to be taught what is required and how to carry out their duties after they have been recruited. As their employment is only for a short period of time this may present serious difficulties of finding suitable people and in arranging an adequate training programme.

The responsibility for completing a census return may be placed directly upon certain members of the community such as the householder or head of an institution, or alternatively, the responsibility may be placed upon the enumerator to complete the reports by interrogating the persons concerned. The reports by the householder can be easily completed on one single night by all concerned, but if the enumerators are to complete the forms they will have to spread the work over several weeks according to the work load and hence omissions or duplications may arise.

It is necessary to ensure that travellers, night workers, nomads and persons without a fixed home on the night of the census are also fully enumerated, but these usually present a small problem except in special areas such as refugee camps, ports and in certain cities such as Calcutta where large populations will be found without fixed homes. In such cases special arrangements have to be made to cover such problems.

The details which can be obtained are strictly limited for several reasons; the team themselves have much work to do in a short period of time, there is limitation in the general level of literacy or education of the population, and it is necessary to retain their willing co-operation for if this is lost refusals may occur with increasing frequency.

The type of question asked will also be affected by the main purpose of the census which will be directed towards social and economic factors rather than towards health and vital statistics. But some of the following items can be obtained with a reasonable degree of accuracy in most communities, but not necessarily all of it in the same census.

1) Identification of the person and of the household
2) Administrative area within which they were enumerated
3) " " " normally reside
4) Sex of each person
5) Age
6) Marital status
7) a. Nationality
   b. Ethnic/Racial Group
8) Place of birth
9) Family or domiciliary grouping.
10) Occupation of those employed
11) Cultural information for each person such as education, religion or language
12) Economic information for each household such as income, size and type of house, water supply, drainage, electricity, radio, motor vehicle, etc.
13) Fertility patterns of adult females
14) Specific questions related to subjects of special interest such as physical or mental handicap.

The reliability of this information will depend upon the training and competence of the census workers and upon the education and willing co-operation of the individuals in the population. This problem is discussed in detail in the UN recommendation for the 1970 Censuses and here some warnings are offered which might be important in the analysis or mortality.
1) Identification. Two problems arise concerning identification, first the enumerator must be able to account for all the separate households in his district, and secondly, the householder must be able to identify each individual person in the household. The enumerator must make sure that this is done accurately.

2) Area. The area of enumeration in a census will consist of a small census tract as described on page 6 and the way in which these small areas are aggregated together to form small communities, towns, cities, or provinces, will present no serious difficulties as this will be determined by the administrative structure of the country concerned. There may, however, be difficulties in interpreting the area of a census tract on the ground itself, the enumerator should check carefully that he is able to define the precise boundaries of his area and that his interpretation of these agrees with those of all his colleagues in adjacent areas to ensure that there is no duplication nor omissions. Special difficulties and problems will arise where there are recent developments, where there are tower blocks with multiple occupancy and in areas where there are long distances between houses.

3) Area of legal residence. The area of normal or legal residence for each individual in the census must be carefully defined so that the census data will correspond to the area in which any future birth, death or marriage would be registered. This will often be difficult with lodgers or visitors or people who may be residing in one house whilst retaining a residence elsewhere; students and young soldiers are one such problem, the sick and the residents in institutions such as hospitals, prisons, mental hospitals or residential schools are another. Foreigners, whether temporary visitors or migrants, must also be considered. These problems and their solution will differ between various countries.

4) Sex. Enumeration of sex will present little difficulty, but in some communities even though a census is confidential there may be special reason for accidental or deliberate omissions of one specific sex or age group.

5) Age. Age is the most important variable in mortality analysis and it is sometimes difficult to obtain with accuracy and consistency. The UN recommendation is that the age of all individuals at their last birthday should be recorded, but some people do not know their birthdays or do not remember how old they are. Newborn infants are frequently omitted and infants of 10 or 11 months are reported as 1 year old; the age of school children may be reported inaccurately to conform to legislation concerning school attendance; middle aged persons tend to state their age to the nearest decade so that there are marked peaks in the number of persons reported at age 40, 50 and 60; elderly persons tend to exaggerate their ages and often report their age at next birthday rather than their attained age. Women may be reluctant to disclose their age to the census takers or to the head of the household.

The enumerators must be alert to these problems and ensure that each householder understands how to interpret the requirements.

In countries where birth certificates are required to obtain children's allowances or to enter school, or to obtain professional certificates, or driving licenses, or to register as a voter, some of these difficulties are reduced, but they are never eliminated. In the case of the elderly when it is necessary to produce birth certificates to prove age for retirement benefits, this seems to improve the accuracy of reporting of age.

In some cultures people advance in age by 1 year at the New Year. The mean age of such populations would be reported as half a year older than that of persons reporting attained age at last birthday, and if they are regarded as age 1 at birth they may be on average 1½ years older.

Some census offices evade the question of age by asking for the actual date of birth and some census offices ask for both, but there are few comparative studies to demonstrate which of these methods is in fact more accurate.

Although there are these problems in determining age, it is always possible to make use of the information available to obtain valid estimates of mortality rates which should not vary from year to year or from region to region within one country which has a relative homogeneous culture and system of administration.

6) Marital Status. Marital status depends upon the social, legal and religious back-
ground of the population, but reporting this on a household form may present difficulties. A rigid stable marital pattern is characteristic of certain countries, but other patterns exist and many differences are found throughout the world.

For accurate enumeration it is essential to define the various types of civil status relative to marriage so that they are clearly understood by the population. The usual categories enumerated are, single, married, married but separated, marriage terminated by divorce, and widowed. Some census will enquire of all persons ever married, how many times they have been married. Common law marriages will usually be recorded as legal marriages in census returns.

7) a. Nationality. All countries wish to identify their own nationality as distinct from foreigners within their country, for purposes of taxation, military service, or provision of social services and to calculate vital statistics. But there are serious problems of interpreting the laws of nationality in a simple way on a census form. Special difficulties will arise with immigrant populations, with persons with dual nationality, with refugees and in areas where national boundaries have been changed.

b. Racial/Ethnic. In addition to nationality which is entirely a legal concept, many countries wish to distinguish racial or ethnic characteristics of their people if the state is a pluralistic society. But these concepts do not usually have the force of a legal definition and in some countries these differences are of a social or linguistic nature rather than a legal one. Such classification can never be precise as there will always be a mixture of these characters in many persons within the population and there will be a tendency for the census returns to be modified according to social or economic esteem.

8) Birth Place. Place of birth is frequently used as an index of migration within or between different populations. It is usually only necessary to identify the city or province or country of birth if different from the place of enumeration. It is a useful concept in the analysis of vital events. It is not difficult to obtain, but it may be misinterpreted by the individuals as the area in which they spent their childhood.

9) Household Size. The census returns themselves are usually grouped together in household groupings so that the pattern of household size is easy to obtain, but a household is not a family and it is most difficult to use household information to estimate average family size. It is almost impossible to reconstruct family data from census returns and a census is not the most useful way of obtaining this information. The household in the context of the census is an economic concept of people living together not a biological one which is the family.

Institutions with many inhabitants such as hotels, hospitals, prisons, barracks, religious institutions and homes for the handicapped or aged, are all special problems which must be considered and definitions of the appropriate household size must be given. Lodgers or servants living within the house of the family may or may not be defined as belonging to the household according to whether they share financial or other responsibilities with the household. Here, local custom must determine the definitions but the effect of these must be considered in subsequent analysis.

10) Employment and Social Class. Occupation is an important concept in the study of vital events as the work that a person does, not only exposes him to specific environmental hazards, but it also establishes his place in society and can be a measure of his social status. But the accurate enumeration of occupations is very complex, which is discussed in depth in International Standard Classification of Occupations published by the ILO. It is necessary to distinguish between those in economic activity, those unemployed, those who are physically active but not engaged in economic activity such as school children and students, and those who have retired. It is necessary to distinguish the industry in which the person is employed, and within the industry the type of occupation and the level at which he is working. It is also necessary to define whether the census is to enumerate the current employment, however short, the last employment for unemployed or retired, or the usual employment if the current employment is temporary. The occupation of a woman in tending the home is not usually classified as an economic activity as it does not produce an external source of income. Despite these many problems occupation has proved to be one of the most useful ways of analysing vital events.
11) **Cultural Characters.** Cultural information such as education, religion or language are certainly important determinants of vital and health events, but they have proved most difficult to use from census data. In some countries, education or years of schooling, have proved to be valuable determinants of mortality. The principles to be used in determining educational levels are laid down by the International Standardisation of Educational Statistics published by UNESCO.

The classification of religion or of language will be entirely dependent upon local experience, but it is often difficult to distinguish a nominal allegiance to a religion or a language from the effective practice of it.

A useful measure of social or economic status used in studies of mortality is the number of persons living in each room of a house or the number of persons per square metre which are measures of over-crowding. These can be determined for each census tract and then related to the vital events that occur in these tracts. A useful measure of economic status would be the family income but it is usually impossible to obtain this at a census as the population would fear that the information might be used for taxation purposes. If each house is assessed for taxes, however, this value will be known to the administrator and the average value for each census tract is a measure of the economic status of the tract.

12) **Economic Factors.** Economic factors in the home such as the availability of water supply, sewage disposal, electrical power, are again important determinants of health and vitality, but they have proved difficult to use because although easy to obtain on a census report it is difficult to relate these to the vital events of birth and death registration.

13) **Fertility.** The fertility pattern experienced by adult women in a population are of fundamental importance to demographers, economists, health administrators and politicians, and questions on these topics are frequently added to census reports, either on a basis of a total count or of a sample. Questions may be addressed to all women within defined ages or may be restricted to married women only. The enquiries may be directed at the total number of pregnancies or of live births, they may ask details of spacing between births and of age and duration of marriage of women at the time of delivery. Valuable information may be obtained even if enquiries are limited to women who have had a delivery during the year preceding the date of the census.

14) **Other Enquiries.** A census is an opportunity to undertake a survey of characteristics concerning the whole population and it is particularly useful for a survey of rare events which would be expensive to enumerate by sampling techniques. Medical conditions which might be added to a census schedule are the number of physically or mentally handicapped persons in a household or persons suffering from a rare disease which can be accurately identified by the population such as blindness or mental handicap.

I.1.c. **Population Sampling Methods**

Sampling survey methods have an important role to play in the study of population statistics and further details are given in the Handbook of Population and Housing Census Methods, Vol. V. UN.

An important but simple technique is to sample within the context of a total census. When a complete census is organised it may be too expensive to ask all members of the population for details of all questions, it is then possible to enumerate the total population, but to ask certain specific questions in greater detail from a sample of the population. This should not present serious difficulties as the census itself will provide the sampling frame and recognised methods of random sampling can be applied to obtain a probability sample which can be used to estimate the mean values in the community. For example, to ask for a detailed occupational history of all persons over a period of years would be impossible, but it might be very useful to study the changing pattern of employment in a sample of the population. Another example might be a study of fertility from detailed obstetric histories which could be recorded by specially trained interviewers or even by medical staff.

A similar problem might arise in preparing tabulations. The coding and mechanical recording of census data is an immense task which can take many months. If a random sample of returns is selected for early coding and punching, it is possible to publish a reasonably
accurate return at an early date. For example a 1% or 10% probability sample would give an acceptable measure of total numbers, age and sex, areal distribution within a matter of a few weeks.

A rather more difficult problem is to complete a survey of a random sample of the population especially as it is invidious to impose legal obligations to complete a census on a selected sample of the people. But the problems are the same as that of any sample survey method; a complete frame must be established to allow a random sample to be drawn from it, which can usually be done from housing lists or from the previous census tracts. To provide useful information which could be relevant in local areas or for detailed cross tabulations, the probability sample should be between 1% and 10% dependent upon the frequency of the characteristics being enumerated and the precision with which estimates are required, as the total number of recorded events will determine the validity of the sample.

Sampling methods are of great value to estimate changes which are occurring in the total population. The expense of a complete census necessarily means that it can only be taken at irregular intervals and so it is an important part of the work of the Census Office to estimate the changing pattern of the population. To some extent this can be done from vital registers which provide exact numbers of births, deaths and marriages and divorces. Emigration is usually more difficult to estimate accurately, especially the detailed characteristics of the emigrants, and changes in economic and social patterns in a population are difficult to estimate. A continuing sample survey census is a most useful tool to solve these difficulties. A permanent stratified sampling frame is established and at regular intervals, frequent random samples are made. By this method it is possible to use quite small samples and to aggregate the results from several surveys to obtain valid estimates of the changing pattern within the community. Monthly samples of 1 in 10,000 or even smaller are used and the U.S. Health Survey consists of monthly samples of about 3,000 in a population of 200 million persons. Examples of topics which can be studied by such sample surveys are the prevalence of morbidity, or the use of the health or social services during a period of time, or the occurrence of accidents, or they may be used for economic studies such as employment. Such permanent samples are limited, however, in their use as they can never give an adequate picture of changing patterns in area.

Where it is impossible to make a complete census of a whole country, sampling methods have an important contribution to make, but this should not in any way be used as an excuse for not making a regular and complete census. A census is most desirable even if nothing more than the numbers of persons of each sex are enumerated. When a census cannot be taken it is difficult to establish a sampling frame which must be based upon areas in which the probability of inclusion in the sample should be proportioned to the population of the area. It will be necessary to stratify the whole country into areas based upon prior knowledge or urban/rural, pastoral/agricultural, or literate/illiterate population and within each of these distinct strata to draw true random samples of definite areas and to enumerate certain characteristics of the population within these sample census. It will then be possible to make valid estimates of known precision concerning the distribution of these characteristics in the total population.

If the whole country cannot be included in the frame for selection of sampling areas it may still be useful to take specific areas with effective administration to ensure that some estimates of mortality are available, but it must be emphasised that such results cannot be projected to relate to the whole country.

I.1.d. Registers of Population

Some countries, particularly in Scandinavia, Netherlands and Eastern Europe, maintain a register of their population for administrative purposes. This was originally accepted as it was introduced by the Church to maintain parochial registers and has been extended for state purposes to cover taxation and domiciliary permits. These registers can be kept up to date by recording all changes of legal residence and of marital status; in Sweden this is done annually.

In parts of South East Asia it is also the practice to issue identity cards to the population when they attain a certain age. Although the purpose of these documents is primarily intended for accounting purposes they can be used as a sampling frame for population studies.
It is essential that a population register should be checked regularly by a complete census to ensure that it has not become obsolete for errors in such registers are cumulative and duplication of individuals becomes possible.

They do, however, provide valuable information concerning internal migration and they enable the registrars to check the personal information provided on birth and death registers.

One technique of data analysis which may be useful in the future is the establishment of a computerised data bank of all vital events and of selected health records such as hospital discharges which are linked together. These are valuable tools in the study of mortality or subsequent morbidity after hospital discharge. Unfortunately, such data banks, however, are often unpopular politically as suspicion may be aroused that they could be used for other methods of surveillance.

I.2. Concerning Registration

I.2.a. Registration Methods

"Everyone has the right to a nationality", Declaration of Human Rights, 1948, Art.15, implies that everyone must have the opportunity to prove his nationality and from this springs the need for all modern states to have an effective system of vital registration.

Similarly, there is an individual need for every person to be able to prove his or her marital status for legal, economic or social purposes. A modern State must also see that all deaths are registered legally to ensure that bodies are disposed of, death duties paid and inheritances transmitted.

The need for such certificates and registers of birth, marriage and death spring from the legal, economic and social needs of the individuals and only secondarily from the need of the state for statistical information concerning its own inhabitants. Although the need for such registers arises from the individual yet the form in which the registers are kept is strongly influenced by the need for adequate statistical information.

The mortality analyst will not be responsible for the large administrative apparatus required to ensure adequate registration of vital events, but the way in which these events are recorded are of fundamental importance to his own work and hence he must thoroughly understand the methods used and their limitations and sources of error. As birth registration is closely involved with infant and maternal mortality he will need to understand the methods of birth registration as well as those of death registration.

I.2.b. Birth Registration

All births occurring in a modern state must be registered within a defined legal time. A live birth is defined as:

Live Birth is the complete expulsion or extraction from its mother of a product of conception irrespective of the duration of pregnancy which, after its separation, (1) breathes, or (2) shows any other evidence of life such as (a) beating of the heart, (b) pulsation of the umbilical cord, or (c) definite movement of voluntary muscles whether or not the umbilical cord has been cut or the placenta is attached. Each product of such a birth is considered live born.

This definition seems to be very explicit, but difficulties arise in ensuring that all live born infants who die shortly after birth are registered and also to ensure that all infants who died before complete extraction from the mother are not recorded as live born. These criteria of life which are so general, (1) breathes, or (2) shows any other evidence of life, have to be interpreted by physicians, by midwives, or unqualified attendants at birth. There will be conflicting pressures if the baby dies quickly to reassure the mother that the infant never lived, or alternatively, if stillborn, to claim that it was alive when higher social benefits are drawn for a live born child.

The responsibility for registering a birth must be placed upon the parents, but it will be the responsibility of the attendant whether medically qualified or not to notify the authorities that a birth has taken place so that they can ensure that registration does take
place. It is unusual to require a medical certificate of birth, but this is sometimes introduced as a method of validation or to obtain additional medical information relating to the pregnancy or birth.

In certain countries either by legislation or by custom infants born alive but who die before the birth has been legally registered are treated as a separate group. In France this is up to three days and in Spain 24 hours after birth. This group is held separately in the statistics, sometimes being included, but mostly being excluded from the numbers (see Chapter V). It needs careful attention by the consumer of the statistics to be sure which method of allocation has been done in a special case.

When the event is registered, various items of information may also be recorded according to the local legal requirements:-

Concerning the event:  
1) Date of birth  
2) Place of birth: (a) geographical place  
   (b) type of institution (home, hospital, etc.)  
3) Single or multiple birth  
4) Attendant at birth: medical, midwife or untrained.

Concerning the infant:  
5) Name  
6) Sex  
7) Birth weight  
8) Gestational age  
9) Presence of any medical or physical defect

Concerning the parents:  
10) Age of father and of mother  
11) a. Whether married  
    b. Duration of marriage  
12) Nationality of father and of mother  
13) Ethnic or racial group  
14) Normal place of residence  
15) Economic characteristics such as employment of the legitimate father of child, or if illegitimate, of the mother  
16) Cultural characteristics, education religion or language  
17) Fertility experience of mother:  
    a. Previous live births and foetal deaths  
    b. Time interval since the last birth

It will be clear that the quality of this information will vary. The date, place and multiplicity of the birth can be expected to be accurately recorded, birth weight will often be well recorded, but clinical details such as gestational age may be less accurate. The sex of the infant may occasionally be misreported especially of premature infants.

The information concerning the parents should be of reasonable quality for the parents are providing information concerning themselves. Although legitimacy is in most societies a very sensitive matter, it may be difficult to conceal this fact from a registrar.

The economic and cultural information may be less accurate than the other personal details, but should be of better quality than that obtained on a census report.

The record of a congenital malformation or disease is a most useful tool in the study of these conditions which constitute an important part of the causes of perinatal mortality, but it is obviously undesirable from the social and political point of view for this to be recorded on a public document. If this information is to be recorded, some system of recording it in a confidential manner should be devised.

I.2.c. Foetal Death Registration.

The definition of a foetal death is given by the W.H.O. as:- "Foetal death is death prior to the complete expulsion or extraction from its mother of a product of conception, irrespective of the duration of pregnancy; the death is indicated by the fact that after such separation the foetus does not breathe or show any other evidence of life, such as beating of the heart, pulsation of the umbilical cord or definite movement of voluntary muscles".
It will be observed that this definition is the converse of the definition of a live birth, but so broad that it includes every product of conception which fails to be delivered alive from the uterus, and thus it includes all pregnancies which are terminated whether legally or not and all spontaneous abortions at any stage of gestation. It is clearly impossible that all such cases should be registered completely.

The primary purpose of the registration of foetal deaths is to obtain a detailed statistical analysis of the medical and social conditions affecting perinatal deaths. It is the practice, therefore, to divide foetal deaths into 3 groups according to the duration of gestation:-

Group I. Early foetal deaths are deaths with less than 20 completed weeks of gestation and a fourth group may be formed to include cases of unknown or uncertain gestation.

Group II. Mid foetal deaths are deaths after 20 completed weeks of gestation but less than 28 weeks.

Group III Late foetal deaths are deaths of 28 completed weeks of gestation or more.

Specific legislation in a particular country will then determine which events are registered and the method of registration, but in many countries only late foetal deaths are registered by law. Although the accepted international definition of the stages of foetal death are based upon gestational age, this age is not accurately known but is often estimated from the last reported menstrual period of the mother, but the clinician certifying a foetal death may also use other clinical information such as the length or weight of the foetus.

The International Classification of Causes of Death defines both an infant death and a foetal death as premature on the evidence of the weight of the infant or foetus at delivery being less than 2500 g.

This information obtained at the registration of a foetal death should be the same as that which would be obtained for a live birth, but in addition it is recommended that there should be a medical certificate of foetal death signed by a medically qualified certifier who states to the best of his knowledge the underlying cause of the foetal death. This medical certificate is similar to the normal death certificate but the cause of death may be attributed to disease in the mother, disease in the foetus, or to external factors.

If it is difficult to determine the stage of development at which conceptions should be consistently included within the definition of foetal deaths, there is less difficulty in determining which should be included in the live born. Any infant which shows any form of independant life after separation from the mother should be included in the live born, no matter how short its expectation of life might be. In some countries, however, local legislation is contrary and allows deaths of infants in the first day to be accounted as foetal deaths.

The main interest in this subject from the medical point of view is the loss of life associated with the perinatal period as a foetus at 28 weeks gestation can be expected to survive and an infant who survives one week after delivery can be expected to have surmounted the hazards of birth. The causes of deaths of infants before delivery and in the first week of life are similar. Hence perinatal mortality rates are requested by W.H.O. which relate to all late foetal deaths and to infant deaths up to the seventh day of life and these should be analysed into the separate causes of death.

These topics are discussed further in Chapters III and V.

I.2.d. Legal Termination of Pregnancy.

Many countries have introduced legislation concerning the legal termination of pregnancies, but no international recommendations have been made concerning the regulation or registration of these events.

It is clear that such events are of considerable importance to the demographer in studies concerning natality, but they are also of relevance in the study of mortality as any selective elimination of the products of conception at an early stage will greatly influence the perinatal mortality.
Such events should, therefore, be registered and their tabulation and analysis should form part of the vital statistics system.

The items to be recorded relate:

(a) to the conceptus
(b) to the mother
(c) to the termination

Little information concerning the conceptus can be obtained except the gestational age, as name, weight and sex will not be known. The information relating to the mother, however, should be in similar detail to that obtained at birth registration, as her age, marital state, area of residence, racial or social group and fertility are all very relevant to subsequent studies of fertility or mortality.

The information concerning the termination should be analogous to the death certificate and should be issued by a medically qualified certifier, it should give the time and place, the reasons and indications for the termination and any complications encountered.

No international recommendations concerning tabulations of termination of pregnancies are available.

1.2.e. Death Registration

A complete register of deaths is required to ensure that the disposal of all bodies is under the control of the civil authority. To dispose of a human body is a difficult procedure and in almost all societies is associated with religious ritual that must be performed publicly. Legally, there is the need to be able to prove the death of a person to inherit property, to claim insurance, to claim social benefits, to detect deaths from violence or for the surviving spouse to be free to remarry. Thus a registration system is needed for the benefit of the individual and not for statistical purposes, but once the system is established the statistician is able to recommend such definitions as are of importance to his studies.

The legal definition of death for registration purposes can be simple as it involves only the fact of death and does not involve the delicate ethical and moral problems which a physician has to consider when he is deciding the time of death. The recommended definition is "Death is the permanent disappearance of all evidence of life at any time after live birth has taken place". The only problem is how to define foetal deaths which are specifically excluded and which will be considered separately in Chapter V. The problem of defining the cause of death, however, is much more difficult and this is discussed at length in Chapter VIII.

When a death comes to be registered there is an important fact which distinguishes it from census records; the person concerned is not there to provide the relevant information.

Two people provide information concerning a death, the physician who provides a medical certificate of death with details of the cause of death which will be discussed later, and the "informant" who is usually the nearest relative available but may be an administrator or a nurse, who provides the information concerning the deceased.

The statistical information obtained at death registration can be divided into that related to the event and that concerned with the individual:

Concerning the events:
1) Date of death
2) Place of death: a) Geographical b) Type of Institution, (Home, Hospital, etc.)
3) Medical attendance
4) Medical cause of death
5) Medical reports available

Concerning the deceased:
6) Sex
7) Age or date of birth
8) Citizenship
9) Place of residence
10) Marital status
11) National, ethnic or racial group
12) Professional. Industry, occupation, status
13) Cultural data. Education, religion, language
14) Fertility data
15) Special information on certain groups such as children or occupational groups.

The fact of death, the date and place of death are fundamental data which will generally be accurately recorded.

The information provided by the physician is subject to many reservations, but this aspect of the problem will be examined in greater detail when we consider mortality according to the different causes of death in Chapter VIII.

The quality of the information concerning the deceased will be variable because the informant may not know the information accurately, and may be old and be disturbed by the fact of bereavement.

Sex will be accurately reported, but age may not be so, especially when sons are reporting the deaths of aged parents and do not have birth certificates or identity documents to support their information. This becomes even more difficult when an elderly person dies with no near relative. This information becomes more accurate, however, if proof of age is required for insurance or social security benefits.

Citizenship and place of usual residence should be well reported. But for purposes of analysis the place of residence in the period before death may not be the place in which the deceased spent the greater part of his life in which he incurred the hazards that lead to his death. But a detailed history of residence is not possible in all cases, but may be sought in a selected sample.

There will obviously be a bias by an informant to report the marital status in the most favourable light and hence there will be a tendency to over report married and widowed compared to single and divorced, and there will be a similar tendency to exaggerate the social prestige of any ethnic/racial or linguistic group if this is in doubt. This is discussed in Chapter VII.

Severe difficulties arise, however, over the problem of occupation; most people survive until retirement age and even those who do not, may die after some years of increasing ill health in a secondary occupation. Moreover, the informant may not know the occupation of the deceased, nor the appropriate status and again there may be a tendency to exaggerate these. A method of linking this information with the census information could be a useful technique to validate the accuracy of these reports.

Education, language and religion, if recorded, should be much more easily verifiable, but fertility patterns will be most difficult to establish for deceased elderly women. The great majority of women will survive their husbands and their death will be reported by a son or by a friend not a member of the family; it is most improbable that they will be able to record accurately details of births which may have taken place 50 or 60 years previously, especially if these had resulted in infant or foetal deaths.

Deaths of all women of potential fertile age should also report whether the woman was pregnant at the time of death or had recently been delivered and whether this pregnancy had any relevance to the death.

Sudden deaths and deaths due to accident or trauma will be subject to special legal enquiries to determine the cause and possible responsibilities for the death. The death certificates in such cases should include information on the physical causes leading to death which are required for the international classification of external agents causing death. Provision for registration of deaths due to major accidents such as, earthquake, civil war, transport accidents, or mining and industrial accidents, may require special attention to ensure that they are included in the routine statistical records.

I.2.f. Sampling Methods for Vital Events
enumeration of all births and deaths is an expensive operation and postulates a certain degree of literacy in the total population of the country concerned and a developed administrative structure. As a result there may be many countries who are unable to provide a complete registration service and must look for an alternative system to obtain valid estimates of birth and death rates.

In a country with a complete system of registration for births and deaths, there can be no place for sampling the births and deaths themselves, for these should be registered for legal and social reasons for the benefit of the individual person and the statistical information which is to be obtained from the registration is only secondary to this main purpose, but there may often be some advantage in such countries to sample the births, or deaths, for more detailed studies of social, obstetric or medical conditions relevant to vital events.

In such studies the registers of live births or deaths would form the sampling frame and a fixed random sample would be selected for further detailed studies. A detailed questionnaire can then be used to investigate specific problems.

In a few countries with complete registration, the expense of coding and punching data cards for all live births has been considered excessive and a sample of births have been processed to prepare the tabular reports on natality. Such a method should be based upon a sound probability sample of the events for valid estimates of rates to be made. As the major cost of any system is always the clerical procedures of registration there is little advantage to be obtained by economies on data processing.

It is important that death registration should be as complete as possible to regulate the consequential legal rights of survivors and to regulate the disposal of the dead so there can be no place in a developed country for sampling methods related to total registration of deaths.

Death certification, however, may often provide an excellent sampling frame for many investigations in which controlled studies can be organised. Thus if an aetiological agent is suspected to be an important element in the causation of a certain type of death, a sample of these deaths together with matched controls can be analysed in depth to establish whether the suspected agent is more prevalent in the sample than in the controls. Many such studies have been successful in elucidating the etiology of a disease. The sample may be selected by any of the characteristics of registration so the study may relate to a specific age and sex group, to a specific area or occupation, or it may relate to the cause of death. In each case the relevant principles of random sampling must be applied both to obtain the sample of cases and of the controls required for comparative purposes.

In some countries, however, where complete registration cannot be established, sampling methods may be introduced to obtain a valid system of estimation for death and birth rates. The minimum objective of any such sampling system should be to obtain statistical estimates of the crude birth and death rates within defined confidence intervals.

The purpose of such samples is to obtain measures not only of the absolute values of the crude birth and death rates, but also measures of age and sex specific rates and of any possible trends over time which may be occurring within these. Hence the samples should be large enough to allow the rates to be estimated within confidence limits which are smaller than the random annual fluctuations known to occur.

The confidence limits of a vital rate are the limits within which the true value, \( \bar{d} \), of the rate in the total population can be estimated to lie with a confidence level of \( \alpha \) where this level is arbitrarily chosen. These limits are given by the standard formula:

\[
\begin{align*}
\text{Lower Limit} &= \bar{d} - t(\bar{d}/n)^{1/2} \\
\text{Upper Limit} &= \bar{d} + t(\bar{d}/n)^{1/2}
\end{align*}
\]

Where \( \bar{d} \) is the mean rate found in the sample,

\( n \) is the total sample size

\( t \) is the normal deviate which corresponds to a probability \( 1 - \alpha \)
In practice the difficulty lies in that most vital rates are events with low frequency of the order of .01 hence a confidence limit of $\pm 10\%$ with a confidence level of .95 requires a sample of 40,000 persons and a confidence limit of $\pm 2\%$ would require a sample of 1 million.

If age specific mortality rates are required, then of course, the sample size will have to be larger still to give reasonable estimates of the individual rates and again it is possible to obtain a formula for the confidence intervals based upon the age specific rates.

These systems of confidence intervals are based upon the assumption of simple random sampling of the events, but in most demographic practices it is very uneconomic to sample individual persons and the method usually employed is multiple stage sampling by which at the first stage a random sample of administrative areas is drawn from the whole country and then within these areas cluster sample of small groups of households near to one another are chosen at random and fully enumerated. The theoretical effect of this two stage cluster sampling is to widen the confidence intervals for a given sample size, but the economy obtained is such that a larger number of persons can be enumerated by the cluster sampling method for the same cost and hence the ultimate rates should be more accurate.

These sampling techniques for vital events are very different in kind to registration. A registrar stays in his basic administrative centre and obtains information which is brought to him by the people concerned. A sample survey must be a completely active process by which the surveyor goes out into the population and makes active contact with the population and obtains information which he himself has sought, as a consequence there are marked differences in the results between the two methods.

When a surveyor has selected a population for detailed study there are two principles by which he can obtain information concerning vital events. He can rely upon the memory of the individuals selected to give him information concerning events which have occurred during the previous year, or some other interval which has been selected. Alternatively, he can enumerate a sample of households or of individuals and return to that sample after a period of time, re-identify the persons in it and record how many of these have had a birth and how many of those whom he cannot enumerate have died, how many have moved away. This second method is known as a "routine round" and has proven to be quite effective in many primitive conditions. The surveyor should return at relatively frequent intervals of three or six months and should, of course, pay special attention to women during the years of fertility. In many studies on the first round he will ask each woman if she is pregnant and it is the experience of all such studies that this question does not usually cause embarrassment to the women concerned. This helps the surveyor to identify women who may be delivered of a stillbirth or who may have suffered an infant death before his routine return.

I.3. Concerning Migration

I.3.a. Definition of Migration.

The third important influence on population changes is migration by which a population may increase rapidly, may be depleted, or may change its character over a period of time. Important as this matter is, however, few countries have adequate information concerning migrants within their boundaries and almost none have detailed statistics of the characteristics of emigrants who have left the country.

Although almost all countries insist on the exhibition of passports on entering the country, little else except a total head count is possible by a frontier service. Work permits and police registration serve to control the numbers of immigrants, but these do not provide a meaningful demographic analysis. Workers first move across frontiers as temporary workers and may retain a home in the old country for many years, they may bring their families with them only after they have become economically established.

A migrant, however, is defined as a person who crosses an international frontier with the intention of remaining in a new country for at least 12 months. This definition excludes almost all tourists and the great majority of casual migrants who follow a harvest or a seasonal employment such as the hotel industry in tourist areas. It has the disadvantage however that it relies upon the stated intent of the migrant to remain away for 12 months which he may not be prepared to admit even to himself at the beginning of his journey. An alternative method is to enumerate and register all migrants when they arrive in the country.
and to keep a constant record of their residence in the form of a population register. This is done in the United States.

I.3.b. Registration of Vital Events Amongst Immigrants.

When a birth occurs to an immigrant family, this family is subject to the same regulations concerning registration as are the local population and so the event will be registered as occurring in the country of registration. There is usually no method by which these events can be systematically transferred to the country of origin of the parents and so they should be recorded and published as vital events in the country of registration unless the parents do not intend to remain in the country for more than a year. Such events are comparatively rare and make little contribution to the total birth rate, but cumulatively they may become a large sub population within the boundaries of a state.

In the case of a death of an immigrant, the death has to be registered within the country in which it occurred and should be included in the vital statistics. If possible, however, it is important that these deaths be identified separately as migrants will usually suffer a death rate which differs from that in their host country and from that in their homeland, and they may be subject to special diseases.

The problem of registration of vital events in immigrant communities is a most difficult one as the migrants are often living in slums without any house registration and hence be unable to register vital events. In such cases a specific project may be required to study the problems concerned. Conversely, in a population which has been subject to a long period of emigration, it is possible to demonstrate from the age and sex pattern of the census data for the remaining population that there have been important alterations in certain groups, but it will be difficult to demonstrate that it is the fitter, more energetic and better educated people who have left behind a population that is older, sicker and less capable of adapting to changes in the environment.
CHAPTER II

ANALYSIS OF POPULATION DATA

II.1. Population Composition

a. Sex and Age
b. Marital Status
c. Area of Residence
d. Economic and Social Characteristics

II.2. Types and Sources of Error in Population Data

a. Planning Errors
b. Incomplete Enumeration
c. Double Counting
d. Inaccuracies of Information
e. Errors of Computation
f. Errors of Printing or Publication

II.3. Detection and Correction of Errors

a. Planning
b. Post Censal Survey
c. Inter Censal Comparisons
d. Internal Consistency Checks
e. Linkage with Vital Statistics
f. U.N. Evaluation of Reliability

II.4. Estimates of Populations

a. Introduction to Estimation
b. Inter Censal Estimation: i) mathematical models
   ii) component methods
c. Post Censal Estimation: i) mathematical models
   ii) component methods
   iii) proportional methods

II.1. Population Composition

A mortality analyst must be in a position to understand the basic structure of a population and how the important demographic characteristics affect the function of a population before he is able to analyse competently the patterns of mortality experienced by the population. This chapter will consider the distribution within populations of characteristics such as sex, age, marital status, place of residence, and socio-economic variables which will have some influence on mortality and it will pay particular attention to the possible sources of error that can arise, how these may be detected and how amended.

II.1.a Sex and Age

Sex is the most easily observed and recorded element of population composition; its classification into a simple dichotomy presents no difficulties in coding, or in preparing tabulations. The ratio of males to females is used as the measure of the sex ratio. In almost all populations this ratio at birth is greater than 1.00, but less than 1.10, but again in most societies the male mortality is greater than the female mortality at all ages, so that the sex ratio starts to fall from a high level at birth until it may become as low as 0.30 in people over 75 years of age. It is fortuitously in balance at 1.00 at some age during the reproduction years between age 15 and 45.
Fig. II.1.

Age and Sex Pyramid for the population of Great Britain in 1841, 1891, 1939 and 1971.
If the sex ratio at birth is reported greater than 1.10 there may be under-reporting of female births. A sudden change in sex ratio at certain ages may be evidence of inaccurate reporting of ages in one sex. For example at ages 15 to 19 or 20 to 24, males may report ages lower than reality to defer military services, whereas, women may increase their ages if there is minimum legal age at marriage. Conversely, when approaching age 40, males may exaggerate their age and females understate it again altering the sex ratio.

The overall sex ratio for all ages is not a statistic of much value for it is entirely dependent upon the age distribution of the two sexes, and therefore, age specific sex ratios should be calculated.

Age is also a character that is universally recognised but its precision is difficult to estimate and although it is theoretically a continuous variable its classification has to be arbitrary. When age distributions are examined by single years of age there is evidence of over reporting of ages ending in 0 or 5 and deficiencies at 1, 7 and 9. There is also a tendency to misrepresent ages at puberty and in old age. The number of infants under 1 year recorded by a census is frequently less than those known to have been born and survived during the year preceding the census. Consequently a Census Office will often make adjustment to a reported age distribution in order to correct these types of error and the mortality analyst must be aware of these problems.

The UN recommends that all age tabulations should be in 5 year age groups starting from age 5 up to aged 84 and that people above that age should be grouped together as 85 and over. Children under the age of 5 should be classified into the single years of age or into under 1 and 1-4 where under 1 is the first year of life and age 1 is the second year of life.

This five year age grouping avoids many of the difficulties of determining the accuracy with which age is recorded, but it still means that there are 36 sex and age groups to be tabulated separately and for many purposes a ten year grouping is adequate. When this is done, children from 0-4 should be shown as the first group and then the ten year groups should start from 5-14 up to 73-84 with an open ended group 85 and over. Populations with few elderly persons or in which the age of the elderly is difficult to determine accurately, should make the final grouping 75 and over.

An effective and interesting technique which is often used to analyse the age and sex structure of a population is to draw the age and sex pyramid. The method is to tabulate the population in five year age groups for each sex and then to calculate the proportion per thousand of the total population in each sex and age group. The proportional frequency histograms are then drawn horizontally with the males drawn in the negative direction on the left and the females drawn positively on the right. Thus an area on the graph represents one thousandth part of population and the total area of each pyramid is the same, equal to 1,000 units.

An example of this technique is given in Figure II.1 which shows the changes in the population structure of Great Britain over 130 years between 1841 and 1971.

**Figure II.1 here**

The age and sex structure of a population is a product of the birth and death rates together with immigration and emigration during the previous 90 years. The pyramid is a historical picture of the end result of many influences. The most important determinant is the birth rate. If this has been in excess of the replacement rate then the pyramid will be triangular, if it has been in balance the pyramid will be a pentagon with two straight sides surmounted by a triangular superstructure. If it has been below replacement the pyramid will be undercut at the base. The death rate determines the shape of the curve at the higher ages, but even here historic factors play a large part in the shaping of the pyramid.

A war will remove many adult males so that the male/female ratio is disturbed, but its influence on a population may be even greater by the sudden drop in the birth rate as husbands join the Army and this is followed by a sharp peak in the birth rate as couples try to complete their families when peace is restored. This effect is seen in the 1939 Great Britain population in the two age groups between 15 and 24 who were born in the years 1919-
Fig. II.2.

Age and Sex Pyramid for 4 Countries -
Mexico 1973, France 1971, Japan 1972,
Ghana 1970,
1925 and also in the 1841 population in the age group 25–34 who were born in the years following 1815. A famine will affect more severely young children under 5 years and both sexes at these ages will be affected equally. An economic depression will reduce the birth rate by postponement of the age at marriage. This effect can be seen in Figure II.1 in the 1939 population aged 0–14 and in the 1971 population aged 30–44.

The changes that have occurred in the sex and age structure of the British population illustrate the various stages in the development of a population. The structure in 1841 was that of a typical developing population with high natality (35 per 1000) and high mortality (22 per 1000); by 1891 the shape of the pyramid structure had not materially altered although the total population had nearly doubled because both natality and mortality had remained high. After 1891, the mortality improved rapidly and after 1921 the natality rates declined steeply below replacement levels so that the population structure in 1939 was badly deformed and the pyramid was standing on an inverted base. From 1940 onwards the birth rate rose and mortality continued to decrease slowly so the 1971 pyramid appears to be more stable than it had been in 1939.

In the modern world there are countries at all stages of this cycle in the development of a population. Figure II.2 illustrates this from 4 countries drawn from the four major continents.

Figure II.2 here

Mexico is typical of a Latin American country which has a very high birth rate, has had a recent reduction in mortality rates and in which the population is expanding rapidly. The situation in Ghana is similar. In both countries the health services would be directed to paediatric and obstetric problems and the social services to education and to family health. The mortality pattern will be dominated by infectious diseases, trauma, and malnutrition, with few degenerative or neoplastic deaths.

Japan is a country which has had a rapid expansion with high birth rate and reduced death rate, but is now embarking on industrial expansion with a recent reduction in the birth rate. Its mortality pattern will be that of middle age together with some cardiovascular and neoplastic deaths. A special feature in the Japanese age pyramid is the very high proportion of persons aged 25 to 29 in 1970 who were born in the years immediately following the end of the war.

The population of France is typical of an advanced industrial country that has suffered a severe loss of population due to wars and has had an increase in the birth rate during the last generation which has caused a recent expansion. In France the mortality pattern will be dominated by cardiovascular and neoplastic causes of death.

If a human population were stable in its structure, and the death rate had remained constant for 80 years, if also the birth rate were at replacement level and there was no migration, then the population pyramid would be predetermined entirely by the sex ratio at birth and by the subsequent mortality from the life table for each sex. But such a situation seldom occurs and human populations do experience large fluctuations in births and in mortality over short periods of time which have long lasting effects on the age and sex structure. An increase in the birth rate will not only increase the population, but will alter the population structure as well and this mark in the population will remain so long as that generation or cohort of persons remain alive.

It is not the purpose of this section to discuss the problems of the rapid increase in the world population which is occurring, but to demonstrate that it is essential to understand these structures of a human population before statistics of mortality and natality can be analysed in a meaningful way.

II.1.b Marital Status

Marital status is the third primary item of information available for all persons enumerated at a census which is of considerable importance in the analysis of the composition of the population. It has a direct bearing on the estimates of reproductive capacity of the
population although it must be recognised that different cultures attach varying stigma to illegitimacy and that in a few societies the majority of births take place outside marriage. The proportion of married women in each five year age group from 15-45, has in the past been a useful index of the reproductive potential in the population. In modern societies, however, more complex measures are needed to measure reproductive potential. This is determined not only by the proportion of fertile women who are married but also by the concepts of a desirable family size which are held by a society, by religious and economic pressures and by the availability of contraceptive advice and services. All of these will require sophisticated sociological methods of analysis which are discussed in the Manual on Fertility Analysis.

The stability of marriage in a community has an important influence on mortality as it can be shown that the death rate is lower in married persons than in single persons and increases again with dissolution of marriage due either to widowhood or divorce. This is clear evidence that the family is an important determinant of health care. But a comparison of marital status between populations with differing cultures is not usually a useful exercise due to the problems of ensuring consistent definitions of marital status which are accurately recorded. Even within a national population there may be marked differences between different social or religious groups in the accuracy of recording marital status.

Thus although marital status is obviously an important determinant of both natality and mortality, the difficulties of using it are such that it will not often be used in analysis of mortality.

II.1.c Area of Residence

Place of residence is a factor in the analysis of population structure that is easy to handle because it is recorded not by the respondent to a census, but by the enumerator who knows precisely which area he is enumerating. Hence this information is usually reliable. The principal difficulty is that it is necessary to transfer details of people living away from home on the night of a census who are counted "de facto" in the area in which they were found. It is necessary, however, to transfer such people to their true residential area as it is in that area that births and deaths will be registered. The major difficulty is to ensure that the same standards and definitions are used both by the Census Office and by the registration service in allocating residence so that deaths can be related to the relevant population at risk.

The location of the residence is highly associated with many other factors which determine social and economic status. The most important of these is whether the place of residence is in an urban or rural area because the risks to health, the provision of educational and medical services, the social security benefits, the economic standard of life itself, will vary considerably between urban and rural and between other geographic or economic areas of a country. A further difficulty arises in the definition of urban and rural areas as many of the worst urban slums may be established in the former rural areas surrounding a city and frequently large industries, or mining areas may be established in areas which are defined as rural.

In some countries the mortality rate in rural areas is lower than that in the urban areas; as early as 1842, Dr. Farr was able to demonstrate that the then new industrial towns of England, of Liverpool and Manchester, had a mortality experience that was almost twice that of rural areas in England. These large differences were responsible for intense political pressure for legislation to improve sanitation during the next few decades. Although this was successful and the mortality did fall, it fell in the rural areas also, so even after 130 years there is still a wide margin in favour of the rural areas.

In the United States the geographical areas of the individual states have remained constant from their establishment and the contrast in mortality between the states is even greater than that between most countries in Western Europe. Despite great national reductions in mortality the relative position between the states remains almost unchanged.

In other countries the mortality in rural areas is higher than in urban areas especially infant mortality and deaths from the enteric infections.

It is important to realise that each area will have its own particular age and sex
structure and that before valid comparisons can be made between areas adjustments must be made to the calculated rates as will be described in Chapter III.

II.1.4 Economic and Social Characteristics

Economic and social characteristics of a population are rather more difficult to analyse than are the geographic patterns. It is almost impossible to classify complete families by social characteristics although this was done in the early Swedish censuses which distinguished noblemen, prominent burghers, tradesmen and others.

One of the important determinants of the social position of a family unit is the source of income of the head of the family so in population analysis the economic variables tend to dominate the social variables. Economic activity can be divided vertically into levels of responsibility which are usually linked with levels of reward, or can be analysed horizontally by division into the types of economic activity undertaken, such as agriculture, mining, manufacturing, transport or services which each have their own particular hazards. There may be an important difference between the number of persons registered in an occupation and those reported by a census. The register records the professional qualifications of a person whereas the census will record the major economic activity, thus people unqualified will be included in the census and qualified persons practising another profession will be excluded. These remarks are particularly relevant to health professions where the number of doctors, nurses, or pharmacists economically employed may differ from the numbers registered.

In comparisons between countries of different cultures or economic levels, there may be entirely different methods of analysing this type of data to be useful and meaningful and it is necessary to try to standardise these. The International Labour Office have made recommendations concerning the methods of classification to be used, but the actual interpretation of these are dependent entirely upon local conditions. Thus the transport industry in North America will differ completely from the same industry in S. East Asia and the mining industry in S. Africa differs completely from the mining industry in Western Europe.

II.2 Types and Sources of Error

II.2.a Planning

The planning and preparation for a census is fundamental to the success of its outcome, any errors or deficiencies at this stage will be irretrievable and will affect every aspect of the reports. The basic mapping and survey work must be accurately done so that there is no overlap between territories with no omissions and boundaries shown on the map can be easily interpreted in the field. The date of the census must be carefully selected to avoid major holidays when people may be away, or times of heavy work such as harvest when people may be too busy. It should not be on the first or last day of a month; preferably it should be close to the 30th June as this would give the mid year populations required for mortality rates. It should be at a time of the year when transport is easy. The population must be carefully prepared by publicity and be reassured about the absolute confidentiality of census data and they should become involved so that they realise its importance. The enumerators must be carefully chosen and trained which will be a difficult problem due to the large numbers involved and the short period for which they are employed. The definition of terms used in the census schedule should be carefully and precisely defined so that there is the minimum of confusion and misunderstanding. When second languages have to be used the validation of the exact translation is especially important. The schedule itself must be pre-tested in small samples to ensure that the population interpret it correctly and that it is feasible to complete it within a reasonable time. Any sampling methods which are to be used in conjunction with the complete census must be carefully defined and listed.

The only way to prevent such errors in planning is for the Census Office to complete all its planning some considerable time before the census data and to leave adequate time for training the enumerators in the use of the census forms. It will not be difficult to detect the effect of errors at this stage such as failure to understand a definition, or special sensitivity by the population concerning a particular question, but it will be highly impossible to amend these matters if it has been left too late.

The Census Office should always consult the department responsible for mortality
analysis at an early stage in the planning because many difficulties in the definition of terms and in the detailed phrasing of questions may arise which will have different implications when deaths are being registered and will thus influence mortality analysis. Conversely, the mortality analysts must make their views known to the census office.

II.2.b Incomplete enumeration.

Failure to obtain complete cover in a Census will never be random as certain subgroups within the population will react differently from others. For example, there was marked underenumeration of the Southern States of the U.S.A. in 1870 after the Civil War and considerable omissions of Moslem communities in the Indian census of 1961; in England and Wales 1971, there were omissions of immigrant populations due to unfortunate publicity. As deaths occurring in such communities will still be registered adequately the consequence of such omissions is an apparent increase in calculated mortality rates.

Within households certain persons may be omitted such as new born children or infants up to the age of school; resident servants also are frequently omitted. Young adult males may be omitted if the census is believed to be associated with conscription, but they may also be omitted because they are much more likely to be away from their home at the time of the census than are other groups of the population. Similarly, elderly people or dependent relatives residing in a house may be omitted.

The homeless, travellers, nomads and vagrants, also present difficulties and it is essential that some system be devised to ensure that the majority if not all these persons are counted. In countries where there are nomadic tribes it is essential to contact the tribal leaders before the census and to ensure that a reasonable estimate of their number is obtained.

II.2.c Double counting

If it is not possible to complete the census in one day it is always possible that some people may be counted twice and steps should be taken to prevent this. In some societies people may have multiple rights at different places of residence and if they are enumerated at each residence it is essential that their returns should be transferred to one principal place of residence at which a birth or death certificate is likely to be registered.

One method to detect double counting is to select a sample of persons all of the same name and to examine the personal details of each individual such as sex, date of birth, place of birth, to see whether duplication has occurred.

The consequence of such over reporting of census populations could be an apparent reduction in mortality rates. Usually, however, errors of over enumeration will be much less than those of under-reporting.

II.2.d Inaccuracies of information

The most important variable for purposes of mortality analysis is the age, yet it is difficult to ensure that this is accurate. Infants who are under 12 months are often called one year old; there is an increase in the number of children reported just before the legal school age and there tends to be an increase of reporting at the age of majority; at all ages there is an excess of people whose ages end in 0 or 5 and some preference for ages terminating in even numbers rather than odd numbers. There is an over reporting of persons at the age at which pensions become payable and there is some over statement of age amongst the elderly. Some ages, of course, are not stated. These inaccuracies will effect different races and classes in different ways according to the literacy or independence of the group.

An important source of error is the simple human inability to write down a number correctly. So a bishop may be described as aged 65, whereas, it is obvious to the enumerator that he is approximately 60. A simple check question will decide that he intended 61. Enumerators should, therefore, be instructed to be alert for such errors.

It is also difficult to ensure completely accurate reporting of marital status. All monogamous societies enumerate all married men. Common law marriages may
be reported or not, persons who are living apart, but not divorced, may report themselves single, married or even divorced. Widows living alone may report that they are married.

Occupation is very difficult to define and to classify accurately. There will always be a tendency to exaggerate the importance of a position held and the householder making the census report may not know the details for individual members of the household.

People who hold more than one employment, or who are working part time may have difficulty to decide which occupation to report. People who are temporarily unemployed will tend to report their last employment, especially if it offers some social status. People who trained to a skilled profession, but are temporarily employed outside this profession, will often report their professional status rather than their employment.

Similar problems of ensuring the accuracy of social, religious, educational and economic data arise. The problems that arise will usually differ in different societies according to the social environment that prevails. It is not possible in a manual intended for many types of communities to be aware of, or to describe, the types of bias that might arise. It is the responsibility of a mortality analyst to be aware of the conditions in the society that he is examining and to be sensitive to the possibility of errors of reporting.

II.2.e Errors of data processing

All census data have to be translated into machine compatible form. This usually implies that the data have to be coded into numeric form. To do this requires a complete army of clerks who perform dull routine coding exercises. Very careful training and supervision of these personnel is required. Routine checking of all their work must be done. The methods of detecting errors and correcting at this stage is the same for all forms of data processing. In Census data, however, the size of the problem presents special difficulties.

Similar problems arise when the data has to be punched on to cards and finally there are possibility of machine errors within the computer system.

II.2.f Errors of printing and publication

There are of course, always errors of printing and careful proof reading is required which again is a dull routine job, but very necessary. Errata list will be published and readers should always check to see whether any results are altered by such lists. With recent advances of direct print out from a computer there should be considerable improvement in the accuracy of printing.

II.3. Detecting and Correcting Errors.

II.3.a Planning and enumeration

In any survey work which involves two human persons, one making a statement and the other recording it, there are bound to be many omissions and discrepancies and this is particularly true of a census when an attempt is made to survey a whole population. No such work can ever be free of these errors, but it is necessary to ensure that they are as infrequent as it is possible to achieve and that some estimate of their magnitude and direction must be made. In the original planning of the census some such scheme for error detection and correction must be incorporated.

The precensal tests which have already been discussed in Chapter II, Section 2, will form a part of this quality control system as they will have shown which parts of the census schedule have proven difficult for the subjects to understand and it will also have indicated that some particular section of the population have more difficulty in responding than do others. The planning system will then be directed towards modifying the schedule, but if it is not possible to eliminate a question, there will always remain some difficulty in replying to it, so particular attention will be paid to the wording of this question on the schedule and to the training of the enumerators.

If a particular section of the community is foreseen to have difficulties because of
linguistic, or educational, or housing problems, again particular attention will be paid to
the organisation in these areas; additional enumerators might be required and special training
given to them.

The first line of defence and the most important against errors is the skill of the
enumerators. The enumerator must be sure that he had complete coverage of all living persons
in his census tract on the night of the census. He is reliant on the good will of the
individual householders so he cannot lose their confidence, but nevertheless he must prompt
them to ensure that all new born children, any possible deformed or demented relatives hidden
in the house, any elderly people, any persons temporarily staying overnight, are all included
in the schedule. He should enquire of the householder what are the exact limits of his
territory and reconcile these with all the neighbouring returns to ensure there are no omissions.

The enumerator must also be the guarantor of the quality of the data. He will see
that every question required has been answered for each person, and draw the attention of the
householder to omissions. He will see that major discrepancies are detected at once, such as
a conflict between name and sex, or age and marital status, or economic status and education.
The planners will have provided him with a list of discrepancies which may occur and which he
should verify.

II.3.b Post census surveys

Every census should be followed by a post census sample survey. Post census surveys
are required for two purposes; to estimate the completeness of enumeration and to test the
accuracy of the information obtained. It is desirable that the responsibility for the post
census surveys should not be undertaken by the census organisation, but should be done
independently, but if done by the Census Office it should be independent of the main
organisation. This is equivalent to the general principle that the auditors of a system
should not also be the accountants.

To establish completeness, a random sample of a number of census tracts will be
selected by some appropriate method which may be stratified according to some relevant
parameters and these tracts will then be re-enumerated a second time with only a short delay
between the census and the survey. As this sample need only be a small proportion of the
total, a more intensive study can be directed towards ensuring completeness of enumeration
and if possible a system for detecting the households within the census tract different from
that used at the census should be used.

To establish accuracy the same random sample may be used as is used to establish
completeness, or alternatively, a second sample can be drawn solely for this purpose. The
population concerned in the sample for accuracy must be advised that the information obtained
will be compared with the original data only for statistical purposes and that although the
confidentiality of the census returns will be technically broken, this will not in any way be
referred back to the individual concerned. The enumeration for the accuracy check will be
carried out by face to face interview with if possible each one of the persons on the schedule
rather than with the householder. It will thus provide a measure not only of the accuracy of
the returns, but also of their validity.

A problem in theoretical sampling technique arises here; the sampling frame for the
post census check is the census tract, and the enumeration is based upon households, but the
ultimate tabulation is based upon persons which may give rise to a biased estimate. If the
characteristics of an error which relates to persons varies according to the size of a house-
hold, there will be a bias in the estimation of the mean value and appropriate weighting of
the error will be required. As an example this is a serious problem in immigrant groups who
fail to understand the census schedule and who are living in larger households than the rest
of the population and hence have a reduced probability of being included in a sample frame.

II.3.c Inter census comparisons

A census office will usually be in a position to make an accurate forecast of the
total population of an area based upon previous information, the vital statistics of the area
and the knowledge of changes in migration. It will thus be in a position to validate its own
census data from its own estimates. Obviously there is a serious danger of a circular
argument here and agreement does not necessarily validate either method, but a discrepancy
does draw attention to an area in which special study is required.
Similarly the characteristics of a previous census may suggest that there have been errors or omissions in enumeration of certain characteristics. For example if the sex ratio in the previous census ten years ago was 1.05 at ages 5-9, but ten years later, allowing for increased male mortality, there is a deficiency of males 15-19, there may be under-enumeration of males, there may be mis-statement of age, there may be differential migration out of the country or there may be evasion of military service.

These intercensal comparisons are discussed further in Chapter II.4. where estimation methods are dealt with in detail.

II.3.d Internal consistency checks

Consistency checks take two forms, first internal consistency of individual returns, and secondly, consistency of tabulated results.

The internal consistency of the census data will itself act as an estimate of the quality of the data. In the same way that the enumerator has checked the internal consistency of the individual returns so will the computer be programmed to check for internal inconsistencies in the data on the punched cards and will exclude all impossible codes. Each card with an apparent inconsistency will then be compared with the original household return to discover whether the error has arisen in coding and punching, or whether it was there originally. It will not be possible at this stage to interrogate the enumerator or the householder and the coding clerk will have to make some arbitrary amendment to the schedule. For example if a fisherman is shown as female, but had the name John, the coder would amend the sex of the person rather than change the occupation to fish wife.

Automatic discrepancy coding can be programmed so that if such a discrepancy arises the computer is instructed to accept the mean value of all other similar cases. So if a bishop is shown as aged 16 years, the computer would reject this value and substitute the mean value of the age of all other bishops, whereas, if this error had been noticed by the enumerator he could have corrected it precisely to 61.

A further check of internal consistency can be done from the preliminary tabulations, especially where there is the possibility of examining a continuous function such as the sex ratio which should start at a level of about 1.05 and decline through life as males die more rapidly than females. If there is a reversal of this ratio then there is either an accumulation of males in a certain age group, or a deficiency of females, or a mis-statement of age. This often happens around the age of maturity or the age of retirement.

A similar example is the well known number preference shown for ages ending in 5 or 0 which can be measured by Whipple’s Index which is five times the ratio of the sum of the number of persons whose ages terminate in 0 or 5 between the ages 23 to 62 to the total number of persons in that age range. If this index departs widely from 1.00 there is a need to graduate the ages to obtain a smooth progress or else to operate only in ten year age groups.

II.3.e Linkage with vital statistics

Census data may be linked with data derived from vital statistics. A random sample of births registered in the year preceding the census can be drawn and the infant sought for in the census returns. First the infant itself may be identified and its sex and age at census determined, but also the personal characteristics of the parents provided at registration can be compared with that provided at the census. Similarly a sample of deaths in the year following the census can be examined and compared with the census returns; this latter exercise is particularly necessary in the case of studies of occupational mortality.

Marriage certificates both before and after the census date can also be checked for comparability of data. To ensure the security of the census data, however, all these comparisons must be kept internal to the census and should not be used to correct data, only to estimate internal inconsistencies of the census data.

The published data on vital statistics and infant mortality can also be used to check the completeness of the returns for the newborn which will usually be deficient. The monthly returns of births and infant deaths will allow an accurate estimate to be made of the number of infants under 1 year in the community at the date of the census. If there is a discrepancy
the vital statistics are likely to be more accurate.

II.3.f U.N. evaluation

The U.N. Demographic Year Book gives estimates of the presumed degree of accuracy of population estimates based upon four major elements:

1) The nature of the base measurement of the population
2) The time elapsed since that measurement
3) The methods used to adjust the estimates
4) The quality of the adjustments.

These elements are analysed according to the code listed below:

1) The nature of the base measurement of the population:
   a) a complete census
   b) a sample survey
   c) a partial census or register
   d) conjecture of base not determined

2) The time elapsed since base measurement.
   Stated time in years.

3) Method of adjusting population estimates:
   a) a continuous population register
   b) balance of Birth, Deaths and Migration
   c) assumed rate of increase
   d) no adjustment but base figure held constant
   x) sum of regional estimates
   .... Method of adjustment not determined

4) Quality of adjustments of estimates of types 3(a) or 3(b):
   a) Population balance adequately accounted for
   b) Population balance assumed to be adequately accounted for
   c) Population balance not adequately accounted for.

Any analyst who is using data from populations with which he is not familiar should examine these estimates carefully and consider the affect they may have upon mortality rates.

II.4 Estimates of Population

II.4.a Introduction

A census is taken at infrequent intervals which gives great detail concerning the population once every decade. For purposes of analysis of vital statistics, it is essential to have details for each year concerning the age and sex structure of the population for the whole state and for the principal cities and districts and social groupings of a state. This is done by estimation.

It will be these estimated figures of the population which will be used in mortality analysis to calculate rates and hence it is most important that the mortality analyst understands how the demographer obtains them.

II.4.b Intercensal estimation

The simplest method of estimation is to derive the size of the population and its distribution as a point of time between two censuses, intercensal interpolation. Two methods of estimation may be used, either a pure mathematical approach or a use of the components effecting the population mathematical models.

Mathematical models. The mathematical approach is simple and easy, but it assumes that the changes taking place during an intercensal period are uniform. Two possible models
are available either that changes have occurred in a linear or arithmetic fashion, or that they have been exponential or geometric. If there has been an annual exponential rate of growth of \("r\) for the total population between two censuses with enumerated total populations \(P_1\) and \(P_2\) and which had a time interval of \(T\) years between them, then the exponential rate of growth can be estimated by the equation:

\[ P_2 = P_1 (1 + r)^T \quad \text{(1)} \]

which gives,

\[ r = \left( \frac{P_2}{P_1} \right)^{1/T} - 1 \quad \text{(2)} \]

Then the population at a specific time \(t\) is estimated by

\[ P_t = P_1 (1 + r)^t = P_1 \left( \frac{P_2}{P_1} \right)^{t/T} \quad \text{(3)} \]

\[ T \log P_t = (T-t) \log P_1 + t \log P_2 \quad \text{(4)} \]

This formula can then be applied to obtain values for \(r(x)\) for each age and sex or racial group and for each city or region within the state where \(r(x)\) is the relevant annual rate of growth of the specific section of the population being studied.

In the case of the linear model with an annual arithmetic growth assumed to be \("a\) per annum, the annual increment is obtained from the equation:

\[ P_2 = P_1 (1 + aT) \quad \text{(5)} \]

which gives,

\[ a = \frac{(P_2 - P_1)}{(P_1\cdot T)} \quad \text{(6)} \]

then for the population \(P_t\) at an interim time, \(t\) years after \(P_1\) was enumerated, the estimation will be:

\[ P_t = P_1 (1 + at) \]

and similarly specific values \(a_x\) can be obtained for each age, sex, social or regional grouping that is required.

Component Methods. A more satisfactory method when full details are available is to use the total birth \((B)\) and total death \((D)\) registration together with net immigration \((M)\) statistics during the interval of \(T\) years between the two censuses to estimate the changes in population. Then the estimate population at the time of the second census should be:

\[ \hat{P}_2 = P_1 + B - D + M \quad \text{(7)} \]

This value will not agree exactly with the actual population at the second census \(P_2\). There will always be a discrepancy \(E\) in the formula which is called the error of closure, \(E\).

\[ E = P_2 - \hat{P}_2 = P_2 - P_1 - B + D - M \quad \text{(8)} \]

This error will usually be attributable to discrepancies in the migration statistics, but it may be due either to under-registration of births or less commonly to under-registration of deaths. Sometimes it may be due to differences in the methods and standards applied in enumerating the two censuses.

The assumption is then made that the closing error was cumulative during the whole of the time \(T\) and, therefore, it is necessary for each single interior population estimate to include an appropriate error estimator.

Thus,

\[ P_t = P_1 + (B_t - D_t + M_t) + E(t/T) \quad \text{(9)} \]

Errors of closure will differ markedly at different ages and may even be of different signs at certain ages. It will be necessary to apply the component estimation to each age and sex group and to any regional, ethnic, or social group if component details are available.

Clearly such an error of closure cannot be calculated until after the second census has been taken and therefore the use of this technique will be retro-active, but in countries with considerable changes in migrant populations it is important that such errors should be corrected. For this reason in some countries historic series of statistics may differ from currently published data.
II.4.c Post censal estimation

The most important estimation problem that a Census Office has to solve for the purposes of analysis of mortality statistics is the constitution and size of a population in the years immediately following a census where no subsequent census is available, and these estimates will form the population at risk used in the calculation of most rates in vital statistics. The mathematical model for this is based upon the previous intercensal estimates and from two previous censuses with an interval of T years between them, age and sex, specific values for \( r_x \) or \( a_x \) are obtained and then \( P_t \) the current population when \( t>T \) is estimated as

\[
P_t = P_2(1 + a(t - T)) \quad \text{for the arithmetic model}
\]

or,

\[
P_t = P_2(1 + r)^{t-T} \quad \text{for the geometric model}
\]

and this estimated for each age and sex group.

The component method of post censal estimation is a little more difficult as it cannot be assumed that the error of closure which was discussed in Chapter II.4b. will operate outside the intercensal period and no method of estimating this error directly is available without a subsequent census.

The census data should be carefully analysed into single years of age with careful adjustments for the first five years of life and particular attention being paid to smooth the tabulation of stated ages at the decades and quinquennia of age as all human populations have a number preference for these ages. For the first year of experience all births should be added then the deaths subtracted from each year of age which will give an estimate of the first year of experience and then all ages should be increased by one year and the steps repeated annually.

Migration presents serious difficulties as the age distribution of migrants is seldom known accurately; some estimate has to be made, however, of the age pattern of both immigrants and emigrants separately and these added or subtracted from the total.

One special and important post censal estimation that is always required is the age distribution of the population as at the mid year June 30th/1st July in the year of the Census. The census will usually have been taken at some convenient date other than June 30th and it is necessary to adjust all ages and all population movements for the few months that intervene. This should be done by the component method and by assuming that all birthdays of persons within one year of age are equally distributed throughout the year so that an exact population proportional to the number of days between the Census and the mid year are promoted to the higher age and births and deaths are amended as required.

A proportional method is sometimes used for estimates of age and sex structures within small regional populations or occupational groups. It may be simple to estimate the total population of a group from economic or housing statistics when the specific vital changes are not known. Having estimated the population total the age and sex distribution is assumed to be the same as at the time of the Census. There are obvious dangers in the use of this method and it should only be used if other methods are not possible.

In summary the mortality analyst will not usually be responsible for the Census, for analysing the Census data, nor for making estimates of population at mid year intervals. But this information forms the very foundation upon which all analysis of mortality is dependent and hence it is of the utmost importance that attention should be paid to all details of how this work has been done.
CHAPTER III

MEASURES OF MORTALITY

1. Introduction

Page

2. General Statistical Measures:
   a. Ratio
   b. Proportion and Prevalence
   c. Rate and Incidence
   d. Weighted Average and Index Numbers

Page 35

3. Primary Measures of Mortality:
   a. Crude Death Rate
   b. Age Specific Death Rate
   c. Probability of Dying
   d. Infant Mortality Rate
   e. Neonatal Mortality Rate
   f. Poetal Death Rate
   g. Perinatal Mortality Rate
   h. Maternal Mortality Rate
   i. Disease Specific Mortality

Page 37

4. Adjusted Measures of Mortality:
   a. Age adjusted rates
   b. Direct method of adjustment
   c. Indirect method of adjustment
   d. Standardized mortality ratio
   e. Adjustment for factors other than age
   f. Adjustment for several factors at once

Page 41

5. Standard Errors of Mortality Rates:
   a. The concept of a standard error of mortality
   b. The standard error of the probability of dying
   c. The standard error of the mid year death rate
   d. The standard error of the age adjusted rates
   e. Confidence limits of a mortality rate
   f. Tests of Significance between mortality rates
   g. Standard errors of sample mortality rates

Page 49

6. Other Measures Related to Mortality

Page 56

III.1 Introduction

The activities of the census and of the vital registration system produce information concerning a community in the form of numbers. For some purposes of planning health and social services, the number is all that is required, for example, the number of school places will relate to the number of live births, the number of jobs to the number of adult males age 15-65, the provision of health facilities will relate to the total population with special weighting for the elderly. Other factors may be measured rather than counted and in particular the expenditure of resources on a Health Service will be measured in terms of money expenditure.

But to give real meaning to these numbers they must be closely connected with one another so that differences in various sectors of the community can be examined and to do this it is necessary to develop certain statistical measures.
The simplest measures which are used in vital statistics are derived from four kinds of mathematical expressions: ratio, proportion, rate and index. Very often the number of vital events in a specified sector of the community forms the numerator and the relevant population adjusted as necessary forms the denominator. Let us first examine these mathematical expressions to see how they are defined and calculated.

III.2 General Statistical Measures

III.2.a Ratio: A ratio is the basic measure of the relative magnitude of two numbers. All other expressions used in vital statistics, such as proportion, rate and index, are special cases of a ratio; a ratio summarises the arithmetical relationship between two characteristics A and B which can be counted in a population; when the numbers of occurrences are a and b respectively then the ratio, r, is defined as:

\[ r = \frac{a}{b} \quad \text{or} \quad a : b \]  

To eliminate decimal points this simple ratio will often be multiplied by a power of 10. Thus the ratio becomes

\[ r = \frac{a}{b} \times 10^k \]  

where \( k \) is a non-negative number and if \( k = 0 \) the expression is a simple ratio with unity as the basis of comparison, if \( k = 2 \) the expression is a percentage, if \( k = 3 \) it is per mille, and if \( k = 6 \) the basis is a million. It is important, however, that such a ratio should not be calculated to a spurious accuracy of significant figures. If the number \( b \) is only of the order of 100 for example, only 2 figures are justified.

The calculation of a ratio is intended to provide further meaningful information concerning the behaviour of the characters A and B. By calculating a ratio the numerator is in some ways adjusted with respect to the denominator which becomes a reference value and leads to an expected value for the numerator.

Example: If in a hospital the operations (A) recorded in one year are 21,000(a) and the number of anaesthetists (B) working in the hospital are 30 (b), the crude numbers alone tell something about the work load of the hospital and about the staffing, but an administrator who wished to consider the expansion of the hospital and how many additional anaesthetists he might need would have to do further analysis. He could start by calculating the ratio

\[ r = \frac{a}{b} = \frac{21,000}{30} = 700 \text{ operations per anaesthetist per year.} \]

This would suggest that one anaesthetist was required for 700 operations; in this case \( k = 0 \) (\( 10^0 = 1 \)) and the ratio relates to the work of one anaesthetist.

Conversely, he might be interested in comparing different hospitals with different numbers of operations, then in order to adjust for these different numbers he would calculate the ratio:

\[ r = \frac{b}{a} = 30/21,000 = .00143 \text{ anaesthetists per operation} \]

but the figure .00143 anaesthetists per operation is rather meaningless so he would put \( k = 4 \) (\( 10^4 = 10,000 \)) and thus:

\[ r = \frac{b}{a} \times 10^k = 14.3 \text{ anaesthetists per 10,000 operations} \]

and this value of \( r \) is now a useful tool for comparing different hospitals with each other, adjusting for the differences in workload. But a ratio is not limited to the relationship between two counts but extends to all measurable variables. Example, if the mean height of a sample of women is 165 cms and the mean height of a similar sample of males is 175 cms, then the male to female ratio of height is

\[ r = 175/165 = 1.06 \text{ which implies that on average men are 6 per cent taller than women.} \]

III.2.b Proportion and prevalence. A proportion is a ratio of a special kind in which the number (a) of persons with a characteristic A is part of the total population (P); thus, the population is of the form \( P = a + \text{remainder} \) and the characteristics of the remainder are not necessarily specified. Hence, the proportion, \( p \), is defined as

\[ p = \frac{a}{P} \]

where there are only two characteristics A and B with frequency a and b this becomes

\[ p = \frac{a}{a+b} \]
Prevalence is a widely used concept in epidemiology and is frequently referred to as a prevalence rate; it is not a rate as defined below, but it is simply the proportion of individuals who exhibit a specific characteristic in a defined population. Prevalence may be defined as either the proportion of individuals with the characteristic at a given point of time (point prevalence) or else as the proportion which manifests the characteristic during a period of time (period prevalence).

If a population can be divided into the number of units exhibiting particular manifestations A, B, C, etc. of a certain characteristic which do not overlap and which include the complete population, then a tabulation of the total population distinguished according to this variable is a frequency distribution and the proportions making up the total is a proportional distribution.

Examples: 1) Area of residence of citizens 2) Marital status of all adult women 3) Causes of death for all deaths 4) Parity of mothers delivering a child 5) Age groups in a population

Proportions also lead to the concept of probability in a defined population whether theoretical or real, and the frequency concept of probability depends upon the estimation of proportions.

2.c Rate and incidence. A rate is a measure of the frequency or rapidity with which events such as births, deaths, onset of illness, occur in a community in a defined interval of time.

Thus: \[ \text{Rate} = \frac{\text{Events per unit of time}}{\text{(Relevant Population)}} \times 10^k \]  

A rate is truly a ratio in which the events to be included are defined by a period of time:

There are four elements in a rate: 1) the number of events (a) of characteristic (A); 2) the defined period of time (t); 3) the relevant population (P); and 4) the parameter of scale k.

Then: \[ \text{Rate} = \frac{a/t}{P} \times 10^k \]

Usually the period of time over which the events are measured, e.g. one year, will be the same interval as that for which they are reported in which case \( t \) will be 1 and may even be omitted from the formula. It is important, however, to remember that it is present because a rate should always be a measure in time. In some tabulation, however, it might be desirable to calculate the equivalent annual rates from monthly or quarterly data in which cases, \( t \), would be 1/12 or 1/4. Or perhaps an annual rate may be calculated from quinquennial records when \( t \) would be 5.

The denominator of a rate should contain - as closely as possible - the population who are "at risk" of incurring the event or character which is to be counted in the numerator as this gives the best measure of what is really occurring in the community and of how this may change as the population changes. But in some demographic rates a large part of the population which has never been at risk is counted in the denominator, for example, the crude birth rate per 1,000 population per annum where all males, children and elderly females are not at risk, or the marriage rate where those already married or under the age of marriage are included. Theoretically, such crude rates are badly constructed and to obtain more meaningful information it is necessary to refine them by qualifying the population "at risk"; but as first estimates, they may provide a valuable reference frame.

Example: A maternal mortality rate, should be a measure of the risk of death associated with pregnancy, child birth and puerperium. Thus the numerator will include the number of reported deaths ascribed to causes related to these conditions. The denominator should include the number of people at risk of death due to one of these causes; the only people who are at such risk are those women who are pregnant, or in childbirth or in the puerperium at any time during the year. But as it is quite impossible to count or even estimate the number of women
in a population who are pregnant at any given time, it is clearly impossible to measure the quantity that should go into the denominator. What is done in practice is to enter in the denominator the number of live births reported in the time interval in question, which gives an impression of the number of women at risk and is a good reference group. See III.3.b.

The usual unit of time used for calculating rates in demographic studies is one year, but if information is available for a different period such as five years, or one month, or a week, then the observed events must be adjusted according to the time interval in which they were observed (t). Particular attention should be paid to monthly rates, however, which should be adjusted by the factor d/365 where d may range between 28 and 31 according to the month.

In epidemiological studies incidence rates are calculated for the number of new cases of a disease when it is first recognised in a person in a defined community. The denominator is usually the total population in spite of the fact that many persons may not be susceptible to the disease as they are immune or are already sick and, hence, it may be necessary to refine these rates. Incidence, however, is a true rate.

Incidence and prevalence are often confused with each other, but should not be so as they set out to measure quantities which are inherently different. Prevalence indicates the number of events present in a population; incidence indicates the speed at which they are occurring over a period of time.

A disease such as tuberculosis may have a relatively high prevalence but a low incidence; whereas, measles may have a high incidence but a low prevalence. Other diseases such as rheumatoid arthritis or diabetes may have a high prevalence but the true incidence cannot be determined because the onset of the disease cannot be accurately determined.

2.4 Weighted averages and index numbers. A weighted average is an average of a set of quantities to which has been attached a series of weights in order to make an appropriate allowance for their relative importance; for example a weighted arithmetic mean (W.A.) of \( x_1, x_2, x_3, \ldots, x_n \) with weights \( w_1, w_2, w_3, \ldots, w_n \) is given by

\[
W.A. = \frac{\sum_{j=1}^{n} W_j x_j}{\sum_{j=1}^{n} W_j}
\]

A weighted average may be used to obtain an estimate of the value of some function which is not in practice susceptible to direct observation or measurement itself. Such measures are frequently used in economic studies in agriculture and in sociology, they are used in mortality analysis to adjust for changes in age or other demographic characteristics of populations.

When such weighted averages have been obtained, their value increases if they can be measured and compared over a period of time or between several different areas. To do this it is necessary to keep the set of weights constant; the weighted average of a set of items of the current year or relevant area is used as the weighted average, using the same set of weights on the same standard items from a former reference time or standard area. This is then an Index Number which is a ratio of two weighted averages.

\[
\text{Index} = \frac{\text{W.A. (Study)}}{\text{W.A. (Standard)}}
\]

Such a method is always required where within a large community there are many sub-populations each of which is subject to different and changing influences. Then to compare the different communities or to compare the same community with itself at different times it is necessary to observe what would happen in a standard community in which the proportions of the items are held constant.

III.3 Primary Measures of Mortality.

The analysis of mortality data calls for the use of all these types of statistical measures and it is necessary for an investigator to remind himself frequently whether a particular measure is a ratio, a proportion, a rate or an Index Number, or whether it is only an approximation to one of these.
III.3.a Crude death rate:

If the number of deaths recorded is D in a defined population P during one year, then the crude mortality rate (m) is defined as follows:

\[ m = \frac{D}{P} \times 10^k \]  

(Note: m is a ratio and also a rate but it is not a proportion).

A Dictionary of Statistical Terms (1954), W.G. Kendall and W.R. Buckland, defines a death rate as:

"Death rate: The number of deaths in a given period divided by the population exposed to the risk of death in that period. For human populations the period is usually one year and if the population is changing in size of the year the divisor is taken as the population at the mid-year'.

The denominator should be, as closely as possible, the population "at risk" because this gives the best reference value. This is obtained by using the estimate of the mid-year population which assumes that the effects of births, deaths, and migration will cancel each other during the course of the year if these are changing at a relatively constant rate during the year.

III.3.b Specific death rates:

Clearly rates referring to the whole population give only limited information because they neglect the large differences between sub groups in the population. Separate death rates may be calculated for the male and female populations, but if these are done for all ages they still remain crude rates. The value of the information provided by a death rate increases with increasing homogeneity of the sub groups and the most important of such groups are the age and sex groups. Each sex is divided into decennial or quinquennial age groups and the mortality rate for each is calculated; then these rates are known as the age-sex specific rates.

If \( D_x \) is the number of deaths occurring in one year all of whom are aged \( x \) at death and \( P_x \) is the mid-year population all of age \( x \) precisely, then the age specific mortality rate at age \( x \) is:

\[ m_x = \left( \frac{D}{P} \right)_x \times 10^k \]

Similarly, if \( (D_{x:t}) \) is the number of deaths registered in one year at ages \( x \) to \( x+t-1 \), and \( P_{x:t} \) is the mid-year population of ages \( x \) to \( x+t-1 \), then the age specific mortality rate for the age group \( x \) to \( x+t \) is:

\[ m_{x:t} = \left( \frac{D_{x:t}}{P_{x:t}} \right) \times 10^k \]

Similarly, many other specific rates can be defined for special groups such as racial groups, marital groups, occupational groups, but as each of these will have very different age structures it will usually be necessary to consider the age-sex specific rates within these sub groups.

Example: The age structure of a group of widows will be very different from that of a group of married women and hence a comparative study of mortality by marital state will need to allow for this difference.

III.3.c The probability of dying:

The probability of dying is defined as the proportion of a population of a specific age and sex group who die during one year. Death rates as defined above are not true proportions although they are ratios and they are rates. But when it is necessary to calculate the probability of dying it is necessary to obtain true proportions. If \( N \) is the total living population at the beginning of a year and \( D \) the number of deaths observed out of \( N \) in the following year, then the probability of dying (q) during the year is defined as:

\[ q = \frac{D}{N} \]

(Note: q is a true proportion)

The relationship between q and the mortality rate, m, can be determined by calculating the mid-year population \( P \), which is:

\[ m = \left( \frac{D}{P} \right) \times 10^k \]
where \( a \) is the proportion of the deaths that had occurred by the mid-year. If \( a \) is assumed to be 0.5, then:

\[
P = N - 0.5D
\]

and hence,

\[
m = \frac{D}{(N - 0.5D)} \quad m = \frac{D}{1-0.5} \quad m = \frac{q}{1-0.5q} = \frac{2q}{2-2q}
\]

and conversely,

\[
q = \frac{m}{1+0.5m} = \frac{2m}{2+m}
\]

This definition of \( q \) holds for any population irrespective of the age structure of \( N \) and in some studies such as a group exposed to a highly toxic hazard their age distribution may be irrelevant. In human mortality studies, however, it is usual to concentrate on the values of \( q \) for a narrow age range.

The extreme restriction would be that the population \( N \) should consist of persons all of the same age exactly at the beginning of the year and this is discussed in more detail in Chapter IV, but in statistical data collection it is an unrealistic one. In a real population with migration and with persons entering and leaving the age group during the year, the definition is different. The deaths belong to a certain age group who were aged \( x \) on the day of their death when it is recorded, even though they may have been aged \( x-1 \) at the beginning of the year. The persons in the population who form the denominator belong to this age group, \( x \), for so long a period as they are exposed to the risk of dying at that age of \( x \).

In a population which has an equal distribution of birthdays throughout the year the number of persons alive in the age group, \( x \), remains constant although the individuals concerned will alter during the year and, hence, the equations above remain unaltered and, hence, the formulae for \( m \) and \( q \) are the same.


The neonatal mortality rate is the number of deaths occurring in the first 28 days of life per 1,000 live births. This is sometimes divided into early neonatal death rate which comprise deaths in the first seven days of life, days 0-6, and the late neonatal death rate which comprise deaths from day 7 - day 27. In both cases the denominator to be used is all live births occurring during the year of report.

III.3.d Infant Mortality Rate:

The infant mortality rate (IMR) is the measure of mortality of infants between birth and the exact age of one year. If in a defined community in one year the number of deaths occurring to infants under the age of 1 year is \( D_o \) and \( B \) is the number of live born in the same year in the same community, then the infant mortality rate is defined as:

\[
IMR = \frac{D_o}{B} \times 10^k \text{ where } k \text{ is 3}
\]

Note: that this is not truly a rate nor a proportion, it is a ratio for which the time over which it is recorded need not be specified so long as births and infant deaths are recorded simultaneously.

Strictly speaking, this definition does not provide an accurate estimate of the risk of death in the first year of life because some of the deaths occur to infants who were born in the preceding year and none of the births except those born on the first day of the year are followed up to their first anniversary date. Where the birth rate and the infant mortality rate are not changing rapidly this conventional method above provides a good measure of mortality in infancy and is easily calculated because both figures will be available. If, however, the rates are changing rapidly it may be necessary to make some adjustment which will be described in Chapter V. below.

III.3.f Foetal death rate.

The foetal death rate is an attempt to measure the mortality before birth and is dependent upon a complete and accurate register of foetal deaths. Let the number of foetal
deaths in one year be $F$. The population at risk of foetal death must be considered carefully; it is all live births $B$ together with the foetal deaths themselves, hence the foetal death rate (FDR) is defined as:

$$F.D.R = \frac{F}{(B+F)} \times 10^k \quad (k = 3)$$

This of course is influenced by possible under registration of both foetal deaths and of births and is subject to the problems of definition which will be discussed in Chapter V.

In some countries the estimated gestation of the foetus at the time of delivery is recorded and then an attempt is made to analyse the foetal deaths by three stages:

1. Foetal deaths born after 28 weeks gestation are late foetal deaths, frequently called still births ($S$).
2. Foetal deaths born after 20 weeks gestation but not more than 28 weeks are middle foetal deaths, and
3. Foetal deaths before 20 weeks gestation are early foetal deaths.

The completeness of registration will decline as attempts are made to include earlier events. There are also serious difficulties in the use of early foetal deaths as this is an open ended group.

III.3.g Perinatal mortality rate.

The perinatal mortality (PNMR) is an estimate of the loss of life around birth and is defined as the number of late foetal deaths and early neonatal deaths ($E$) from 1,000 births whether live or still:

$$PNMR = \frac{(S + E)}{(S+B)} \times 10^k \quad (k=3)$$

This is a useful measure as it avoids many of the inherent difficulties of determining whether a child was alive or dead at the time of delivery and also because the causes of death during the last trimester of pregnancy and during the first week of life are similar so it is convenient to deal with these two groups as one entity.

To summarise these rates of mortality during gestation and the first year of life we have:

<table>
<thead>
<tr>
<th>Time Interval</th>
<th>Type of Death</th>
<th>Mortality Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conception - 19 weeks</td>
<td>Early Foetal</td>
<td>Poetal Death Rates</td>
</tr>
<tr>
<td>20 - 27 weeks</td>
<td>Middle Foetal</td>
<td></td>
</tr>
<tr>
<td>28 - Birth</td>
<td>Late Foetal</td>
<td>Perinatal Mortality</td>
</tr>
<tr>
<td>Birth - 6 days</td>
<td>Early Neonatal</td>
<td>Neonatal &quot;</td>
</tr>
<tr>
<td>7 days - 27 days</td>
<td>Late Neonatal</td>
<td></td>
</tr>
<tr>
<td>28 days - 1 year</td>
<td>Post Neo-natal</td>
<td>Infant Mortality</td>
</tr>
</tbody>
</table>

Note: All of these measures of loss of life in the uterus or first year of life, are truly ratios and not rates as defined above.

III.3.h Maternal mortality rate.

(MMR) is the measure of the mortality of women attributed to puerperal causes during a year. If $D_m$ is the number of such women who died from maternal or puerperal causes during a year and $B$ is the number of live births during the year then:

$$MMR = \frac{(D_m/B)}{10^k} \quad \text{(where } k \text{ is usually } 4)$$

(Note: This is a ratio).

The population at risk should be theoretically all women ever pregnant during the year, but this is an elusive figure which it is difficult to obtain. The total number of
births should be reduced by the number of excess multiple births, but it should in turn be
increased by the number of foetal deaths. Moreover, some of the deaths attributed to
pregnancy may relate to births in the previous year and some of the women pregnant in the
current year will not give birth till the following year and hence will be recorded in that
year. Despite these objections the conventional mortality rate as defined above is a
reasonably useful measure of maternal mortality.

III.3.i. Disease specific mortality.

In the case of age and sex or area specific rates, the specificity of the rates
apply both to the numerator and the denominator used in their calculation, so this demands
that all the deaths for a given age and sex group are related to the population at risk of
mortality who are of the same age and sex or area. For disease specific rates the situation
is rather different and two uses of the term must be carefully distinguished.

Death rates may be calculated in which only those deaths attributed to a specific
cause of death such as tuberculosis or cancer are included in the numerator, but the
denominator remains the total population at risk. Such death rates will be known by their
specific names such as "the tuberculosis mortality" or "the cancer mortality" and so they will
be specific for a cause of death but will not relate to a specific sub-population. Within
these disease rates the same types of analysis by age and sex groups which are used for total
mortality can be applied. Such rates for a single cause of death will be smaller than those encountered in the total mortality and hence it is usual to calculate them per hundred thousand at risk or sometimes per million, (i.e. k=5 or k=6).

The rates in World Health Statistics Annual are an excellent example of disease
specific rates by age and sex for each country based upon hundred thousand population at risk
(k=5).

The term disease specific mortality may however be used in an entirely different
sense in which all deaths occurring to patients with a specific disease are used to calculate
a disease specific mortality rate.

This implies that the numerator of the rate is the total number of deaths, but the
denominator is specific to those persons who have a stated disease. In this sense "the cancer
mortality" means the deaths from whatever cause of patients who have cancer. The analysis of
such mortality rates must be done with considerable care as there is an inherent difficulty in
defining the true population at risk and the time from which they are to be included in the
population at risk. Such rates should only be used in a strict cohort sense as described in
Chapter VI in which the date of entry to the diseased population can be defined accurately.
It will usually be possible to do this by using the date of notification for communicable
diseases or the date of first registration of cancer patients who can then be followed to
death. In such cases the disease specific mortality rates are likely to be much higher than
expected in a healthy population and hence these rates are often published as per cent (i.e.
k=2).

As an example, disease specific rates are published for all cancer sites in "Cancer.
A Handbook of Epidemiology and Prognosis", J.A.H. Waterhouse, Churchill Livingstone, London and

III.4 Adjusted measures of mortality.

III.4.a Age adjusted rates. The crude death rate is quite misleading if it is used to make
comparisons between the mortality in two populations of differing age and sex proportional
distributions, and the age specific rates do not separately give a complete picture of the
mortality experience. Hence it is necessary to develop summary statistics which will allow
valid comparisons to be made between communities.

Example: The crude death rates in 1960 for two Florida Counties were:

- Pinellas Co. Population 374,665 Deaths 5,726 Crude death rates 15.3 per 1,000
- Dade Co. Population 935,049 Deaths 8,322 Crude death rates 8.9 per 1,000

The observed difference in mortality between these two counties is very large and
the ratio of the mortality in Pinellas to that in Dade is 1.72, from which it appears that
the risk of death in Pinellas Co. is about 70% higher than in Dade Co.
But Table III.1 shows that the difference in age structure of the two populations is very marked and that the age specific rates are higher in Dade Co. at every age, except 1-14. The observed difference in the crude rates is an expression of the large proportion of elderly people in Pinellas, 40% of the population are over 55 years of age, whereas in Dade Co. only 20% are over this age.

**TABLE III.1**

Proportional Age Distribution per 100,000 Persons and Age Specific Death Rates per 1,000 persons in two Florida Counties.

| Age Group | Population per 100,000 | Age Specific Death Rates Per 1,000 Persons | Ratio  
|-----------|------------------------|--------------------------------------------|--------
|           | Pinellas (ii) | Dade (iii) | Pinellas (iv) | Dade (v) | (iv/v = vi) |
| 0 - 1     | 1,500 2,000 | 28.2 28.8 | 0.98 |
| 1 - 4     | 5,900 8,000 | 1.4 1.2 | 1.17 |
| 5 - 14    | 13,900 17,400 | 0.6 0.4 | 1.50 |
| 15 - 24   | 8,700 11,600 | 0.8 1.2 | 0.67 |
| 25 - 34   | 9,000 13,300 | 1.4 1.7 | 0.83 |
| 35 - 44   | 11,100 15,000 | 3.0 3.3 | 0.91 |
| 45 - 54   | 11,100 12,600 | 7.7 8.2 | 0.94 |
| 55 - 64   | 13,900 16,000 | 15.9 16.7 | 0.95 |
| 65 - 74   | 17,600 7,300 | 28.9 31.1 | 0.93 |
| 75+       | 7,300 2,800 | 82.5 84.2 | 0.98 |
| ALL AGES  | 100,000 100,000 | 15.3 8.9 | 1.72 |

The problem to be solved is how to obtain a valid summary expression of the differences and there are various method of obtaining a solution.

The method of adjustment for death rates is essentially similar to that of obtaining any weighted average in which a series of weights are derived from a standard population and are then applied to the age specific rates of a number of separate study populations, Chapter III.2.d. The mortality experience of the study populations can then be compared one with another as all have the same weights attached to them.

In all methods of adjustment a choice has to be made of the weights which are applied to the age specific death rates in the study population. There are two different methods of approaching this problem which are referred to as the direct method of adjustment and the indirect method of adjustment.

For the various forms of direct adjustment the weights are proportional to the sex and age distribution known to exist in some standard population. Usually the "standard" population will be a real population such as a national population at the date of a census. Sometimes it may be a pooled population obtained by adding together several other populations. Sometimes it may even be a theoretical population which does not exist at all. For the equivalent average death rate the weights are proportional to the length of the age intervals used to calculate the age specific rates. For the life table method of direct adjustment the weights are proportional to the $\hat{L}_x$ of the standard population life table.

The notation required for the description and analysis of these different methods of adjustment become rather complex and the method used in this Chapter is to attach a prefix to symbols derived from a standard population, but to omit these where they refer to a study population. In the case of both standard and study populations a suffix is attached to indicate that age specific values are being considered but when total values are being considered these are omitted. These notations are summarized in Table III.2.

**TABLE III.2**

For the indirect methods of adjustment the weights are based upon the sex and age specific death rates experienced by some standard population in a standard period of time and the actual weights used are proportional to the age specific death rates in the standard population multiplied by the age distribution of the study population.
### TABLE III.2

Summary of the notation used for adjusted death rates

<table>
<thead>
<tr>
<th>Notation</th>
<th>In Study Population</th>
<th>In Standard Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age specific population</td>
<td>$P_x$</td>
<td>$S_P x$</td>
</tr>
<tr>
<td>Age specific deaths observed</td>
<td>$D_x$</td>
<td>$S_D x$</td>
</tr>
<tr>
<td>Age specific mortality rate</td>
<td>$m_x = D_x/P_x$</td>
<td>$S_m = S_D / S_P x$</td>
</tr>
<tr>
<td><strong>Age specific &quot;Expected&quot; Deaths:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1) Standard mortality in Study Population</td>
<td>$E_x = P_x \cdot S_m x$</td>
<td>-</td>
</tr>
<tr>
<td>2) Study mortality in Standard Population</td>
<td>-</td>
<td>$S_E = S_P \cdot m_x$</td>
</tr>
<tr>
<td>Total Population</td>
<td>$P = \sum P_x$</td>
<td>$S_P = \sum S_P x$</td>
</tr>
<tr>
<td>Total Deaths Observed</td>
<td>$D = \sum P_x \cdot m_x$</td>
<td>$S_D = \sum S_D x$</td>
</tr>
<tr>
<td>Total Mortality Rate</td>
<td>$m = D/P$</td>
<td>$S_m = S_D / S_P$</td>
</tr>
<tr>
<td>Total &quot;Expected&quot; Deaths (i)</td>
<td>$E = \sum P_x \cdot s_m x$</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(ii)</td>
<td>$S_E = \sum S_P x \cdot m_x$</td>
</tr>
</tbody>
</table>

### III.4.b Direct method of age adjustment

(1) **Use of a standard population.** The first method of adjustment for rates to make allowances for changes in the proportional frequency of the age groups is the Direct Method. Using this method the age specific rates for the study populations to be compared are all calculated and then these rates are multiplied by the number in each of the age groups in the standard population. The sum of these products is the total number of "expected" deaths if the study death rates had been operating on the standard population. Dividing these "expected" deaths by the total standard population gives the Direct Adjusted Death Rate.

Let there be two mid-year populations, a standard one which is denoted as $S_P$ and the study population which is denoted $P$ then the number of persons in the age group $x$ in the standard population is $S_P x$ and in the study population is $P x$ then the direct adjusted death rate $m_{dir}$ is given by the following equation,

$$m_{dir} = \frac{\sum (S_P \cdot m_x)}{S_P} \quad \ldots \ldots \quad 19$$

This may be compared with the general form of a weighted average where the weights are the population of specific age in the standard population as in equation (7).

$$W.A. = \frac{\sum (W_x \cdot m_x)}{\sum W_x} \quad \ldots \ldots \quad 20$$

The only remaining problem is to select the standard population and it is customary to use the national population for this purpose adjusted proportionately to a total population of 1 million persons. Table III.3 then shows how this is applied to Pinellas Co. and Dade Co.

### TABLE III.3. here

The second column of Table III.3 is the proportional frequency distribution by age of the U.S. Population in 1960 for 1 million persons. Columns (iii) and (iv) are the age specific rates in the two counties and columns (v) and (vi) are the number of deaths that would have occurred to the standard population of 1 million if these rates had been incurred. The overall rates would have then been 8.2 per 1,000 in Pinellas and 8.6 per 1,000 in Dade, and the ratio of the adjusted rates is 0.95, an advantage of 5 per cent to Pinellas. This agrees with the suspicion aroused from the age specific rates in Table III.1 that Dade may be the less healthy area.
## TABLE III.3

<table>
<thead>
<tr>
<th>Age Group (Year)</th>
<th>1960 U.S. population (distribution per million)</th>
<th>Age specific death rates (per 1,000 pop.) (from Table 3.)</th>
<th>Expected deaths in 1960 U.S. population using county age specific rates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(i)</td>
<td>(ii)</td>
<td>(iii)</td>
</tr>
<tr>
<td>&lt; 1</td>
<td>23,000</td>
<td>28.2</td>
<td>Pinellas</td>
</tr>
<tr>
<td>1 - 4</td>
<td>90,000</td>
<td>1.4</td>
<td>28.8</td>
</tr>
<tr>
<td>5 - 14</td>
<td>199,000</td>
<td>0.6</td>
<td>Dade</td>
</tr>
<tr>
<td>15 - 24</td>
<td>134,000</td>
<td>0.8</td>
<td>1.2</td>
</tr>
<tr>
<td>25 - 34</td>
<td>127,000</td>
<td>1.4</td>
<td>1.7</td>
</tr>
<tr>
<td>35 - 44</td>
<td>134,000</td>
<td>3.0</td>
<td>3.3</td>
</tr>
<tr>
<td>45 - 54</td>
<td>114,000</td>
<td>7.7</td>
<td>8.2</td>
</tr>
<tr>
<td>55 - 64</td>
<td>87,000</td>
<td>15.9</td>
<td>16.7</td>
</tr>
<tr>
<td>65 - 74</td>
<td>61,000</td>
<td>28.9</td>
<td>31.1</td>
</tr>
<tr>
<td>75+</td>
<td>31,000</td>
<td>82.5</td>
<td>84.2</td>
</tr>
<tr>
<td>TOTALS</td>
<td>1,000,000</td>
<td>15.3</td>
<td>8.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.164</td>
<td>8,567</td>
</tr>
</tbody>
</table>

| Age adjusted rates (per thousand) | 8.2 | 8.6 |

There is clearly a loss of sensitivity in the use of the age adjusted rates instead of using all the age specific rates, but where there are more than two or three comparisons to be made it is not possible to compare the separate rates. In the USA for instance there are over 3,000 counties hence an age adjusted rate is essential for meaningful comparisons.

(ii) Use of a pooled population. If only a few comparisons are to be made and especially if two only are to be contrasted, it is sometimes considered inappropriate to use a standard population which may have quite a different age structure from either of the two to be compared. If this is the case a special standard population is constructed in which both populations are added together so that:

\[
\bar{s}_x = \frac{l_p}{x} + \frac{2p}{x} \\
\]

The technique is then precisely the same as before; the age specific rates in both populations are applied to the pooled standard population and a total adjusted rate is calculated.

When this was done for the example above the age adjusted rates were Pinellas 10.4 per 1,000 and Dade 10.9 per 1,000. The ratio is still 0.95. Again Dade is less healthy.

(iii) Use of an artificial population. One simple method of constructing the standard population is to assume that there are the same number of persons for each single age of life from 0 to 80, then the weights to be attached to each age specific rate are simply the age intervals involved in the grouping. This technique is known as the equivalent average death rate (E.A.D.R.). This is done in Table III.4. The ratio is 0.96. Again Dade is less healthy.
### TABLE III.4

Mortality in Two Florida Counties, 1960. Direct method of adjustment using hypothetical population with constant weights for each year as standard

<table>
<thead>
<tr>
<th>Age Group (Years)</th>
<th>Hypothetical Population</th>
<th>Age specific death rates (per 1,000 pop.) (from Table 2.)</th>
<th>Expected deaths in hypothetical population using county age specific rates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(i)</td>
<td>Pinellas (iii)</td>
<td>Dade (iv)</td>
</tr>
<tr>
<td>1</td>
<td>10,000</td>
<td>28.2</td>
<td>28.8</td>
</tr>
<tr>
<td>1 - 4</td>
<td>40,000</td>
<td>1.4</td>
<td>1.2</td>
</tr>
<tr>
<td>5 - 14</td>
<td>100,000</td>
<td>0.6</td>
<td>0.4</td>
</tr>
<tr>
<td>15 - 24</td>
<td>100,000</td>
<td>0.8</td>
<td>1.2</td>
</tr>
<tr>
<td>25 - 34</td>
<td>100,000</td>
<td>1.4</td>
<td>1.7</td>
</tr>
<tr>
<td>35 - 44</td>
<td>100,000</td>
<td>3.0</td>
<td>3.3</td>
</tr>
<tr>
<td>45 - 54</td>
<td>100,000</td>
<td>7.7</td>
<td>8.2</td>
</tr>
<tr>
<td>55 - 64</td>
<td>100,000</td>
<td>15.9</td>
<td>16.7</td>
</tr>
<tr>
<td>65 - 74</td>
<td>100,000</td>
<td>28.9</td>
<td>31.1</td>
</tr>
<tr>
<td>75+</td>
<td>50,000</td>
<td>82.5</td>
<td>84.3</td>
</tr>
<tr>
<td><strong>TOTALS</strong></td>
<td><strong>800,000</strong></td>
<td><strong>15.3</strong></td>
<td><strong>8.9</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>Adjusted rates</strong></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**

1. **Use of a specific population.** If each group of age specific rates are weighted according to the population distribution of their own population, then the adjusted total rate is the crude rate. It is, therefore, clear that the reason why crude rates cannot be compared one with another is because each of them uses a different method of weighting of the age specific rates to obtain a total rate and hence they do not use a standard base of reference. Indeed a crude death rate is a rate weighted according to the age distribution of the population itself. If one county, however, wishes to compare its own mortality rate with that of other countries it may, of course, use its own population distribution as the standard against which to compare the mortality rates for all other countries. Thus in the example above, Pinellas County could use its population distribution as standard and apply the death rates experienced by Dade County, this would give an adjusted rate for Dade County of 16.0 and the ratio of Pinellas to Dade is 0.96.

2. **A comparison of the direct methods of adjustment.** Although each of the four standard populations used above have given almost exactly the same ratio of adjusted rates for the two Counties, it is apparent that the absolute level of the adjusted rates is very different. The USA 1960 population had a young structure and hence the adjusted rates are low, less than 9 per 1,000. The pooled population has an older structure because Pinellas has an old structure and the two adjusted rates are intermediate between the two crude unadjusted rates. The third or linear population attached heavy weighting to the older age rates and hence the adjusted rates are both large, being over 13.5 per 1,000. The fourth population which is that of Pinellas is very heavily weighted by the elderly. It is thus apparent that the adjusted rate cannot be used in isolation, but should be used only for comparative purposes in the context of the particular standard population that was used.

III.4.c **Indirect Age Adjusted Rates.**

Sometimes the total number of deaths for each of a set of populations may be known, but the distribution of the ages at death is not known or possibly the population may be a
small one in which the proportional random variations of the age specific rates is expected to be large. In such cases an alternative method of adjustment may be used.

This method adopts not a standard population distribution, but a standard set of age specific death rates and postulates that the age distribution of the study population is known. If the standard age specific rates are denoted by $s_m$ and the crude death rate in the standard population is $s_m$ and the age distribution of the study population is given by $p_x$ and the total study population is $P$, then the total number of "expected" deaths, $E$, in the study population if the standards rate prevailed is:

$$E = E \left( \frac{s_m}{p_x} \right)$$  \hspace{1cm} 22

the adjusted rate by the indirect method $m_{ind}$ is defined as

$$m_{ind} = \frac{s_m}{D/E}$$  \hspace{1cm} 23

where $D$ is the actual number of deaths in the study population, $E$ is the number of deaths that would be "expected" if the standard rates prevailed in the study population, and $s_m$ is the observed crude death rate in the standard population.

Although this indirect age adjusted rate is not a true "weighted average" of the form

$$\frac{\sum w_x m_x}{\sum w_x}$$  \hspace{1cm} 24

nevertheless it has certain advantages as an adjusted measure of mortality. First, it does not require any knowledge of the age distribution of the deaths in the study population; second, the adjusted rate which is derived can be compared directly to the standard rate; third, it may be used for small areas or small groups of persons when the number of expected deaths in one age group might be very small; fourth most of the calculations can be done before the mortality rate is known.

**TABLE III.5**

Mortality in Two Florida counties 1960. Indirect method of adjustment 1960 United States age specific death rates as standard

<table>
<thead>
<tr>
<th>Age Group (Years)</th>
<th>Death rates (per 1,000) standard U.S. 1960</th>
<th>Population</th>
<th>Expected deaths in county using U.S. specific rates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(i)</td>
<td>(ii)</td>
<td>(iii)</td>
</tr>
<tr>
<td>0 - 1</td>
<td>27.0</td>
<td>5,674</td>
<td>18,819</td>
</tr>
<tr>
<td>1 - 4</td>
<td>1.1</td>
<td>22,167</td>
<td>74,554</td>
</tr>
<tr>
<td>5 - 14</td>
<td>0.5</td>
<td>51,932</td>
<td>162,633</td>
</tr>
<tr>
<td>15 - 24</td>
<td>1.1</td>
<td>32,565</td>
<td>108,310</td>
</tr>
<tr>
<td>25 - 34</td>
<td>1.5</td>
<td>33,877</td>
<td>126,938</td>
</tr>
<tr>
<td>35 - 44</td>
<td>3.0</td>
<td>41,633</td>
<td>140,768</td>
</tr>
<tr>
<td>45 - 54</td>
<td>7.6</td>
<td>41,670</td>
<td>118,013</td>
</tr>
<tr>
<td>55 - 64</td>
<td>17.4</td>
<td>51,985</td>
<td>93,058</td>
</tr>
<tr>
<td>65 - 74</td>
<td>38.2</td>
<td>65,783</td>
<td>67,994</td>
</tr>
<tr>
<td>75+</td>
<td>108.0</td>
<td>27,379</td>
<td>25,960</td>
</tr>
<tr>
<td><strong>TOTALS</strong></td>
<td><strong>9.5</strong></td>
<td><strong>374,665</strong></td>
<td><strong>225,047</strong></td>
</tr>
</tbody>
</table>

**Expected crude death rates** 18.8  \hspace{1cm} 9.9

**Local Comparability Factor** 0.51  \hspace{1cm} 0.96
In Table III.5 the age specific mortality for the U.S. in 1960 is used as the standard set of death rates and these are multiplied by the number of persons in each age group in the study population. This product is the number of deaths that would have occurred in the study population if the national mortality rates had prevailed locally and hence these products are referred to as the "Expected Number of Deaths".

The total "Expected deaths" in Pinellas would be 7,052 and in Dade 9,264, dividing these by the relevant populations gives "expected death rates" of 18.82 and 9.91, a ratio of 1.90. Hence even if the standard US mortality had prevailed equally in both counties the expected crude death rate in Pinellas would have been almost twice that in Dade.

The indirect adjusted rate in the local area is then obtained by multiplying the ratio of the Actual death rate to the Expected Deaths by the crude standard rate.

<table>
<thead>
<tr>
<th>Local Indirect adjusted rate</th>
<th>=</th>
<th>Standard rate</th>
<th>x</th>
<th>Actual Local Rate</th>
<th>=</th>
<th>Expected Local Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

but clearly this can be rearranged in the form

<table>
<thead>
<tr>
<th>Local Indirect adjusted rate</th>
<th>=</th>
<th>Actual Local Rate</th>
<th>x</th>
<th>Standard Rate</th>
<th>=</th>
<th>Expected Local Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that the ratio (Standard rate/Expected local rate) in Equation 26, is not dependent upon the number of deaths in the local area and that all the information for this calculation is available before the deaths are registered. Thus it is possible to calculate this ratio beforehand in a central office for all areas in a country and then all that is required to obtain the local indirect adjusted rate is to multiply this ratio by the crude local rate which is a simple procedure. This ratio is known as the Local Comparability Factor (L.C.F.)

In the case of Pinellas the local comparability factor is 9.5/18.8 = 0.51 and for Dade 9.5/9.9 = 0.96.

The local comparability factor when a standard set of mortality rate is being used is entirely dependent upon the proportional age distribution of a population. A community with a similar age distribution to the standard population will have a L.C.F. equal to 1.0, a community with a higher proportion of aged persons will have L.C.F. <1.0 and a community with a higher proportion of young persons will have a L.C.F. >1.0. Clearly Pinellas is a very elderly community but even Dade is older than the U.S. standard population.

This now gives the indirect age adjusted mortality rate for Pinellas 15.3 x 0.51 = 7.8, and for Dade 8.9 x 0.96 = 8.5 and the ratio Pinellas/Dade = 0.92.

The conclusion from Table III.5 is the same as that derived from the direct methods of adjustment that Pinellas has a better mortality experience than Dade, but this method also has the advantage that it demonstrates that both counties do better than the national experience.

III.4.d Standardised Mortality Ratio.

The indirect method of adjustment can be simply related to the age specific mortality rates for the national experience. This suggests that rather than calculate an adjusted rate for every area within the national boundaries an Index Number should be calculated in which the national experience is taken as the reference value and this is obtained by dividing each of the area adjusted rates, by the National crude death rate. This Index Number is given the name of the Standardised Mortality Ratio, (S.M.R.), which is expressed by the following Equation (27).

\[ SMR = \frac{\text{Indirect Adjusted Rate}}{\text{Standard Rate}} \]

and then directly from equation 23 this becomes

\[ SMR = \frac{D}{E} \]

where D is the total number of actual deaths recorded, E is the "expected" number of deaths if the standard rates prevailed and P is the local total population.
Thus the Standardised Mortality Ratio is both the ratio of the local indirect adjusted mortality rate to the all ages standard rate and also the ratio of the actual number of deaths to the expected number of deaths in the local population.

The use of the Standardised Mortality Ratio (SMR) is very simple. A County or a City will know what is its mid-year population age distribution, it will know the national age specific standard rates to be used for adjustment so it can calculate the "expected" number of deaths in its area before the end of the year and then immediately the last death of the year has been registered it will know what is its SMR by the simple calculation of the ratio $\text{SMR} = \frac{\text{Actual Number of Deaths}}{\text{Expected Number of Deaths}}$, and will not require any computer or Central Statistical Office advice to do so. Alternatively, the Central Statistical Office when it makes the mid-year estimates of population for each area can calculate the "Expected" number of deaths in every area and use these to adjust rates in all reports.

**Summary.** The results of these various methods of age adjustment are summarised in Table III.6.

<table>
<thead>
<tr>
<th>Method of Adjustment</th>
<th>Standard Population</th>
<th>Age adjusted Death Rates</th>
<th>Ratio Pinellas / Dade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct</td>
<td>U.S. 1960</td>
<td>8.2</td>
<td>8.6</td>
</tr>
<tr>
<td>Direct</td>
<td>Two Co's pooled</td>
<td>10.4</td>
<td>10.9</td>
</tr>
<tr>
<td>Direct</td>
<td>Weights equal to age groupings</td>
<td>12.9</td>
<td>13.5</td>
</tr>
<tr>
<td>Direct</td>
<td>Weights equal to Pinellas</td>
<td>15.3</td>
<td>16.1</td>
</tr>
<tr>
<td>Indirect</td>
<td>U.S. 1960</td>
<td>7.8</td>
<td>8.5</td>
</tr>
<tr>
<td>Indirect</td>
<td>SMR U.S. 1960</td>
<td>0.82</td>
<td>0.89</td>
</tr>
<tr>
<td>Unadjusted</td>
<td>Crude Rates</td>
<td>15.3</td>
<td>8.9</td>
</tr>
</tbody>
</table>

Each method of adjustment gives a slightly different result from the others and this is inevitable as they are using different sets of weights. What is important is that the same system can be selected both nationally and internationally and be used consistently.

**III.4.e Adjustment for factors other than age.**

Age adjustment is not the only way in which these basic principles can be applied, other factors such as sex, race, marital status, occupational group, can all be used as basis for adjusting and refining groupings. It is also possible to standardize for combination of factors and this will usually be necessary because the age pattern within other social or cultural groups will be different. For example, if it were necessary to compare two regions in which two races were living but the proportions were different, it would be necessary to adjust for race, for sex and for age before a completely adjusted comparison could be made.

To simplify matters the example which is discussed below relates to infant mortality where an age factor is irrelevant and sex differences are ignored.

**Example:** In 1960 the infant mortality rate in Alabama was 32.4 per 1,000 and in Kentucky 27.9 per 1,000, a ratio of 1.16 to the disadvantage of Alabama, thus it would appear that Kentucky were the healthier state for infants. But it is clear from Table III.7 that in Alabama infant mortality of white children was better than that of white children in Kentucky and the same applies for non-white children. What is important is that non-white children have a higher infant mortality rate than white children and that there are more of them in Alabama.

**TABLE III.7**

It is apparent that some method of standardization is necessary, the problem arises what weights should be attached to the different infant mortality rates. Should a national population be used, should a pooled population be used, should an artificial population be
used or should an indirect method be used? The choice is arbitrary but a national population hardly seems relevant in this case and what is done in Table 8. is to use an arbitrary population in which 75% of the population is white, this is an intermediate experience between that of Kentucky (88%) white and Alabama (63%) white.

**TABLE III.7**


<table>
<thead>
<tr>
<th>ALABAMA</th>
<th></th>
<th>KENTUCKY</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>White</td>
<td>Non-white</td>
<td>TOTAL</td>
</tr>
<tr>
<td>Live births</td>
<td>50,828</td>
<td>30,018</td>
<td>80,846</td>
</tr>
<tr>
<td>Infant deaths</td>
<td>1266</td>
<td>1,350</td>
<td>2,616</td>
</tr>
<tr>
<td>Infant deaths per 1,000 live births</td>
<td>24.9</td>
<td>45.0</td>
<td>32.4</td>
</tr>
</tbody>
</table>

**TABLE III.8**


Direct Method - Arbitrary Standard

<table>
<thead>
<tr>
<th>Race</th>
<th>Arbitrary Standard</th>
<th>Putative Infant Deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ALABAMA</td>
<td>KENTUCKY</td>
</tr>
<tr>
<td>White</td>
<td>7,500</td>
<td>187</td>
</tr>
<tr>
<td>Non-White</td>
<td>2,500</td>
<td>112</td>
</tr>
<tr>
<td>TOTAL</td>
<td>10,000</td>
<td>299</td>
</tr>
<tr>
<td>Race adjusted rate</td>
<td>29.9</td>
<td></td>
</tr>
</tbody>
</table>

**III.5**

Standard Errors of Mortality Rates.

III.5.a The concept of a standard error of mortality.

A death rate is usually calculated from the mortality experience of an entire population and only in special cases is it determined from a sample, consequently it is sometimes suggested that there is no sampling error in the estimate of the rate and hence the standard error if it exists should be ignored. There are, however, circumstances in which the concept of a standard error is required because a single death rate for one country for one year is not a statistic that can be considered in isolation without regard to the experience of other communities or of other years. Such a single rate will always be considered in the context of a series of rates over a number of years or of separate individual rates occurring in different cities or provinces during the same year. From this point of view a human society experiences a particular pattern of mortality and the specific expression of this underlying pattern in a unique community is subject to the ordinary laws of probability.

The total mortality rate for a community is the sum of the experience of all these individuals comprising that community, who at the beginning of a year all started together exposed to the pattern of mortality which will vary in some way between all the individuals so that at the end of the year a selected sample will have died. Without in any way judging the philosophic or religious implications of determinism or indeterminism, it is not possible to know who will be selected by death and our knowledge concerning the distribution will be that the selection for death appears as if it were randomly distributed.

It is in this sense that a death rate is subject to random variation even though it is based upon the total number of deaths occurring in an entire population.

Random variation will not, of course, account for all or even most of the fluctuations that will be observed in a country or a region from year to year. This variation will comprise two parts, first the true environmental influences such as seasons, epidemics, famine or war, environmental pollution or industrial and economic hazards, and secondly, imposed upon this a random variation due to unrecognised or chance factors.
III.5.b Derivation of the standard error of the probability of dying, \( q_x \).

Every year in a population there are a certain number of persons who attain their \( x \)th birthday and who will die before attaining their \((x+1)\)th birthday. So each year there may be considered to be a trial to estimate the true value of the probability of dying at age \( x \).

Let \( N_x \) be the number of persons reaching age \( x \) in year \( i \), let \( D_x \) be the number of persons who die before reaching their next birthday having attained age \( x \) in year \( i \), and let \( q_x \) be the true probability of dying which is assumed to be constant but is unknown.

Then in any one year the standard application of the binomial theorem gives the unbiased estimations of \( q_x \) and of the variance of \( q_x \) as

\[
E(q_x) = \hat{q}_x = \frac{D_x}{N_x} \\
\text{Var}(\hat{q}_x) = \frac{\hat{q}_x(1-\hat{q}_x)}{N_x}
\]

30

31

The central limit theorem then states that if \( N_x q_x = D_x \) is sufficiently large and a variable \( Z \) is defined as

\[
Z = \frac{\hat{q}_x - q_x}{\sqrt{\text{Var}(\hat{q}_x)}}
\]

32

then \( Z \) is normally distributed with zero mean and unit variance and hence confidence limits can be determined within which \( q_x \) is known to lie.

But in most mortality studies the observed value of \( \hat{q} \) will be small so the value of \( p_x \) will be close to unity when the Poisson approximation is valid.

\[
\text{Var} \hat{q}_x = \frac{\hat{q}_x}{N_x}
\]

33

Also we have

\[
\hat{q}_x = \frac{D_x}{N_x}
\]

34

hence,

\[
\text{Var} \hat{q}_x = \frac{\hat{q}_x^2}{D_x}
\]

35

Also by definition the probability of surviving the year is \( \hat{p}_x = 1 - \hat{q} \) and hence,

\[
\text{Var} \hat{p}_x = \text{Var} \hat{q}_x
\]

36

Example: In England in 1971 the number of live births in the city of Bristol was 6,306 and the number of infant deaths under 1 week was 62. What is the unbiased estimate of the probability of dying in the first week of life and what is its variance. What is the variance of the observed number of deaths?

Answer: If the probability of dying in the first week of life in the City of Bristol is assumed to be constant over a period of years then the unbiased estimation of \( q_o \), the probability of death, is the proportion found in the sample, hence

\[
\hat{q}_o = \frac{62}{6,306} = 0.00983
\]

37

\[
\text{Var} (\hat{q}_o) = \frac{(0.00983) (0.99017)}{6,306} = 1.54 \times 10^{-6}
\]

38

Standard Error of \( \hat{q}_o \) = \( \text{Var} (\hat{q}_o)^{\frac{1}{2}} \) = \( 1.24 \times 10^{-3} \)

39

Var (D) = \( (0.00983) (0.99017) (6306) \) = 61.379

40

Standard Error of D = \( (\text{Var} D)^{\frac{1}{2}} \) = 7.834

41

Poisson Standard Error of D = \( (0.00983 \times 6306)^{\frac{1}{2}} \) = 7.873

42

The standard error of \( q_o \) is 0.00983 and the standard error of the observed number of first week deaths is 7.83. Thus over a period of years the number of first week deaths might be expected to vary around 62 deaths per year with a standard deviation of \( \pm 7.83 \). Not all individual infants are at equal chance of death as some will be premature and some will be injured at birth, but statistically these patterns of risk will follow similar patterns from year to year and an average risk has a valid meaning.
III.5.c The standard error of the mid year death rate

The proportion of deaths \( \bar{d} \), was defined in paragraph III.3.c above as the proportion of the dead out of the original number of individuals \( N \) who started a trial of which the outcome could only be death (D) or survival (N-D). In practice death rates are not calculated from a population of individuals all of whom start the year alive, but are based upon the mid-year estimate of the population \( P \) and the number of deaths is determined separately and independently from the population and certain deaths will already have occurred before the population is enumerated. This estimate of mortality based upon the mid-year population is known as \( m \). The mid-year death rate \( m_x \) is defined as the ratio of the number of deaths \( D_x \) to the total number of years lived \( L_x \) by the population during the reference year between the age interval \( x \) and \( x+1 \) and where the proportion of the year for which the deceased had survived, assumed to be a half.

\[
\frac{m}{x} = \frac{D_x}{N_x - \frac{1}{2}D_x}
\]

and hence,

\[
m_x = 2q_x / (2-q_x)
\]

Where \( m \) is estimated directly from the mid year estimates of the population \( P_x \) the unbiased estimator of \( m_x \) is

\[
E(m_x) = m_x = \frac{D_x}{P_x}
\]

and assuming that \( m_x \) is small the variance of \( m_x \) is

\[
\text{Var}(m_x) = m_x / P_x = \frac{m_x^2}{P_x}
\]

III.5.d Standard errors of age adjusted death rates.

In Section III.3 several methods of computing an age adjusted death rate were discussed. Although each method was developed for a particular purpose they all assume the form of a weighted mean of the age specific death rates. The notation to be used was summarised in Table III.2.

To derive the variance of an adjusted rate it is first necessary to identify clearly which are the random variables concerned. Clearly the age specific death rates \( m_x \) for a study population are random variables. The population proportions both of the study population \( P_x/P \) and of the standard Population \( P_x/P \) are treated as fixed data not subject to random variation. The age specific death rates for the standard population \( m_x \) are considered to be random variables in the sense that they are estimates of the underlying pattern of mortality which is given expression in the study populations.

We therefore reformulate the adjusted rates as linear functions of the basic random variables which are the age specific death rates in a study population.

The adjusted rate \( m_{adj} \) is given by a formula of the type

\[
m_{adj} = \frac{\Sigma(W_x m_x)}{\Sigma W_x}
\]

where the \( W_x \) are the weights tabulated in Table III.10 and the \( m_x \) are the age specific mortality rates.

The general rule for the variance of a sum may be applied so that

\[
\text{Var}(m_{adj}) = \left( \Sigma \frac{W_x^2}{x} \text{Var}(m_x) \right) + 2 \Sigma \frac{W_x}{x} \Sigma \frac{W_x}{x} \text{Cov}(m_x, m_z)
\]

If it is assumed that the random changes in the mortality rate in the sense described above for a single age interval are not associated with a change in the proportion of deaths in the other age groups, then the covariances in (30) are zero and since the variance of \( m_x \) is \( m_x / P_x \)

\[
\text{Var}(m_{adj}) = \frac{\Sigma (W_x^2 m_x^2)}{\Sigma W_x} \text{Var}(m_x) = \frac{\Sigma (m_x^2)}{P_x m_x}
\]

1) Direct method of adjustment. In the case of the direct method of adjustment the weights are \( S_{P_x/P} \) and hence the variance of the direct adjusted rate is given by:

\[
\text{Var}(m_{dir}) = \frac{S_{P_x m_x^2}}{S_{P_x} m_x} \frac{1}{S_{P_x m_x}} = \frac{(S_{P_x m_x})^2}{(S_{P_x m_x})} \frac{1}{S_{P_x}^2}
\]

which is equivalent to:

\[
\text{Var}(m_{dir}) = \frac{\text{Standard Expected Deaths}^2}{\text{Actual Deaths}} / \text{Std. Pop.}^2
\]

and hence the standard error of \( m_{dir} \) for a single study population is:
\[ \text{Std error of } (m_{dir}) = \left( \frac{(\text{Std Expected Deaths})^2}{\text{Actual Deaths}} \right)^{\frac{1}{2}} / \text{Std. Pop.} \]

and where \( k \) is a scaling factor the standard error of \( (k.m_{dir}) \) is

\[ \text{Std error of } k.m_{dir} = k \left( \frac{(\text{Std ExpectedDeaths})^2}{\text{Actual Deaths}} \right)^{\frac{1}{2}} / \text{Std. Pop.} \]

The calculation of this is relatively easy as is shown in the example below.

**Example:** It is required to calculate the standard errors of the direct age adjusted death rates using the U.S. national population in 1960 as standard for Pinellas and Dade County.

The actual deaths experienced in the two counties and the deaths expected in a population of one million with the U.S. 1960 proportional age distribution are given below in Table III.9. For each age the ratio \( D/A \) is calculated and these are then added together. The square root of the sum is obtained and this is then divided by the standard population and adjusted for scale by multiplying by \( k \).

**TABLE III.9**

Computation of standard error of the direct adjusted death rate in Pinellas and Dade Co. U.S. 1960 population as standard

<table>
<thead>
<tr>
<th>PINELLAS COUNTY</th>
<th>DADE COUNTY</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age Interval</strong></td>
<td><strong>Actual</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Deaths</strong></td>
</tr>
<tr>
<td>(i)</td>
<td>(ii)</td>
</tr>
<tr>
<td>0 - 1</td>
<td>160</td>
</tr>
<tr>
<td>1 - 4</td>
<td>30</td>
</tr>
<tr>
<td>5 - 14</td>
<td>30</td>
</tr>
<tr>
<td>15 - 24</td>
<td>26</td>
</tr>
<tr>
<td>25 - 34</td>
<td>47</td>
</tr>
<tr>
<td>35 - 44</td>
<td>124</td>
</tr>
<tr>
<td>45 - 54</td>
<td>320</td>
</tr>
<tr>
<td>55 - 64</td>
<td>279</td>
</tr>
<tr>
<td>65 - 74</td>
<td>1,901</td>
</tr>
<tr>
<td>75+</td>
<td>2,259</td>
</tr>
<tr>
<td><strong>TOTALS</strong></td>
<td>5,726</td>
</tr>
</tbody>
</table>

Pinellas County Std. error of \( m_{dir} \) per 1,000 population =
\[ (15,305)^{\frac{1}{2}} \times \frac{1,000}{1,000,000} = 123.7/1,000 = .1237 \]

Dade County Std. error of \( m_{dir} \) per 1,000 population =
\[ (8,967)^{\frac{1}{2}} \times \frac{1,000}{1,000,000} = 94.7/1,000 = .0947 \]

ii) **Indirect method of adjustment.** In the case of the indirect method of adjustment the weights to be attached to the age specific mortality rates are

\[ w_x = \frac{E_x}{P_x} \]

and hence the variance of the indirect adjusted mortality rate \( m_{ind} \) is

\[ \text{Var } m_{ind} = \sum \left[ \frac{(m/E)_x}{P_x} \right]^2 \cdot \frac{m}{P_x} \]

\[ = \left[ \frac{(m/E)_x}{P_x} \right]^2 \cdot \sum \frac{m}{P_x} \]

\[ = \left[ \frac{(m/E)_x}{P_x} \right]^2 \cdot D \]

\[ = \left[ \frac{(m/E)_x}{P_x} \right]^2 \cdot D \]

\[ = \left[ \frac{(m/E)_x}{P_x} \right]^2 \cdot D \]

\[ = \left[ \frac{(m/E)_x}{P_x} \right]^2 \cdot D \]

\[ = \left[ \frac{(m/E)_x}{P_x} \right]^2 \cdot D \]
### Table III.10

Summary of the formulae, the weights used and the variances for adjusted death rates

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Basic Formulae of adjusted rate</th>
<th>Weights of $m_x$</th>
<th>Formulae for Variance</th>
<th>Variance of adjusted Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>i) Crude Death Rate</td>
<td>$\Sigma (P_x \cdot m_x / P)$</td>
<td>$P_x / P$</td>
<td>$i) \Sigma (P_x / P)^2 \cdot (m_x / P_x)$</td>
<td>$= D/P^2$</td>
</tr>
<tr>
<td>ii) Direct Adjusted rate</td>
<td>$\Sigma (n_x \cdot m_x / n_x)$</td>
<td>$n_x / \Sigma n_x$</td>
<td>$ii) \Sigma \left[ (n_x / \Sigma n_x)^2 \cdot (m_x / P_x) \right]$</td>
<td>$= \Sigma (n_x / D) / (n_x / P_x)^2$</td>
</tr>
<tr>
<td>iii) Equivalent Average</td>
<td>$\Sigma (n_x \cdot m_x / \Sigma n_x)$</td>
<td>$n_x / \Sigma n_x$</td>
<td>$iii) \Sigma \left[ (n_x / \Sigma n_x)^2 \cdot (m_x / P_x) \right]$</td>
<td>$= \frac{1}{n^2} \Sigma (m_x / P_x)$</td>
</tr>
<tr>
<td>Death Rate</td>
<td></td>
<td></td>
<td>where there are $n$ age groups</td>
<td></td>
</tr>
<tr>
<td>iv) Standardised Mortality</td>
<td>$\Sigma (P_x \cdot m_x) / \Sigma (P_x \cdot s_x)$</td>
<td>$P_x / \Sigma (P_x \cdot s_x)$</td>
<td>$iv) \Sigma \left[ (P_x / \Sigma P_x \cdot s_x)^2 \cdot (m_x / P_x) \right]$</td>
<td>$= D/E^2$</td>
</tr>
<tr>
<td>Ratio</td>
<td>$\Sigma (P_x \cdot m_x) / \Sigma (P_x \cdot s_x) \cdot s_x$</td>
<td>$(P_x \cdot s_x) / (\Sigma P_x \cdot s_x)$</td>
<td>$v) \Sigma \left[ (P_x \cdot s_x / \Sigma P_x \cdot s_x)^2 \cdot (m_x / P_x) \right]$</td>
<td>$= (s^2) \cdot D/B^2$</td>
</tr>
<tr>
<td>v) Indirect Adjusted Rate</td>
<td>$\Sigma (P_x \cdot s_x) / \Sigma (P_x \cdot s_x)$</td>
<td>$(P_x \cdot s_x) / (\Sigma P_x \cdot s_x)$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
where E is the total expected deaths in the study population if the standard age specific rates applied and D is the total actual deaths that did occur.

Hence the standard error of \( m_{\text{ind}} \) is 
\[
\sigma = \sqrt{\frac{s}{E}} \cdot \left( \frac{E}{D} \right)^{\frac{1}{2}}
\]

and the standard error of \( k \cdot m_{\text{ind}} \) is 
\[
k \cdot \left( \frac{s}{E} \right) \cdot \left( \frac{E}{D} \right)^{\frac{1}{2}}
\]

In the example in Table 5 above we have for Pinellas County -

\[
SE(m_{\text{ind}}) = \frac{(1000 \times 0.0095 \times 5726^{\frac{1}{2}})}{7052} = 0.102
\]

and for Dade County -

\[
SE(m_{\text{ind}}) = \frac{(1000 \times 0.0095 \times 8322^{\frac{1}{2}})}{9264} = 0.094
\]

III.5.d **Standardised mortality ratio.**

In the case of the standardised mortality ratio the weights are -

\[
W_x = \frac{P_x}{(P_x \cdot \frac{m_x}{E})} = \frac{P_x}{E}
\]

Hence the variance of SMR is -

\[
Var(S.M.R.) = \left( \frac{P_x}{E} \right)^2 \cdot \frac{m_x}{P} = \frac{(P_x \cdot m_x)}{P} = \frac{D}{E} \cdot \frac{P_x}{E} = \frac{D}{E} \cdot \frac{P_x}{E} = \frac{(P_x \cdot m_x)}{P} = \frac{SMR^2}{D}
\]

hence Std. Error of SMR = \( \frac{SMR}{D^{1/2}} \)

It is therefore very easy to calculate the standard error of an SMR, it is the SMR itself divided by the square root of the number of deaths observed in the study population.

For Pinellas this is 
\[
(0.81)/(5726)^{1/2} = 0.11
\]

and for Dade this is 
\[
(0.90)/(8322)^{1/2} = 0.10
\]

These various methods of calculating the variances of adjusted rates are all summarised in Table 10.

**TABLE III.10 HERE**

III.5.e **Confidence limits of a mortality rate.**

One important use of the standard error of an observed mortality rate is to determine boundaries within which the "true" or underlying rate can be expected to lie with a known probability. This is the method of confidence limits.

If the underlying mortality rate is \( \mu \) and the observed rate in a single year is \( m \), and the sampling variance of \( m \) is \( s^2 \) then the central limit theorem may be used to determine upper and lower boundaries within which \( \mu \) can be expected to lie,

\[
Pr(m - t \cdot s \leq \mu \leq m + t \cdot s) = 1 - \alpha
\]

where \( t \) is the normal deviate and \( \alpha \) is the probability of the normal deviate being greater than \( |t| \)

It is customary to select \( \alpha = 0.05 \) and hence \( t = 1.968 \) so equation becomes -

\[
Pr(m - 1.968 \cdot s \leq \mu \leq m + 1.968) = 0.95
\]

or \( \mu \) lies within limits of plus or minus 1.968 with a confidence of 95 per cent
Example: For Pinellas County in 1960, the direct adjusted death rate based on the U.S. national mortality rates was 8.168 with a standard error of 0.1237. Now the experience of a unique year will not give an exact measure of the risks to life existing in this community over a period of time, it is a single sample of these risks. Can we say with any degree of certainty that the underlying rate lies within certain limits?

It can be said that it is true with a probability of 0.95 (or that the statement will be true in 19 cases out of 20) that the underlying risk lies within limits which are precisely defined, the upper limit being

\[ 8.168 + (1.96)(0.1237) = 8.41 \]

and the lower limit being

\[ 8.168 - (1.96)(0.1237) = 7.93 \]

It is important to appreciate that the underlying mortality is assumed to be constant and is not assumed to fluctuate but that the information available only provides a degree of confidence to assert the limits within which it may lie.

III.5.f Tests of significance for mortality rates.

Another classical application of the methods of statistical estimation is to test the significance of the probability that two observed mortality rates might have been obtained by random sampling from populations with no difference between them. (Null Hypothesis).

Then if a statistic \( z \) is calculated where

\[
z = \frac{m_1 - m_2}{(\text{Var}(m_1)^{\frac{1}{2}} + \text{Var}(m_2)^{\frac{1}{2}})\frac{1}{2}}\]

\( z \) is assumed to be normally distributed with mean zero and unit variance and hence the probability of obtaining values equal to or greater than \( |z| \) can be estimated.

Example. From Table III.9 we have for Pinellas the direct adjusted rate was 8.168 with a standard error of 0.1237 and for Dade it was 8.567 with a standard error of 0.0947.

Hence,

\[
z = \frac{8.567 - 8.168}{(0.1237^2 + 0.0947^2)^{\frac{1}{2}}} = \frac{0.399}{0.156} = 2.56
\]

The probability of \( |z| \geq 2.56 \) occurring by chance is about one in a 100 trials.

The way in which we use this estimate of probability must be done with careful thought. There are approximately 3,000 counties in the U.S. and hence in any one year the number of pairs of counties which could be compared one with another is 3,000 \times 2,999/2 which is about 4.5 million pairs.

Thus a probability as small as 0.01 is adequate to reject the Null Hypothesis that there is no difference between Pinellas and Dade yet there will always be a large number of pairs of counties which would yield a value of \( |z| \) as large as, or even larger than, 2.56, even if there were no real differences in mortality experience between them.

In mortality analysis the problems of identifying possible demographic, medical, and social economic influences upon mortality will always be a greater challenge to the analyst than that of determining the statistical significance of differences. Usually in demography there are many opportunities to repeat observations that are of marginal statistical significance so as to refute or confirm the hypotheses that are being considered.

III.5.g Standard error of sample death rates.

It should be emphasised that although \( D_a \) and \( D_d \) in the previous discussion in Section 5.c and d, both refer to the number of people in a total population yet if they are the population and deaths observed in a sample of a population the formulae of the sample variances of \( q_x \) and \( m_x \) hold equally for the sample population. If the sample population is obtained by an accurate probability sampling technique from a frame based upon a total population when the rates for that total population can be estimated using the standard sampling methods and confidence limits can be calculated.
III.6 Other Measures Associated with Mortality.

III.6.a Proportions of mortality.

Proportional mortality may be a useful summary of the distribution of mortality between different categories of disease, different age groups or different sexes. It is defined as the proportion of all deaths which are assigned to one particular characteristic or cause. Clearly it is not a rate but it is a ratio in which the denominator is the total number of deaths. It is simple to calculate because it is based upon the total number of deaths from a specific cause, or those which occur to a certain age or sex group divided by the total number of deaths for all causes or for specific age or sex groups. It is thus much easier to calculate than a rate as it does not require knowledge of the population at risk, it does not even require complete registration of all deaths so long as there is no serious bias in the under-registration, and it does not require exact specification of the period of time during which the deaths were recorded so long as the period for the total deaths was the same as that for the proportion to be analysed.

Thus:

\[
\text{Proportional Mortality} = \frac{\text{Number of deaths within a specified group}}{\text{Total Number of deaths within the total population}}
\]

The proportional mortality possesses several advantages over other more complex indices:

1. It is simple to calculate
2. It is readily available in a large number of countries
3. It allows international comparisons to be made despite varying quality of statistical information if population structures are known to be similar.

It has the serious disadvantage that it ignores completely the population structure and that rates cannot be calculated from it and hence absolute values cannot be known.

III.6.b Proportion of causes of death.

One form of proportional mortality is the proportion of total deaths attributed to one group of causes. In particular the proportion of total deaths which are attributed to infections may be used as a measure of the efficiency of health services in developing countries and similarly the proportion of deaths attributed to cancers or to road traffic accidents is a measure of the failure of health services in more developed countries. In cancer mortality the proportional distribution of the different cancers is a useful indicator of different risks in communities which may have similar demographic characteristics but have different environments. For example the proportion of breast cancer amongst Japanese cancer is lower than amongst European women, whereas stomach cancer is higher.

III.6.c Proportion of deaths by sex.

The proportion of deaths who are males is sometimes used as a measure because although all men and women must ultimately die the rate at which they do so varies considerably according to the social and economic environment in which they live. It is characteristic of many economically advanced societies that the proportion of male deaths increases over the years as women live longer.

III.6.d Proportion of deaths by age.

The proportion of deaths above a certain specified age may also be a useful discriminant of different patterns of mortality. The particular age at which the deaths are divided to obtain a proportional mortality will be a matter of arbitrary decision depending upon the purposes of a particular study whether it is directed towards childhood mortality or to the pattern of death in the elderly. One useful index is the proportion of deaths under 5 years of age which is a measure of the importance of childhood mortality. At the opposite extreme the proportion of deaths at 75 years or over is a measure of the efficiency of health services. But all such measures are dependent upon the age distribution of the population as portrayed by the age and sex pyramids as shown in Fig. II.2, and comparisons should not be made between countries who are at different stages of demographic development.

Swaroop (1960) suggested the term "proportional mortality indicator" P.M.I., should be used for the proportional mortality at ages 50 years and over. He found that this varied in different countries and at different periods of time from over 0.80 to 0.20. It effectively discriminated countries with advanced economy and strong health services in which the P.M.I. was high, from those countries with an undeveloped economy and weak health services in which the P.M.I. was low. It should be noted, however, that this "indicator" is largely a factor of the past history of the birth rate in a population rather than a function of the current death rate.
III. 6.e Case-fatality ratios. For many diseases it is important to characterise the severity by the risk of death. According to the data available there are several approaches:

1. If all cases are known from their incidence and if it is possible to follow them up for a certain time then we could calculate the proportion of deaths among the persons at risk. This proportion is an empirical approximation to the underlying conditional probability of dying from the disease if one got the disease.

   This longitudinal approach is possible for severe infectious diseases, when nearly all cases are notified. It is applicable for traffic accidents, etc. It is inappropriate if there is incomplete notification, especially if cases are only recognised after or shortly before death, because these cases suggest that there have been other unrecognised cases.

2. For a frequent disease a cross-sectional approach has to be preferred. For a certain period the deaths assigned to this disease and the new cases (incidence) could be connected by a ratio. But evidently this estimate is useful only for an endemic disease with constant balance over a longer period of time. In the beginning or at the end of an acute epidemic it may be strongly biased and give misleading estimates. But even in a constant phase the estimate depends on the equal completeness of the notification of incidence and death.

3. If the true prevalence of a disease at a certain date can be estimated then it is possible to count the number of deaths occurring in a period of time before and after the prevalence was estimated and hence construct a case fatality date. This procedure has the same errors as method (2) above and an additional uncertainty resulting from the neglect of the duration of the disease.

   The usefulness of calculating case-fatality-ratios is limited. It should be done only for contagious diseases of short duration and for external lesions. But if the follow-up must cover a longer period of time there are special statistical difficulties which are discussed in Volume II of the Manual (follow-up studies).
### CHAPTER IV

THE LIFE TABLE AND ITS CONSTRUCTION

<table>
<thead>
<tr>
<th>IV.1.</th>
<th>Introduction</th>
<th>...</th>
<th>...</th>
<th>58</th>
</tr>
</thead>
<tbody>
<tr>
<td>IV.2.</td>
<td>Description of the Life Table</td>
<td>...</td>
<td>...</td>
<td>59</td>
</tr>
<tr>
<td>a.</td>
<td>Measures of time</td>
<td>...</td>
<td>...</td>
<td>59</td>
</tr>
<tr>
<td>b.</td>
<td>Definition of terms used</td>
<td>...</td>
<td>...</td>
<td>60</td>
</tr>
<tr>
<td>c.</td>
<td>Relationship between $m_x$ and $q_x$</td>
<td>...</td>
<td>...</td>
<td>62</td>
</tr>
<tr>
<td>d.</td>
<td>Relationship between $m_{x,t}$ and $q_{x,t}$</td>
<td>...</td>
<td>...</td>
<td>63</td>
</tr>
<tr>
<td>e.</td>
<td>Probability of surviving the first year of life</td>
<td>...</td>
<td>...</td>
<td>64</td>
</tr>
<tr>
<td>f.</td>
<td>The last interval of life table</td>
<td>...</td>
<td>...</td>
<td>65</td>
</tr>
<tr>
<td>IV.3.</td>
<td>Construction of the Abridged Current Life Table</td>
<td>...</td>
<td>...</td>
<td>65</td>
</tr>
<tr>
<td>IV.4.</td>
<td>Other Methods of Constructing an Abridged Current Life Table</td>
<td>...</td>
<td>...</td>
<td>67</td>
</tr>
<tr>
<td>a.</td>
<td>Reed and Merrel</td>
<td>...</td>
<td>...</td>
<td>67</td>
</tr>
<tr>
<td>b.</td>
<td>Greville</td>
<td>...</td>
<td>...</td>
<td>67</td>
</tr>
<tr>
<td>c.</td>
<td>U.S. Abridged Life Tables</td>
<td>...</td>
<td>...</td>
<td>68</td>
</tr>
<tr>
<td>d.</td>
<td>England and Wales Abridged Life Tables</td>
<td>...</td>
<td>...</td>
<td>69</td>
</tr>
<tr>
<td>e.</td>
<td>Weisler</td>
<td>...</td>
<td>...</td>
<td>69</td>
</tr>
<tr>
<td>f.</td>
<td>Chiang</td>
<td>...</td>
<td>...</td>
<td>69</td>
</tr>
<tr>
<td>g.</td>
<td>Keyfitz</td>
<td>...</td>
<td>...</td>
<td>70</td>
</tr>
<tr>
<td>h.</td>
<td>Summary</td>
<td>...</td>
<td>...</td>
<td>70</td>
</tr>
<tr>
<td>j.</td>
<td>Life Table death rate</td>
<td>...</td>
<td>...</td>
<td>70</td>
</tr>
<tr>
<td>IV.5.</td>
<td>The Variance of the Life Table</td>
<td>...</td>
<td>...</td>
<td>71</td>
</tr>
<tr>
<td>a.</td>
<td>Standard error of $d_{x,t}$ - mean age at death</td>
<td>...</td>
<td>...</td>
<td>71</td>
</tr>
<tr>
<td>b.</td>
<td>Standard error of $e_0$ - expectation of life</td>
<td>...</td>
<td>...</td>
<td>73</td>
</tr>
<tr>
<td>c.</td>
<td>Scale of bias introduced by different methods</td>
<td>...</td>
<td>...</td>
<td>74</td>
</tr>
<tr>
<td>IV.6.</td>
<td>Value of Life Tables and Difficulties In Their Use</td>
<td>...</td>
<td>...</td>
<td>74</td>
</tr>
<tr>
<td>IV.7.</td>
<td>Generation or Cohort Life Table</td>
<td>...</td>
<td>...</td>
<td>76</td>
</tr>
<tr>
<td>a.</td>
<td>A total community</td>
<td>...</td>
<td>...</td>
<td>76</td>
</tr>
<tr>
<td>b.</td>
<td>A closed community</td>
<td>...</td>
<td>...</td>
<td>77</td>
</tr>
<tr>
<td>IV.8.</td>
<td>Life Tables from Incomplete Data</td>
<td>...</td>
<td>...</td>
<td>77</td>
</tr>
<tr>
<td>a.</td>
<td>Life Tables based upon census records only</td>
<td>...</td>
<td>...</td>
<td>77</td>
</tr>
<tr>
<td>b.</td>
<td>Life Tables based upon death records only</td>
<td>...</td>
<td>...</td>
<td>79</td>
</tr>
</tbody>
</table>

### IV.1 Introduction

The life table is largely a product of actuarial science, but its application is not limited to the computation of insurance premiums; the life table is also a valuable analytical tool for demographers, epidemiologists, physicians, and research workers in other areas of public health.

The current life table, as the name implies, gives a cross-section view of the mortality and survival experience of all ages in a population during one short period of time, usually for a single year, but sometimes for a period of three, five or ten years. It is dependent entirely on the age-specific death rates prevailing in the year or years, for which it is constructed. Such tables project the life span of each individual in a hypothetical cohort on the basis of the actual death rates in a given population. When we speak of the life expectancy of an infant born in a current year, for example, we mean the life expectancy that would be obtained if he were subjected throughout his life to the same age-specific
mortalities as those prevailing in the current year. The current life table is thus a fictitious pattern reflecting the mortality experience of a real population during a calendar year. However, it is an effective means of summarizing mortality and survival experience of a population and is a sound basis for making statistical inference about the specific population under study for that calendar year.

A current life table may be based on the deaths occurring over three, or more calendar years instead of one. For each age group the average number of deaths per year is then divided by the corresponding population size of the middle of the three years to obtain the age specific death rate. Usually, the middle year is a census year, so that population figures are available and more accurate and these are adjusted to the mid year population. The advantage of such a table is to reduce the possible abnormalities in mortality pattern which may exist in a single calendar year.

Data for constructing life tables are sometimes refined by graduation or other methods for smoothing or reducing the effect of extreme values, or of clustering deaths, or population, at certain ages due to misstatements of age, these techniques for refinement of life table data were developed by actuarial scientists. While refinement of data has merit in smoothing data, it is difficult to make proper statistical inference of life table functions which are based on such information and these smoothing techniques are not discussed in detail in this chapter.

Life tables may be either complete or abridged. In a complete life table the functions are computed for each year of life; an abridged life table differs only in that it deals with age intervals greater than one year, except possibly the first year or the first five years of life. Both, however, will always relate to the mortality experience of the current calendar year, or years. A mortality analyst will not be interested in the detailed complete table, nor will he be in a position to compute one. What he requires is a simple summary of mortality at certain ages which he can obtain from the abridged table.

This chapter will describe the general form of a life table with interpretations of its various functions and present a method of constructing a current abridged life table. Theoretical aspects of life table functions will be discussed in detail in a separate Manual.

IV.2. Description of the Life Table

IV.2.a Measures of time

The essential difficulty in the preparation of a life table, is that it is attempting to examine the mortality function in two dimensions of time at once. The data available is recorded in calendar time, whereas the life table itself is recorded in time of ageing of the individuals in the population.

Consider Figure IV.1 which is a bivariate graph with calendar time, \( y \), along the vertical axis and age of an individual \( x \) along the horizontal axis. (This was first used by Lexis (1840)). The vertical axis is clearly infinite in both directions, the horizontal axis is completely finite starting at zero and never exceeding say 120. The life line of an individual man can be represented by a diagonal line starting at age zero on the calendar date of his birth and his age will always be exactly determined by the calendar date along a line at 45° to both axes. A population will be a complete set of diagonal lifelines starting from varying points on the vertical axis and moving diagonally up and to the right until terminated by death.
The diagram illustrates the concept of a calendar year and a census population. The vertical line AB represents age 60, and the census population is shown as the horizontal line through K. The data is counted from the 1st January to the 31st December of the year 1970.

Data relating to HK is available in a census or mid-year population analysis by age but is not available concerning AB. To count the number of deaths that occur at age 60 in the year 1970, it is necessary to count all the lifelines that terminate in the rectangle ABCD, some of whom would not have been yet 60 at a mid-year census, and hence would not cross the line HK. In contrast, to count the number of deaths that occur to men who become age 60 in the year 1970, and die before they are 61, it is necessary to count all life lines that terminate in the rhomboid ABEC some of whom will die in the year 1971. Data concerning ABCD is available from the age-specific mortality rates but no data is available concerning ABEC.

**IV.2.b Definition of terms used**

Now let us define the terms to be used for life table analysis, the conventional notation used by actuaries and demographers is forbidding to other professionals and in this chapter the notation has been simplified. In general, two suffixes will be used for each symbol, the first denoting the lower limit of the age group, and the second the width of the age interval. Thus \( m_{x} \), \( m_{x+t} \) is the mortality rate in the age interval \( x \) to \( x+t \). When the mortality for a single year of age is required, the second suffix is omitted and thus \( m_{x} \) is the mortality rate at age \( x \).

- **Calendar year**, \( y \). This is the normal calendar year from 1st January to 31st December.
- **Calendar interval**, \( i \). The calendar interval is the period of years from \( y \) to \( y+i \) during which a study is made.
- **Age**, \( x \). Persons are defined as being of age \( x \) from the precise anniversary date of their birth \( x \) years after this had occurred and remain aged \( x \) until the \( x+1 \) anniversary.
- **Age interval**, \( t \). The age interval, \( x \) to \( x+t \), is the interval of age between the \( x \)th anniversary date of birth and the \( (x+t) \)th anniversary date. Thus the \( x \)th birthday defines the lower boundary and the \( (x+t) \)th birthday the upper boundary.

For the final age interval in a life table the lower boundary is referred to as \( w \) and the upper boundary is undetermined. A typical set of age intervals for a life table is 0-1, 1-5, 5-10, 10-15, etc. 80-85, 85 and over. This notation differs from the practice in population statistics where age limits would be described as 0, 1-4, 5-9, etc. 80-84, 85 and over. Despite obvious difficulties this is the standard practice.
Number of deaths recorded, $D_{x,t}$. This is the number of deaths recorded during a calendar interval (usually one year, but it may be more) of persons who were aged $x$ to $x+t$ at the time of death.

Mid year population, $P_{x,t}$. This is the number of persons estimated as alive at the middle of the calendar interval (usually at 30th June) who were then aged $x$ and not yet aged $x+t$.

Birth Date population, $N_x$. This is the number of persons who become aged $x$ during a calendar interval.

Age specific mortality, $m_{x,t}$. This is the mortality rate for the specific age group $(x,x+t)$. It is the ratio of the deaths recorded during a year to the mid year population.

Probability of death at age $x$, $q_{x,t}$. This is the proportion of those alive at age $x$ who die during the age interval, $x$ to $x+t$.

Probability of surviving age $x$, $P_{x,t}$. This is defined as the population of those alive at age $x$ who survive to age $x+t$.

Number alive at age 0, $l_0$. This is the number of persons who are assumed to enter the life table alive at birth. It is an arbitrary number, usually 100,000 and is referred to as the radix.

Number alive at age $x$, $l_x$. This is defined as the number of persons who would survive to precise age $x$ out of the original radix $l_0$ if the life table mortality had operated upon them. Note $l_{x+t}$ will be the number of survivors at age $x+t$.

Life table deaths, $d_{x,t}$. The number of deaths that would occur to the $l_x$ persons of age $x$ in the interval $x$ to $x+t$ if exposed to the probability of death $q_{x,t}$.

Fraction of the last interval of life, $a_{x,t}$. This is the proportion of the last interval of life lived by the $d_{x,t}$ persons who die during the interval, $x$ to $x+t$.

Years lived in the interval $t$, $L_{x,t}$. This is the total number of years of life experienced by the $l_x$ persons aged $x$ during the age interval $x$ to $x+t$.

Years lived beyond age $x$, $T_x$. This is the total number of years of life experienced by the $l_x$ persons aged $x$ until all of them have died.

Expectation of life, $e_x$. This is the mean number of years of life experienced by the $l_x$ persons aged $x$ until all of them have died.

These definitions give rise to some algebraic relationships:

$$m_{x,t} = \frac{D_{x,t}}{P_{x,t}}$$  \hspace{1cm} 1

$$q_{x,t} = \frac{d_{x,t}}{N_{x,t}}$$  \hspace{1cm} 2

$$p_{x,t} = 1 - q_{x,t}$$  \hspace{1cm} 3

$$l_{x,t} = l_x \cdot p_{x,t}$$ \hspace{1cm} x=0,t,2t,...,w-1 \hspace{1cm} 4

$$L_{x,t} = t(l_x - (1-a_{x,t})d_{x,t})$$ \hspace{1cm} 5

$$T_x = \sum_{x=x}^{x=w} L_{x,t}$$ \hspace{1cm} 6

$$e_x = \frac{T_x}{l_x}$$ \hspace{1cm} x=0,1,...,w,... \hspace{1cm} 7

Clearly if $l_0$ is arbitrarily determined and if the $a_{x,t}$ are known, it is possible to progress directly from equation 2 to equation 7, but unfortunately the set of $q_{x,t}$ are not known because the deaths recorded are the $D_{x,t}$ and the population is $P_{x,t}$; $d_{x,t}$ and $N_{x,t}$ are not recorded. Thus we can calculate $m_{x,t}$ but not $q_{x,t}$.

Therefore, the relationship between $P_{x,t}$ and $q_{x,t}$ is required.
IV.2.c Relationship between $m_x$ and $q_x$

Let us first examine the relationship for a single year of life $x$ and a single calendar year $y$. Figure IV.3 represents a period of 3 calendar years $y-1$, $y$, $y+1$ and 1 single age group.

\[\text{Figure IV.3}\]

If a census were taken at the middle of year $y$, the persons enumerated at aged $x$ would be represented by all the life lines crossing the line $HK$ and the deaths during the year $y$ at age $x$ are all deaths recorded in the rectangle $ABCD$.

Thus,

\[m_x = \frac{\text{Deaths in ABCD}}{\text{Lines crossing HK}}\]

All the persons who become aged $x$ during the year $y$ would be represented by the number of lines crossing $AB$ and the number of deaths of these persons at age $x$ would be the deaths recorded in the rhomboid $ABEC$.

Thus,

\[q_x = \frac{\text{Deaths in ABEC}}{\text{Lines crossing AB}}\]

Let us make three assumptions:

(i) No change in age mortality rates between years $y$ and $y+1$.

Then the deaths counted in triangle $BEC$ will equal those in triangle $ACD$.

(ii) No change in the enumerated population of age $x$ in the years $y$ and $y+1$. Then the population crossing $AD$ equals the population crossing $BC$ and equals that crossing $HK$.

(iii) The number of deaths is constant throughout the year of age $x$ to $x+1$. Then the population crossing $AB$ equals population crossing $HK$ minus deaths in $ABC$ and deaths in $ABC$ equals half the deaths in $ABCD$.

Thus,

\[q_x = \frac{\text{Deaths in ABEC}}{\text{Population crossing AB}} = \frac{\text{Deaths in ABCD}}{\text{Population crossing HK} + \frac{1}{2} \text{Deaths in ABCD}} \quad \ldots \ldots \quad 8\]

\[= \frac{\text{Deaths in ABCD/HK}}{1 + \frac{1}{2} (\text{Deaths in ABCD/HK})}\]

Hence,

\[q_x = \frac{m_x}{1 + \frac{1}{2} m_x} = \frac{2m_x}{2 + m_x} \quad \ldots \ldots \quad 9\]

and substituting in equation (3)

\[p_x = \frac{2 - m_x}{2 + m_x} \quad \ldots \ldots \quad 10\]
and conversely by manipulation of equation 9,

\[ m_x = \frac{q_x}{1 - \frac{1}{2} q_x} = \frac{2q_x}{2 - q_x} \]

The discussion above brings out clearly that the relationship of equation 9 is dependent upon several assumptions.

The first assumption that there is no time trend in mortality is readily acceptable for two adjacent years unless there is a specific major catastrophe, of an epidemic, a famine, or a war, in one of them. This is therefore usually an acceptable assumption for the construction of a life table for a single calendar year.

The second assumption that there is no trend in the population structure is again acceptable for single years unless there are known to be large differences in births or migration due perhaps to famine or war in the year of birth of a group aged \( x \).

The third assumption that the number of deaths is constant throughout the interval of age, is clearly not valid for the first year of life, 0-1, where many deaths occur in the first month, nor is it valid for the second year of life, 1-2. At older ages, although mortality is increasing with age within one year of life, other random events, such as seasonal changes and epidemics, do make this assumption approximately valid.

Thus by use of equation 9, a complete set of \( q_x \) can be calculated if a set of \( m_x \) exist.

But several difficulties still exist,

(i) There is not usually a set of \( m_x \) available but a set of \( m_{x.5} \)
(ii) The problem of \( q_{0.5} \) has not been determined.
(iii) The problem of \( q_{x.5} \) has not been determined.

**IV.2.d Relationship of \( m_{x.t} \) and \( q_{x.t} \)**

Consider Figure IV.4, which portrays a population aged \( x \) to \( x+5 \) for 6/years, \( y \) to \( y+5 \). A mid year census is taken in year \( y \) represented by all lifelines crossing HK, all deaths over \( x \) and under \( x+5 \) are enumerated, represented by all lines terminating in ABCD. The probability of those aged \( x \) in year \( y \) dying before age \( x+5 \) is the proportion of lifelines which cross AB and terminate in ABXY.
As before to solve the relationship between these, several assumptions are required:

if, 1) no change in mortality between year y and y+5
and if, 2) no change in population structure

then the number of deaths at ages x, x+1, x+2, x+3, x+4 in the years y, y+1 to y+5 will remain constant and hence the deaths in ABCD will equal the deaths in ABXY,

and if, 3) the deaths in ABCD occur constantly throughout the age interval x to x+5

then, population at AB will equal 1/5th population at HK plus half the deaths in ABCD.

Thus,

\[
\begin{align*}
\frac{m_{x,5}}{Population_{on\ HK}} &= \frac{Deaths_{in\ ABCD}}{Deaths_{in\ ABXY}} = \frac{Deaths_{in\ ABCD}}{1/5\ HK + \frac{1}{2}\ Deaths_{in\ ABXY}} \\
q_{x,5} &= \frac{\frac{2m_{x,5}}{1/5 + \frac{1}{2}m_{x,5}}}{\frac{2 - 5m_{x,5}}{2 + 5m_{x,5}}} \\
p_{x,5} &= \frac{\frac{2 - t m_{x,t}}{2 + t m_{x,t}}}{1 + \frac{1}{2}m_{x,t}}
\end{align*}
\]

This gives,

\[
\begin{align*}
q_{x,5} &= \frac{m_{x,5}}{1/5 + \frac{1}{2}m_{x,5}} \\
p_{x,5} &= \frac{2 - 5m_{x,5}}{2 + 5m_{x,5}}
\end{align*}
\]

and in general,

\[
\begin{align*}
q_{x,t} &= \frac{m_{x,t}}{1/5 + \frac{1}{2}m_{x,t}} \\
p_{x,t} &= \frac{2 - t m_{x,t}}{2 + t m_{x,t}}
\end{align*}
\]

The first two assumptions may still be acceptable if the interval t is relatively short and it is usually 5 years in most methods of calculating abridged tables. The third assumption is, however, clearly invalid over age 10, as the deaths at age x+4 will be greater than those at age x, but for the moment we accept this assumption and will examine the effect of the bias introduced later.

IV.2.e The probability of surviving first year of life, \(p_0\)

It is notorious that all censuses omit some children in the first year of life and hence it is not valid to estimate \(m_0\) from the population. It is necessary to use the number of births registered, \(B_0\), in the year and the number of infant deaths, \(D_0\). The age at death of infant deaths will be recorded in narrower intervals than 1 year.

If the number of deaths in the first month is \(d_{1/12}\), then the probability of dying in the first month is, \(q_{1/12}\), where

\[
q_{1/12} = \frac{D_{1/12}}{B_{0}}
\]

and the proportion of survivors to one month of age is

\[
p_{1/12} = 1 - q_{1/12}
\]

and the number surviving to age one month is

\[
f_{1/12} = B_0 p_{1/12}
\]
This can then be repeated for age intervals 1 - 3 months, 3 - 6 months, 6 - 9 months
and 9 - 12 months, to obtain the proportion surviving to one year, $\ell'_1$.

Then by definition $p_0$ the probability of the first survival to the first birthday is

$$p_0 = \ell'_1/\ell'_0$$

and $q_0 = 1 - p_0$ .......................... 19

A similar method is used to obtain $p_{1.4}$ from the single years of life 1, 2, 3 and 4.
Modifications in the calculation of $p_0$ are discussed in more detail in Chapter V.

IV.2.f  The last interval of life table

If the ages of the population were accurately known up to the highest age it would be
possible to calculate $m_x$ up to an age group when there are no further persons alive. It is
clear, however, that the accuracy of reported ages at advanced age is very inadequate and it
is unusual to have verification of the date of birth at these ages. Consequently a pragmatic
way of closing the life table is adopted which gives an estimate of the years of life expected
to be lived beyond age 85.

Let the lower boundary of the upper age group be $w$, then the corresponding values
$D_w$, $P_w$, $m_w$, $q_w$, $p_w$, $\ell_w$, $L_w$, $T_w$, and $e_w$, all refer to this open ended group of $w$ and over which
theoretically is an infinite interval, but is in fact very limited.

There are by definition no survivors of this interval hence $p_w = 0$ and $q_w = 1$.

Every person alive at age $w$ will die during this interval consequently $\ell_w = d_w$,
but also the mortality rate at any age is defined as the number of persons who die at that age
divided by the persons years lived at that age, therefore,

$$m_w = d_w/L_w = \ell_w/L_w$$

hence,

$$L_w = \ell_w/m_w$$ .......................... 20

$T_w$ is defined as the total number of years lived beyond age $w$ until all are dead.
All the $\ell_w$ persons die in this interval.
Therefore,

$$T_w = L_w$$ .......................... 21

and thus the life table can be closed.

We are now in a position to construct an abridged life table.

IV.3.  Construction of an Abridged Current Life Table.

The method of constructing an abridged current life table using only a standard desk
calculator is shown in Tables IV.1 and 2 for the male population of Trinidad and Tobago.
Although many computer programs are available to do this work a student will obtain a much
better appreciation of the use of the life table if he is confident that he can calculate one
himself and knows precisely how each figure has been obtained.

| TABLE IV.1 HERE |

The first column shows the usual age groups used for a five year age grouping of an
abridged table.

Columns 2 and 3 record the number of males estimated as resident at the middle of the
year and the number of deaths registered during the year. It will have been necessary to
allocate persons of unstated age according to some agreed formula.

Column 4 is the ratio of Col. 3 to Col. 2. and is the age specific mortality rate per
unit of population. It is usually recorded to five or six decimals even though the denominators
may not justify such accuracy because subsequent calculation will be extensive and
cumulative errors may occur if these figures are rounded off too early in the exercise.

Columns 5 and 6 are straight calculations of $2 - t.m_{x,t}$ and $2 + t.m_{x,t}$. They are not
required for age 0 - 1 and for age 85+ as $p_0$ is obtained by other means and $p_{85+}$ is assumed
to be 0.00000.
### TABLE IV.1

**TRINIDAD AND TOBAGO MALES, 1963**

5-year Abridged Life Table Calculation of $p_{x.t}$

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Population $P_{x.t}$</th>
<th>Deaths $D_{x.t}$</th>
<th>Mortality $m_{x.t}$</th>
<th>2-t.m $2t_{x.t}$</th>
<th>2+t.m $2+t_{x.t}$</th>
<th>Survival $P_{x.t}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>x to x+t</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>0 - 1</td>
<td>16,900</td>
<td>789</td>
<td>.04669</td>
<td>(a)</td>
<td>(a)</td>
<td>(a)</td>
</tr>
<tr>
<td>1 - 4</td>
<td>61,550</td>
<td>178</td>
<td>.00289</td>
<td>1.98844</td>
<td>2.01156</td>
<td>.98856</td>
</tr>
<tr>
<td>5 - 10</td>
<td>66,750</td>
<td>46</td>
<td>.00069</td>
<td>1.99565</td>
<td>2.00435</td>
<td>.99561</td>
</tr>
<tr>
<td>10 - 15</td>
<td>54,350</td>
<td>48</td>
<td>.00088</td>
<td>1.99560</td>
<td>2.00440</td>
<td>.99561</td>
</tr>
<tr>
<td>15 - 20</td>
<td>47,050</td>
<td>50</td>
<td>.00106</td>
<td>1.99470</td>
<td>2.00053</td>
<td>.99471</td>
</tr>
<tr>
<td>20 - 25</td>
<td>36,250</td>
<td>58</td>
<td>.00160</td>
<td>1.99200</td>
<td>2.00800</td>
<td>.99203</td>
</tr>
<tr>
<td>25 - 30</td>
<td>30,300</td>
<td>54</td>
<td>.00178</td>
<td>1.99110</td>
<td>2.00890</td>
<td>.99114</td>
</tr>
<tr>
<td>30 - 35</td>
<td>25,550</td>
<td>54</td>
<td>.00211</td>
<td>1.98945</td>
<td>2.01055</td>
<td>.98951</td>
</tr>
<tr>
<td>35 - 40</td>
<td>23,200</td>
<td>82</td>
<td>.00353</td>
<td>1.98235</td>
<td>2.01765</td>
<td>.98250</td>
</tr>
<tr>
<td>40 - 45</td>
<td>23,500</td>
<td>117</td>
<td>.00498</td>
<td>1.97310</td>
<td>2.02490</td>
<td>.97541</td>
</tr>
<tr>
<td>45 - 50</td>
<td>19,650</td>
<td>145</td>
<td>.00730</td>
<td>1.96350</td>
<td>2.03650</td>
<td>.96415</td>
</tr>
<tr>
<td>50 - 55</td>
<td>17,300</td>
<td>195</td>
<td>.01112</td>
<td>1.94365</td>
<td>2.05635</td>
<td>.94519</td>
</tr>
<tr>
<td>55 - 60</td>
<td>12,650</td>
<td>265</td>
<td>.02095</td>
<td>1.89255</td>
<td>2.10475</td>
<td>.90046</td>
</tr>
<tr>
<td>60 - 65</td>
<td>10,050</td>
<td>310</td>
<td>.03085</td>
<td>1.84575</td>
<td>2.15425</td>
<td>.85879</td>
</tr>
<tr>
<td>65 - 70</td>
<td>5,880</td>
<td>363</td>
<td>.06205</td>
<td>1.68975</td>
<td>2.31025</td>
<td>.73141</td>
</tr>
<tr>
<td>70 - 75</td>
<td>4,350</td>
<td>301</td>
<td>.06920</td>
<td>1.65400</td>
<td>2.34600</td>
<td>.70503</td>
</tr>
<tr>
<td>75 - 80</td>
<td>2,200</td>
<td>252</td>
<td>.11455</td>
<td>1.42725</td>
<td>2.57275</td>
<td>.55476</td>
</tr>
<tr>
<td>80 - 85</td>
<td>1,150</td>
<td>166</td>
<td>.14435</td>
<td>1.27825</td>
<td>2.72175</td>
<td>.46964</td>
</tr>
<tr>
<td>85+</td>
<td>750</td>
<td>186</td>
<td>.24800</td>
<td>(b)</td>
<td>(b)</td>
<td>(b)</td>
</tr>
<tr>
<td>ALL AGES</td>
<td>459,550</td>
<td>3,659</td>
<td>.00796</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a) $p_{0.1}$ is obtained by other methods
b) $p_{85+}$ is by definition 0.00000

### TABLE IV.2

**TRINIDAD AND TOBAGO MALES, 1963**

5-year Abridged Life Table Calculation of expectation of life $e_x$

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Living $l_x$</th>
<th>Survival $p_{x.t}$</th>
<th>Years Lived $L_{x.t}$</th>
<th>Total Years $T_x$</th>
<th>Expectation $e_x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>x to x+t</td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
</tr>
<tr>
<td>0 - 1</td>
<td>100,000</td>
<td>.95510 (a)</td>
<td>96,181 (a)</td>
<td>6,368,811</td>
<td>63.69</td>
</tr>
<tr>
<td>1 - 5</td>
<td>95,510</td>
<td>.98856</td>
<td>379,854</td>
<td>6,272,630</td>
<td>65.67</td>
</tr>
<tr>
<td>5 - 10</td>
<td>94,417</td>
<td>.99656</td>
<td>471,275</td>
<td>5,892,776</td>
<td>62.41</td>
</tr>
<tr>
<td>10 - 15</td>
<td>94,093</td>
<td>.99561</td>
<td>469,433</td>
<td>5,421,501</td>
<td>57.62</td>
</tr>
<tr>
<td>15 - 20</td>
<td>93,680</td>
<td>.99471</td>
<td>467,160</td>
<td>4,952,068</td>
<td>52.86</td>
</tr>
<tr>
<td>20 - 25</td>
<td>93,184</td>
<td>.99203</td>
<td>464,062</td>
<td>4,484,908</td>
<td>48.13</td>
</tr>
<tr>
<td>25 - 30</td>
<td>92,441</td>
<td>.99114</td>
<td>460,158</td>
<td>4,020,846</td>
<td>43.50</td>
</tr>
<tr>
<td>30 - 35</td>
<td>91,622</td>
<td>.98981</td>
<td>455,708</td>
<td>3,560,888</td>
<td>38.86</td>
</tr>
<tr>
<td>35 - 40</td>
<td>90,661</td>
<td>.98250</td>
<td>449,340</td>
<td>3,104,980</td>
<td>34.25</td>
</tr>
<tr>
<td>40 - 45</td>
<td>89,075</td>
<td>.97541</td>
<td>439,898</td>
<td>2,655,640</td>
<td>29.81</td>
</tr>
<tr>
<td>45 - 50</td>
<td>86,884</td>
<td>.96415</td>
<td>426,632</td>
<td>2,215,742</td>
<td>25.50</td>
</tr>
<tr>
<td>50 - 55</td>
<td>83,769</td>
<td>.94519</td>
<td>407,368</td>
<td>1,789,110</td>
<td>21.36</td>
</tr>
<tr>
<td>55 - 60</td>
<td>79,178</td>
<td>.90046</td>
<td>376,188</td>
<td>1,381,742</td>
<td>17.45</td>
</tr>
<tr>
<td>60 - 65</td>
<td>71,297</td>
<td>.85679</td>
<td>330,958</td>
<td>1,005,554</td>
<td>14.10</td>
</tr>
<tr>
<td>65 - 70</td>
<td>61,086</td>
<td>.73141</td>
<td>264,412</td>
<td>674,596</td>
<td>11.04</td>
</tr>
<tr>
<td>70 - 75</td>
<td>44,679</td>
<td>.70503</td>
<td>190,448</td>
<td>410,084</td>
<td>9.18</td>
</tr>
<tr>
<td>75 - 80</td>
<td>31,500</td>
<td>.65476</td>
<td>122,438</td>
<td>219,736</td>
<td>6.98</td>
</tr>
<tr>
<td>80 - 85</td>
<td>17,475</td>
<td>.48964</td>
<td>64,205</td>
<td>97,298</td>
<td>5.57</td>
</tr>
<tr>
<td>85+</td>
<td>8,207</td>
<td>.00000</td>
<td>33,093 (b)</td>
<td>33,093 (b)</td>
<td>4.03</td>
</tr>
</tbody>
</table>

a) $L_{0.1}$ is obtained by other means
b) $L_w = \frac{l_{85}}{m_{85}}$
Column 7 is the ratio of Column 5 and Column 6 and this is the value of the proportion of survivors of the interval \( t \). \( P_{x,t} \), \( P_0 \) and \( P_{85} \) are obtained by other methods.

A new table is now started using these \( P_{x,t} \) values. Table IV.2. Column 2 is the number of persons alive at age \( x \) to \( x+t \), \( l_x \). The first item \( l_0 \) the radix is assumed to be 100,000 then each successive value of \( l_{x+t} \) is obtained by multiplying \( l_x \) by \( P_{x,t} \) in accordance with equation 4.

Column 4 is the years of life lived in the interval \( t \), \( L_x \), and as \( a_{x,t} \) is assumed to be 0.5 this is \( t(\frac{l_x + l_{x+t}}{2}) \). \( L_0 \) is obtained by special means and \( L_{85} \) is \( l_{85}/a_{85} \).

Column 5 is the years lived beyond age \( x \), \( T_x \), and is simply a successive summation of column 4 starting at the bottom where \( T_{85} = L_{85} \).

Column 6 is the expectation of life at age \( x \), \( e_x \), and is the ratio of column 5 to column 3. This is rounded off to one or two decimals according to the size of the population and the accuracy of the original data. The second decimal is one hundredth part of a year or about half a week and the third decimal is about 8 hours in a life time. Obviously such a degree of accuracy is not required in the analysis of a public health problem.

The abridged life table is now complete. All of this can be done on a standard desk calculating machine in approximately one hour according to the experience of the calculator. Packaged computer programs are available for most computer systems and many of these will allow other more complex formulae to be used if these are required.

IV.4 Other Methods of Constructing Abridged Life Tables

The methods described above for constructing life tables do not graduate the mortality to obtain smoothed or interpolated values, but construct a life table which is a direct summary of the specific mortality reported in a stated population in a known period of time subject to all the imperfections of the recording procedures and the hazards of particular epidemics or fluctuations in mortality.

There are many other ways of completing the abridged life table and most methods derive a formula to link the age specific death rates, \( m_{x,5} \) for 5 year intervals of age with the probability of survival \( p_{x,5} \) for the 5 years from the beginning of the interval. This section will discuss some of these methods in current use by health statisticians.

IV.4.a Reed and Merrel (1939) sought a functional relationship between \( m_{x,5} \) and \( p_{x,5} \) which was a second order polynomial in \( m_x \) related to log \( p_{x,5} \). In the case of a five year interval their formula was:

\[
p_{x,5} = \exp\left(-5m_{x,5} + 53a m^2_{x,5}\right)
\]

and,

\[
\log p_{x,5} = -(5m_{x,5} + 3a m^2_{x,5})
\]

where \( a \) is a constant to be determined from a standard life table for each population. (Note: This, \( a \), is not to be confused with Chiang's \( a_x \) the mean interval lived during \( x \) to \( x+n \) by those who die during the interval).

They calculated the survival rates for the first 5 years of life from the birth and death registration and then using the \( p_{x,5} \) values derived from equation 22 the \( l_x \) value can be determined from the relationship \( l_{x+5} = l_x \cdot p_{x,5} \).

(Note: Although King's method is the fundamental method taught to actuarial students it is not used in public health work and is unnecessarily difficult for them, hence I omit King's method).

IV.4.b Greville (1943) estimated \( q_{x,5} \) the probability of death in a five year period from \( m_{x,5} \) the five year interval mortality rate by considering Gompertz's Law of Mortality which is of the form:

\[
m_x = B e^{c x}
\]

or,

\[
\log m_x = \log B + x \log c
\]

which is a linear relationship between \( \log m_x \) and \( x \) with the slope of this line being \( \log c \).
Experience shows that variations in the value of log \( c \) has little effect at the younger ages and at higher ages most life tables follow Gompertz's relationship closely with a value of log \( c \) which is approximately:

\[
0.080 < \text{log } c < 0.110
\]

He then showed that

\[
q_{x,t} = \frac{m_{x,t}}{(1/t + m_{x.t}(1/2 + t/12(m_{x.t} - \text{log } c))}
\]

He found from U.S. life tables that it was convenient to assume that log \( c = 0.090 \), and then for a 5 year interval

\[
q_{x.5} = \frac{m_{x.5}}{(0.2 + 0.4625 m_{x.5} + 0.4167 m_{x.5}^2)}
\]

He used exact methods to determine \( q_0 \) to \( q_4 \) and hence \( \ell_5 \) and then was able to complete the abridged life table in the usual way by calculating the survivors at each 5 year interval from age 5 by

\[
\ell_{x+5} = \ell_x(1 - q_{x.5})
\]

He then assumed that the life table death rate was the same as the population death rate, so that

\[
m_{x.5} = \ell_{x.5} q_{x.5}
\]

where \( L_{x.5} \) is the number of years lived in the interval \( x \) to \( x+5 \),

thus,

\[
L_{x.5} = \frac{\ell_{x.5}}{m_{x.5}}
\]

To close the table he assumed in the usual way that for the final age group \( L_w = \ell_w/m_w \) as in equation 20. He then obtained \( T_x \) and \( e_x \) in the usual way.

IV.4.c The U.S. Abridged Life Tables (1947), were calculated by Greville using a different technique. This method does not postulate that a Gompertz relationship exists but it assumes that there is a standard life table that is related to the specific population for which an abridged life table is required. This is appropriate when calculating life tables for regions of a country when a full national life table is available or when calculating annual tables for a country in years immediately following the tabulation of a complete table.

The method calculates the ratios \( r_{x.5} \), of \( q_{x.5} \) to \( m_{x.5} \) for the standard table and assumes that these ratios remain constant in the population for which an abridged life table is required.

Thus,

\[
r_{x.5} = \frac{q_{x.5}}{m_{x.5}}
\]

for the standard table,

and hence,

\[
1_{q_x} = r_{x.5} \frac{m_{x.5}}{m_{x.5}}
\]

for the specific table where \( 1_{q_{x.5}} \) and \( 1_{m_{x.5}} \) are the values in the specific population for which an abridged life table is required.

Special calculations are made to obtain values for \( q_0 \) to \( q_4 \) and thus obtain an accurate value for \( 1\ell_5 \), for the specific population.

To obtain the values of \( L_{x.5} \) for the specific table, this method then calculates for the standard table the mean number of years of life lost during the interval \( x \) to \( x+5 \) by those who died in that interval which is designated \( g_x \), and assumes that this value can be used for the specific table.

Thus from the standard table,

\[
g_x = \frac{5 \ell_x - L_{x.5}}{\ell_x q_{x.5}}
\]
and then in the specific table this value of \( q_x \) is used to calculate \( l_{x,5} \).

\[
l_{x,5} = \frac{1}{x} \left( 5 - \frac{1}{q_{x,9} K_x} \right)
\]

\( l_x \) and \( q_x \) are then obtained in the usual way.

(Note: There is a close relationship between \( g_x \) used in this technique and the \( a_x \) used by Chiang).

IV.4.d Method in England and Wales

The Registrar General assumed that mortality in the interval between \( x \) and \( x+5 \) is constant and hence that the years of life lost by those dying in the interval is one half of the interval and thus equation 14 can be used to calculate \( p_{x,5} \),

\[
p_{x,5} = \frac{2-5m_x}{2m_x}
\]

He calculates \( p_0 \) and \( p_{1.4} \) exactly and closes the table by assuming \( q_w = 1/n_w \).

IV.4.e Wiesler (1951) suggested that for an abridged table the relationship \( \ell_{x+1}/\ell_x = p_x \) could be modified by the introduction of a parameter \( t \) such that over an interval of \( n \) years

\[
\frac{\ell_{x+n}}{\ell_x} = (p_x)^t
\]

he calculated the values of \( t \) for 5 year intervals for seven countries, male and female, and showed that, \( t \), could be assumed to be equal to 5 from age 5 years up to 44 for men and up to 49 for women, but after that, \( t \), was greater than 5. For ages below 5 he also calculates a value for, \( t \), but it is obviously preferable to calculate the individual values of \( p_0, p_1, p_2, p_3, p_4 \), in the usual way.

IV.4.f Chiang (1968), introduced the concept \( a_x \) the mean proportion of the last year of life lived in the case of the complete life table and \( a_{x,t} \) the mean proportion of the last interval of life lived for the abridged table. This concept is both simple and powerful when the information is available. What is required is that the date of birth be available on the death certificate then it is possible to calculate immediately the number of days elapsed between the \( x \)th birthday and the date of death. This is a parameter that lies along the original Lexis lifeline and hence no manipulation of calendar intervals to equal age intervals is required.

Chiang then showed that the proportion of the population aged \( x \) dying in the interval \( x \) to \( x+1 \) is given by

\[
q_x = \frac{D_x}{N_x}
\]

with only one assumption that there is no calendar trend in mortality, but

\[
m_x = \frac{D_x}{P_x}
\]

and,

\[
p_x = \frac{N_x - D_x}{x} + a_{x,x} x
\]

Thus,

\[
m_x = \frac{D_x/N_x}{1 - (1-a_x)D_x/N_x} = q_x
\]

and thus,

\[
q_x = \frac{m_x}{1 + (1-a_x)m_x}
\]

and similarly for age intervals of \( t \)

\[
q_{x,t} = \frac{m_{x,t}}{1/t + (1-a_{x,t})m_{x,t}}
\]

If \( a_x = \frac{1}{2} \), then equation 38 is identical with equation 9 and if \( a_{x,t} = \frac{1}{2} \), equation 39 is identical with equation 12.

Chiang investigated the distribution of \( a_x \) for the Californian life table 1960 and was able to show that except for ages 0 and 1, \( a_x \) did not depart significantly from 1/2. It
is, therefore, possible to use equation 9 for all ages over 2 years for the complete life table. He then showed that if \( a_x \) is 1/2 and \( q_x \) follows the life table pattern \( a_{x+t} \) will not be equal to 0.5 but will in general approach 0.52. He was able to show that the results were similar for 7 different countries.

Hence for the abridged life table it might be preferable to use equation 39 rather than equation 12, but the effect on the total expectation of life at birth is of the order of an addition of one week, which is usually less than the standard error of this statistic.

These adjustments and refinements are not of importance to the mortality analyst who is looking for factors having large effects upon mortality.

Chiang's method is, however, also applicable to the first few years of life where the date of birth of the infant deaths will be easily obtained. Using this information he obtained the following values for California, 1970.

\[
\begin{align*}
  a_0 & = 0.09 \\
  a_1 & = 0.43 \\
  a_2 & = 0.45 \\
  a_3 & = 0.47 \\
  a_4 & = 0.49
\end{align*}
\]

It cannot be assumed that these values can be transferred to any other population, but the method is simple and quick when the basic data is available.

IV.4.g Keyfitz (1968) tackled the difficulty that arises because of assuming that the population in successive cohorts remains constant when in fact it is clear that in most countries in the world the population is increasing.

He introduced a parameter, \( r \), which is the rate of increase of successive annual cohorts. He was able to demonstrate that an annual increase of 0.01 in \( r \), would lower the computed value of \( e_0 \) by about 1 week of life. The maximum rate of natural increase in a national population is of the order of 4 per cent and thus at the most this adjustment will make a difference of the order of 1 month in the expectation of life at birth.

IV.4.h Although these various methods are based upon very different assumptions, when applied to actual mortality rates they do not result in significant differences of importance to mortality analysis. But any analysis based upon life table technique must state the methods used to calculate the table in order that valid comparisons can be made. In all cases it is important that \( p_0 \) the probability of surviving the first year of life, be calculated with great care and it is preferable that individual values are obtained for \( p_1, p_2, p_3, p_4 \), which will enable \( \ell_5 \) to be determined accurately and then abridged approximation methods can be used from 5 up to the closing age of the table. The details of calculating \( p_0 \) will be discussed in Chapter V.

IV.4.j Life table death rate

It is natural to ask what would be the crude death rate for the life table population which was subject to the age specific death rates \( m_x \) and adjusted or weighted by the population \( \ell_x \). Let this crude death rate be \( m_L \), then

\[
m_L = \frac{\Sigma \ell_x}{\Sigma \ell_x} \quad \ldots \ldots \ldots \quad 40
\]

But by the method of construction all the population must die and hence \( \Sigma \ell_x \) will be the radix \( \ell_0 \) and similarly the total number of years lived by the whole population will be \( \ell_x = T_0 \).

Hence,

\[
m_L = \frac{\ell_0}{T_0} \quad \ldots \ldots \ldots \quad 41
\]

but

\[
\frac{T_0}{\ell_0} = e_0 \quad \ldots \ldots \ldots \quad 42
\]

hence

\[
m_L = \frac{1}{e_0} \quad \ldots \ldots \ldots \quad 43
\]

So the crude life table death rate is by definition the reciprocal of the expectation of life at birth and hence does not contribute any additional information. For this reason it is seldom used.
IV.5. The Variance of the Life Table

The concept of the variance and of the standard error applied to mortality analysis was discussed in detail in Chapter III.5 where the variances of $m_x$, $q_x$, and $p_x$ were derived and the use of these in determining confidence intervals for estimates of total mortality or for tests of significance of the differences between rates were discussed. The same principles and similar formulations are equally valid in the study of a life table.

The life table is essentially no more than a method of analysis of the deaths that actually have occurred at stated ages in a defined population, although it is a lengthy and complex method. These deaths may be considered to be the result of a stochastic process which is operating with different intensity at different ages and in different sexes; the intensity of the mortality is observed in various strata of sex and age groups and then the mean value and the variance can be obtained by the standard methods of stratified sampling.

IV.5.a The standard error of mean age at death

If the life table is considered, not in terms of the survivors $\ell_x$, but rather in terms of the number of deaths, $d_{x,t}$ which are calculated to occur between age $x$ and $x+t$, then the $\ell_x$ column is seen to be the sum of the $d_{x,t}$ column starting the summation at the highest age group. The $l_{x,t}$ column is a simple adjustment to allow for scale factors and the $T_x$ column which is the cumulant of $l_{x,t}$ is closely related to the second cumulant of $d_{x,t}$. If the $T_x$ column is itself now summed from the highest ages, the sum of the $T_x$ column is related to the third cumulant of $d_{x,t}$. It is well known that the mean and the variance of a frequency distribution may be obtained from the first three cumulants of the distribution. Hence the mean age at death and the variance of the frequency distribution of age at estimated death are available.

In general, if $K_1$, $K_2$, $K_3$, are the first three cumulants of a frequency distribution which is tabulated in successive unit intervals, $x$; for $i = 0$ to $n$ with cell frequencies $f_i$, then,

$$K_1 = \sum_{i=0}^{n} f_i$$  

$$K_2 = \sum_{i=1}^{n} x_i f_i$$  

$$K_3 = \sum_{i=2}^{n} (x_i)(x_i-1) f_i/2$$

then the total frequency of events, $N$, is given by

$$N = K_1$$

and the mean of the distribution $\bar{x}$ is given by

$$\bar{x} = K_2/K_1$$

and the variance of the distribution $\text{Var}(x)$ is given by

$$\text{Var} x = \left( \frac{2K_3 + K_2}{K_1} \right) - \left( \frac{K_2}{K_1} \right)^2$$

Example: In the case of Trinidad males, Table IV.2 shows the calculated deaths in five year intervals and the cumulants of the frequency distribution of $y_i$ for $i = 0$ to 17, where

$$5y = x \text{ and } f_i = d_{x,t}$$

Then,

$$K_1 = 100,000$$

$$K_2 = 1,223,248$$

$$K_3 = 7,729,706$$
TABLE IV.3

TRINIDAD AND TOBAGO (MALES) 1963

Use of cumulants of calculated deaths to estimate mean and variance of age at death

<table>
<thead>
<tr>
<th>Working Value</th>
<th>Age Group x, x+t</th>
<th>Central Age z</th>
<th>Deaths d_{x.t}</th>
<th>1st Sum</th>
<th>2nd Sum</th>
<th>3rd Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0 - 5</td>
<td>2.5</td>
<td>5,583</td>
<td>100,000</td>
<td>1,223,248</td>
<td>7,729,706</td>
</tr>
<tr>
<td>1</td>
<td>5 - 10</td>
<td>7.5</td>
<td>324</td>
<td>94,417</td>
<td>1,128,831</td>
<td>etc.</td>
</tr>
<tr>
<td>2</td>
<td>10 - 15</td>
<td>12.5</td>
<td>413</td>
<td>94,093</td>
<td>1,034,738</td>
<td>&quot;</td>
</tr>
<tr>
<td>3</td>
<td>15 - 20</td>
<td>17.5</td>
<td>496</td>
<td>93,680</td>
<td>941,058</td>
<td>&quot;</td>
</tr>
<tr>
<td>4</td>
<td>20 - 25</td>
<td>22.5</td>
<td>743</td>
<td>93,184</td>
<td>847,874</td>
<td>&quot;</td>
</tr>
<tr>
<td>5</td>
<td>25 - 30</td>
<td>27.5</td>
<td>819</td>
<td>92,441</td>
<td>755,433</td>
<td>&quot;</td>
</tr>
<tr>
<td>6</td>
<td>30 - 35</td>
<td>32.5</td>
<td>961</td>
<td>91,622</td>
<td>663,811</td>
<td>&quot;</td>
</tr>
<tr>
<td>7</td>
<td>35 - 40</td>
<td>37.5</td>
<td>1,586</td>
<td>90,661</td>
<td>573,150</td>
<td>&quot;</td>
</tr>
<tr>
<td>8</td>
<td>40 - 45</td>
<td>42.5</td>
<td>2,191</td>
<td>89,075</td>
<td>484,075</td>
<td>&quot;</td>
</tr>
<tr>
<td>9</td>
<td>45 - 50</td>
<td>47.5</td>
<td>3,115</td>
<td>86,884</td>
<td>397,191</td>
<td>&quot;</td>
</tr>
<tr>
<td>10</td>
<td>50 - 55</td>
<td>52.5</td>
<td>4,591</td>
<td>83,769</td>
<td>313,422</td>
<td>&quot;</td>
</tr>
<tr>
<td>11</td>
<td>55 - 60</td>
<td>57.5</td>
<td>7,881</td>
<td>79,178</td>
<td>234,244</td>
<td>&quot;</td>
</tr>
<tr>
<td>12</td>
<td>60 - 65</td>
<td>62.5</td>
<td>10,211</td>
<td>71,297</td>
<td>162,947</td>
<td>&quot;</td>
</tr>
<tr>
<td>13</td>
<td>65 - 70</td>
<td>67.5</td>
<td>16,407</td>
<td>61,086</td>
<td>101,861</td>
<td>&quot;</td>
</tr>
<tr>
<td>14</td>
<td>70 - 75</td>
<td>72.5</td>
<td>13,179</td>
<td>44,679</td>
<td>57,162</td>
<td>33,889</td>
</tr>
<tr>
<td>15</td>
<td>75 - 80</td>
<td>77.5</td>
<td>14,025</td>
<td>31,500</td>
<td>25,682</td>
<td>8,207</td>
</tr>
<tr>
<td>16</td>
<td>80 - 85</td>
<td>82.5</td>
<td>9,268</td>
<td>17,475</td>
<td>8,207</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>85+</td>
<td>87.5</td>
<td>8,207</td>
<td>8,207</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

\[ K_1 = 100,000 \quad K_2 = 1,223,248 \quad K_3 = 7,729,706 \]
Hence,

\[ N = 100,000 \]

\[ \bar{y} = 12.23248 \]

\[ \text{Var}(y) = 17.1930 \]

\[ \text{Std. dev.}(y) = 4.140 \]

Hence,

\[ \text{mean age at death}, \bar{x} = 63.66 \text{ years} \]

\[ \text{std. dev. of ages at death}, \text{s.d.}(x) = 20.73 \text{ years} \]

The mean age at death of the calculated deaths should equal the mean expectation of life at birth of the radix \( e_0 \). The difference in \( \bar{x} = 63.66 \) from Table IV.3 and \( e_0 = 63.69 \) in Table IV.2 is entirely due to the incorrect weighting of the first and last age groups in Table IV.3, which has been done for simplicity to demonstrate the basic method.

**TABLE IV.3. HERE**

There is in practice no need to calculate the working values for this exercise, the values are already available in the standard life table itself. From Table IV.3 we have

\[ K_1 = 100,000 \]

\[ K_2 = T_x = 6,368,811 \]

\[ \text{Mean age}(\bar{x}) = 63.69 \text{ years} \]

and, \( K_3 \) can be obtained by summation of \( T_x \) from the highest age to age 5 and multiplying this by the interval width \( t \)

\[ K_3 = (41920362) \ t \]

Hence,

\[ \text{Var}(x) = 454.32 \]

\[ \text{S.d.}(x) = 21.31 \]

To obtain these values the only additional calculation has been to sum the \( T_x \) column.

Thus the life table for Trinidad is a frequency distribution of the age at death with mean \( \bar{x} = 63.69 \) and variance \( \text{Var}(x) = 454.32 \) standard deviation 21.31 years. What is the sampling variance of the mean?

If the deaths had occurred at random in a theoretical stable life table population then the sampling variance

\[ \text{Var}(\bar{x}) = \frac{\text{Var}(x)}{N} \]

where \( N \), the number of events sampled, is the total number of deaths at all ages \( N = D_{x,t} = 3659 \).

Hence if the deaths were a random sample from a stable population the sampling variance would be

\[ \text{Var}(\bar{x}) = 454.32/3659 = .1242 \]

and Std. error of \( \bar{x} = .35 \)

The 95% confidence limits of \( \bar{x} \) would therefore be at least 63.69 ± .70.

**IV.5.b Standard error of the expectation of life**

It is clear, however, that the deaths actually recorded are not a random sample drawn from the stable life table population, but are a sample drawn from the existing population which has a different distribution to the stable population and hence the actual variance will in fact be greater than that obtained by simple random sampling. Moreover, deaths will occur in clusters due to climate, accidents, infectious diseases and socio-economic factors and also due to the errors discussed in Chapter II, all of which will tend to increase the observed variance; thus the standard error of the mean age at death and hence of the expectation of life at birth in fact will be larger than that estimated above.

In most population mortality studies the standard deviation of the mean age at death in the stable population will be greater than 15 years, but less than 25 years. A country or state with a population of about 20 million will have approximately 100,000 recorded deaths in
one sex in a year and hence the standard error of the mean age at death will be at least of the order of .05 year to .08 year. In such a country the 95 per cent confidence limits would therefore be at least of the order of .10 year to .15 year, i.e. one or two months of life.

In smaller countries the confidence limits would of course be wider in inverse proportion to the square root of the number of recorded deaths.

IV.5.c Scale of bias introduced by various methods

This enables some estimate to be made of the importance of the various conflicting factors in methods of computing the abridged life table.

The five year abridged table computed by the actuarial method gives a biased result because it assumes that deaths occur at a constant rate during an interval $t$, and hence that the central age may be assumed to be the mean age at death of those dying in the interval. Although this appears to be a perfectly valid assumption for an interval of one year where other sources of variation completely swamp the effect of increasing mortality with age, it is not true for an interval of 5 years.

Chiang's studies of seven countries would suggest that $a_{x:t}$ the proportion of the last interval of life which is experienced is approximately 0.52 rather than 0.5 and hence the bias in abridged tables of 5 year intervals compared to the single years of life table, is of the order of about 0.1 years, or about one month.

Similarly for a ten year interval the value of $a_{x:t}$ should be approximately 0.53 and hence the bias is of the order of 0.3 year compared with the single years of life table or 0.1 years compared with the five year table. Table IV.4 is the ten year interval life table for Trinidad where the expectation of life in ten year intervals is compared with the expectation derived from the 5 year table. Here it is seen that the difference between the two methods of estimating $e_x$ remains relative constant from $e_0$ to $e_{55}$.

| TABLE IV.4. HERE |

Chiang's method, based upon the known age at death in days, is an exact method and makes no assumptions concerning the distribution of mortality in the interval $t$. The methods of Reed and Merrel, of Greville and of most other workers in this subject all make assumptions concerning the shape of the mortality curve and make adjustments accordingly to their basic formula for $a_{x:t}$. These adjustments are all of the same order as the adjustments suggested by the Chiang method and for most national populations based upon total populations of 20 million these adjustments are less than the confidence limits of the expectation of life.

The method of Keyfitz, however, is a completely different technique which by re-iterative approximations on a computer makes an adjustment to allow for the bias which arises from the current population steadily changing with each successive cohort. Once again this adjustment is relatively small being of the order of .02 years in $e_0$ for each 1 per cent change in annual cohort size. It is very unusual for a population to increase by more than 3 per cent per cohort and hence the greatest adjustment made by this method is of the order of .06 years.

Although the analysis above uses the technique of confidence intervals to suggest the order of magnitude of the possible random variation in $e_x$, it would nevertheless be unwise to use these other than as a guide to levels of accuracy as it is clear that the distributions do not approximate to normality and are bimodal hence the standard methods of estimation are not justified. If tests of significance between the mortality of different populations are required these should be based upon one of the methods of adjusted rates discussed in Chapter III.

IV.6 Value of the Life Tables and Difficulties In Their Use.

The expectation of life at birth, $e_0$, is a summary measure of the total mortality experience of a population which is little affected by any pattern of the age and sex pyramid or by the pattern of the birth rate or of the history of migration. It is a summary statistic which is not dependent upon a standard population nor a standard mortality experience and does not depend upon the size of the radix used for the calculation of the life table. It is a summary measure which has real meaning to the layman, the politician and the physician, who consider that they can understand the concept of the number of years that a newborn may be expected to live.
<table>
<thead>
<tr>
<th>Age</th>
<th>Death Rate</th>
<th>Survival</th>
<th>Alive</th>
<th>Years Lived</th>
<th>Total Years</th>
<th>Expected Life t = 10 yrs</th>
<th>Expected Life t = 5 yrs</th>
<th>Increase in bias</th>
</tr>
</thead>
<tbody>
<tr>
<td>x x+t</td>
<td>m_{x+t}</td>
<td>P_{x+t}</td>
<td>\ell_x</td>
<td>L_{x+t}</td>
<td>T_x</td>
<td>e_{x,10}</td>
<td>e_{x,5}</td>
<td></td>
</tr>
<tr>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
<td>(8)</td>
<td>(9)</td>
</tr>
<tr>
<td>0 - 1</td>
<td>.04669</td>
<td>.95510</td>
<td>100,000</td>
<td>96,181</td>
<td>6,380,977</td>
<td>63.81</td>
<td>63.69</td>
<td>.12</td>
</tr>
<tr>
<td>1 - 5</td>
<td>.00289</td>
<td>.98856</td>
<td>95,510</td>
<td>379,854</td>
<td>6,284,796</td>
<td>65.80</td>
<td>65.67</td>
<td>.13</td>
</tr>
<tr>
<td>5 - 15</td>
<td>.00078</td>
<td>.99223</td>
<td>94,417</td>
<td>940,505</td>
<td>5,904,942</td>
<td>62.54</td>
<td>62.41</td>
<td>.13</td>
</tr>
<tr>
<td>15 - 25</td>
<td>.00130</td>
<td>.98708</td>
<td>93,684</td>
<td>930,785</td>
<td>4,964,437</td>
<td>52.99</td>
<td>52.86</td>
<td>.15</td>
</tr>
<tr>
<td>25 - 35</td>
<td>.00193</td>
<td>.98088</td>
<td>92,473</td>
<td>915,890</td>
<td>4,033,652</td>
<td>43.62</td>
<td>43.50</td>
<td>.12</td>
</tr>
<tr>
<td>35 - 45</td>
<td>.00426</td>
<td>.95829</td>
<td>90,705</td>
<td>888,135</td>
<td>3,117,762</td>
<td>34.37</td>
<td>34.25</td>
<td>.12</td>
</tr>
<tr>
<td>45 - 55</td>
<td>.00915</td>
<td>.91250</td>
<td>86,922</td>
<td>831,190</td>
<td>2,229,627</td>
<td>25.65</td>
<td>25.50</td>
<td>.15</td>
</tr>
<tr>
<td>55 - 65</td>
<td>.02533</td>
<td>.77517</td>
<td>79,316</td>
<td>704,000</td>
<td>1,398,437</td>
<td>17.63</td>
<td>17.45</td>
<td>.18</td>
</tr>
<tr>
<td>65 - 75</td>
<td>.06510</td>
<td>.50886</td>
<td>61,484</td>
<td>463,855</td>
<td>694,437</td>
<td>11.29</td>
<td>11.04</td>
<td>.25</td>
</tr>
<tr>
<td>75 - 85</td>
<td>.12478</td>
<td>.26239</td>
<td>31,287</td>
<td>197,480</td>
<td>230,582</td>
<td>7.37</td>
<td>6.98</td>
<td>.39</td>
</tr>
<tr>
<td>85+</td>
<td>.24800</td>
<td>.00000</td>
<td>8,209</td>
<td>33,102</td>
<td>33,102</td>
<td>4.03</td>
<td>4.03</td>
<td>.00</td>
</tr>
</tbody>
</table>

\[ P_{x+t} = \frac{2 - t}{2 + t} \cdot m_{x+t} \]
It provides public health workers, demographers, medical scientists with a tool for making unbiased comparisons between different countries, between sexes within a country, between regional communities, and between economic and occupational groups. It allows valid comparisons to be made in assessing trends of mortality over time.

The ratio \( \frac{L_6}{L_2} \) is a convenient measure to compare the survival rate between two age segments in different populations: for example one might want to know if Swedish women who survive to age 20, have as good a chance of surviving to age 45 as do Italian women aged 20, by comparing the ratio \( \frac{L_{45}}{L_{20}} \).

The expectation of life has, however, certain disadvantages; the principal one is that although it is possible to compare two life tables and decide which has the higher survival, it is not possible to obtain a linear relationship between them. For example if all age specific mortality rates are doubled the expectation of life may be reduced by only a small amount and will certainly not be reduced by half. For example, if all the age specific mortality rates for the US 1960 males were reduced by one half the expectation of life at birth would be increased by 11 years and the expectation of life between ages 15 and 65 would increase by 1 year.

It also has the disadvantage of any summary statistic that it may obscure very important differences in certain age groups; for example a certain population may have very high mortality in a specific sex and age group, but this would have little effect upon the expectation of life. In mortality analysis we are looking for small trends in age specific mortality rates, these will never be found by life table methods.

Life table estimates have all the disadvantages of any statistical measures based upon the population census and vital records. Data may be incomplete, data may be biased, ages may be misstated, deaths may be under-reported. All the problems discussed in Chapter II remain valid and are not eliminated by statistical expertise.

For the complete life table there are serious problems in tabulating accurate data for single years of life and frequently these are not available. The computations are tedious and time consuming when computer services are not available.

Due to the high mortality in early life the expectation of life may actually increase as a child survives the high risks of early life and becomes a better risk at the age of 2 or 5 than he was at birth. For example \( e_0 \) for males in Sweden 1964 was 71.6 and in Germany FDR 1964, it was 67.0, a difference of 4.6, but \( e_1 \) in Sweden was 71.8 and in Germany 68.0, a difference of 3.8 years. Thus there is a 19% improvement between the two countries over 1 year of age.

Despite these difficulties the life table technique is a useful method available to summarise mortality data and to make valid comparisons between populations if used with discretion.

A mortality analyst must thoroughly understand the techniques used and the implications and difficulties in their interpretation and use.

**IV.7 Generation or Cohort Life Table**

**IV.7.a A total community**

Although the current life table is a valuable summary of the mortality experienced by a population in a short interval of time, there is no group of persons who will actually experience such a pattern of mortality, for the persons who do not die but survive the current life table mortality will in future experience a differing mortality as they age year by year, which will be the successive current life table mortalities. In the same way the individual survivors in a current life table \( L(x) \) did not experience a probability of dying in earlier years equal to that experienced by the younger members of the current population.

It is, however, possible to construct retrospectively a life table from the mortality experience of a group of persons born in the same year or in a group of adjacent years. In Chapter VI.5 there is a description of the methods by which a series of age specific death rates \( z_x \) or age group death rates \( m_{x.t} \) can be obtained which are all relevant to the experience of one birth group or cohort of people.
These mortality rates can then be used in exactly the same way as is described in Chapter IV.3 or IV.4 to construct a complete life table or an abridged life table for the mortality which was really experienced by an identifiable group of individuals. Table IV.5 shows the difference that existed between the actual abridged life table for the cohort of persons born in England and Wales in 1881-1885, and the current life table based upon the mortality rates of persons dying in the period 1881-1885. It will be seen that at the younger ages there is not a great divergence, because the mortality in childhood did not change rapidly in this period, but the differences at age 50 and over are very considerable.

**TABLE IV.5 HERE**

IV.7.b A closed community

In any biological study of a closed community it is customary to enumerate all the persons in the community at the commencement of the study and to count the survivors at serial points of time. In an epidemiological study the births, or a random sample of them, in one year may be enumerated and then these specific individuals are followed for a period of years. This is the classical case of a follow-up study which will be discussed in Pt. II of this manual, and it is, of course, what happens in the simple case of a pharmacological study of animals where given batches of animals are followed up for toxicological experiments to calculate Lethal Dose levels or Median Survival Times; the technique is even more widely used for example in the durability of engineering parts or of electrical lights.

In such studies the fundamental observations available are the proportions of survivors still alive at the beginning of each interval of time; that is there is a direct observation of the $p_x$ distribution. Then the number of persons dying in the closed community is obtained by the formula:

$$d_x = \ell_x - \ell_{x+1} \quad \ldots \ldots \ldots$$

and

$$q_x = d_x/\ell_x = 1 - \ell_{x+1}/\ell_x \quad \ldots \ldots \ldots$$

where $p_x = \ell_{x+1}/\ell_x$ is the proportion of survivors from age $x$ to age $x+1$.

Then exactly as before in Chapter IV.3 we have the total time lived in the period $x$ to $x+1$ is:

$$L_x = \ell_x - (1 - a) d_x \quad \ldots \ldots \ldots$$

where $a$ is the proportion of time the deaths are assumed to have lived in the period $x$ to $x+1$ and if $a$ is assumed to be 0.5

$$L_x = \frac{1}{2}(\ell_x + \ell_{x+1}) \quad \ldots \ldots \ldots$$

and the total years lived is the sum of $L_x$ in the periods of time after $x$ until the group is extinguished so

$$T_x = \sum L_x = \frac{1}{2} \ell_x + \ell_{x+1} \quad \ldots \ldots \ldots$$

and hence the expectation of life at time $x$ is

$$e_x = T_x/\ell_x = 0.5 + (\ell_{x+1})\ell_x \quad \ldots \ldots \ldots$$

The problems that arise when there are entrants and withdrawals to the closed community will be dealt with in detail in Part II of this Manual. This type of analysis is widely used in the analysis of mortality of patients with cancer who have been treated by differing methods of therapy.

IV.8 Life Tables From Incomplete Data.

IV.8.a Life tables based upon census records only.

In some countries it is possible to organise a complete and accurate census of the population at regular intervals, but the administrative cost in money and manpower of organising a permanent system of vital registration is not possible. If in such cases there were no migration, or if the effect of migration could be estimated independently, then the fundamental information available is the age specific structure of the two census enumerations.
TABLE IV.5

England and Wales

Comparison between Current Life Table and Cohort Life Table

Current Life Table 1881-1886

Cohort Life Table 1881-1886 to 1961

<table>
<thead>
<tr>
<th>AGE</th>
<th>MALES</th>
<th>FEMALES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\ell_x$</td>
<td>$\ell_x$</td>
</tr>
<tr>
<td></td>
<td>Cohort</td>
<td>Current</td>
</tr>
<tr>
<td>0</td>
<td>100,000</td>
<td>100,000</td>
</tr>
<tr>
<td>10</td>
<td>73,505</td>
<td>73,142</td>
</tr>
<tr>
<td>20</td>
<td>71,274</td>
<td>70,381</td>
</tr>
<tr>
<td>30</td>
<td>68,044</td>
<td>65,816</td>
</tr>
<tr>
<td>40</td>
<td>63,635</td>
<td>59,319</td>
</tr>
<tr>
<td>50</td>
<td>58,404</td>
<td>50,467</td>
</tr>
<tr>
<td>60</td>
<td>49,430</td>
<td>38,668</td>
</tr>
<tr>
<td>70</td>
<td>34,750</td>
<td>23,265</td>
</tr>
<tr>
<td>80</td>
<td>14,538</td>
<td>8,153</td>
</tr>
<tr>
<td>$e_o$</td>
<td>47.7 years</td>
<td>43.2 years</td>
</tr>
</tbody>
</table>
of the same population $^{1}p_{x}$ and $^{2}p_{x}$ and the chance of survival of an age group $^{1}p_{x}$ aged $x$ at a census taken in the year $t$, to age $x+n$ at a census $n$ years later is:

$$p_{x}p_{x+1} \ldots p_{x+n} = \frac{2^{p}}{x+n}/p_{x}$$

but

$$q_{x.n} = 1 - p_{x}p_{x+1}p_{x+2} \ldots p_{x+n}$$

and thus

$$q_{x.n} = 1 - \left(\frac{2^{p}}{x+n}/p_{x}\right)$$

From this data it is then theoretically possible to construct a life table for the population as at the time of the first census.

There are, however, serious difficulties. All the possible sources of inaccuracy and bias in census taking which were discussed in Chapters I and II will have an effect on the calculations. Because there is no accurate measure of total births and deaths in the interim period between the two censuses there can be no independent estimate of the closing error and hence no comparative estimate of the relative efficiency of the two census. Migration will always be a serious problem in an area without a vital registration system, for unless the country concerned has completely restricted boundaries there can be no control over migration.

However, in certain countries such as India and Brazil where census methods are well controlled, this approach has been used to calculate life table functions.

IV.8.b Life Tables based upon death records only

Sometimes there is only information concerning deaths and no information concerning the population structure. This was the case that faced John Graunt (1662) when he analysed the Bills of Mortality for the plague years in London and he outlined the concept of a life table although he was unable to construct one because he did not have the ages at death. This was also the case of some of the early actuarial tables such as those of Edmund Halley for Breslau (1687-9) or those of Dr. Price in the Northampton Tables (1725-1780). It was Halley the astronomer who first used the data from burial records for the city of Breslau to calculate a life table function and communicated his results to the Royal Society of England (1693). He was aware that this technique was only valid if the population was a stationary one, and had been stationary for some years; he was able to justify that this was so in Breslau in the late 17th century by a comparative study of baptisms and burials in the city over a period of years. Price, however, was misled in his estimates because there was a proportion of Baptists living in Northampton whose baptisms were not registered at birth, but whose burials took place in the Churchyard, hence he overestimated the mortality and underestimated the expectation of life.

If there is a stationary population, and ages at death are known, this provides the fundamental information concerning $d_{x}$ and hence it can be assumed that

$$\ell_{x} = \frac{t}{x}d_{x}$$

and thus the $\ell_{x}$ column of the life table can be constructed and from this $q_{x}$ estimated by the formula

$$q_{x} = \frac{\ell_{x} - \ell_{x+1}}{\ell_{x}}$$

In modern circumstances this method should rarely be used because the underlying assumption of a stationary population can seldom be justified. It is, however, used by archaeologists or zoologists who find skeletons of men or animals from which the age at death of individuals can be estimated and in which population cannot be known, but is assumed to have been stationary.
CHAPTER V

FOETAL, INFANT AND CHILDHOOD MORTALITY

V.1. Introduction

Infant mortality has for a long time been regarded as a true reflection of a country’s socio-economic and health conditions. The rate of loss in the first year of life has attracted particular attention because -

a) mortality is relatively high, the probability of dying in the first year of life often exceeding the values observed in the following fifty to sixty years of life;

b) it has a considerable impact on the average expectation of life and the rate of population growth;

c) it has a disproportionate share in total mortality, and,

d) it is sensitive to environmental and sanitary conditions.

In developed countries in recent decades infant mortality has experienced dramatic reductions through the combined effect of improvements in standards of living, better environmental and personal hygiene, the extension of maternal and child welfare schemes, improved obstetric care, as well as prevention and treatment of diseases in infancy. Yet even in these countries infant mortality may still be regarded as a useful indicator of health in so far as a higher rate in certain areas or specific groups may indicate that there is "something wrong".

This decline in infant mortality has been mainly due to large-scale reductions in mortality after the first week of life at which age the infant is more sensitive to environmental hazards and responds to the provision of health and preventive measures. This change has now led to increasing concern for mortality in the early weeks of life, for foetal loss and to the study of perinatal mortality which is the loss of life between the 28th week of gestation and the end of the first week after delivery.

Childhood mortality, because of its sensitivity to environmental conditions, particularly to infections and to nutritional deficiency, may be regarded as an even more useful indicator of social health hazards than is the infant mortality rate, but frequently it is not
available in sufficient detail for adequate analysis. In many cultures and countries, childhood mortality still represents a serious problem and for substantive as well as analytical reasons all loss in the first five years of life has to be studied as an entity.

V.2. Definitions.

V.2.a. Live birth or foetal death.

W.H.O. has promulgated international standard definitions of "live birth" and "foetal death". They are cited in Chapter I.2.b and I.2.c. The definitions are mutually exclusive and should be strictly observed by any registration system so that comparisons can be validly and consistently made.

There are two major issues involved in the definitions; namely, (a) signs of life, and (b) viability. In the definitions of live birth "any sign of life" is accepted but the terminology employed is subject to differing subjective interpretations. Viability is deliberately excluded from the definition due to difficulties of establishing unequivocal criteria and because of the impact of medical progress in this field. On the other hand, this implies that a patently non-viable foetus must be registered as a live birth and subsequently as an infant death when there is a "sign of life" after such procedures as therapeutic termination of pregnancy in the early weeks of gestation or a spontaneous abortion due to maternal conditions before 28 weeks. Any conception which shows any sign of life when separated from its mother must be included in the live born irrespective of duration of gestation.

The definitions of foetal death, also do not specify any particular period of gestation that is required to qualify for a foetal death and hence in theory at least all terminations not resulting in a live birth should be registered as a foetal death, but this is never the practice. Hence, it is important that an estimate be made of the age of gestation for both live births and foetal deaths.

Foetal deaths should be subdivided as far as possible into the following age groups:-

<table>
<thead>
<tr>
<th>Group</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Less than 20 completed weeks of gestation (Early foetal deaths)</td>
</tr>
<tr>
<td>II</td>
<td>20 completed weeks of gestation but less than 28 (Mid foetal deaths)</td>
</tr>
<tr>
<td>III</td>
<td>28 completed weeks of gestation and over (Late foetal death or stillbirth)</td>
</tr>
<tr>
<td>IV</td>
<td>Gestation period not classifiable in Groups I, II &amp; III</td>
</tr>
</tbody>
</table>

There still exists considerable variation in national definitions, some countries for instance have not accepted the WHO recommendations because of the implications such changes might have on their legal definitions. The foetal deaths before 28 weeks are often referred to as miscarriages or spontaneous abortions.

Differences in definitions may pose serious difficulties for a comparative analysis of foetal and infant mortality data. Occasional studies have been carried out to assess the importance of different definitions for comparability. An investigation was carried out in the mid-fifties in Sweden, where it is the practice to register non-viable deliveries as foetal deaths whether they show signs of life or not. The results are presented in Table V.1.

<table>
<thead>
<tr>
<th>Swedish practice</th>
<th>WHO definition</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>Rate</td>
<td>Number</td>
</tr>
<tr>
<td>Live births</td>
<td>107,960</td>
<td>...</td>
</tr>
<tr>
<td>Perinatal deaths</td>
<td>3,106</td>
<td>28.3</td>
</tr>
<tr>
<td>Stillbirths</td>
<td>1,836</td>
<td>16.7</td>
</tr>
<tr>
<td>Deaths under 1 week</td>
<td>1,270</td>
<td>11.8</td>
</tr>
<tr>
<td>Deaths under 28 days</td>
<td>1,427</td>
<td>13.2</td>
</tr>
<tr>
<td>Deaths under 1 year</td>
<td>1,871</td>
<td>17.3</td>
</tr>
</tbody>
</table>

+ Rates per 1,000 births (liveborn and stillborn)
++ Rates per 1,000 live births

This table attests to the advantage of combining late foetal deaths (stillbirths) and deaths under one week as "perinatal deaths". The concept of perinatal mortality, first proposed by Peller, is a pragmatic means of overcoming the difficulties stemming from differences in the dividing line between live birth and still birth. This, however, is only one side of the picture. The delimitation of still births or late foetal deaths from early and intermediate foetal deaths is by no means universally accepted and varying practices in drawing this dividing line may seriously hamper the comparability of the data. The Statistical Office of the U.N. commented on this problem after a careful evaluation of the scantly available information as follows: "The variation in the criterion of physical viability as measured by utero-gestational age would introduce sizable differences into statistics of stillbirths and, in fact, render them almost completely non-comparable".  

To demonstrate the effect of excluding live births with gestation of less than 28 weeks on infant mortality, the results of a study carried out in the Netherlands are given in Table V.2.

<table>
<thead>
<tr>
<th>Excluding live births</th>
<th>Including live births</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>with gestations under 28 weeks</td>
</tr>
<tr>
<td>Live births</td>
<td>Number</td>
</tr>
<tr>
<td>Perinatal deaths *</td>
<td>245,739</td>
</tr>
<tr>
<td>Stillbirths *</td>
<td>6,004</td>
</tr>
<tr>
<td>Deaths under 1 week ++</td>
<td>3,645</td>
</tr>
<tr>
<td>Deaths under 1 month ++</td>
<td>2,359</td>
</tr>
<tr>
<td>Deaths under 1 year ++</td>
<td>2,732</td>
</tr>
<tr>
<td></td>
<td>3,783</td>
</tr>
</tbody>
</table>

* Rates per 1,000 births (liveborn and stillborn) 
++ Rates per 1,000 live births

The exclusion of live births of less than 28 weeks gestation would have a marked effect of reducing the first week mortality rate from 11.3 to 9.6, a reduction of 15%.

On the whole, however, the effect of alternative definitions on measures of foetal and infant mortality is not absolutely clear and the results presented above can be regarded only as indicators of the magnitudes involved. The table also shows that postneonatal mortality is hardly, if at all, affected by differing definitions.

V.2.b. Definition of time of infant death.

In the case of infant deaths the age of the child at death is asked for and WHO asks for details to be given where possible:

"i) By single days for the first week of life, (under 24 hours, 1, 2, 3, 4, 5, 6 days) 7 to 13 days, 14 to 20 days, 21 to 27 days, 28 days up to but not including 2 months by single months of life from 2 months to 1 year (2, 3, 4, 5, 6, 7, 8, 9, 10, 11 months); or, if this is unpracticable,

ii) Under 24 hours, 1 to 6 days, 7 to 27 days, 28 days up to but not including 3 months, 3 to 5 months, 6 months but under 1 year; or, if this is unpracticable,

iii) Under 7 days, 7 to 27 days, 28 days but under 1 year."

Here again, there are difficulties of boundary conditions, the day of birth is day 0 and ends after 24 hours. When does day 1 end? At 48 hours or at midnight on the second day? Are infants 29 days old always considered to be 1 month old?

Fortunately, these precise distinctions are not necessary for most purposes of comparisons although they were used in Table IV.5. to calculate the duration of life lived in

the first year of life by infants who die in that year. The important boundaries are 7 days and 28 days which define the ends of the perinatal and neonatal periods and here it is clear that the seventh or the twenty-eighth days are the days after the completion of 1 week or of 4 weeks of life.

V.2.c. Definition of childhood mortality.

There is no universally accepted definition of this term. Generally, however, "early childhood mortality" refers to deaths at the ages one to under five, but the term "childhood mortality" is widely used to cover this age group and will be used in this sense in the manual.

In less developed countries the number of deaths at age 1 - 4 sometimes equal or even exceed the number of infant deaths and thus this age group is one of the most important to be studied. Whenever data permit, early childhood mortality should be studied by individual years of age. Such an analysis is important because it is the second year of life which generally dominates the picture of early childhood age. After the second year of life, death rates decline, although with varying promptness. Mortality in the second year of life may be considered as a better index of general social and environmental health than infant mortality as it is less diluted with extraneous genetic or obstetric factors.

V.3. Foetal Mortality.

Efforts to register foetal deaths of gestational age below 28 weeks have so far not proved successful. In fact, inadequate information on the magnitude of foetal loss at less than 28 weeks of gestation constitutes a major gap in knowledge of pregnancy outcome.

Occasional special studies have been carried out with a view to throwing some light on the magnitude of early and intermediate foetal loss such as the study by Shapiro, Levine and Abramovics* based upon a population of 12,000 women covered by the Health Insurance Plan of Greater New York. Life table techniques were used in their study. They arrived at the following tentative estimates of the probability of early or middle foetal death by gestational age up to the 28th week of gestation:

<table>
<thead>
<tr>
<th>from</th>
<th>to 28 weeks</th>
<th>21%</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>weeks</td>
<td>16%</td>
</tr>
<tr>
<td>6</td>
<td>&quot;</td>
<td>13%</td>
</tr>
<tr>
<td>8</td>
<td>&quot;</td>
<td>9%</td>
</tr>
<tr>
<td>10</td>
<td>&quot;</td>
<td>5%</td>
</tr>
<tr>
<td>12</td>
<td>&quot;</td>
<td>3%</td>
</tr>
<tr>
<td>14</td>
<td>&quot;</td>
<td></td>
</tr>
</tbody>
</table>

Other studies had similar estimates, but there may be a remarkable underestimation for the early weeks and some workers estimate that losses in the first 8 weeks of gestation may exceed 80% of all conceptions.

V.4. Induced Abortion.

Induced abortion constitutes a serious health problem in many countries. The nature and size of the problem differs from country to country, the variations being at least partly determined by differences in legislation. With a growing tendency to liberalise abortion laws, statistical information on the number of legally induced abortions is becoming available to a steadily increasing extent. However, legislation on abortion may not necessarily put an end to illegal recourse to abortion at least for some time after liberalization of abortion, as can be seen from the experience of various countries.**


Various formulae are used to measure legally induced abortions. Ideally, one would like to know the proportion of pregnancies terminating in an abortion. The total number of pregnancies, however, is usually not known, therefore, either the total number of live born or deliveries of "known pregnancies" observed for the specified period are used as a denominator. The most frequently encountered formula is $A/B \cdot \frac{1}{10^k}$

* Factors associated with early and late foetal loss, Excerpta Medica International Congress Series 224, Advances in Planned Parenthood IV, Boston, April 1970.

** See for instance: Survey Techniques in Fertility and Family Planning Research Experience in Hungary, Budapest 1969/2, pp.110 ff.
but clearly it would be better to have a more general measure such as \( \frac{A}{(B+A+F)} \times 10^k \). \[ 2 \]

where \( A = \) number of legally induced abortions
\( B = \) number of live born
\( F = \) number of foetal deaths
\( k = \) constant (usually 3)

The data refer generally to the same period of observation although there is a difference in time between the events.

These measures are computed as age specific rates sometimes on the basis of maternal age at conception, but sometimes on the basis of maternal age at pregnancy termination. The latter method may introduce a bias especially for the youngest age group as shown in Table V.3.

<table>
<thead>
<tr>
<th>Age of mother (completed years)</th>
<th>Legal abortion per 1,000 live born by maternal age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Maternal age at termination</td>
</tr>
<tr>
<td>- 19</td>
<td>153</td>
</tr>
<tr>
<td>20 -</td>
<td>128</td>
</tr>
<tr>
<td>25 -</td>
<td>253</td>
</tr>
<tr>
<td>30 -</td>
<td>525</td>
</tr>
<tr>
<td>35 -</td>
<td>987</td>
</tr>
<tr>
<td>40 and older</td>
<td>1,576</td>
</tr>
</tbody>
</table>

The reason for this phenomenon is that woman may be in the group under 19 if they abort but already in the group 20-24 if they delivered at term. This measure may also give a distorted picture if there is a marked change in the annual number of pregnancies.

Another way of analysing the data is to measure the incidence rate of abortion among the general population, or preferably among the female population of reproductive age (05-44 years), calculating for instance:

\[ \left( \frac{A}{P_{FR}} \right) \times 10^k \]

where \( A = \) number of abortions
\( P_{FR} = \) average number of females of reproductive age
\( k = \) constant (usually 3).

These measures of abortion are supplements to the information available on crude birth rates or general fertility rates where the number of live born may be related to the total population or to the female population of reproductive age. The definitions for these two types of measurement varies and it is therefore essential to state unequivocally the "reference population" used in the denominator. The two types of measurement are based on a different emphasis in approach.

For detailed studies of "abortion behaviour" an abortion register can provide useful and detailed information, and would permit linkage of successive abortions and the establishment of the probability of abortion during the fertile lifetime of a woman. However, setting up such a register involves ethical considerations and may be feasible only when and where abortion is accepted by society and legislation has been undertaken.

V.4.b. Illegally induced abortion.

Little information is naturally available on illegally induced abortion. Estimates made in various countries attest more to the courage and philosophy of the researcher than it contributes to assessment of the problem. The recently developed randomised response technique may offer a means of employing survey technique may offer a means of employing survey techniques for estimating illegal abortions.

V.5. Infant Mortality.

Despite the general awareness of the importance of knowledge on infant mortality for demographic and public health purposes, it has been estimated that only 2% of the world's infant deaths are recorded completely and accurately. When even elementary information on the magnitude of the infant mortality problem is either missing or grossly defective in many countries, it is small wonder that more sophisticated information such as causes of death are available for only a minority of the world's infant deaths. It is here that sampling methods and ad hoc surveys can be very useful in indicating the scale of some of the mortality rates and the pattern of the causes. Thus allowing preventive measures to be planned and evaluated.

V.5.a. Measures of infant mortality.

Infancy is defined as the period of life before the first anniversary of birth. The conventional annual infant mortality rate is calculated by relating the number of infant deaths to the number of live born in the same period, i.e.

\[
\text{IMR} = \frac{D}{B} \times 10^k \quad \text{(k usually = 3)}
\]

where \(D\) is the number of deaths between birth and before reaching the exact age of one year which occur in a district during a certain calendar year, and \(B\) is the number of live born during the same calendar year and in the same district.

The infant mortality is not a "rate in the conventional sense as mentioned in Chapter III, Section 2.c, nor is it a proportion nor an estimate of the probability of dying during the first year of life. It is a ratio of its own character, but it is conventionally used and is successful in giving a simple measure for infant mortality by the use of two figures which are commonly available and easily understood.

When the number of births does not change markedly from one year to the next, the conventional "rate" described above gives a sufficiently reliable measure of mortality in infancy. However, when there is a significant downward trend in the number of live-born, the infant mortality rate computed in this way may be biased upwards because many of the infant deaths will have been born in the previous year; similarly a rising birth rate will apparently reduce the infant mortality. Several methods to allow for this weakness are known in the literature. The best procedure is, of course, to cross-classify age and year (or month) of birth, thus permitting the calculation of refined rates of infant loss.

For more refined analyses there are two approaches possible:

1. to estimate the true proportion of deceased children under 1 year of age among an appropriate group of live-born; and

2. to estimate the true rate for infant mortality in accordance with the definition of a rate.

Both of the principles would be justified and would serve different purposes. But in infant mortality statistics, the first method is applied and leads to estimates of the probabilities of dying. These are used for life table calculations where the cohort principle is indispensable for the first year of life. The probability approach dominates in all methodological studies on infant mortality. The deaths in a certain calendar year, occurring at a given age span, belong to two different birth cohorts. Many of the infant deaths registered in a year, occur among infants born in the preceding calendar year and similarly, among the live-born of a specified calendar year, some will die in the same year and some in the following year before completion of the first year of life. Figure V.1. demonstrates by a numerical example the principles of the different calculations.

**Conventional Calculation.** The year of birth is shown as the vertical axis and the year in which an infant death occurs is the horizontal axis. Thus the diagram shows that in the year 1970 there were 100,000 live births and 3,000 of these died before the end of that year, but a further 500 of them died before the age of 1 year, but after the end of calendar year 1970. Similarly 120,000 were born in 1971 of whom 2,800 were infant deaths in 1971 and 400 in 1972, and 130,000 were born in 1972 of whom 2,700 died in the same year and 400 in the following year.

Fig. V.1.

Hypothetical Analysis of Deaths in the first year of life assigned either to the year of birth or to the year of death.
The conventional method of calculating the infant mortality rate for the year 1971 is to take the infant deaths recorded in that year, 500 from the 1970 cohort and 2,800 from the 1971 cohort and then divide the total 3,300 infant deaths by the births in 1971 which were 120,000 to give an infant mortality rate of 27.5 per 1,000 live births. Similarly in 1972 the infant mortality rate would be \((400 + 2,700)/(130,000) = 23.8\) per 1,000.

Note that the rate for 1970 cannot be calculated from the data because the deaths during 1970 from the 1969 cohort of births are not available.

**FIGURE V.1. HERE**

Probability of dying within the first year of life, \(p_o\).

An alternative approach is to consider the deaths that actually occur in each cohort of births and then calculate the true proportion of deaths or the probability of dying in the first year of life.

There were 100,000 live born in 1970, 120,000 in 1971, 130,000 in 1972; out of the number of 100,000 of the birth cohort 1970, there were 3,000 infant deaths in the same year and 500 in the following year (1971). The probability of dying within the first year of life is therefore obtained by dividing \((3,000 + 500)/100,000 = 35.0\) per 1,000. This probability refers to one birth cohort and one year of age but includes part of the mortality risk of two calendar years. For the birth cohort 1971, the probability of dying within the first year is \((2,800 + 400)/120,000 = 26.7\) per 1,000. Note that the probability for the birth cohort 1972 cannot be calculated until the 1973 figures become available.

This kind of calculation is not entirely satisfactory because it relates to two years of observation and does not correspond to the general principles of annual statistical recording. Therefore, modifications are introduced to calculate probabilities of dying in the first year of life based only upon the year of observation - say 1971 in the example.

The method of Boeckh subdivides the probability of surviving the first year of life into two subsequent parts which are to be estimated within one calendar year.

i) the probability of surviving the calendar year of birth;

ii) the probability of surviving the next calendar year until the first birthday.

The product of these is the probability of surviving the first year of life.

In the example above the probability of surviving the calendar year of birth in 1971 is

\[
\frac{120,000 - 2,800}{120,000} = .9767
\]

The probability of surviving up to age 1 in 1971 which is the calendar year subsequent to birth in 1970 is \(97,000 - 500 = .9948\) as 97,000 infants born in 1970 are known to have been alive at the beginning of 1971. The product of these then gives an estimate of the probability in 1971 of surviving the first year of life \(p_o = .9717\). The probability of dying in the first year of life for 1971 is therefore \(q_o = 1 - p_o = 0.0284\), or 28.4 per 1,000.

Similarly for 1972 the probability of survival is .9759 and the probability of dying .0241 or 24.1 per 1,000.

A more extensive discussion of the various methods used to adjust infant mortality rates was written by W.P.D. Logan, Population Bulletin, No. 3, 1953, UN New York.

For Austria, where the birth rate was rising between 1956 and 1965 the values computed by the conventional and the Boeckh methods are compared below:
If the partial probability of surviving are near to 1 it is possible to work with an approximate formula:

\[
q_x = \frac{D_x^x}{B_x} + \frac{D_x^{x-1}}{B_x^{x-1}}
\]

where

\begin{align*}
B_x &= \text{live born in year } x \\
B_{x-1} &= \text{live born in year } x-1 \\
D_x^x &= \text{infant deaths of cohort } B_x \text{ who died in year } x \\
D_x^{x-1} &= \text{infant deaths of cohort } B_{x-1} \text{ who died in year } x
\end{align*}

In the example, we get for 1971

\[
q_x = \frac{2,800}{120,000} + \frac{500}{100,000} = 0.0233 + 0.0005 = 0.0238 = 28.3 \text{ per thousand}
\]

in accordance with the exact calculation. For these calculations, the infant deaths for each year are split in two parts according to the birth year.

Formula (5) may be used approximately even if the subdivision of the deaths according to the year of birth is not available. Then it is sufficient to use the ratio:

\[
f = \frac{D_x^x}{D_x} = \frac{D_x}{D_x^x + D_x^{x-1}}
\]

which may be estimated from former years or from a sample. The formula for the infant mortality rate by "numerator separation" is:

\[
\text{IMR} = \frac{f \cdot D_x^x}{B_x} + \frac{(1-f)D_x^{x-1}}{B_x^{x-1}}
\]

More frequently used is the "denominator separation factor" (r) leading to the formula:

\[
\text{IMR} = \frac{D_x}{rB_x + (1-r)B_{x-1}}
\]

r and f are connected by the formula:

\[
r = \frac{fB_x^{x-1}}{fB_{x-1} + (1-f)B_x}
\]

f and r are identical if the births in successive years are constant (\(B_x = B_{x-1}\)).
f and r depend upon the relation between early and late infant mortality, and this in turn depends to some extent upon the absolute level of total infant mortality. The higher the infant mortality is, in general, the higher the later part of it is, and the lower the factors r and f. An estimate of r is:

<table>
<thead>
<tr>
<th>Infant Mortality (per thousand)</th>
<th>r</th>
<th>Infant Mortality (per thousand)</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.84</td>
<td>80</td>
<td>0.67</td>
</tr>
<tr>
<td>30</td>
<td>0.81</td>
<td>100*</td>
<td>0.63</td>
</tr>
<tr>
<td>40</td>
<td>0.77</td>
<td>150*</td>
<td>0.57</td>
</tr>
<tr>
<td>50</td>
<td>0.74</td>
<td>200</td>
<td>0.54</td>
</tr>
<tr>
<td>60</td>
<td>0.71</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(*In countries with high infant mortality, these estimates could be wrong because of a possible "abnormal age distribution of infant" deaths).

The use of separation factor methods should be confined to situations where there are small but appreciable changes in the birth rates. If birth rates are nearly constant, the conventional method is sufficient; if birth rates have high seasonal fluctuations, quarterly or monthly calculations are necessary.

The use of separation factors is justified only to calculate infant mortality for "all causes"; it may produce distorted results when applied to specific causes, particularly those occurring soon after birth (for instance, birth injury). For causes of death, the conventional method of the number of live-born in the year of observation as denominator is on the whole the best practical procedure.

V.5.b. Quarterly and monthly infant mortality

Infant mortality has in many countries large seasonal variations which are related to different morbidity and mortality risks prevailing often in different ages and stages of development. Therefore, the seasonal analysis is of great importance for the analysis of infant mortality.

For seasonal analysis, a tabulation of the number of births by month or by quarter is necessary and a cross classification of deaths by month or quarter of birth is desirable. For approximate methods, the monthly (quarterly) distribution of deaths is sufficient.

If monthly infant deaths are tabulated by month of birth, the following formula is used:

\[
\text{IMR}_m = \frac{D_m}{\sum a_i B_i} \times 365 \times 1,000
\]

where:

- \(D_m\) = infant deaths in a given calendar month
- \(a_i\) = proportion of infant deaths in a given calendar month who were born in the same month to total number of infant deaths observed in this calendar month.
- \(B_i\) = proportion of infant deaths in a given calendar month who were born in preceding month to total number of infant deaths observed in the given calendar month.
- \(t\) = number of days in calendar month for which computation is carried out (thus allowing for unequal length of calendar months).
The weighting of the number of births in the preceding months is necessary in order to overcome the difficulties possibly resulting from fluctuating birth rates during the former months. Methodologically, this approach does not fit the definition of a proportion (probability) nor that of a rate. The denominator is a weighted average of a monthly number of births and corresponds to the conventional infant mortality rate and to formula (4).

If the monthly deaths are not cross-classified by the month of birth it is possible to use "average weights" for the "a's". These weights vary considerably according to the level of infant mortality. Table V.5 gives weights for levels of about 70 and 20 per 1,000 live-born:

<table>
<thead>
<tr>
<th>Quarter</th>
<th>Weight</th>
<th>Month</th>
<th>Coefficient</th>
<th>Month</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.387</td>
<td>0</td>
<td>0.247</td>
<td>0</td>
<td>0.475</td>
</tr>
<tr>
<td>1</td>
<td>0.266</td>
<td>1</td>
<td>0.162</td>
<td>1</td>
<td>0.164</td>
</tr>
<tr>
<td>2</td>
<td>0.167</td>
<td>2</td>
<td>0.097</td>
<td>2</td>
<td>0.069</td>
</tr>
<tr>
<td>3</td>
<td>0.126</td>
<td>3</td>
<td>0.080</td>
<td>3</td>
<td>0.055</td>
</tr>
<tr>
<td>4</td>
<td>0.054</td>
<td>4</td>
<td>0.069</td>
<td>4</td>
<td>0.045</td>
</tr>
</tbody>
</table>

**IMR = 20^0/oo**
<table>
<thead>
<tr>
<th>Quarter</th>
<th>Weight</th>
<th>Month</th>
<th>Coefficient</th>
<th>Month</th>
<th>Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.607</td>
<td>0</td>
<td>0.045</td>
<td>0</td>
<td>0.025</td>
</tr>
<tr>
<td>1</td>
<td>0.199</td>
<td>9</td>
<td>0.042</td>
<td>9</td>
<td>0.022</td>
</tr>
<tr>
<td>2</td>
<td>0.102</td>
<td>10</td>
<td>0.039</td>
<td>10</td>
<td>0.019</td>
</tr>
<tr>
<td>3</td>
<td>0.067</td>
<td>11</td>
<td>0.036</td>
<td>11</td>
<td>0.017</td>
</tr>
<tr>
<td>4</td>
<td>0.025</td>
<td>12</td>
<td>0.017</td>
<td>12</td>
<td>0.008</td>
</tr>
</tbody>
</table>

**Source**: Les coefficients attachés à un niveau voisin de 70^0/oo on été calculés par P. Vincent (cf. J. Bourgeois, de la mesure de la mortalité infantile, *Population*, no.1, 1946); les autres systèmes de coefficients résultent de calculs personnels.

The method corresponds to the "denominator separation factor r" (formula 8).

**V.5.c Infant Mortality by Age at death**

A subdivision of mortality in the first year of life by shorter age periods is an essential analytical tool for a better understanding of the forces at work and to provide important clues for public health action. This is due to the fact that the underlying cause pattern varies with age and, consequently, the importance to be attached to specific preventive and curative measures is, to a substantial extent, determined by the age pattern of infant mortality. The traditional "minimum" distinction is between "neonatal" and "postnatal" mortality. Neonatal deaths are generally defined as "deaths under 28 days of life" (sometimes also "under one month") whereas postneonatal deaths covers the rest of the first year, i.e. "deaths from 28 days to under one year". The basic reason for this distinction is that postneonatal deaths are particularly sensitive to environmental influences and exogenously determined. In neonatal mortality, however, less tractable endogeneous factors play an important role. Postneonatal mortality is, therefore, more responsive to improvements in the socio-economic environment, and on the basis of present day knowledge and technology easier to bring under control than neonatal mortality.

The formulae for computing these rates are:

\[
P_n^z \times 10^k \quad \text{and} \quad P_p^z \times 10^k
\]

where

\[
P_n^z = \text{neonatal deaths in period } z
\]

\[
P_p^z = \text{postneonatal deaths in period } z
\]

\[
B^z = \text{number of liveborn in same period}
\]

\[
k = \text{constant (usually 3)}
\]
The summing up of the neonatal and the postneonatal mortality rate gives the infant mortality rate. However, convenient as this may be, the denominator for postneonatal mortality is inflated as it includes also those liveborn who died already in the neonatal period. Let us assume that there are 100,000 liveborn, 5,000 neonatal and 15,000 postneonatal deaths. Then the conventional postneonatal mortality rate equals 150 per thousand, whereas excluding the 5,000 deaths in the neonatal rate increases the death rate to 158 per thousand.

With the drastic reductions in postneonatal mortality in low mortality countries, increasing attention has focussed on the first four weeks of life. A distinction between "early neonatal" (i.e. deaths under one week) and "late neonatal" deaths (i.e. deaths from 7 to under 28 days of life) has been found useful. The nearer to the date of birth the less, on the whole, the progress in the struggle to reduce infant loss. For this reason a subdivision by individual days, may provide important information. However, care has to be taken to ascertain whether the tabulation of the data based on the difference between day of birth and day of death or on an accurate computation of the time lived.

If possible, infant mortality should be analyzed on the basis of a cross-tabulation of infant deaths by month of birth, age in completed time and month of death. Infant mortality by age can be analyzed by applying life table techniques; it allows also an investigation into mortality patterns by month of birth, a subject found to be of considerable interest in several studies.

V.5.d Infant mortality by gestational age and birth weight

Premature born children have the highest risk of dying within the first year of life. The medical and social care for the premature infant in countries with low infant mortality is the main problem in the struggle to improve further neonatal mortality.

Prematurity may be measured by gestational age, commonly counted as weeks or months after the last menstrual period ("Lmp"). But in many countries, reliable data about the true gestational age cannot be expected, partially due to irregularities of the menstrual period and persistent bleeding, or sometimes due to psychological inhibitions or due to inaccurate recording.

Therefore, it is more convenient to use length and weight at birth to distinguish between premature and mature newborn. In the ICD, death on prematurity is defined by a weight of less than 2500 g. But it is better to subdivide the weight at birth into more than two classes, for instance by classes of 500g or 250 g.

It is necessary to have the frequency distribution of the birth weights of all newborn and for the deceased children in the first year of life. To get this, a special question in the questionnaire completed at birth is necessary.

To calculate the infant mortality for all weight groups needs no special methods. The conventional formula is sufficient to get "weight-specific" infant mortality rates. They may be refined by subdivision according to age (including late foetal deaths).

Table V.6 shows the large differences by birth weight for the German Democratic Republic, 1971;

<table>
<thead>
<tr>
<th>Birth weight</th>
<th>Birth weight</th>
<th>All birth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Birth weight</td>
<td>&gt; 2500 g.</td>
<td>&gt; 2500 g.</td>
</tr>
<tr>
<td>Birth weight</td>
<td>&lt; 2500 g.</td>
<td>&gt; 2500 g.</td>
</tr>
<tr>
<td>Infant mortality rate per 1,000</td>
<td>157.2</td>
<td>9.3</td>
</tr>
<tr>
<td>Proportion among live births</td>
<td>5.9%</td>
<td>94.1%</td>
</tr>
<tr>
<td>Proportion among infant deaths</td>
<td>51.3%</td>
<td>48.7%</td>
</tr>
</tbody>
</table>

V.5.e Age of mother at time of infant death.

In computing infant mortality specific for age of mother, age of mother should refer to the age she attained at birth of the child. Where such studies are carried out on the basis of linkage of information derived from birth and death certificate, this does not pose a problem. However, when age of mother is reported on the death certificate, there is a danger of reporting age of mother at death of infant which may distort the mortality rates at the two extremes of childbearing age.
As a rule, it has therefore to be emphasised that enumerator and denominator should, on principle, correspond.

V.5.f Legitimacy and infant mortality.

Special difficulties may arise in calculating infant mortality rates specific for legitimacy status. Illegitimately born children may be reported as legitimate at time of death when they have been legitimised in the interval between birth and death, thus distorting the relationship between numerator and denominator (illegitimate mortality is in such a case too low, and mortality rates for legitimates, too high). The reliability of reporting may likewise vary especially when illegitimacy is not accepted by society (more cases may then be concealed in reporting of death of an infant than in reporting live birth), also tending to reduce the mortality of illegitimately born infants.

V.5.g Factors affecting accuracy and comparability of data.

For a proper assessment of foetal, infant and childhood mortality data, the following list sites some of the important factors to be examined before drawing inferences:

1. The definition of vital events and the application of these definitions to a legal system within a particular country.
2. Registration practices and completeness of registration.
3. Geographical coverage of registration.
4. Administrative arrangements for collecting and processing of data.
5. Tabulation practices.
6. Methods of measurement.

Despite long sustained efforts by national and international authorities and indisputable progress achieved over past decades, the interpretation of differences within a country, between countries and trends in time still poses many difficulties and pitfalls such as:

a) In some countries infants who were born alive but died before registration are counted as stillbirths, but social security benefits may be greater if the child is recorded as live born. Inspection of data on stillbirths and infant deaths in the first week of life, especially if these are analysed by individual days of life, may provide important clues.

b) In some countries there is no compulsory registration of stillbirths, thus encouraging parents who want to spare themselves trouble and expenses to consider deaths "shortly" after birth as "stillborn".

c) Delays in reporting may cause a discrepancy between the number of occurrences and the number reported within a given period. The Christmas and New Year holidays are such a possible source of error.

d) Geographical assignment of both vital events and the corresponding population should be based on the same principle, i.e. preferably on the "resident population concept". However, the definition and application of the term "resident" may vary from country to country.

e) Age at death may be tabulated on a calendar day basis or on completed time. An illustration of the importance of this factor is obtained from a special investigation in the Netherlands:

<table>
<thead>
<tr>
<th>Year</th>
<th>Deaths occurring within 24 hours of birth</th>
<th>Deaths occurring on calendar day of birth</th>
<th>Ratio (1)/(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1958</td>
<td>1439</td>
<td>1084</td>
<td>1.33</td>
</tr>
<tr>
<td>1959</td>
<td>1529</td>
<td>1088</td>
<td>1.41</td>
</tr>
</tbody>
</table>

In France and Spain live-born babies dead before registration of their birth are not included in the statistics of live-born, nor in the number of deaths. They are counted as a special group of stillborn. Therefore, they are excluded from the statistics of causes of death. The numbers in 1970 are:
<table>
<thead>
<tr>
<th></th>
<th>Absolute</th>
<th>% of</th>
<th>Ratio to 100 deaths</th>
<th>Ratio to 100 deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Stillborn</td>
<td>under 1 week</td>
<td>under 1 month</td>
</tr>
<tr>
<td>France</td>
<td>2,598</td>
<td>17.7</td>
<td>42.6</td>
<td>-</td>
</tr>
<tr>
<td>Spain</td>
<td>4,963</td>
<td>29.5</td>
<td>-</td>
<td>75.4</td>
</tr>
</tbody>
</table>

In the statistics published by WHO, this group is transferred to the live-born and to infant deaths.

In other countries, parents seek to have infants dying shortly after birth registered as stillborn because of financial reasons (expenses for funeral, for instance). However, the reverse may also occur, i.e. parents may for religious, cultural or social benefit reasons prefer the registration of a stillbirth as a live birth and subsequent death.

V.6. Perinatal Mortality.

With the reduction of post neonatal mortality the concept of "perinatal mortality" becomes more important. Perinatal mortality is measured by relating the sum of the number of stillbirths and of deaths under one week to the number of live-born. This is called the "perinatal mortality rate", though "perinatal mortality ratio" is frequently encountered. It has been internationally recommended but it is an illogical measure and should be replaced by the more appropriate formula used in some countries where the denominator is the sum of the number of live and stillborn (see Chapter III). It is, therefore, necessary for an analyst to be alert when comparing data to be certain which denominator has been used in every case.

However, the perinatal mortality is subject to one important limitation, which concerns variations in the defined boundaries between "abortions" and "stillborn" and "liveborn". It is usual to confine perinatal mortality to those late foetal deaths which occur after 28 weeks of gestation but all live born irrespective of the gestational period are included. Reporting rules and procedures for foetal deaths under 28 weeks do vary considerably between countries. For example, if a foetus shows "any sign of life" and dies in the first hour of life, it may be registered as liveborn and infant death in one country or it may be discarded as an abortion in the other, especially if the foetus does not meet certain specified criteria of viability (e.g. minimum length or minimum weight at birth, which in some countries are included in the definition of live-born). Maternal benefits may play a role in reporting procedures when benefits vary for abortions, stillbirths and livebirths. This was discussed in detail above.

A further problem concerns the interpretation of the perinatal mortality rate. It has been usual to regard the perinatal mortality rate as an index of the quality of obstetric care. However, in recent years, increasing criticism has been raised against this interpretation. One serious point raised is the fact that progress in obstetric care may lead to a postponement of deaths; for instance, a foetus which might have been prematurely delivered at 24 weeks of gestation would be discounted as an abortion; improvement of obstetric care may then produce a survival into the stillbirth period at over 28 weeks and then increase the perinatal mortality rate. In fact, the improvement in the quality of obstetric care would then result in an apparent increase of perinatal mortality and tend to conceal a real decline in perinatal mortality.

An inspection of the figures for late foetal deaths, deaths on the first day of life and the ratio of early neonatal mortality to late neonatal mortality, as well as the ratio of neonatal to postneonatal mortality, provide important clues for the evaluation of the reliability of the data. An instructive comparison of stillbirth rates and deaths and mortality under 24 hours of life is presented in Table V.8.

<table>
<thead>
<tr>
<th>TABLE V.8. HERE</th>
</tr>
</thead>
</table>

This shows not only the expected variation between countries with low death rates and those with relatively high mortality, but also shows a marked difference in the ratio of stillbirths to deaths on the first day of life which fluctuates between 0.9 in Czechoslovakia to 10.7 in France. Clearly the separation of live births from still births is not satisfactory in international comparisons.
**TABLE V.8.**

Stillbirth Ratio and Death Rate at Under 1 day of Age in 21 Countries in 1971. Both per 1,000 Live Births.

<table>
<thead>
<tr>
<th>Country</th>
<th>Stillbirth ratio (1)</th>
<th>Death Rate Under One Day (2)</th>
<th>Ratio (1)/(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mauritius</td>
<td>40.4</td>
<td>7.0</td>
<td>5.7</td>
</tr>
<tr>
<td>Mexico</td>
<td>20.2</td>
<td>3.2</td>
<td>6.3</td>
</tr>
<tr>
<td>Japan</td>
<td>14.1</td>
<td>2.1</td>
<td>6.6</td>
</tr>
<tr>
<td>Israel</td>
<td>10.7</td>
<td>5.6</td>
<td>1.9</td>
</tr>
<tr>
<td>Thailand 1)</td>
<td>1.3</td>
<td>0.5</td>
<td>2.7</td>
</tr>
<tr>
<td>Belgium</td>
<td>10.8</td>
<td>6.9</td>
<td>1.6</td>
</tr>
<tr>
<td>Bulgaria</td>
<td>10.2</td>
<td>2.3</td>
<td>4.4</td>
</tr>
<tr>
<td>Czechoslovakia</td>
<td>7.4</td>
<td>8.1</td>
<td>0.9</td>
</tr>
<tr>
<td>Denmark</td>
<td>8.2</td>
<td>4.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Finland</td>
<td>8.1</td>
<td>4.1</td>
<td>2.0</td>
</tr>
<tr>
<td>France 2)</td>
<td>13.0</td>
<td>1.2</td>
<td>10.7</td>
</tr>
<tr>
<td>Federal Republic of Germany</td>
<td>9.9</td>
<td>9.8</td>
<td>1.0</td>
</tr>
<tr>
<td>Italy</td>
<td>14.6</td>
<td>7.6</td>
<td>1.9</td>
</tr>
<tr>
<td>Netherlands</td>
<td>10.3</td>
<td>3.2</td>
<td>3.2</td>
</tr>
<tr>
<td>Poland</td>
<td>9.6</td>
<td>6.1</td>
<td>1.6</td>
</tr>
<tr>
<td>Portugal</td>
<td>22.2</td>
<td>7.5</td>
<td>2.9</td>
</tr>
<tr>
<td>Romania</td>
<td>12.1</td>
<td>1.8</td>
<td>6.8</td>
</tr>
<tr>
<td>Sweden</td>
<td>7.9</td>
<td>3.6</td>
<td>2.2</td>
</tr>
<tr>
<td>Switzerland</td>
<td>8.6</td>
<td>5.3</td>
<td>1.6</td>
</tr>
<tr>
<td>England and Wales</td>
<td>12.6</td>
<td>5.9</td>
<td>2.1</td>
</tr>
<tr>
<td>Australia</td>
<td>9.1</td>
<td>6.9</td>
<td>1.3</td>
</tr>
</tbody>
</table>

1) As an example for incomplete registration.
2) Deaths before registration of the birth are not recorded.

**V.7. Early Childhood Mortality.**

The WHO has recommended the tabulation of deaths by individual years of age (1, 2, 3 and 4). However, often the lack of information on the population at risk makes it necessary to combine these data in one group.

In less developed countries (but to a certain extent also in developed countries) deaths in the 2nd year of life are numerically preponderant. In rural Punjab (1957 - 59) more than 70% of early deaths occurred in the 2nd year of life. In Rumania the proportion was more than 50% (1963) and even in Sweden it was 35% (1963). The following table compares the risk of dying by single years of age:

**TABLE V.9.**

Probability of dying in the first five years of life \(10,000q_x\) (males).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>407</td>
<td>245</td>
<td>442</td>
<td>353</td>
</tr>
<tr>
<td>1</td>
<td>24</td>
<td>20</td>
<td>98</td>
<td>23</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>15</td>
<td>45</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>12</td>
<td>27</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>9</td>
<td>20</td>
<td>9</td>
</tr>
</tbody>
</table>
As Hammoud* showed, inaccurate age statements may seriously affect the accuracy of the death rates for these ages. He found a tendency to round the age at death to months and years, together with some preference for even over odd numbers. The misreporting of age of infant at death resulted in an underestimation of the infant mortality rate in the study area by more than 5%, the impact being concentrated on postneonatal mortality which was underestimated by almost 10%. Mortality at age one and two were overestimated in compensation. His findings are the more interesting as in the UAR date of birth and date of death are recorded, the misclassification being not only to faulty reporting but also, to a substantial extent, to negligent clerical handling in the processing stage.


A survey of current levels of foetal, infant and childhood mortality is seriously hampered by lack of available data, and the unreliability of much of the available information. It is, therefore, necessary to restrict the following to selected major points of interest.

V.8.a Foetal mortality.

There is no information on a national scale available on early and intermediate foetal mortality. Data on late foetal deaths are published in the UN Demographic Yearbook and the World Health Statistics Annual of the WHO (Vol. I). There are wide differences in the reported frequency of late foetal deaths, for instance in Europe the presently recorded late foetal death ratio (per 1,000 liveborn) varies between 8 and 32, certainly not a reflection of the true margin of variation but more of reporting rules and procedures.

Although foetal mortality has been relatively resistant to efforts for improvement, nevertheless some progress has been recorded in the majority of countries with low mortality. In Austria, for instance, the late foetal mortality ratio declined from 33 (1960) to 11 (1969) per 1,000 liveborn, i.e. a reduction by 69%.

V.8.b Infant mortality.

The situation is better when we consider data concerning infant mortality both regarding the quantitative and the qualitative aspect of availability of data. Estimates of infant mortality cover 2767 countries with a population of 3892, i.e. 98% of the world population. The distribution is as follows:

<table>
<thead>
<tr>
<th>Level of Infant Mortality (per 1,000 liveborn)</th>
<th>Country Population (millions)</th>
<th>Number</th>
<th>Absolute Population</th>
<th>% of world population</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 20</td>
<td>24</td>
<td>589</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>20 -</td>
<td>51</td>
<td>544</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>50 -</td>
<td>33</td>
<td>1273</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>100 -</td>
<td>29</td>
<td>1144</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>150 and more</td>
<td>24</td>
<td>343</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

It has to be borne in mind that countries for which no estimates are provided are probably biased towards high infant mortality levels (except for certain non-member countries of the UN for which, for various reasons, no information is available). With reservations, one may take the absence of data on infant mortality as a crude indicator of the level of infant mortality. Infant mortality recorded large scale reductions since the turn of the century due to improved standards of living and sanitary conditions as well as wider availability and better accessibility of health services. As an illustration, Table V.11. gives data on infant mortality in Austria which indicate the pace of this reduction (Table V.11).

Since 1900 neonatal mortality decreased by 80%, postneonatal mortality, however, decreased by 95%. The rates of decrease observed in the past two decades differ by age. Postneonatal mortality showed an annual rate of decrease of 8%, neonatal mortality a decrease of 4% and perinatal mortality a decrease of 3%.
TABLE V.11.
Infant Mortality in Austria 1871 - 1970

<table>
<thead>
<tr>
<th>Years</th>
<th>Per 1,000 Liveborn</th>
<th>Ratio to 1871/75</th>
<th>Ratio to Previous Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>1871/75</td>
<td>287.2</td>
<td>100</td>
<td>-</td>
</tr>
<tr>
<td>1876/80</td>
<td>258.9</td>
<td>90</td>
<td>90</td>
</tr>
<tr>
<td>1881/85</td>
<td>264.1</td>
<td>92</td>
<td>102</td>
</tr>
<tr>
<td>1886/90</td>
<td>256.8</td>
<td>89</td>
<td>97</td>
</tr>
<tr>
<td>1891/95</td>
<td>245.5</td>
<td>85</td>
<td>96</td>
</tr>
<tr>
<td>1896/1900</td>
<td>224.3</td>
<td>78</td>
<td>91</td>
</tr>
<tr>
<td>1901/05</td>
<td>211.5</td>
<td>74</td>
<td>94</td>
</tr>
<tr>
<td>1906/10</td>
<td>197.4</td>
<td>69</td>
<td>93</td>
</tr>
<tr>
<td>1911/15</td>
<td>183.8</td>
<td>64</td>
<td>93</td>
</tr>
<tr>
<td>1914/18</td>
<td>191.2</td>
<td>67</td>
<td>104</td>
</tr>
<tr>
<td>1919/20</td>
<td>156.6</td>
<td>55</td>
<td>82</td>
</tr>
<tr>
<td>1921/25</td>
<td>140.0</td>
<td>49</td>
<td>89</td>
</tr>
<tr>
<td>1926/30</td>
<td>117.2</td>
<td>41</td>
<td>84</td>
</tr>
<tr>
<td>1931/35</td>
<td>99.0</td>
<td>34</td>
<td>84</td>
</tr>
<tr>
<td>1936/38</td>
<td>88.2</td>
<td>31</td>
<td>89</td>
</tr>
<tr>
<td>1939/43</td>
<td>85.9</td>
<td>30</td>
<td>97</td>
</tr>
<tr>
<td>1946/50</td>
<td>75.6</td>
<td>26</td>
<td>88</td>
</tr>
<tr>
<td>1951/55</td>
<td>51.3</td>
<td>18</td>
<td>68</td>
</tr>
<tr>
<td>1956/60</td>
<td>41.0</td>
<td>14</td>
<td>80</td>
</tr>
<tr>
<td>1961/65</td>
<td>30.9</td>
<td>11</td>
<td>75</td>
</tr>
<tr>
<td>1966/70</td>
<td>26.3</td>
<td>9</td>
<td>85</td>
</tr>
</tbody>
</table>

The course of infant mortality by age of infant at death is well illustrated by Table V.12 where the trends for England and Wales between 1906 and 1970 are shown for each age of infant. Over the period first day deaths fell to less than 50% and first week deaths to about a third but with increasing age there were dramatic falls and the mortality at ages over 6 months had fallen by 95% over the period.

In the first week of life the major influences are prematurity, congenital defects, respiratory distress and birth trauma which are less tractable factors, whereas in the second half of the first year of life, infectious diseases, gastroenteritis, and nutritional deficiencies are important causes of death and these are much more susceptible to control by social and medical advances, by improved education of mothers and by rising standards of hygiene.

The effect is even larger than appears in the tables because the post neonatal mortality rate is calculated in the conventional way as a ratio of total births whereas some infants have already died before reaching this age. Thus, if in a population there are 100,000 births 5,000 neonatal deaths and 3,000 post neonatal deaths, the conventional post neonatal rate is 3,000/100 = 30 per thousand, whereas it is preferably calculated as 3,000/95 = 31.6 per thousand. The greater the neonatal rate the greater this effect will be. It is however usually neglected in such studies.
### TABLE V.12

**England and Wales 1906-1970**

**Infant Mortality by age at death.**

**Rates per 1,000 Live Births**

<table>
<thead>
<tr>
<th>Period</th>
<th>Under 1 day</th>
<th>1-6 days</th>
<th>7-27 days</th>
<th>4 weeks -2 months</th>
<th>3 months -5 months</th>
<th>6 months to under 1 year</th>
<th>Infant Mortality</th>
</tr>
</thead>
<tbody>
<tr>
<td>1906-1910</td>
<td>11.5</td>
<td>13.0</td>
<td>15.7</td>
<td>22.8</td>
<td>22.0</td>
<td>32.1</td>
<td>117.1</td>
</tr>
<tr>
<td>1911-1915</td>
<td>11.4</td>
<td>12.7</td>
<td>14.9</td>
<td>20.2</td>
<td>19.6</td>
<td>30.0</td>
<td>108.7</td>
</tr>
<tr>
<td>1916-1920</td>
<td>11.0</td>
<td>12.4</td>
<td>14.7</td>
<td>16.5</td>
<td>14.6</td>
<td>22.8</td>
<td>90.9</td>
</tr>
<tr>
<td>1921-1925</td>
<td>10.4</td>
<td>11.3</td>
<td>11.7</td>
<td>12.8</td>
<td>11.3</td>
<td>17.5</td>
<td>74.9</td>
</tr>
<tr>
<td>1926-1930</td>
<td>10.3</td>
<td>11.5</td>
<td>9.9</td>
<td>10.8</td>
<td>9.5</td>
<td>15.4</td>
<td>67.6</td>
</tr>
<tr>
<td>1931-1935</td>
<td>10.7</td>
<td>11.7</td>
<td>8.0</td>
<td>9.9</td>
<td>8.5</td>
<td>12.1</td>
<td>61.9</td>
</tr>
<tr>
<td>1936-1940</td>
<td>10.4</td>
<td>11.2</td>
<td>7.7</td>
<td>8.8</td>
<td>7.8</td>
<td>9.4</td>
<td>55.3</td>
</tr>
<tr>
<td>1941-1945</td>
<td>9.3</td>
<td>9.5</td>
<td>7.2</td>
<td>8.9</td>
<td>7.7</td>
<td>7.2</td>
<td>49.8</td>
</tr>
<tr>
<td>1946-1950</td>
<td>7.9</td>
<td>8.4</td>
<td>4.9</td>
<td>5.8</td>
<td>5.0</td>
<td>4.4</td>
<td>36.3</td>
</tr>
<tr>
<td>1951-1955</td>
<td>7.5</td>
<td>7.5</td>
<td>3.9</td>
<td>3.4</td>
<td>3.0</td>
<td>2.5</td>
<td>26.9</td>
</tr>
<tr>
<td>1956-1960</td>
<td>7.5</td>
<td>6.3</td>
<td>2.4</td>
<td>2.6</td>
<td>2.1</td>
<td>1.8</td>
<td>22.6</td>
</tr>
<tr>
<td>1961-1965</td>
<td>7.2</td>
<td>5.2</td>
<td>1.9</td>
<td>2.5</td>
<td>2.2</td>
<td>1.7</td>
<td>20.6</td>
</tr>
<tr>
<td>1966-1970</td>
<td>6.3</td>
<td>4.4</td>
<td>1.7</td>
<td>2.5</td>
<td>2.0</td>
<td>1.5</td>
<td>18.4</td>
</tr>
<tr>
<td>Ratio 1966-70/1906-10</td>
<td>0.55</td>
<td>0.34</td>
<td>0.11</td>
<td>0.11</td>
<td>0.11</td>
<td>0.05</td>
<td>0.16</td>
</tr>
</tbody>
</table>
V.8.c Childhood mortality.

The distribution of countries by the level of early childhood mortality was estimated as follows:

<table>
<thead>
<tr>
<th>Level of childhood mortality (per 100,000 population 1 - 4)</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 100</td>
<td>29</td>
</tr>
<tr>
<td>100 -</td>
<td>11</td>
</tr>
<tr>
<td>200 -</td>
<td>16</td>
</tr>
<tr>
<td>500 -</td>
<td>18</td>
</tr>
<tr>
<td>1,000 or higher</td>
<td>23</td>
</tr>
</tbody>
</table>

Here the same remarks hold as for infant mortality concerning the probable level of childhood mortality in countries for which even rough estimates were not possible.

Mortality at this age recorded even more impressive declines in today's low mortality countries than infant mortality. The decrease recorded for Austria since 1900 is more than 95%.

An analysis by single years of age is presented below in Table V.13.

Age (completed year) 10,000qx

<table>
<thead>
<tr>
<th>TABLE V.13</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>1901/05</th>
<th>1959/61</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>Females</td>
</tr>
<tr>
<td>--------</td>
<td>---------</td>
</tr>
<tr>
<td>1</td>
<td>579</td>
</tr>
<tr>
<td>2</td>
<td>279</td>
</tr>
<tr>
<td>3</td>
<td>176</td>
</tr>
<tr>
<td>4</td>
<td>123</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ratio 1959/61 to 1901/05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>.05</td>
</tr>
<tr>
<td>.06</td>
</tr>
<tr>
<td>.07</td>
</tr>
<tr>
<td>.08</td>
</tr>
</tbody>
</table>
MORTALITY ANALYSIS IN TIME

VI.1. Trend Analysis ...

VI.2. Long-term trends by age and sex.
   a. Crude mortality rates ...
   b. Age and sex specific trends ...
   c. Age adjusted rates ...
   d. Concomitant changes in methodology ...
   e. Expectation of life at birth - international comparisons of trend ...

VI.3. Short-term Trends ...
   a. The decade ...
   b. The year ...
   c. The season ...

VI.4. Trends by Cause of Death ...

VI.5. Cohort Analysis ...
   a. The basic concepts ...
   b. The complete cohort table ...
   c. The abridged cohort table ...
   d. Errors specific to cohort analysis ...

VI.1 Trend Analysis.

In all demographic studies the passage of time is one of the most important variables that can be observed because the medical, physical, social and economic environment in which a community exists will change with time and will have an effect upon the mortality of the population which can be observed and recorded. The mortality experienced by a community is one of the basic measures of public health, but it only has real meaning in the context of the environment, the culture and the climate experienced by that unique community. Within short periods of time these social and environmental factors may be considered constant and hence a serial analysis of changes of mortality rates within a population over a short period of time, such as a decade, is a valid measure of changes in public health. Over longer periods if there are accurately recorded changes in the environment or in the social or medical care of a population then a study of mortality rates over time will indicate whether the recorded changes have been associated with changes in mortality and, if they are so associated, they will indicate what types of people were affected, what diseases have been altered and what was the delay in time between the environmental changes and the changes in mortality.

An example of a close relationship in time between an environmental change and mortality is the classic study of the effect of the London fogs in 1952 by A.B. Hill, who was able to demonstrate a rapid rise in mortality in London after only 24 hours of fog and he was able to show how the mortality changed in relation to the fog over a period of two weeks. The fog particularly affected elderly people who already had some respiratory disease. A different type of association is the rise in mortality from lung cancer in males in many countries of the world which was seen to follow the increase in smoking cigarettes after a lapse of 20 years or more and the rise particularly affects males who smoke over 20 cigarettes a day.
VI.2. Long-term Trends by Age and Sex.

VI.2.a Crude mortality rates.

During the last 30 or 40 years almost all countries have experienced considerable falls in mortality rates. In the United States the total mortality rate has fallen from 10.9 per 1,000 persons in the period 1931-1935 to 9.5 in the period 1961-65; in Sweden the rate in 1931-1935 was 12 and in 1961-65 it was 10; in Japan the fall was from 18 in 1931-35 to 8 in 1961-65 and in Chile from about 28 in 1931-35 to 11 in 1963-65. Only in England and Wales has the total mortality remained relatively constant at 12.0 per 1,000 persons in 1931-35 and 11.8 in 1961-65.

It is clear that these changes in rates cannot be accepted uncritically at their absolute values; the structure of the populations themselves will have altered completely during the 30 years between 1931 and 1961. Thirty years is less than half the expectation of life for people living in many communities, but nevertheless it is a long time in the social, medical and demographic changes which influence society.

It is, therefore, necessary to examine methods of adjustment of mortality so that a true comparison of trends may be obtained. There are several ways in which this can be done.

VI.2.b Age and sex specific trends.

To ensure complete uniformity the mortality rate for a single sex and age group in one country might be examined in detail over a long period of time. Table VI.1 shows the mortality rates for males at each age for the period 1931-1970 in England and Wales. The table demonstrates that there has been a fall in the infant mortality rate over 40 years from 70 per 1,000 to 21 per 1,000, a fall of 70%; for boys age 1-4 the fall has been even more dramatic from 6.9 to 0.86, a fall of 88%, but over age 55 the falls have been smaller at about 10%. But the crude mortality at all ages falls completely to show the underlying trends as this has only fallen by about 3% from 12.7 to 12.3 in the 40 years. What has happened is that as the age specific mortality rates have fallen the proportion of the population in the higher age groups has increased, thus tending to increase the crude mortality rates and that in this population, quite by chance, the two opposite trends have held the crude mortality rates in equilibrium. It is thus always undesirable to use the crude mortality rates to study mortality trends unless it can be demonstrated that the population concerned has retained the same age and sex distribution.

When a specific rate has been selected for a study of trends the standard statistical methods of analysis are used, but in the case of mortality trends it is customary to analyse not the calculated rates themselves, but the logarithm of the rate for this has the great advantage that the logarithm of a rate which is changing by a constant proportion each year will be a linear function of time and also that the difference between the logarithm of the rates for two age groups will remain constant for groups if both rates change by a constant proportion. The slope of the regression line of the logarithm of the rate will be the best estimator of the rate of change of the rate over time. For these reasons it is customary for studies of mortality trends over time to be based upon the logarithm of the rates.

**FIGURE VI.1 HERE**

An illustration of the use of a semi-logarithmic graph to plot the trends of mortality for specific age groups is shown in Figure VI.1 which uses the rates for England and Wales from 1931-71. A horizontal axis of the graph is time in years, the vertical axis is the logarithm of the mortality rates. It is a feature of human mortality that the mortality rates decline during childhood from the day of birth until about 10 years and that after this age of maximum vitality they start to increase again at a rate which is proportional to the exponential of age. This means that in the semi-logarithmic graph the mortality rates for one sex above the age of 10 can be shown clearly without confusion for a long period of time. The experience of the two sexes should be shown separately, but side by side and the rates for children under 10 should be shown on another graph so that they do not confuse the figure by crossing over the trend lines for the older ages.

Figure VI.1 shows all the relevant information about trends of age and sex specific mortality for one country for 40 years and if possible it is advisable to prepare such graphs to enable the trends to be studied.

But if trends for several countries have to be compared it is necessary to use a summary statistic for the mortality experience of a whole country.

VI.2.c Adjusted rates.

Two ways of measuring the total effect of changes of mortality are to use a method of
Fig. VI.1
England and Wales. Age and Sex Specific Death Rates per 1000 1931-35, by five year intervals to 1966-70
adjustment or to use the methods of the life table. As described in Chapter III, the method of adjustment is to estimate either the effect of changing rates on a standard population (direct method) or the effect of standard rates on a changing population (indirect method). The life table technique described in Chapter IV uses the age specific mortality rates to calculate the proportion of persons who would survive at the end of each year until all have died, and then calculates the total number of years of life that can be expected for the total population. The expectation of life at birth is usually used as the most important characteristic of a mortality table when studying trends. In England and Wales it has been customary to study the trends in mortality by the indirect method of adjustment using the standardized mortality ratio (England and Wales 1950-52 mortality rates taken as standard).

The trend of mortality amongst males in England and Wales between 1931-1970 using both the standardized mortality ratio and the expectation of life is shown in Table VI.1. The standardized mortality ratio using the average rates for the three years 1950-52 as standard have fallen from 134 in 1931-35 to 88 in 1966-70, a fall of 33%. In contrast, however, the expectation of life at birth has increased from 58.7 years to 68.7 years, an increase of 17%. or the life table death rate has fallen from 19.0 to 14.5. The reason for this marked difference in the estimate of the improvement of mortality is that the improvements in expectation of life is very insensitive to large changes in mortality rates.

VI.2.d Concomitant changes in methodology.

It is also important to ensure that there have not been changes in the methods of recording data or of preparing rates derived from these.

TABLE VI.1

Table VI.1 and Figure VI.1 also demonstrate the effect of two such possible changes in England and Wales during the period 1939-49; mortality rates for males were calculated for the civilian population only and the deaths of all military personnel were excluded. This had the surprising effect of increasing the age specific rates for males aged 13-24 and aged 25-34 because the civilian population at these ages were selected men who were unfit for military service and who in consequence included many men who were at high risk of death. These rates for these years are, therefore, not a true measure of trend.

In the period 1946-50, new social security benefits for the elderly were introduced which mean that after 1950 all persons claiming these benefits had to produce a birth certificate, hence the accuracy of the statements of age at death was improved and it is probable that the apparent increase in the mortality rates over age 85 in the period 1951-60 was due to this cause and not to a true decline in the health of these men. In many newly established states and in countries in which boundaries have changed recently, these changes will have considerable effect on the character of the population and it will be important to identify such changes in any study of trends.

One of the most important methodological changes is the decennial revision of the International Classification of Causes of Death, which will be discussed later in Chapter VIII (section 4), but these changes will not of course influence the measurement of mortality from all causes of death.

VI.2.e Expectation of Life at Birth, e_o - International comparisons of trend.

Many countries now endeavour to calculate an accurate estimate for e_o but only a few countries in the world have a series of estimates which show the trends over a number of years. Table VI.2 shows the trend in e_o for a selected number of countries with records which cover the period 1920-1970. For a few countries it is possible to obtain values of e_o which go back to the beginning of the 19th century, but such information is only of historic interest. The countries in Table VI.2 are selected to represent both the pre-industrial countries and the industrial countries of North America and Europe.

TABLE VI.2

The trend in every country has been for the expectation of life for males and for females to increase during the most part of the half century; in most countries the life expectancy of women is higher than for men and the increase in expectancy has been greater for women than for men. There are, however, a few exceptions.

The expectancy of life was lower for women than for men in both India and Sri Lanka
### TABLE VI.1

Trends in Age Specific Death Rates

**England and Wales.**

<table>
<thead>
<tr>
<th>PERIOD</th>
<th>AGE</th>
<th>0-1</th>
<th>1-4</th>
<th>5-9</th>
<th>10-14</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65-74</th>
<th>75-84</th>
<th>ALL AGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1931-35</td>
<td></td>
<td>70</td>
<td>6.9</td>
<td>2.3</td>
<td>1.4</td>
<td>2.8</td>
<td>3.3</td>
<td>5.4</td>
<td>11.2</td>
<td>23.6</td>
<td>56.7</td>
<td>135</td>
<td>12.7</td>
</tr>
<tr>
<td>36-40</td>
<td></td>
<td>62</td>
<td>5.0</td>
<td>2.0</td>
<td>1.3</td>
<td>2.6</td>
<td>3.1</td>
<td>5.0</td>
<td>11.0</td>
<td>25.0</td>
<td>56.8</td>
<td>140</td>
<td>13.5</td>
</tr>
<tr>
<td>41-45(2)</td>
<td></td>
<td>56</td>
<td>3.7</td>
<td>1.7</td>
<td>1.2</td>
<td>3.6</td>
<td>4.2</td>
<td>4.8</td>
<td>9.8</td>
<td>23.1</td>
<td>51.7</td>
<td>122</td>
<td>15.1</td>
</tr>
<tr>
<td>46-50(2)</td>
<td></td>
<td>41</td>
<td>1.9</td>
<td>0.88</td>
<td>0.69</td>
<td>1.5</td>
<td>1.9</td>
<td>3.2</td>
<td>8.6</td>
<td>22.4</td>
<td>51.6</td>
<td>119</td>
<td>12.8</td>
</tr>
<tr>
<td>51-55</td>
<td></td>
<td>30</td>
<td>1.2</td>
<td>0.55</td>
<td>0.48</td>
<td>1.0</td>
<td>1.4</td>
<td>2.7</td>
<td>7.9</td>
<td>22.5</td>
<td>54.6</td>
<td>127</td>
<td>12.5</td>
</tr>
<tr>
<td>56-60</td>
<td></td>
<td>25</td>
<td>1.0</td>
<td>0.49</td>
<td>0.40</td>
<td>1.0</td>
<td>1.2</td>
<td>2.4</td>
<td>7.4</td>
<td>21.9</td>
<td>53.7</td>
<td>123</td>
<td>12.3</td>
</tr>
<tr>
<td>61-65</td>
<td></td>
<td>23</td>
<td>0.94</td>
<td>0.47</td>
<td>0.41</td>
<td>1.0</td>
<td>1.1</td>
<td>2.4</td>
<td>7.4</td>
<td>21.8</td>
<td>54.2</td>
<td>123</td>
<td>12.4</td>
</tr>
<tr>
<td>66-70</td>
<td></td>
<td>21</td>
<td>0.86</td>
<td>0.42</td>
<td>0.39</td>
<td>0.96</td>
<td>1.0</td>
<td>2.3</td>
<td>7.1</td>
<td>21.2</td>
<td>53.6</td>
<td>119</td>
<td>12.3</td>
</tr>
</tbody>
</table>

**FEMALES**

<table>
<thead>
<tr>
<th>PERIOD</th>
<th>AGE</th>
<th>0-1</th>
<th>1-4</th>
<th>5-9</th>
<th>10-14</th>
<th>15-24</th>
<th>25-34</th>
<th>35-44</th>
<th>45-54</th>
<th>55-64</th>
<th>65-74</th>
<th>75-84</th>
<th>ALL AGES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1931-35</td>
<td></td>
<td>54</td>
<td>6.2</td>
<td>2.1</td>
<td>1.4</td>
<td>2.5</td>
<td>3.1</td>
<td>4.3</td>
<td>8.0</td>
<td>17.0</td>
<td>42.8</td>
<td>108.9</td>
<td>11.4</td>
</tr>
<tr>
<td>36-40</td>
<td></td>
<td>48</td>
<td>4.4</td>
<td>1.7</td>
<td>1.1</td>
<td>2.2</td>
<td>2.7</td>
<td>3.8</td>
<td>7.5</td>
<td>16.5</td>
<td>41.8</td>
<td>107.1</td>
<td>11.6</td>
</tr>
<tr>
<td>41-45(2)</td>
<td></td>
<td>44</td>
<td>3.3</td>
<td>1.3</td>
<td>1.0</td>
<td>2.1</td>
<td>2.5</td>
<td>3.3</td>
<td>6.4</td>
<td>14.0</td>
<td>36.0</td>
<td>93.5</td>
<td>11.1</td>
</tr>
<tr>
<td>46-50(2)</td>
<td></td>
<td>32</td>
<td>1.6</td>
<td>0.64</td>
<td>0.54</td>
<td>1.3</td>
<td>1.8</td>
<td>2.6</td>
<td>5.5</td>
<td>12.8</td>
<td>34.4</td>
<td>93.2</td>
<td>10.9</td>
</tr>
<tr>
<td>51-55</td>
<td></td>
<td>23</td>
<td>1.0</td>
<td>0.39</td>
<td>0.34</td>
<td>0.60</td>
<td>1.1</td>
<td>2.1</td>
<td>4.9</td>
<td>11.8</td>
<td>33.1</td>
<td>92.4</td>
<td>10.9</td>
</tr>
<tr>
<td>56-60</td>
<td></td>
<td>20</td>
<td>0.82</td>
<td>0.33</td>
<td>0.27</td>
<td>0.45</td>
<td>0.81</td>
<td>1.8</td>
<td>4.5</td>
<td>10.9</td>
<td>30.7</td>
<td>86.4</td>
<td>10.9</td>
</tr>
<tr>
<td>61-65</td>
<td></td>
<td>18</td>
<td>0.78</td>
<td>0.32</td>
<td>0.25</td>
<td>0.43</td>
<td>0.73</td>
<td>1.8</td>
<td>4.4</td>
<td>10.6</td>
<td>29.8</td>
<td>83.6</td>
<td>11.2</td>
</tr>
<tr>
<td>66-70</td>
<td></td>
<td>16</td>
<td>0.70</td>
<td>0.28</td>
<td>0.25</td>
<td>0.42</td>
<td>0.65</td>
<td>1.7</td>
<td>4.3</td>
<td>10.3</td>
<td>28.0</td>
<td>77.5</td>
<td>11.2</td>
</tr>
</tbody>
</table>

(1) Infant Mortality Rate per 1,000 live births

(2) Deaths in Civilian Population per 1,000 Civil Population
<table>
<thead>
<tr>
<th>PERIOD</th>
<th>Mexico</th>
<th>U.S.A.</th>
<th>Jamaica</th>
<th>Chile</th>
<th>Sri Lanka</th>
<th>Japan</th>
<th>India</th>
<th>Denmark</th>
<th>France</th>
<th>Italy</th>
<th>Sweden</th>
<th>England and Wales</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MALES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1920 - 29</td>
<td>31.0</td>
<td>55.5</td>
<td>35.9</td>
<td>-</td>
<td>32.7</td>
<td>42.1</td>
<td>26.9</td>
<td>60.3</td>
<td>52.2</td>
<td>49.3</td>
<td>61.0</td>
<td>55.6</td>
</tr>
<tr>
<td>1930 - 39</td>
<td>38.9</td>
<td>57.7</td>
<td>-</td>
<td>40.4</td>
<td>-</td>
<td>46.9</td>
<td>-</td>
<td>62.0</td>
<td>55.9</td>
<td>53.8</td>
<td>63.8</td>
<td>58.7</td>
</tr>
<tr>
<td>1940 - 49</td>
<td>43.8</td>
<td>61.6</td>
<td>51.2</td>
<td>40.9</td>
<td>46.8</td>
<td>42.6</td>
<td>32.4</td>
<td>67.8</td>
<td>61.9</td>
<td>-</td>
<td>69.1</td>
<td>66.4</td>
</tr>
<tr>
<td>1950 - 54</td>
<td>48.1</td>
<td>65.6</td>
<td>55.7</td>
<td>49.8</td>
<td>56.4</td>
<td>56.2</td>
<td>-</td>
<td>69.8</td>
<td>63.6</td>
<td>63.8</td>
<td>70.5</td>
<td>66.5</td>
</tr>
<tr>
<td>1955 - 59</td>
<td>52.3</td>
<td>66.6</td>
<td>-</td>
<td>-</td>
<td>58.1</td>
<td>63.9</td>
<td>41.9</td>
<td>70.4</td>
<td>65.0</td>
<td>65.7</td>
<td>71.2</td>
<td>67.5</td>
</tr>
<tr>
<td>1960 - 64</td>
<td>-</td>
<td>66.6</td>
<td>62.6</td>
<td>54.4</td>
<td>61.9</td>
<td>65.3</td>
<td>47.5</td>
<td>70.3</td>
<td>67.2</td>
<td>67.2</td>
<td>71.6</td>
<td>68.3</td>
</tr>
<tr>
<td>1965 - 69</td>
<td>61.0</td>
<td>66.8</td>
<td>NA</td>
<td>60.5</td>
<td>NA</td>
<td>67.7</td>
<td>NA</td>
<td>70.2</td>
<td>67.8</td>
<td>67.9</td>
<td>71.8</td>
<td>68.3</td>
</tr>
<tr>
<td>FEMALES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1920 - 29</td>
<td>33.2</td>
<td>57.4</td>
<td>38.2</td>
<td>-</td>
<td>30.7</td>
<td>43.2</td>
<td>26.6</td>
<td>61.9</td>
<td>56.1</td>
<td>50.8</td>
<td>63.2</td>
<td>59.6</td>
</tr>
<tr>
<td>1930 - 39</td>
<td>41.9</td>
<td>61.0</td>
<td>-</td>
<td>41.3</td>
<td>-</td>
<td>49.6</td>
<td>-</td>
<td>63.8</td>
<td>61.6</td>
<td>57.5</td>
<td>66.1</td>
<td>62.9</td>
</tr>
<tr>
<td>1940 - 49</td>
<td>44.6</td>
<td>65.9</td>
<td>54.6</td>
<td>43.2</td>
<td>44.7</td>
<td>51.1</td>
<td>31.7</td>
<td>70.1</td>
<td>67.4</td>
<td>-</td>
<td>71.6</td>
<td>71.5</td>
</tr>
<tr>
<td>1950 - 54</td>
<td>51.1</td>
<td>71.0</td>
<td>58.9</td>
<td>53.9</td>
<td>54.8</td>
<td>59.6</td>
<td>-</td>
<td>72.6</td>
<td>69.3</td>
<td>67.2</td>
<td>73.4</td>
<td>71.2</td>
</tr>
<tr>
<td>1955 - 59</td>
<td>55.1</td>
<td>72.7</td>
<td>-</td>
<td>-</td>
<td>57.1</td>
<td>68.4</td>
<td>40.6</td>
<td>73.8</td>
<td>71.1</td>
<td>70.0</td>
<td>74.7</td>
<td>73.0</td>
</tr>
<tr>
<td>1960 - 64</td>
<td>-</td>
<td>73.1</td>
<td>66.6</td>
<td>59.9</td>
<td>61.4</td>
<td>70.2</td>
<td>47.5</td>
<td>74.4</td>
<td>73.8</td>
<td>72.3</td>
<td>75.7</td>
<td>74.1</td>
</tr>
<tr>
<td>1965 - 69</td>
<td>63.7</td>
<td>73.7</td>
<td>NA</td>
<td>66.0</td>
<td>NA</td>
<td>72.9</td>
<td>NA</td>
<td>74.7</td>
<td>75.0</td>
<td>73.4</td>
<td>76.5</td>
<td>74.4</td>
</tr>
</tbody>
</table>
during the period 1920–1960 but since that time the two sexes are about the same in both countries; in Mexico where the expectation of life has doubled in 50 years, the female rate has remained almost constantly 2 years longer than the male rate; in certain industrial countries, particularly the USA, Denmark, Sweden and England and Wales the male expectation of life has not increased during the period from 1955.

Each of these exceptions is the possible topic of a detailed study of the age and sex specific rates to elucidate more clearly what are the true influences affecting the trends.

In such studies particular attention should be paid to the methodology used to obtain the life table estimates. For instance, in some countries where adequate registration is not possible, estimates of the probability of surviving $\hat{p}_x = (1 - \hat{q}_x)$ may be obtained by the method of differences between the censuses which will introduce very different biases from the standard methods of estimating $q_x$ from the mortality rates, $m_x$ (See Chapter IV.8.)

VI.3 Trends in Shorter Interval of Time.

VI.3.a The decade.

When the trends are to be studies over a short period of time it is possible to make the assumption that the technique of age adjustment will not greatly influence any conclusions to be drawn from trend lines and consequently the labour of calculating an adjusted rate or an expectation of life may be avoided. This is a particular advantage if disease specific rates are to be studied.

Table VI.3 here

Table VI.3 examines the trends of certain selected causes of death as defined by the ICD 8th revision (see Chapter VIII) in England and Wales over the decade 1961–1971 by the use of the crude death rates per million. This table illustrates some of the types of patterns of trend that can be discerned.

The intestinal infectious diseases were already at a very low rate in 1961 and there has been a very slow drift for mortality to decline; there has been no epidemic causing a sudden increase in deaths during this period to be detected by this crude rate, although there were several outbreaks of typhoid causing some deaths in this period. This suggests that detailed trends by areas should be studied. Tuberculosis mortality has declined steadily year by year during the decade, with two hesitations until the rate has fallen by 60% at the end of the decade. This again suggests that there may be interesting age or sex specific patterns that should be examined in more detail.

The cancer death rates show two trends, mortality from cancer of the colon and rectum has remained very constant during the whole period, but cancers of the lung are increasing steadily.

The trend for ischaemic heart disease is clearly a rising one, but there are marked fluctuations in individual years which would suggest that there are at least two factors influencing these deaths, one a long-term trend and the other a short-term effect which operates in certain particular years. Further studies to identify these would be suggested.

The influenza mortality is typical of that of an epidemic type disease with differences of the order of tenfold between years and with the epidemic years appearing almost at random. The fluctuations in the mortality from pneumonia are very similar to those from influenza and they occur in the same years which suggest that pneumonia is influenced by two factors, one of which is associated with influenza and one which is a long-term trend.

The deaths from asthma are particularly interesting – they demonstrate a steady rise from 1962 reaching a peak in 1966–1967, followed by a fall. Detailed studies to investigate this have been undertaken and suggest that one specific method of therapy used during this period may have been the cause and that when this was withdrawn the mortality rate from asthma fell again.

The deaths from trauma show yet a different pattern – they remain constant until the end of 1966, when they show a sudden drop during 1967 and then remain constant again at a lower level for the next four years. In 1967 the 8th Revision of the ICD replaced the 7th Revision, and it might be possible that this change in level could be attributed to an
### TABLE VI.3

Trends in Selected Causes of Death

Death Rates Per Million

England and Wales 1961-1971

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MALES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intestinal Infections</td>
<td>001-009</td>
<td>20</td>
<td>22</td>
<td>22</td>
<td>21</td>
<td>19</td>
<td>20</td>
<td>18</td>
<td>17</td>
<td>17</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>Tuberculosis</td>
<td>010-019</td>
<td>105</td>
<td>99</td>
<td>94</td>
<td>79</td>
<td>69</td>
<td>71</td>
<td>61</td>
<td>63</td>
<td>56</td>
<td>47</td>
<td>43</td>
</tr>
<tr>
<td>Cancer Colon</td>
<td>153</td>
<td>164</td>
<td>163</td>
<td>156</td>
<td>157</td>
<td>156</td>
<td>164</td>
<td>165</td>
<td>172</td>
<td>168</td>
<td>172</td>
<td>172</td>
</tr>
<tr>
<td>Rectum</td>
<td>156</td>
<td>130</td>
<td>130</td>
<td>128</td>
<td>128</td>
<td>126</td>
<td>126</td>
<td>130</td>
<td>132</td>
<td>131</td>
<td>130</td>
<td>132</td>
</tr>
<tr>
<td>Lung</td>
<td>162</td>
<td>869</td>
<td>894</td>
<td>909</td>
<td>933</td>
<td>958</td>
<td>969</td>
<td>1002</td>
<td>1012</td>
<td>1040</td>
<td>1045</td>
<td>1060</td>
</tr>
<tr>
<td>Ischaemic Heart Disease</td>
<td>410-414</td>
<td>2974</td>
<td>3106</td>
<td>3212</td>
<td>3093</td>
<td>3241</td>
<td>3231</td>
<td>3191</td>
<td>3374</td>
<td>3405</td>
<td>3392</td>
<td>3501</td>
</tr>
<tr>
<td>Influenza</td>
<td>470-474</td>
<td>156</td>
<td>66</td>
<td>63</td>
<td>21</td>
<td>17</td>
<td>70</td>
<td>16</td>
<td>79</td>
<td>102</td>
<td>154</td>
<td>13</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>480-486</td>
<td>634</td>
<td>644</td>
<td>743</td>
<td>603</td>
<td>619</td>
<td>689</td>
<td>619</td>
<td>768</td>
<td>790</td>
<td>804</td>
<td>732</td>
</tr>
<tr>
<td>Accidents and Trauma</td>
<td>800-999</td>
<td>611</td>
<td>610</td>
<td>618</td>
<td>620</td>
<td>604</td>
<td>603</td>
<td>565</td>
<td>535</td>
<td>547</td>
<td>532</td>
<td>528</td>
</tr>
</tbody>
</table>

**NOTE:** 1961 - 1966 The 7th Revision of I.C.D. was in use

1967 - 1971 " 8th " " " " " " " 
artifact of coding, but this was not the case – the fall was almost entirely due to a reduction in mortality from motor vehicle traffic accidents as a result of the introduction of severe laws against drivers using alcohol. The importance of changes in the ICD revisions will be discussed in Chapter VIII.

Mortality trends may, therefore, be of all types; first they may be constant; second they may show a linear trend either rising or falling; third, they may be fluctuating with cyclic variations or with random variations, or finally, they may show sudden discontinuities. The total mortality from all causes is, therefore, the effect of the combination of many different causes operating in different ways.

Each particular pattern must be examined carefully by demographic and epidemiological methods and it will seldom be found useful to calculate a particular mathematical formula to fit a trend as this can seldom be extrapolated beyond the recorded experience.

VI.3.b Trends within a year.

The acute epidemics of the past, plague, cholera, or influenza, announced themselves by a sudden increase in the numbers of deaths in a community, and communities took good care to report the total number of deaths to the authorities so that precautions might be taken. It was from such measures of the trend of mortality that the science of epidemiology was born.

It is still an important function of any health information system to report deaths and causes of death quickly so that an analysis of mortality trends by day, by week or by season can be undertaken speedily as part of a public health control system. Although a notification system for certain diseases is obviously more sensitive and effective as a method of control, yet such systems are often incomplete so that mortality analysis remains a necessary control system.

The most virulent epidemics that have swept the world during the last 30 years have been due to influenza. It is a disease that may not be accurately diagnosed especially if there are other chronic respiratory diseases endemic in the area, but its case fatality rate is high and the immediate increase in mortality, especially amongst the elderly, can be large. Figure VI.2 shows the weekly deaths from pneumonia and from influenza in England and Wales during the years 1968, 1969, and 1970, when there were two outbreaks of influenza. It is clear that a study of the trends of the total number of deaths alone would accurately define the dates of the epidemics and moreover, that the severity of the epidemics were such that they could be identified by total deaths from all causes.

In studies of short term effects there is no necessity to use any sophisticated calculations at all, but simply to record the actual numbers of deaths occurring.

VI.3.c The season.

The influence of seasons is another factor that can be examined from studies of mortality trends. In the temperate regions of the Northern hemisphere mortality increases during the winter months, but in sub-tropical regions mortality may increase either during the hotter seasons of the year, or during the rainy season. In some areas a cold wind or a desert storm may cause sudden but regular rises in mortality. This is an association known from ancient times and was reported in general descriptive terms by Hippocrates. In modern times, however, such effects should be measured accurately and if prophylactic methods are employed to reduce the excess seasonal mortality, it is essential that the trends in mortality should be accurately recorded to demonstrate the efficacy of the methods.

The analysis of deaths by week or month can only be done if the deaths are recorded according to the date of occurrence of the event because the date of registration will depend upon the number of working days in a month and will be severely affected by holidays.

TABLES VI.4 and 5 HERE

Tables VI.4 and VI.5 show the number of deaths occurring each month in England and Wales in 1971. The first point to remember is that the number of days in a calendar month differs and an adjustment must be made for this. The most convenient way to do so is to adjust the number of deaths as if each month has 31 days and deaths in February are increased by 11% and those in April, June, September and November by 3%.

Table VI.4 illustrates the cyclical pattern of mortality in the Northern hemisphere;
### TABLE VI.4

Trends by Month of Death

England and Wales 1971

<table>
<thead>
<tr>
<th>Month: ALL AGES</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>VI</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
<th>IX</th>
<th>X</th>
<th>XI</th>
<th>XII</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AGE 15 - 19</td>
<td>M</td>
<td>128</td>
<td>99</td>
<td>126</td>
<td>107</td>
<td>158</td>
<td>130</td>
<td>168</td>
<td>151</td>
<td>125</td>
<td>110</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>50</td>
<td>43</td>
<td>63</td>
<td>61</td>
<td>45</td>
<td>52</td>
<td>72</td>
<td>50</td>
<td>36</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>AGE 75 - 79</td>
<td>M</td>
<td>4143</td>
<td>3425</td>
<td>3803</td>
<td>3485</td>
<td>3233</td>
<td>3052</td>
<td>3025</td>
<td>2908</td>
<td>2887</td>
<td>3135</td>
<td>3245</td>
</tr>
<tr>
<td></td>
<td>F</td>
<td>4749</td>
<td>3915</td>
<td>4284</td>
<td>3916</td>
<td>3679</td>
<td>3540</td>
<td>3474</td>
<td>3305</td>
<td>3745</td>
<td>3893</td>
<td>4448</td>
</tr>
</tbody>
</table>

### TABLE VI.5

England and Wales. Males only by month.

<table>
<thead>
<tr>
<th>Cause</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>VI</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
<th>VIII</th>
<th>IX</th>
<th>X</th>
<th>XI</th>
<th>XII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer of Lung</td>
<td>2158</td>
<td>1790</td>
<td>2232</td>
<td>2063</td>
<td>2076</td>
<td>2047</td>
<td>2122</td>
<td>2078</td>
<td>2041</td>
<td>2225</td>
<td>2111</td>
<td>2188</td>
</tr>
<tr>
<td>Ischaemic Heart Disease</td>
<td>7976</td>
<td>6799</td>
<td>7807</td>
<td>7088</td>
<td>6766</td>
<td>6502</td>
<td>6237</td>
<td>5955</td>
<td>5990</td>
<td>6662</td>
<td>7487</td>
<td>7601</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>2211</td>
<td>1621</td>
<td>1732</td>
<td>1552</td>
<td>1308</td>
<td>1175</td>
<td>1129</td>
<td>1065</td>
<td>1035</td>
<td>1258</td>
<td>1426</td>
<td>1869</td>
</tr>
<tr>
<td>Motor vehicle traffic accidents</td>
<td>420</td>
<td>353</td>
<td>328</td>
<td>344</td>
<td>354</td>
<td>352</td>
<td>395</td>
<td>394</td>
<td>377</td>
<td>401</td>
<td>384</td>
<td>408</td>
</tr>
</tbody>
</table>
deaths of males fluctuate from 21,070 per 31 days in August to 29,083 per 31 days in January, an increase of 38% and there is an exactly similar pattern for women. The table, however, also demonstrates that the pattern need not necessarily be the same at all ages or for both sexes. The mortality rate for males aged 15-19 is highest in July and lowest in November due to sport and traffic accidents.

It is, therefore, important once again to analyse the data in depth.

Similarly Table VI.5 demonstrates that certain diseases will fluctuate widely so that deaths from pneumonia in January are double those in August, whereas, cancer of the lung remains constant throughout the year. Ischaemic heart disease follows a cyclical pattern and it is important to note that motor vehicle traffic accidents are higher in the winter than the summer although traffic is higher in the summer.

VI.4 Trends by Cause of Death.

In any study of trends the diagnosis of the cause of death is of primary importance for any public health, epidemiological or even for socio-economic studies. For if there is an improvement in the health of the community due to improved health services or effective therapy, or if there is a decline in health standards due to a new infection or to worsening environmental conditions these will manifest themselves in some specific type of mortality which can be identified and diagnosed.

There are, however, serious difficulties in handling this information concerning causes of death which are discussed in greater detail in Chapter VIII and the problems of trend analysis by cause of death will be examined in greater depth in that chapter.

VI.5 Cohort Analysis.

VI.5.a The basic concepts.

Age and sex specific rates which have been calculated from the mortality experience of a population for a particular period of one year are a cross section of the experience of people who were born at different times and have experienced different hazards throughout life.

There is no inherent reason why the mortality of group of persons aged say 60 in one year should be the same as that experienced by a group of men a year older than themselves in the preceding year.

An alternative method of examining the rates is to consider the mortality experience of a single group of people born in one year as they go through life and to compare their life time experience with that of another group born in a different year. Such groups are known as cohorts from the description of the formation of Roman soldiers who marched in file one behind another into battle. These generation or cohort mortality tables represent what actually has happened in life, but they are difficult to calculate and to use because a long series of accurate rates are required over 60 or 70 years and the results are only completed when the whole cohort has passed away.

But there are many important differences between the life time cohort experience and the cross sectional official statistics. Derrick (1925) presented a cohort analysis of England for the period 1776 to 1916 and Kermack, McKendrick and McKinley (1934) did a similar analysis for Scotland from 1860 and for Sweden from 1751, more recently Moriyama and Gustavus (1972) completed an analysis for the United States for four cohorts starting in 1896-1900. Frost 1939 applied this technique to disease specific rates for tuberculosis in the United States and Picken (1940) did the same for the English tuberculosis experience and Springett (1950) completed a fuller analysis. Other workers have applied this technique to death from cancer sites, Stocks (1953) Haenszel (1956), Cutler (1954). Case (1956) reviewed the use of these techniques; and concluded that the existing ideas of laws of mortality based upon cross sectional or "age contour" death rates could be quite misleading and that cohort analysis should be an essential part of any detailed study of national mortality.

VI.5.b The complete cohort table.

The concept of the method is basically simple. It is to identify a whole series of cohorts which are formed by the population born in a single year or a few adjacent years, then
to follow them and their mortality experience throughout the remainder of their lives. The tabulations produced by a Central Statistical Bureau, however, will not be in a suitable form for immediate use because the annual reports published will give for each calendar year an age specific mortality rates for each single year of age for all causes of death, these will not be available for specific causes.

For any one calendar year there is an adequate cross sectional or "age contour" study of mortality which can be compared with similar contours in other years, and by comparing successive age specific rates over a series of years there is an adequate calendar contour for each age of life. But to obtain a cohort analysis it is necessary to consider the mortality for an increasing age group in each successive calendar year report. The principle is best illustrated by a diagram. Table VI.6 is a complete matrix for a population of all the age specific mortality rates ($m_x$) from age 0-79 in single years of age for a series of 80 calendar years from year $t$ to year $t + 79$. Age is denoted by $x$, calendar year by $t$, and a specific cohort by C.

If this table is examined along the rows then each row will give the trend of mortality over the 80 years for one specific age group and it is this type of trend that has been discussed in Chapter VI.3. If the columns of the table are examined these give the pattern of mortality ($m_x$) for all ages from 0-79, which is the way in which the data are examined in Chapter VII.

If, however, the diagonals of this table are examined they indicate the mortality experience of one group of persons all born in one year for each age of their life. A cohort of persons born in year $t_0$ are indicated in the table by the letter $C_0$. These same people are 1 year old in year $t + 1$, they are $x$ years of age in year $t + x$ and 79 years of age in year $t + 79$ and hence their mortality can be identified throughout their life.

It is this diagonal analysis that we propose to develop. Various methods have been used. The first problem is that national mortality rates are not usually available for single years of life for specific diseases, but are published in quinquennial age groupings and thus to proceed to the full form of analysis in Table VI.6 it is necessary to use some method of interpolation to devise the single age rates. But in doing this it is not the object to smooth the data in such a way that irregularities which may be great interest to epidemiologists or medical statisticians are eliminated; such as for instance a sudden increase in motor traffic accidents at 18 years, or of suicide at age of retirement. This, therefore, implies that some of the more sophisticated techniques used for actuarial purposes are not entirely suitable. Moriyama and Gustavus (1972) used the method developed by Beers (1945) which minimises the fourth differences. No general consensus is available to recommend any specific method, but a simple difference formula is all that is required. The chosen interpolation method should then be applied to obtain single age specific mortality rates for every year to be studied.

A more difficult method is to use the figures for the population and the total number of deaths from each cause of death to be studied and interpolate these separately and then recalculate the mortality rates. Such a refinement does not, however, seem to be justified unless there are special reasons why the published mortality rates cannot be relied upon.

The deaths that occur during one calendar year $t + x$ of persons aged $x$ completed years at the time of death, consist of persons who were born in the year $t$ and year $t + 1$. But this difficulty is avoided in the first stages of the analysis by the approximation that all members of a cohort are assumed to be at the same age at the mid year.

When these single age mortality rates have been derived over a long series of years, it is possible to construct a complete table of the form of Table VI.6 and hence pick out the cohort mortality experience of individual years of birth. These then are the basic cohort mortality rates.

But just as in age specific mortality analysis it has been found that for biological or epidemiological studies it is preferable to work with quinquennial age groups, so in cohort studies it is advisable to work with aggregates of five cohorts forming successive five year groups by year of birth from $t - 2$ to $t + 2$ onwards. No firm recommendations exist as to which years should be taken for the centre of such cohorts, but Moriyama and Gustavus (1972) used 1896, 1908, 1918 and 1928 as their centre points, and Case (1956) used 1853, 1873, 1893,
### TABLE VI.6

**Construction of a cohort table of mortality**

<table>
<thead>
<tr>
<th>Current Age (Ax)</th>
<th>CAL ENDAR YEAR (Yt)</th>
<th>( t_0 )</th>
<th>( t+1 )</th>
<th>( t+2 )</th>
<th>( t+x-1 )</th>
<th>( t+x )</th>
<th>( t+x+1 )</th>
<th>( t+79 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td>( C_0 )</td>
<td>( C+1 )</td>
<td>( C+2 )</td>
<td></td>
<td></td>
<td></td>
<td>( C+79 )</td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>( C-1 )</td>
<td>( C_0 )</td>
<td>( C+1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>( C-2 )</td>
<td>( C-1 )</td>
<td>( C_0 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
</tr>
<tr>
<td>( x-1 )</td>
<td></td>
<td>( C-x+1 )</td>
<td></td>
<td></td>
<td>( C_0 )</td>
<td>( C+1 )</td>
<td>( C+2 )</td>
<td></td>
</tr>
<tr>
<td>( x )</td>
<td></td>
<td>( C-x+1 )</td>
<td></td>
<td></td>
<td>( C-1 )</td>
<td>( C_0 )</td>
<td>( C+1 )</td>
<td></td>
</tr>
<tr>
<td>( x+1 )</td>
<td></td>
<td>( C-x+1 )</td>
<td></td>
<td></td>
<td>( C-2 )</td>
<td>( C-1 )</td>
<td>( C_0 )</td>
<td></td>
</tr>
<tr>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
<td>( \ldots \ldots )</td>
</tr>
<tr>
<td>79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>( C_0 )</td>
</tr>
</tbody>
</table>

\( C_j = Y_t - A_x \)

### TABLE VI.7

**Cohort Grouping in Quinquennial Age Groups for 5 Calendar Years**

<table>
<thead>
<tr>
<th>Current Age</th>
<th>CAL ENDAR YEAR</th>
<th>( t_0+x-2 )</th>
<th>( t_0+x-1 )</th>
<th>( t_0+x )</th>
<th>( t_0+x+1 )</th>
<th>( t_0+x+2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( x-2 )</td>
<td>( C_0 )</td>
<td>( C_1 )</td>
<td>( C_2 )</td>
<td>( C_3 )</td>
<td>( C_4 )</td>
<td></td>
</tr>
<tr>
<td>( x-1 )</td>
<td>( C-1 )</td>
<td>( C_0 )</td>
<td>( C_1 )</td>
<td>( C_2 )</td>
<td>( C_3 )</td>
<td></td>
</tr>
<tr>
<td>( x )</td>
<td>( C-2 )</td>
<td>( C-1 )</td>
<td>( C_0 )</td>
<td>( C_1 )</td>
<td>( C_2 )</td>
<td></td>
</tr>
<tr>
<td>( x+1 )</td>
<td>( C-3 )</td>
<td>( C-2 )</td>
<td>( C-1 )</td>
<td>( C_0 )</td>
<td>( C_1 )</td>
<td></td>
</tr>
<tr>
<td>( x+2 )</td>
<td>( C-4 )</td>
<td>( C-3 )</td>
<td>( C-2 )</td>
<td>( C-1 )</td>
<td>( C_0 )</td>
<td></td>
</tr>
</tbody>
</table>
The mean value of 5 calendar years for one specific 5 year age group is used to represent the mortality of a five year cohort.

<table>
<thead>
<tr>
<th>Current Age</th>
<th>CALENDAR YEAR</th>
<th>Mean of Five Years</th>
<th>Relevant Cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 4</td>
<td>( t_0 + x ) ( t_0 + x-1 ) ( t_0 + x ) ( t_0 + x+1 ) ( t_0 + x+2 )</td>
<td>( c_0 ) ( c_{-5} ) ( c_{-10} ) ( c_{-15} ) ( c_{-20} ) ( c_{-25} ) ( c_{-75} ) ( c_{-80} )</td>
<td>C-2 to C+2 C-7 to C-3 C-12 to C-8 C-17 to C-13 C-22 to C-18 C-27 to C-23 C-77 to C-73 C-82 to C-78</td>
</tr>
<tr>
<td>5 - 9</td>
<td></td>
<td>( c_{-10} ) ( c_{-15} ) ( c_{-20} ) ( c_{-25} ) ( c_{-75} ) ( c_{-80} )</td>
<td></td>
</tr>
<tr>
<td>10 - 14</td>
<td></td>
<td>( c_{-15} ) ( c_{-20} ) ( c_{-25} ) ( c_{-75} ) ( c_{-80} )</td>
<td></td>
</tr>
<tr>
<td>15 - 19</td>
<td></td>
<td>( c_{-20} ) ( c_{-25} ) ( c_{-75} ) ( c_{-80} )</td>
<td></td>
</tr>
<tr>
<td>20 - 24</td>
<td></td>
<td>( c_{-25} ) ( c_{-75} ) ( c_{-80} )</td>
<td></td>
</tr>
<tr>
<td>25 - 29</td>
<td></td>
<td>( c_{-75} ) ( c_{-80} )</td>
<td></td>
</tr>
<tr>
<td>etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75 - 79</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>80 - 84</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The mortality rates at a single age for five adjacent cohorts are obtained from an unweighted mean of the five cohorts as representative of the quinquennial cohort.

VI.5.c Abridged cohort table.

The computation required for a complete cohort table for a number of causes of death which would cover 80 or more years of age and for a long series of years is very laborious and time consuming and even with advanced computer methods would be expensive. It is also doubtful whether such detailed analysis is required to study the effect of trends which are likely to be influencing mortality slowly over many years. Consequently it is recommended that a truncated form of cohort analysis should be constructed using quinquennial age groups and quinquennial cohorts.

The difficulty in doing this directly from the published data is that the members of a five years age grouping will vary for each calendar year, so that the membership of a cohort will differ year by year.

TABLE VI.7 HERE

This is illustrated in Table VI.7 where it will be seen that the quinquennial age group $x - 2$ to $x + 2$ during the period of calendar year $t - 2$ to $t + 2$ will consist of members drawn from 9 separate cohorts $C - 4$ to $C + 4$. The two extreme cohorts both contribute one year of experience to the central matrix, and the next inner cohorts $C - 3$ and $C + 3$ contribute two years each to the experience, the central cohort $C_0$ contributes 5 years experience to this block and does not appear in any adjacent quinquennial block.

If, therefore, the five successive annual values of the death rates between ages $x - 2$ and $x + 2$ are averaged, this pooled estimate of the mortality for $m_{x-2:x+2}$ will be derived from 25 cohort-years of experience; 19 of these cohort years will be contributed by the quinquennial cohort from $C - 2$ to $C + 2$, but 3 of the years are drawn from the preceding older cohort, who were born in years $C - 4$ and $C - 3$ and three from the following younger cohort born in $C + 3$ and $C + 4$.

If there is a trend in the long term pattern of a disease, then it is probable that the bias of introducing the experience of the two adjacent cohorts into estimate for the experience of the quinquennial cohort $C - 2$ to $C + 2$ will cancel each other and hence that the pooled average of the death rate for the five years is an unbiased estimation of the experience of the cohort $C - 2$ to $C + 2$. This is illustrated in Table VI.8, this simplifies matters greatly for all that is now required is to establish tables

TABLE VI.8 HERE

of the quinquennial age specific mortality rates, add them together in blocks of five and take a mean, which is a paper and pencil exercise and does not require extensive computation. It is equally applicable to mortality from individual causes of death as it is to total mortality experience.

An even more abbreviated form of the cohort table can be constructed by observing that the membership of the age specific group $x - 2$ to $x + 2$ in year $t_0$ consists of the same group of cohorts as the membership of the age group $x + 3$ to $x + 7$ in the year $t_5$. In practice it will be found that certain causes of death do not become at all prominent before a certain specific age. For example death due to cancer of the lung is rare before age 35 even in countries with a high incidence hence a series followed for 40 years would give a completely adequate cohort study for one cohort from age 35 to 74. Hence by an appropriate cross tabulation of quinquennial age specific death rates once every five years, a very simple estimate of cohort experience is obtained. This is a useful, valid and quick method of testing whether it is worthwhile to do a more detailed study. It does, however, reject 80% of the information available and hence should only be used as a preliminary technique.
VI.5.d  

Errors specific to cohort analysis.

The basic assumption of cohort analysis is that a single group of persons can be studied throughout their life time and if this assumption does not hold the technique is not relevant. It is, therefore, necessary that a series of tabulations should be available over many years, this period need not be as long as the average duration of life, but a shorter interval is useful when an overlap of several cohorts at different ages can contribute useful information. The minimum useful period, however, would be about 20 years.

The problem of immigration and emigration is a complication which is difficult to solve as the population of a country receiving immigrants is altered by this factor and a country loosing emigrants cannot know what happens to them and will have a selected survivor population left behind. One particular form of emigration, however, which can have profound effects on cohort analyses is the migration of military personnel in foreign wars who go abroad and only the survivors return, but the deaths of the fallen are not recorded in the records of the vital statistics in the home country. With two major world wars in the last 60 years this is a very relevant factor.

Another difficulty is change of national boundaries as a result of war, national independence or treaties of federation where the whole structure of a population may have changed.

Administrative difficulties such as changes in methods of registration or census methods, or changes in the methods of coding or data processing effect cohort analysis just as they do age specific analysis, but with cohort analysis the effect of such changes will be seen at different ages in different cohorts thus producing a more confused effect.

Nevertheless, although the difficulties are considerable, any of these types of problems which are liable to effect cohort studies will equally effect any studies of time series and any analyses of age specific rates. Hence the difficulties are inherent in the population itself rather than in the technique.

REFERENCES:


SPRINGETT, V.H., (1950), J. Hygiene, 481, 361.
